# Electronic Navigation Systems

## 3rd edition

# Electronic Navigation Systems

**Laurie Tetley** IEng FIEIE
Principal Lecturer in Navigation and Communication Systems

and

**David Calcutt** PhD MSc DipEE CEng MIEE
Formerly Senior Lecturer, Department of Electrical and Electronic Engineering,
University of Portsmouth

BUTTERWORTH
HEINEMANN

OXFORD AUCKLAND BOSTON JOHANNESBURG MELBOURNE NEW DELHI

Butterworth-Heinemann
Linacre House, Jordan Hill, Oxford OX2 8DP
225 Wildwood Avenue, Woburn, MA 01801–2041
A division of Reed Educational and Professional Publishing Ltd

A member of the Reed Elsevier plc group

First published *Electronic Aids to Navigation* 1986
Reprinted 1988
Second edition published as *Electronic Aids to Navigation: Position Fixing* 1991
Third edition 2001

**British Library Cataloguing in Publication Data**
Tetley, L. (Laurence), 1941–
Electronic navigation systems. – 3rd ed.
1. Electronics in navigation
I. Title II. Calcutt, D. (David), 1935–
623.8'504

**Library of Congress Cataloguing in Publication Data**
A catalogue record for this book is available from the Library of Congress

ISBN 0 7506 51385

Composition by Genesis Typesetting, Laser Quay, Rochester, Kent
Printed and bound in Great Britain

# Contents

# Preface

This new edition of *Electronic Navigation Systems* has been extensively rewritten to provide navigators with a detailed manual covering the principles and applications of modern systems.

The past decade has been witness to huge advances in technology and no more so than in maritime navigation and position fixing. As you might expect, spearheading this technological advance has been the computer. It has become as common on board ships as in our normal lives where it now influences virtually everything that we do. A new generation of ship's officer has been trained to use computers, trained to understand how they work and, more importantly, how they can be made to assist in the business of safe and precise navigation. But it would be a serious error to assume that the technology is perfect. All the systems currently used for navigation and position fixing are as near perfect as they can be, but it would be foolhardy to ignore the human link in the electronic chain of action and reaction. In the end, it is a ship's captain who bears the ultimate responsibility and the navigating officer who, with pride, safely brings his ship into port.

Readers will find that this new expanded edition includes many new systems and techniques whereas some older, now obsolete systems have been deleted. The hyperbolic systems, which once formed the backbone of global position fixing, have been decimated by the continuing expansion of the Global Positioning System (GPS).

The hyperbolic systems Decca and Omega have gone, but Loran-C, the one terrestrial network providing extensive coverage, remains as the designated back-up system to the GPS. By Presidential order, on 1 May 2000, Selective Availability, the method by which GPS accuracy was downgraded for civilian users, was set to zero. This significant event means that submetre accuracy position fixing is now available for all users, a factor that will have a major impact on GPS equipment and subsystems over the next decade.

Whilst the GPS is the undisputed king amongst satellite systems, it is by no means the only one. GLONASS, created and maintained by the Russian Federation, also provides users with accurate position fixes and the European Community is actively considering another system to be totally independent of the other two.

Although position fixing by satellite is of paramount importance there are other systems essential to safe navigation. Speed logging, depth sounding, and automatic steering systems are equally as important as they were decades ago and even that most traditional of all systems, the gyrocompass, has been digitized and refined. But essentially, system parameters remain unchanged; it is the collecting, processing and display of data that has been transformed.

Computerization and continuing development of large-scale integration (LSI) technology have been directly responsible for most of the changes. The large-scale manufacture of microchips has enabled the production of low-cost equipment with capabilities that could only have been dreamed about a decade ago. This reduction in size and cost has also brought sophisticated navigation equipment within reach of small-boat owners.

*Electronic Navigation Systems* has been written to support the training requirements of STCW-95 and consequently the book is an invaluable reference source for maritime navigation students. As with previous editions, each chapter opens with system principles and then continues with their application to modern equipment. Some sections, typically gyrocompass and automatic steering, still contain valid descriptions of analogue equipment but these have been further strengthened with the introduction of new digital technology. Wherever possible we have described the systems and equipment that you, the reader, are likely to meet on board your craft whether it is large or small.

The Global Maritime Distress and Safety System (GMDSS) is a subject which no mariner can ignore and consequently it has been outlined in this book. For extensive details about the principles and applications of this global communications system, see our book *Understanding GMDSS*.

Radar and Automatic Radar Plotting Aids (ARPA) are obviously essential to safe navigation and indeed are now integrated with other navigation systems. They are discussed in depth in the companion volume to this publication, *Electronic Aids to Navigation (RADAR and ARPA)*.

Laurie Tetley and David Calcutt
2000

# Acknowledgements

A book of this complexity containing leading edge technology must inevitably owe much to the co-operation of various individuals, equipment manufacturers and organizations. To single out one or more organizations is perhaps invidious. In many cases we have had no personal contact with individuals but despite this they gave freely of their time when information was requested.

We are extremely grateful for the assistance that the following companies and organizations gave during the writing of this book. We are particularly indebted to the organizations that permitted us to reproduce copyright material. Our sincere thanks go to the following.

COSPAS-SARSAT Secretariat
Det Norske Veritas (DNV)
Furuno Electric Co. Ltd
Garmin Industries
ICAN
The INMARSAT Organization
The International Maritime Organization (IMO)
Kelvin Hughes Ltd
Koden Electronics Co. Ltd
Krupp Atlas Elektronik
Litton Marine Systems
The NAVTEX Coordinating Panel
PC Maritime
SAL Jungner Marine
S G Brown Ltd
Sperry Marine Inc.
Thomas Walker & Son Ltd
Trimble Navigation Ltd
UK Hydrographic Office (UKHO)
Warsash Maritime Centre

The following figures are from the IMO publications on GMDSS and *The Navtex Manual*, and are reproduced with the kind permission of the International Maritime Organization, London: Figure 11.1, page 370; Figure 11.3, page 374; Figure 11.4, page 376; Figure 11.7, page 381; Figure 11.8, page 382; Figure 11.10, page 384; Figure 11.11, page 385.

# Chapter 1

# Radio wave propagation and the frequency spectrum

## 1.1 Introduction

This chapter outlines the basic principles of signal propagation and the radio frequency spectrum used by the navigation systems likely to be encountered on board merchant ships. The use of radio waves for terrestrial global communications and navigation causes major problems, particularly in the areas of frequency allocation and interference. Consequently, for safe and efficient working practices to be maintained on the restricted radio frequency spectrum, it is essential that this limited resource is carefully policed.

Radio waves cannot and do not respect international boundaries and, consequently, disputes arise between nations over the use of radio frequencies. The international governing body for radio communications services is the International Telecommunications Union (ITU) which, quite rightly, strictly regulates the allocation and use of frequencies. Any dispute that arises is settled by the ITU through various committees and affiliated organizations. All users of radiocommunications systems must be aware that they are licensed to use only specific frequencies and systems in order to achieve information transfer. It would be chaos if this were not so. Essential services, aeronautical, maritime or land based, would not be able to operate otherwise and lives could well be put at risk.

## 1.2 Maritime navigation systems and their frequencies

Maritime radio navigation requirements have always posed unique problems for the shipboard operator. A ship at sea presents many difficulties to the radio communications design engineer. The ship is constructed of steel which, when floating in salt water, becomes a very effective electromagnetic screen capable of rejecting or reflecting radio waves. In addition, modern ocean-going vessels are streamlined, spelling an end to those sturdy structures, i.e. smoke stacks and masts, that traditionally were used for holding antenna systems. Consequently, shipboard antenna systems tend to be less efficient than was once the case, giving rise to difficulties in both transmission and reception.

Maritime radio navigation and communication systems operate in a number of frequency bands. Listed below is a brief summary.

- Loran-C on the medium frequency 100 kHz.
- Navtex data on 518 kHz.
- Voice, radiotelex and digital selective calling in medium frequency band 1.6–3.4 MHz.
- Voice, radiotelex and DSC in high frequency bands between 3 and 30 MHz
- Voice and DSC in the very high frequency band 30–300 MHz.

- RADAR and SART on the frequency of 9 GHz.
- GPS satellite signals on L-band frequencies.
- INMARSAT communications signals on L-band frequencies.

In each case, the carrier frequency used has been chosen to satisfy two main criteria, those of geographical range and the ability to carry the relevant information. The geographical range of a radio wave is affected by many parameters, but in the context of this book, range may basically be related to the choice of frequency band, which in turn determines the method of radio wave propagation.

## 1.3 Radio wave radiation

The propagation of radio waves is a highly complex natural phenomenon. It is simplified in the following pages to provide an understanding of the subject with a level of knowledge necessary to comprehend modern navigation systems.

Energy is contained in a transmitted radio wave in two forms, electrostatic energy and electromagnetic energy. The radiation of energy from a simple antenna may be described by considering a centre-fed dipole antenna, which is shown electrically in Figure 1.1.

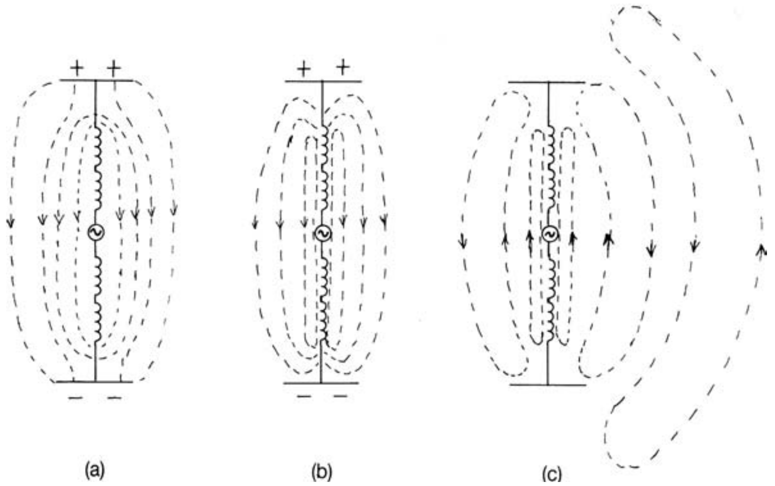


**Figure 1.1** Radio wave radiation from a centre-fed dipole antenna.

The antenna shown is formed of two coils, each end of which is at the opposite potential to the other with reference to the centre point. As a complete unit, the antenna forms a tuned circuit that is critically resonant at the carrier frequency to be radiated. The two plates, one at each end of the coil assembly, form a capacitor. Radio frequency current, from the output stage of a suitable transmitter, shown here as a generator, is applied at the centre of the two coils. One of the basic electrical laws of physics states that whenever an electron has its velocity altered by an accelerating force there will be a detachment of energy. In the case of an antenna system this detachment is the energy that is lost from the transmitter and radiated as electrical energy into the atmosphere.

The diagrams clearly show the distribution of the electric field produced around an antenna when an oscillatory radio frequency is applied to it. In Figure 1.1(a) the top plate of the antenna is instantaneously driven positive with respect to the base plate and the current flow in the wire is zero. At this instant the field produced is entirely electric and the electrostatic lines of force are as shown in the diagram.

After the peak of the signal has passed, electrons will begin to flow upwards to produce a current flow in the wire. The electric field will now start to collapse (Figure 1.1(b)) and the ends of the lines of force come together to form loops of electrostatic energy. After the potential difference (positive top plate to negative base plate) across the two plates of the effective capacitor has fallen to zero, current continues to flow and, in so doing, starts to charge the effective capacitor plates in the opposite direction. This charge forms new lines of force in the reverse direction to the previous field, negative to the top plate and positive at its base. The collapse of the initial electrostatic field lags the change in potential that caused it to occur and, consequently, the new electric field starts to expand before the old field has completely disappeared. The electric fields thus created (Figure 1.1(c)) will be caused to form loops of energy, with each new loop forcing the previous loop outwards, away from the antenna. Thus, radio frequency energy is radiated as closed loops of electrostatic energy.

Because a minute current is flowing around each complete loop of energy, a magnetic field will be created around the loop at 90° to it. Thus, the magnetic lines of force produced around the vertical electric field created by a vertical antenna, will be horizontal. Two fields of energy, in space quadrature, have thus been created and will continue in their relative planes as the radio wave moves away from the transmitting antenna.

The electric and magnetic inductive fields are in both time and space quadrature and are 90°out of phase with each other in time, and at right angles to each other in space. The electric field is of greatest importance to the understanding of radio wave propagation, the magnetic field only being present when current flows around the loop as the electric field changes.

Figure 1.2 shows the relative directions of the electric field (*E*), the magnetic field (*H*) and the direction of propagation. The oscillating electric field is represented by the vertical vector OE, the magnetic field by OH, and the direction of propagation by OD. Another electrical law of physics, Fleming's right-hand rule, normally applied to the theory of electrical machines, applies equally to the direction of propagation of the radio wave.

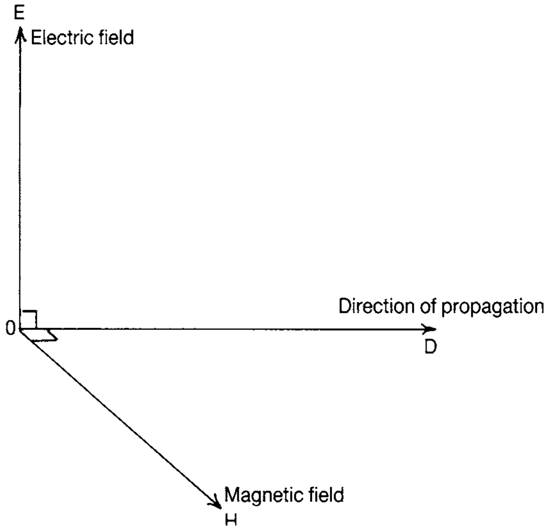


**Figure 1.2** The angular relationship of the E and H fields.

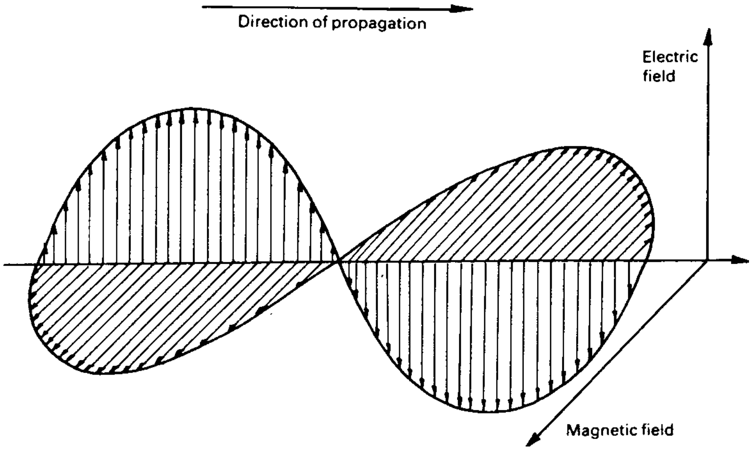


**Figure 1.3** Amplitude variations of the E and H fields.

At any instantaneous point along the sinusoidal wave of the electric field it is possible to measure a minute current flow in the loop of energy. The current will be increasing and decreasing as it follows the rate of change of amplitude of the sinusoidal frequency (carrier wave) of the radio wave (see Figure 1.3). It is this instantaneous change of current which, when in contact with a receiving antenna, causes a current to flow at the receiver input and a minute signal voltage, called an electromotive force (e.m.f.), to appear across the antenna input.

The transmitted signal may now be considered to be a succession of concentric loops of ever-increasing radius, each one a wavelength ahead of the next. Radio waves thus produced will be similar in appearance to the waves caused on the surface of a pond when a rock is tossed into it. Similarly, the radio waves radiate outwards from the source and diminish in amplitude with distance travelled from the transmitter. Each loop moves away from the transmitting antenna at the speed of light in free space, usually approximated to be $300 \times 10^6 \, \text{ms}^{-1}$, and it is common practice to call the leading edge of each loop a wavefront. The distance between each wavefront depends upon the frequency being radiated and is called the wavelength, λ (lambda).

## 1.4 Frequency, wavelength and velocity

Although a variable, the velocity of electromagnetic radio waves propagated in the troposphere, close to the earth's surface, is accepted to be $300 \times 10^6 \, \text{ms}^{-1}$. This figure is important because it enables the wavelength of a transmitted frequency to be calculated and from that a number of other essential parameters can be determined.

$$\text{Wavelength } \lambda = \frac{300 \times 10^6}{\text{Frequency}} \text{ (in metres)}$$

The actual length of one radio wave during one alternating cycle is a measure of the distance travelled, and the number of alternating cycles per second is a measure of the frequency.

## 1.5 Radio frequency spectrum

Table 1.1 indicates how the available frequency spectrum has been divided into usable bands. By referring to this table it is possible to gain some initial idea of the approximate range over which radio waves may be received. For instance, if all other parameters remain constant, the anticipated radio range of signals propagated on the VHF band, or those higher, is effectively that of 'line-of-sight'. Consequently, ship-to-ship communications between a life-raft and a surface vessel could expect to have a range of 2–7 nautical miles depending upon the system installation and the relative heights of the antennae. Because of its line-of-sight nature, VHF radio ranges beyond the horizon can only be achieved by using repeater stations or satellites. Maritime mobile satellite systems use much higher frequencies in what is termed the L band and the C band, each providing a line-of-sight link.

**Table 1.1** The frequency spectrum

| *Abbreviation* | *Band* | *Frequency range* | *Wavelength* |
|---|---|---|---|
| AF | Audio | 0 Hz–20 kHz | ∞ to 15 km |
| RF | Radio | 10 kHz–300 GHz | 30 km to 0.1 cm |
| VLF | Very low | 10–30 kHz | 30 km to 10 km |
| LF | Low | 30–300 kHz | 10 km to 1 km |
| MF | Medium | 300–3000 kHz | 1 km to 100 m |
| HF | High | 3–30 MHz | 100 m to 10 m |
| VHF | Very high | 30–300 MHz | 10 m to 1 m |
| UHF | Ultra high | 300–3000 MHz | 1 m to 10 cm |
| SHF | Super high | 3–30 GHz | 10 cm to 1 cm |
| EHF | Extreme high | 30–300 GHz | 1 cm to 0.1 cm |

### 1.5.1 Spectrum management

Radio waves do not respect international boundaries and an international framework has been established in order to control the use of frequencies, the standards of manufacture and the operation of radio equipment in order to limit the likelihood of interference. The forum for reaching international agreements on the use of the radio frequency spectrum is the International Telecommunications Union (ITU). Membership of the ITU is dependent upon acceptance of the strict convention which exists to uphold the regulations laid down by the various conferences and meetings of the ITU.

The radio spectrum management policies agreed among the signatories of the convention are published by the ITU as international radio regulations. One of these is the international Table of Frequency Allocations, which provides the framework for, and the constraints on, national frequency use and planning. The Table of Frequency Allocations and the radio regulations documents are revised at the World Administrative Radio Conferences (WARC) held at periods of 5–10 years.

The administrative structure established by the ITU convention comprises a Secretariat headed by the Secretary General, an Administrative Council, a registration board for radio frequencies, and the consultative committees for radio and telecommunications.

The International Radio Consultative Committee (CCIR) forms study groups to consider and report on the operational and technical issues relating to the use of radio communications. The International Telecommunications Consultative Committee (CCIT) offers the same service for telecommunications. The study groups produce recommendations on all aspects of radio communications. These recommendations are considered by the Plenary Assembly of the CCIR and, if accepted, are incorporated into the radio regulations. Another subgroup of the ITU, the International Frequency Registration Board (IFRB) considers operating frequencies, transmitter sites, and the location of satellites in orbit. Within Europe, a further body, the Conference of European Telecommunications Administrations (CEPT) assists with the implementation of the ITU radio regulations on a national level. Every country appoints an agency to enact the radio regulations thus laid down. In the United Kingdom for instance it is the Radiocommunications Agency and in the USA, civil use of the radio frequency spectrum is controlled by the Federal Communications Commission.

## 1.6 Radio frequency bands

Radio wave propagation characteristics (see Table 1.2) are dependent upon the frequency used.

**Table 1.2** Radio frequency band characteristics

| *Designation & Frequency* | *Propagation Mode* | *Characteristics* |
|---|---|---|
| Very low frequency 3–30 kHz | Large surface wave | Very high power transmitters and large antennae needed |
| Low frequency 30–300 kHz | Surface wave and some sky wave returns | High power transmitters; limited number of channels; subject to fading |
| Medium frequency 0.3–3 MHz | Surface wave during day. Some sky wave returns at night | Long range at night; subject to fading |
| High frequency 3–30 MHz | Sky waves returned over long distances | Global ranges using ionospheric returns |
| Very high frequency 30–300 MHz | Mainly space wave. Line of site | Range depends upon antenna height |
| Ultra high frequency 0.3–3 GHz | Space wave only | Line of sight; satellite and fixed link |
| Super high frequency 3–30 GHz | Space wave only | Line of sight; radar and satellite |
| Extreme high frequency 30–300 GHz | Space wave only | Not used for mobile communications |

### 1.6.1 VLF (very low frequency) band

VLF radio signals propagate using a combination of both ground and space waves. They require vast amounts of power at the transmitter to overcome earth surface attenuation and can be guided over great distances between the lower edge of the ionosphere and the ground. Because VLF possesses a very long wavelength, huge antenna systems are required. As an example, at 10 kHz the wavelength is 30 km. An efficient antenna, often quoted as ‘a half-wavelength antenna’, needs to be 15 km long and it is only possible to construct one on land, usually slung between mountain peaks.

### 1.6.2 LF (low frequency) band

Communication is mainly by a ground wave, which suffers increasing attenuation as the frequency increases. Range therefore depends upon the amplitude of the transmitted power and the efficiency of the antenna system. Expected range for a given low frequency and transmitter power is between 1500 and 2000 km. At LF the wavelength is reduced to a point where small-size antennae are practicable. Although the sky wave component of LF propagation is small it can be troublesome at night when it is returned from the ionosphere.

### 1.6.3 MF (medium frequency) band

Ground wave attenuation rapidly increases with frequency to the point where, at the higher end of the band, its effect becomes insignificant. For a given transmitter power, therefore, ground wave range is inversely proportional to frequency. Range is typically 1500 km to under 50 km for a transmitted signal, with a peak output power of 1 kW correctly matched to an efficient antenna.

In the band below 1500 kHz, sky waves are returned from the ionosphere both during the day and night, although communication using these waves can be unreliable. Above 1500 kHz the returned sky wave has greater reliability but is affected by changes in the ionosphere due to diurnal changes, seasonal changes, and the sun-spot cycle. From experience and by using published propagation figures it is possible for reliable communications to be achieved up to a range of 2000 km.

### 1.6.4 HF (high frequency) band

This frequency band is widely used for terrestrial global communications. Ground waves continue to be further attenuated as the frequency is increased. At the low end of the band, ground wave ranges of a few hundred kilometres are possible but the predominant mode of propagation is the sky wave.

Because ionization of the upper atmosphere is dependent upon the sun's radiation, the return of sky waves from the ionosphere will be sporadic, although predictable. At the lower end of the band, during the hours of daylight, sky waves are absorbed and do not return to earth. Communication is primarily by ground wave. At night, however, lower frequency band sky waves are returned and communication can be established but generally with some fading. Higher frequency band sky waves pass through the ionized layers and are lost. During the day the opposite occurs. Low frequency band skywaves are absorbed and those at the higher end are returned to earth. For reliable communications to be established using the ionized layers, the choice of frequency is usually a compromise. Many operators ignore the higher and lower band frequencies and use the mid-range for communications.

### 1.6.5 VHF (very high frequency) band

Both ground waves and sky waves are virtually non-existent and can be ignored. Communication is via the space wave which may be ground reflected. Space waves effectively provide line-of-site communications and consequently the height of both transmitting and receiving antennas becomes important. A VHF antenna may also be directional. Large objects in the path of a space wave create blind spots in which reception is extremely difficult or impossible.

### 1.6.6 UHF (ultra high frequency) band

Space waves and ground reflected waves are used with highly directional efficient antenna systems. Signal fading is minimal, although wave polarization may be affected when the wave is ground reflected resulting in a loss of signal strength. Blind spots are a major problem.

### 1.6.7 SHF (super high frequency) band

Frequencies in this band possess very short wavelengths and are known as microwaves. Communication is by space wave only. Because of the minute wavelength, compact and highly directional antennas can be designed. This band is used for maritime radar and satellite communications.

### 1.6.8 EHF (extreme high frequency) band

Communications is by space wave only. Highly directional antennas are used. Scattering and signal loss is a major problem. The band is not currently used for maritime communications.

## 1.7 Radio wave propagation

Whilst all transmitting antenna systems produce one or more of the three main modes of propagation (see Figure 1.4), one of the modes will predominate. If all other parameters remain constant, the predominant mode of propagation may be equated to the frequency used. For the purpose of this explanation it is assumed that the mode of propagation is dependent upon frequency because that is the only parameter that may be changed by an operator. The three modes of propagation are:

- surface wave propagation
- space wave propagation
- sky wave propagation.

### 1.7.1 Surface wave propagation

The surface wave is a radio wave that is modified by the nature of the terrain over which it travels. This can occasionally lead to difficulty in maritime navigation systems where the wave travels from

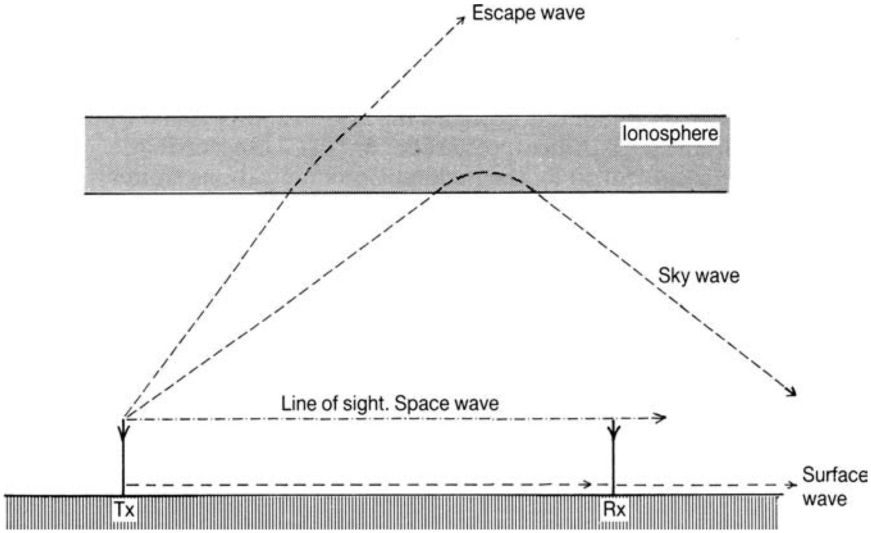


**Figure 1.4** Radio wave modes of propagation.

one medium to another, over a coastline for instance. The refraction caused in such cases is likely to induce errors into navigation systems.

A surface wave will predominate at all radio frequencies up to approximately 3 MHz. There is no clear cut-off point and hence there will be a large transition region between approximately 2 and 3 MHz, where the sky wave slowly begins to have influence.

The surface wave is therefore the predominant propagation mode in the frequency bands VLF, LF and MF. As the term suggests, surface waves travel along the surface of the earth and, as such, propagate within the earth's troposphere, the band of atmosphere which extends upwards from the surface of the earth to approximately 10 km.

### *Diffraction and the surface wave*

An important phenomenon affecting the surface wave is known as diffraction. This term is used to describe a change of direction of the surface wave, due to its velocity, when meeting an obstacle. In fact, the earth's sphere is considered to be a large obstacle to surface waves, and consequently the wave follows the curvature of the earth (Figure 1.5).

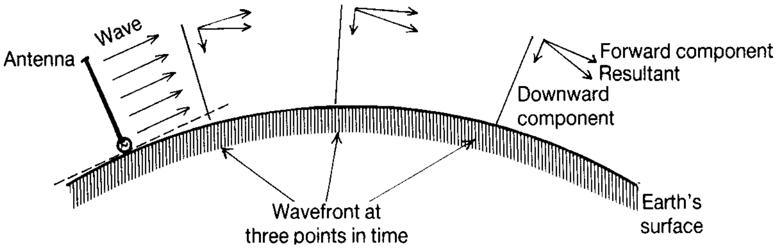


**Figure 1.5** Tilting of the surface wavefront caused by diffraction.

The propagated wavefront effectively sits on the earth's surface or partly underground and, as a result, energy is induced into the ground. This has two primary effects on the wave. First, a tilting of the wavefront occurs, and second, energy is lost from the wave. The extent of the diffraction is dependent upon the ratio of the wavelength to the radius of the earth. Diffraction is greatest when the wavelength is long (the lower frequency bands) and signal attenuation increases with frequency. This means that surface waves predominate at the lower end of the frequency spectrum and, for a given transmitter power, decrease in range as frequency increases.

The amount of diffraction and attenuation also depends upon the electrical characteristics of the surface over which the wave travels. A major factor that affects the electrical characteristics of the earth's surface is the amount of water that it holds, which in turn affects the conductivity of the ground. In practice, seawater provides the greatest attenuation of energy and desert conditions the least attenuation.

The propagation range of a surface wave for a given frequency may be increased if the power at the transmitter is increased and all other natural phenomena remain constant. In practice, however, transmitter power is strictly controlled and figures quoting the radio range are often wild approximations. For instance, NAVTEX data is transmitted on 518 kHz from a transmitter designed to produce an effective power output of 1 kW. This gives a usable surface wave range of 400 miles. But, under certain conditions, NAVTEX signals may be received over distances approaching 1000 miles.

Another phenomenon caused by radio-wave diffraction is the ability of a ground-propagated wave to bend around large objects in its path. This effect enables communications to be established when a receiving station is situated on the effective blind side of an island or large building. The effect is greatest at long wavelengths. In practice, the longer the wavelength of the signal in relation to the physical size of the obstruction, the greater will be the diffraction.

### 1.7.2 Sky wave propagation

Sky waves are severely influenced by the action of free electrons, called ions, in the upper atmosphere and are caused to be attenuated and refracted, possibly being returned to earth.

The prime method of radio wave propagation in the HF band between 3 and 30 MHz is by sky wave. Because under certain conditions, sky waves are refracted from the ionosphere, this band is used extensively for terrestrially-based global communications. Once again, however, there is no clear dividing line between surface and sky waves. In the frequency range between 2 and 3 MHz, surface waves diminish and sky waves begin to predominate.

Sky waves are propagated upwards into the air where they meet ionized bands of atmosphere ranging from approximately 70 to 700 km above the earth's surface. These ionized bands, or layers, have a profound influence on a sky wave and may cause it to return to earth, often over a great distance.

#### *The ionosphere*

A number of layers of ionized energy exist above the earth's surface. For the purpose of explaining the effects that the layers have on electromagnetic radiation it is only necessary to consider four of the layers. These are designated, with respect to the earth's surface, by letters of the alphabet; D, E, $F_1$ and $F_2$, respectively (Figure 1.6). They exist in the ionosphere, that part of the atmosphere extending from approximately 60 km above the earth's surface to 800 km.

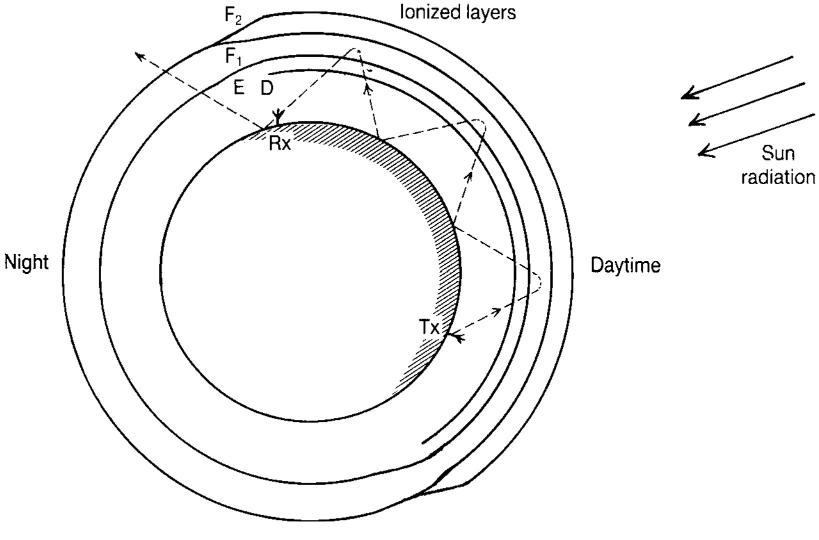


**Figure 1.6** Ionized layers and their effect on long-range communications.

Natural ultraviolet radiation from the sun striking the outer edge of the earth's atmosphere produces an endothermic reaction, which in turn, causes an ionization of atmospheric molecules. A physical change occurs producing positive ions and a large number of free electrons. The layers closer to the earth will be less affected than those at the outer edges of the atmosphere and, consequently, the D layer is less ionized than the $F_2$ layer. Also, the amount of ultraviolet radiation will never be constant. It will vary drastically between night and day, when the layers are in the earth's shadow or in full sunlight. In addition, ultraviolet radiation from the sun is notoriously variable, particularly during solar events and the 11-year sun-spot cycle. During these events, the ionized layers will be turbulent and sky waves are seriously affected.

Whilst it may appear that radio communication via these layers is unreliable it should be remembered that most of the environmental parameters affecting the intensity of an individual layer are predictable. The external natural parameters that affect a layer, and thus the communication range, are:

- the global diurnal cycle
- the seasonal cycle
- the 11-year sun-spot cycle.

## *Radio wave ionospheric refraction*

An electromagnetic radio wave possesses a wavelength, the velocity of which is affected when it passes from one medium to another of a different refractive index, causing a change of direction to occur. This change of direction is called refraction.

As previously stated, the atmosphere is ionized by the sun's radiation. It is convenient to view the ionized region produced by this action as ionized layers. The outermost layer, closest to the sun's radiation, will be intensely ionized, whereas the layer closest to the earth's surface is less ionized (Figure 1.7). Due to the collision of free electrons, an electromagnetic radio wave entering a layer will have its velocity changed causing the upper end of the wavefront to speed up. If, before the wave reaches the outer edge of an ionized layer, the angle of incidence has reached the point where the wavefront is at right angles to the earth's surface, the radio wave will be returned to earth where it will strike the ground and be reflected back into the ionosphere.

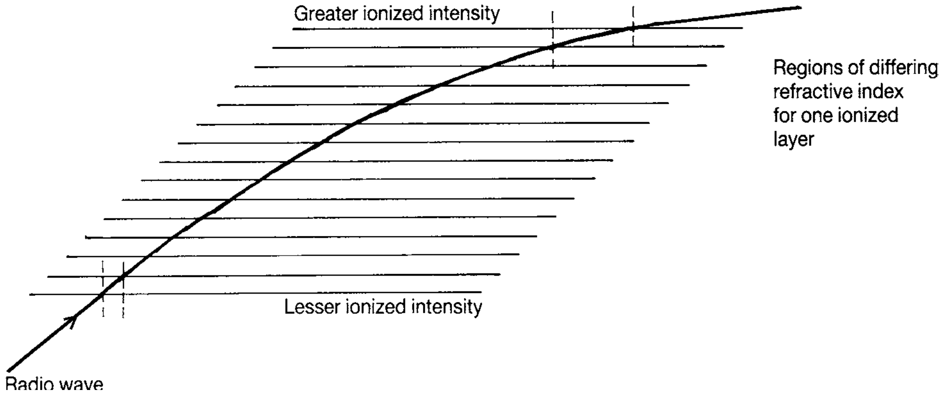


**Figure 1.7** Radio wave refraction due to progressively higher ionization intensity.

The extent of refraction, and thus whether a radio wave is returned to earth, can be controlled and is dependent upon three main parameters:

- the density of the ionosphere
- the frequency of propagation
- the angle of incidence of the radio wave with a layer.

Obviously it is not possible to control the density of the ionosphere, but other parameters may be changed by a shore-based radio station which has control over antenna systems. For a maritime mobile system, however, it is only the frequency that can be changed.

Despite its complexity, it is the phenomenon of refraction that enables terrestrial global communications to be achieved. Radio waves make several excursions between being refracted by the ionosphere and reflected from the earth's surface, with each journey being known as one hop.

### 1.7.3 Space wave propagation

The space wave, when propagated into the troposphere by an earth surface station, is subject to deflection by variations in the refractive index structure of the air through which it passes. This causes the radio wave to follow the earth's curvature for a short distance beyond the horizon making the radio horizon somewhat longer than the visible horizon. Ship's navigators will know the effect whereby the surface radar range extends slightly beyond the horizon. Space waves propagated upwards away from the troposphere may be termed free space waves and are primarily used for satellite communications.

Space waves are rarely returned from the ionosphere because the wavelength of the carrier frequency is reduced to the point where refraction becomes insignificant. Such a wave, when propagated upwards, passes through the ionized layers and is lost unless it is returned by an artificial or natural earth satellite.

If a space wave is propagated along the surface of the earth or at a short height above it, the wave will move in a straight line from transmitting antenna to receiving antenna and is often called a line-of-sight wave. In practice, however, a slight bending does occur making the radio horizon somewhat longer than the visual horizon.

The troposphere extends upwards from the earth's surface to a height of about 10 km where it meets the stratosphere. At the boundary between the two there is a region called the tropopause which possesses a different refractive index to each neighbouring layer. The effect exhibited by the tropopause on a radio space wave is to produce a downward bending action, causing it to follow the earth's curvature. The bending radius of the radio wave is not as severe as the curvature of the earth, but nevertheless the space wave will propagate beyond the visual horizon. In practice, the radio horizon exceeds the visual horizon by approximately 15%.

The actual range for communications in the VHF band and above is dependent upon the height of both the transmitting and receiving antennae. The formula below gives the radio range for VHF communications in nautical miles:

$$R = 2.5\sqrt{h_T + h_R}$$

where $h_T$ and $\mathrm{h_R}$ are in metres.

Given a ship's antenna height of 4 m and a coastal radio station antenna height of 50 m the expected radio range is approximately 23 nmiles. This rises to 100 nmiles for antenna heights of 4 m and 100 m, respectively. Ship-to-ship communications with each ship having a 4-m high antenna gives a range of

10 nmiles. Search and rescue (SAR) communications between a life-raft and another surface vessel may have a range of only 4 nmiles.

It should be noted that VHF space waves cannot pass through, or be diffracted around, large objects, such as buildings or islands, in their path. This gives rise to extensive radio shadow areas behind large structures.

## 1.8 Signal fading

One of the major difficulties encountered when radio waves are propagated via the earth's atmosphere is that of signal fading. Fading is a continual variation of signal amplitude experienced at the antenna input to a receiving system. In practice, fading may be random or periodic but in each case the result will be the same. If the signal input to a receiver falls below the quoted sensitivity figure there may be no output from the demodulator and hence the communications link is broken. If the signal amplitude at the antenna doubles, a large increase in audible output will be produced either causing possible overloading of an automatic system or discomfort for an operator. Steps are taken at the receiver to overcome the problem of signal fading, which may be classified as one of three main types:

- general signal fading
- selective fading
- frequency selective fading.

### 1.8.1 General signal fading

In a global system, fading may occur because of the continually changing attenuation factor of an ionospheric layer. Ultraviolet radiation from the sun is never constant, and consequently, the intensity of the ionization of a layer will continually change. The signal attenuation of a specific layer may cause complete signal fade-out as the intensity of the sun's radiation changes. With the exception of this extreme case, the use of automatic gain control (AGC) circuits in a receiver effectively combats this phenomenon.

### 1.8.2 Selective fading

Selective fading occurs for a number of reasons. Radio waves arriving at an antenna may have travelled over two or more different paths between transmitter and receiver. Each path-length is different and the signals arriving at the receiving antenna produce a combined signal amplitude, which is the phasor sum of the two. The two signals, of the same frequency and the same origin, will be out of time-phase with each other and will therefore produce a resultant signal that is either larger or smaller in amplitude than the original. In most cases the signal path-lengths are unpredictable and often variable, leading again to the need for a good quality AGC circuit in the receiver. This effect can occur, as shown in Figure 1.8, when two sky waves are refracted from the ionosphere over different path-lengths, when a sky wave and a ground wave are received together, or when two ground waves are received over different paths.

### 1.8.3 Frequency selective fading

This occurs where one component of a transmitted radio wave is attenuated to a greater extent than other components. In any wideband communications link a large number of frequencies are contained

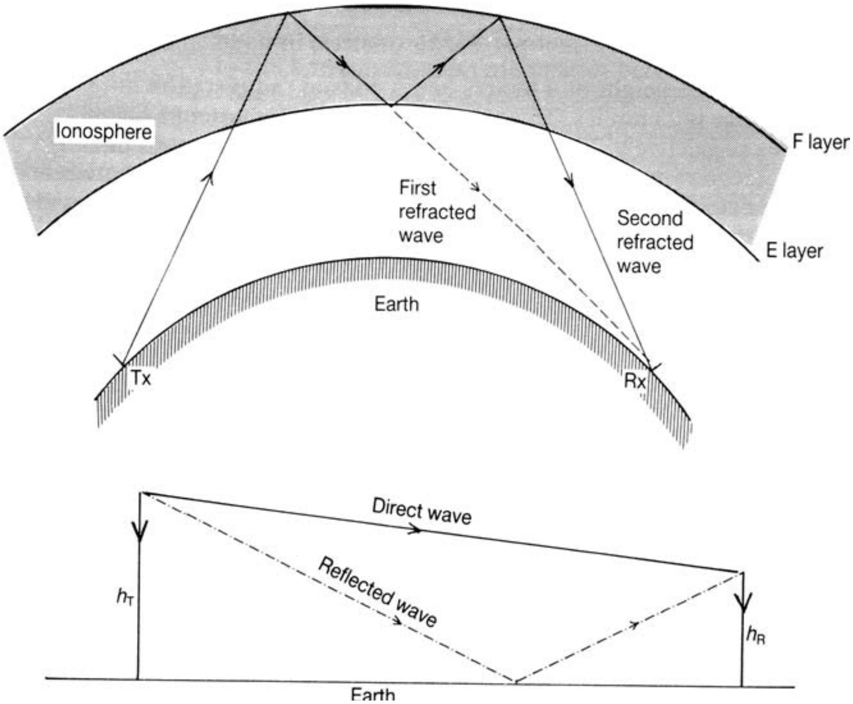


**Figure 1.8** Signal fading caused by multipath propagation.

within the bandwidth of the transmitted signal. The individual frequencies contained in the transmission are those of the fundamental carrier frequency plus the RF frequencies generated by the method of modulation employed. To produce an error-free or distortion-free communications link, all modulation baseband frequencies at the transmitter must be faithfully reproduced at the receiver output. If any of the modulated frequencies are lost in the transmission medium, which may happen when frequency selective fading is present, they cannot be reproduced by the receiver.

More importantly, however, if the carrier frequency is lost in the transmission medium it will be impossible to demodulate the audio intelligence at the receiver, unless specific circuitry is available and the carrier loss is predictable.

Frequency selective fading cannot be cured by the use of AGC circuits in a receiver. Its effects can, however, be limited by using:

- a transmission which radiates one mode only – a carrier frequency or narrow band signal
- single sideband (SSB J3E) fully suppressed carrier transmission telephony
- frequency modulation.

## 1.9 Basic antenna theory

An antenna is arguably the single most critical part of any radio communications system and those used by radio navigation systems are no exception. Unfortunately, however, it is often the part of a radio installation that is less than efficient, not because of deficiencies in antenna design but because of the major problems of antenna siting and installation. As ships become more streamlined, the available antenna space reduces, often to the point where multiple antenna systems simply cannot be fitted.

Radio navigation systems use a variety of antennae, each one designed with individual characteristics to suit operational needs, but whatever the construction, they all operate on similar principles.

Antenna design and construction is a complex area of radio communications theory and the following description is limited to that needed to understand radio navigation systems. Whilst some basic antenna theory is considered, it should be noted that it is only necessary for the reader to understand antennae from an operational and maintenance viewpoint.

An antenna is essentially a piece of wire that may or may not be open at one end. The shortest length of wire that will resonate at a single frequency is one that is critically long enough to permit an electric charge to travel along its length and return in the period of one cycle of the applied radio frequency. This period of one cycle is called the wavelength. The velocity of a propagated RF is that of light waves, i.e. $299\,793\,077\,\mathrm{ms}^{-1}$, which is usually approximated to $300 \times 10^6\,\mathrm{ms}^{-1}$ for convenience. The wavelength in metres of any RF wave is therefore:

$$\lambda = \frac{300 \times 10^6}{f}$$

Because the RF charge will travel the length of the wire and return, it follows that the shortest resonant wire is one half of a wavelength long. In fact many antenna systems are called half-wave or $\lambda/2$. If, as an analogy, the resonant length is assumed to be a trough with obstructions at each end and a ball is pushed from one end, it will strike the far end and return, having lost energy. If, at the instant the ball hits the near end obstruction, more energy is given to the ball it will continue on its way indefinitely. However, it is critically important that the new energy is applied to the ball at just the right time in order to maintain the action. In practice, if the timing is in error the length of the resonant trough may be changed to produce the optimum transfer of energy along the wire. Antennae, therefore, must be constructed to be a critical length to satisfy the frequency of the applied RF energy.

Antennae, exhibit the ‘reciprocity principle’, which means that they are equally as efficient when working as a transmitting antenna or as a receiving antenna. The main difference is that a transmitting antenna needs to handle high power and is usually more substantially built and better insulated than a corresponding receiving antenna. For efficient radio communications, both the transmitting and receiving antennae should possess the same angle of polarization with respect to the earth. Polarization refers to the angle of the transmitted electric field ($E$) and, consequently, if the $E$-field is vertical, both transmitting and receiving antennae must be vertical. The efficiency of the system will reduce progressively as the error angle between transmitting and receiving antennae increases up to a maximum error of 90°.

### 1.9.1 Half-wavelength antenna

An antenna operating at precisely half a wavelength is traditionally called a Hertz antenna. Many antennae do not operate at $\lambda/2$ because they would be excessively long. A $\lambda/2$ antenna is effectively a $\lambda/4$ transmission line with a signal generator, the transmitter, at one end and an open circuit at the other, as shown in Figure 1.9.

Ohm’s Law states that when an open circuit exists the current will be zero and the potential difference (p.d.) across the open circuit will be maximum. Figure 1.10 shows voltage ($E$) and current ($I$) standing waves which indicate this fact.

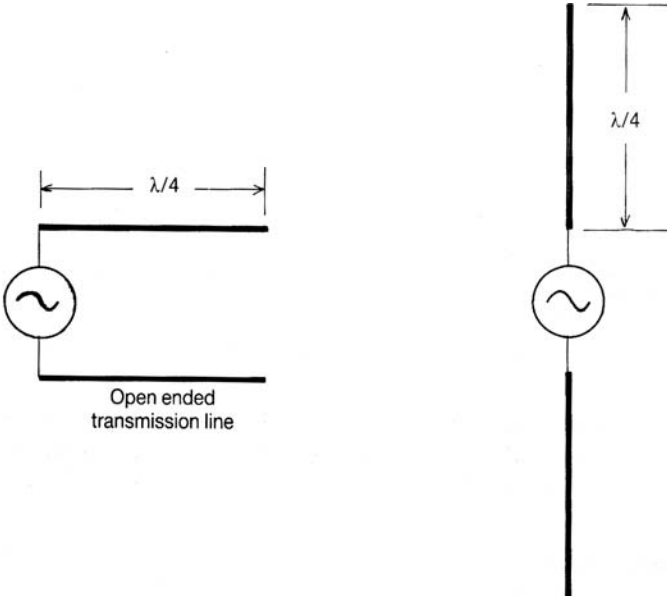


**Figure 1.9** Half-wavelength antenna derived from a quarter-wavelength transmission line.

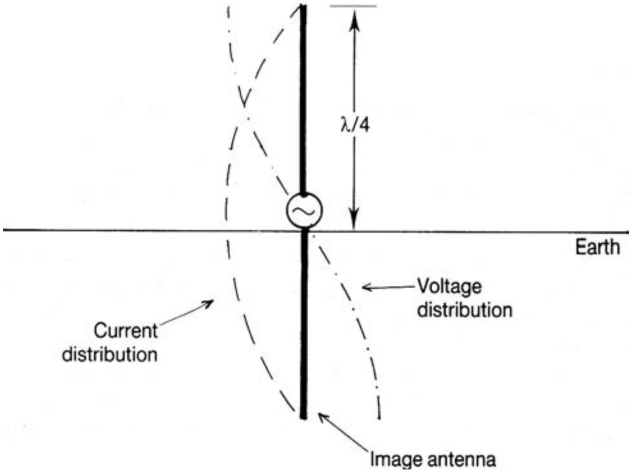


**Figure 1.10** A grounded quarter-wavelength antenna showing the voltage and current distribution curves.

*E* and *I* distribution curves are standard features of antenna diagrams. If the generator (signal source) is λ/4 back from the open circuit, the *E* and *I* curves show minimum voltage and maximum current at the antenna feed point. In most cases this is the desirable *E* and *I* condition for feeding an antenna. If the two arms of the transmission line are now bent through 90°, a λ/2 efficient antenna has been produced.

Ohm's Law also states that the resistance of a circuit is related to the voltage and the current. In this case the impedance of the antenna will be maximum at the ends and minimum at the centre feed point.

Again this is desirable because the centre impedance is approximately 73 Ω, which ideally matches the 75 Ω (or in some cases 50 Ω) impedance coaxial cable used to carry the output of the transmitter or the input to a receiver.

### 1.9.2 Physical and electrical antenna lengths

Ideally, an antenna isolated in free space would follow the rules previously quoted, whereby the actual and electrical lengths were the same. Both are calculated to be λ/2 of the transmission frequency. However, because the velocity of the radio wave along the wire antenna is affected by the antenna supporting system and is slightly less than that in free space, it is normal to reduce the physical length of the antenna by approximately 5%. In practice, the corrected physical length of an antenna is therefore 95% of the electrical length.

Antennae and feeders are effectively 'matched transmission lines', which, when a radio frequency is applied, exhibit standing waves, the length of which are determined by a number of factors outside the scope of this book. However, the waves are basically produced by a combination of forward and reflected power in the system. A measurement of the ratio between forward and reflected power, called the standing wave ratio (SWR), provides a good indication of the quality of the feeder and the antenna. Measurement of the SWR is made using voltage and becomes voltage standing wave ratio (VSWR).

### 1.9.3 Antenna radiation patterns

A graph showing the actual intensity of a propagated radio wave at a fixed distance, as a function of the transmitting antenna system, is called a radiation pattern or 'polar diagram'. Most antenna radiation patterns are compared with that of a theoretical reference antenna called an isotropic radiator. Radiation patterns may be shown as the *H*-plane or the *E*-plane of transmission or reception. Figure 1.11 shows the *E*-plane radiation patterns of an isotropic radiator and a λ/2 dipole antenna.

It should be noted that this is a two-dimensional diagram whereas the actual radiation pattern is three-dimensional. The maximum field strength for the λ/2 dipole occurs at right angles to the antenna and there is very little radiation at its ends. In the horizontal plane, therefore, this type of antenna is directional, whereas an isotropic radiator is omnidirectional. However, a λ/2 antenna can be made omnidirectional when it is vertically polarized.

A second important principle of an antenna is its beamwidth. The radiation pattern is able to illustrate the antenna beamwidth. It is calculated at the 'half-power points' or –3 dB down from the

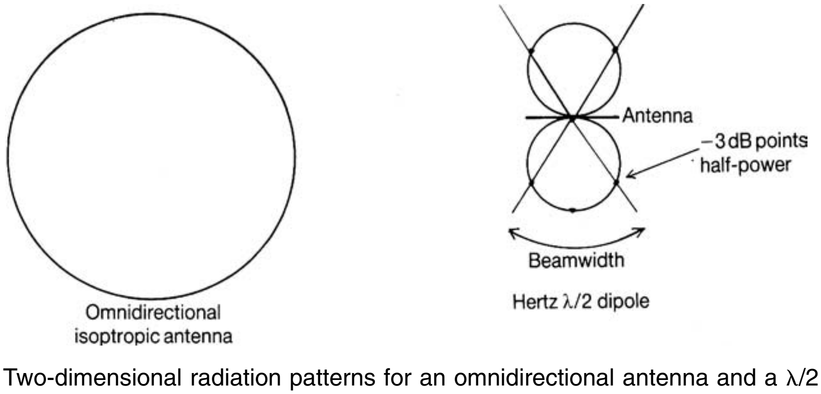


**Figure 1.11** Two-dimensional radiation patterns for an omnidirectional antenna and a λ/2 antenna.

peak point. If the receiving antenna is located within the beamwidth of the transmitting antenna good communications will be made.

Antenna gain patterns for receiving antennas are again called polar diagrams or azimuth gain plots (AGP).

### 1.9.4 Antenna gain and directivity

Antenna gain and directivity are very closely linked. The greater the directivity an antenna exhibits, the greater it will appear to increase the transmitted signal in a specific direction. The λ/2 dipole, for instance, possesses a gain of typically 2.2 dB, on those planes at right angles to the antenna, when compared with an isotropic radiator. As a consequence, zero signals will be propagated along the other two planes in line with the dipole.

Both properties of gain and directivity are reciprocal and apply equally to both transmitting and receiving antennae. In practice it is important to consider the effect of both the transmitter and receiver antenna gains in a complete radio communications system. The formula below provides a simple method of calculating the signal strength at a receiver input.

$$P_r = \frac{P_t G_t G_r \lambda/2}{16\pi^2 d^2}$$

where $P_r$ = power received in watts, $P_t$ = power output of transmitter in watts, $G_t$ = the ratio gain of the transmitting antenna, $G_r$ = the ratio gain of the receiving antenna, $\lambda$ = wavelength of the signal in metres, and $d$ = the distance between antennae in metres.

### 1.9.5 Ground effects

The overall performance of an antenna system is extensively changed by the presence of the earth beneath it. The earth acts as a reflector and, as with light waves, the reflected radio wave leaves the earth at the same angle with which it struck the surface. Figure 1.12 shows the direct and reflected radio waves at a receiving antenna.

Because the surface of the earth is rarely flat and featureless, there will be some directions in which the two waves are in phase, and thus are additive, and some where the two are out of phase, and thus subtractive.

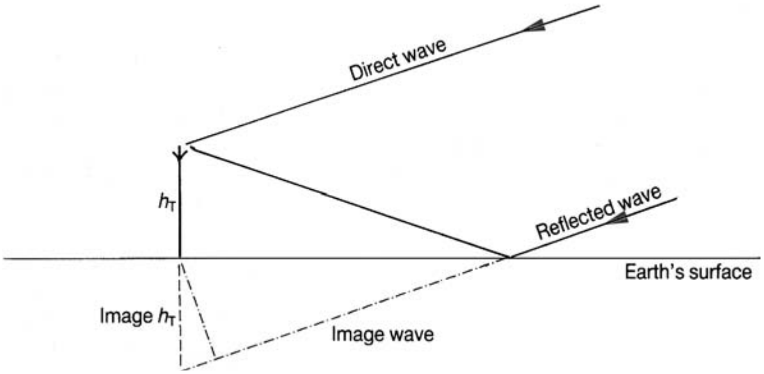


**Figure 1.12** Direct and earth reflected radio waves received by an antenna.

Because the effects of ground wave reflected waves are unpredictable, some antenna arrays are constructed with a ground plane. Reflections from the ground plane are, to some extent, predictable and may be compensated for in the receiving system. Satellite navigation antennas and VHF RDF fixed antennae often use a ground plane to improve sensitivity and limit signal reflections.

### 1.9.6 Antenna efficiency

Antenna efficiency is of particular importance in all communications systems. If the efficiency of an antenna drops to 50%, the maximum radiated signal also drops resulting in a consequent loss of range. It would be rare indeed to find any system that is 100% efficient and antennae are no exception. However, antenna losses are well documented and, consequently, the effective isotropic radiated power (EIRP) figure for a system is usually calculated with reference to known efficiency figures.

The losses leading to inefficiency in an antenna system may generally be classed as dielectric losses affecting the transmission properties of the antenna. Such losses in a transmitting antenna may be produced by arcing effects and corona discharge, and in a receiving antenna they may be produced by bad connections or damaged wiring. Most of these losses can be controlled by careful installation, good positioning of the antenna, and diligent maintenance.

### 1.9.7. Antenna feed lines

Whilst the connection between the transmitter output and the antenna input appears to be made by a simple wire it is, in fact, made by a balanced transmission line that possesses impedance. Usually, the feed line is a correctly terminated coaxial cable specifically designed for the purpose. For most transmitting and receiving antenna systems the feed line possesses an impedance of 50 or 75 Ω. Because of its need to handle more power, a transmitter coaxial cable will be physically larger than a corresponding receiver coaxial line, unless of course both use the same line. The inner copper conductor forms the live feed wire with the screen sheath providing the ground line. The outer sheath should be bonded to ground to prevent inductive pick-up in the centre conductor wire, which would generate interference in the communications link.

Coaxial cables used in a marine environment are double sheathed and occasionally armour plated. They are fully waterproofed and should remain so throughout their life. Moisture ingress into the cable insulation material will cause considerable losses as energy is absorbed and not radiated.

## 1.10 Glossary

The following lists abbreviations, acronyms and definitions of specific terms used in this chapter.

| | |
|---|---|
| **Antenna** | A carefully constructed device for the reception or transmission of radio energy into the air. |
| **Antenna gain pattern (AGP)** | Occasionally also referred to as polar diagrams. These are a graphical representation of the transmitting or receiving properties of an antenna. |
| **CEPT** | Conference of European Telecommunications Administrations. A group that assists with the implementation of ITU radio regulations on a national level. |
| **CCIR** | International Radio Consultative Committee. The body that considers and reports on issues affecting the use of radio communications. |
| **CCIT** | International Telecommunications Consultative Committee. |

| | |
|---|---|
| **Diffraction** | The term describing the 'bending' of a surface radio wave ground large obstacles in its path. |
| **E field** | Radio wave electrostatic energy field. |
| **EHF** | Extreme high frequency, the 30–300 GHz band. Still experimental. |
| **Fading** | The loss of power in a radio wave caused by environmental effects. |
| **FCC** | Federal Communications Commission. The body which polices the civilian use of radio communications in the USA. |
| **Feed line** | The wire connecting an antenna to the communications system. |
| **H field** | Radio wave electromagnetic energy field. |
| **HF** | High frequency, the 3–30 MHz band. Traditionally provides terrestrial global communications using medium power and acceptable antenna lengths. |
| **ITU** | International Telecommunications Union, the radio frequency watchdog. |
| **LF** | Low frequency, the 30–300 kHz band. Requires long antenna and large power input to be useful. Generally ground wave mode only. |
| **MF** | Medium frequency, the 300 kHz to 3 MHz band. Traditionally provides short-range communications using medium power and acceptable antenna lengths. |
| **Refraction** | The 'bending' of a sky wave by the effect of the ionosphere causing it to return to earth. |
| **RF spectrum** | The usable section of the extensive natural frequency spectrum. |
| **SHF** | Super high frequency, the 3–30 GHz band; microwaves. Line of sight communications. Generally used for satellite communications and RADAR. |
| **Sky wave** | A propagated radio wave that travels to the ionosphere from where it may or may not be returned to earth. |
| **Space wave** | A propagated radio wave that travels in a straight line. Used for point-to-point communications |
| **Surface wave** | A propagated radio wave that predominantly travels along the surface of the earth. |
| **UHF** | Ultra high frequency, the 300 MHz to 3 GHz band; microwaves. Line-of-sight transmission. Generally used for satellite communications. |
| **VHF** | Very high frequency. The 30–300 MHz band. Line-of-sight transmission from short antenna using low power. Maritime short-range communications band. |
| **VLF** | Very low frequency, the 10–30 kHz band. Requires huge antenna and great power for long-range communication. |
| **WARC** | World Administrative Radio Conference. The body that produces radio regulations and a Table of Frequency Allocations. |
| **Wavelength** | The physical length in metres between one cycle of the transmitted frequency. A parameter used in the calculation of antenna lengths. |

## 1.11 Summary

- Radio waves travel through free space at approximately $300 \times 10^6\,\mathrm{ms}^{-1}$.
- The frequency, wavelength and velocity of the radiowave are interrelated.
- The radio frequency spectrum is regulated by the International Telecommunications Union (ITU).
- The Table of Frequency Allocations and radio regulatory documents are revised at the World Administrative Conference (WARC).

- The radio frequency spectrum is divided into several bands: they are VLF, LF, MF, HF, VHF, UHF, SHF and EHF.
- A propagated radio wave contains both electromagnetic and electrostatic energy called the magnetic field and the electric field.
- A radio wave propagates from an antenna in one or more of three modes; surface wave, sky wave and space wave.
- Surface waves travel along the ground and consequently the transmitted power is attenuated, thus limiting communication range.
- Sky waves travel to the ionosphere from where they may or may not be returned to the earth. Sky waves provide terrestrial global communications.
- Space waves offer line-of-sight communications. Range is limited by the curvature of the earth, and large objects in the path of the wave will block the signal creating shadow areas.
- Amplitude and/or frequency fading of the signal are a major problem in communication systems.
- Antennae are critically constructed to satisfy frequency, power and environmental requirements.
- Transmitting antenna need to handle large power outputs and are more robust than receiving antenna, although a single antenna may be employed for both purposes.
- Antennas may be directional or not depending upon requirements.
- Antenna feed lines are often called coaxial cables and consist of an inner (signal) wire surrounded by a mesh of copper called the earth (ground) connection.

## 1.12 Revision questions

1. Why does it appear that the radiocommunications range on MF/HF is greater at night than during the day at your location?
2. How is it possible to receive LF radio waves in regions that are radio-shadow areas to VHF radio waves?
3. Unwanted sky wave reception gives rise to errors in some navigation systems, typically Loran-C. Why is the effect more prevalent at night?
4. How may frequency selective fading be minimized in a receiver system?
5. How are the receptive properties and an antenna’s physical length related?
6. What is an antenna azimuth gain plot?
7. If a VHF antenna is remounted higher on the mast of a vessel, radio communications range is increased. Why is this?
8. If a vertical antenna is remounted horizontally at the same height above sea level, radio communications range is severely reduced. Why is this?
9. How are an antenna’s directivity and gain related?
10. By carefully locating some antennas, problems of signal fading, and in the case of GPS, errors in the range calculation can be reduced. Why is this?

# Chapter 2

# Depth sounding systems

## 2.1 Introduction

Sonar (*so*und *n*avigation *a*nd *r*anging) is the acronym identifying those systems that rely for their operation on the transmission and reception of acoustic energy in water. The term is widely used to identify all modern systems that propagate acoustic or electromagnetic energy into seawater to determine a vessel's speed or the depth of water under the keel. This book is not concerned with those specialized sonar techniques that are used for locating submerged objects, either fish or submarines. A navigator in the Merchant Navy is interested only in the depth of the water beneath the vessel, an indication of the speed of his ship and the distance run. See Chapter 3 for a description of speed logging equipment.

The first section of this chapter deals with the characteristics and problems that arise from the need to propagate energy in seawater.

## 2.2 The characteristics of sound in seawater

Before considering the problems of transmitting and receiving acoustic energy in seawater, the effects of the environment must be understood. Sonar systems rely on the accurate measurement of reflected frequency or, in the case of depth sounders, a precise measurement of time and both these parameters are affected by the often unpredictable ocean environment. These effects can be summarized as follows.

- Attenuation. A variable factor related to the transmitted power, the frequency of transmission, salinity of the seawater and the reflective consistency of the ocean floor.
- Salinity of seawater. A variable factor affecting both the velocity of the acoustic wave and its attenuation.
- Velocity of sound in salt water. This is another variable parameter. Acoustic wave velocity is precisely $1505\,\mathrm{ms}^{-1}$ at 15°C and atmospheric pressure, but most echo-sounding equipment is calibrated at $1500\,\mathrm{ms}^{-1}$.
- Reflective surface of the seabed. The amplitude of the reflected energy varies with the consistency of the ocean floor.
- Noise. Either inherent noise or that produced by one's own transmission causes the signal-to-noise ratio to degrade, and thus weak echo signals may be lost in noise.

Two additional factors should be considered.

- Frequency of transmission. This will vary with the system, i.e. depth sounding or Doppler speed log.
- Angle of incidence of the propagated beam. The closer the angle to vertical the greater will be the energy reflected by the seabed.

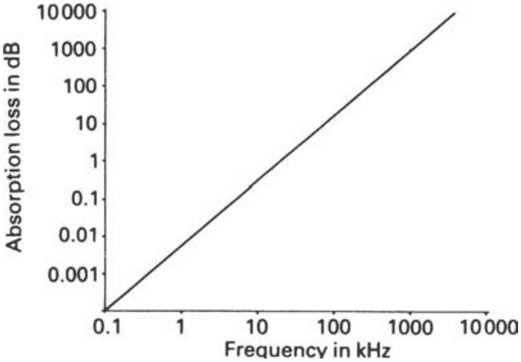


**Figure 2.1** A linear graph produced by plotting absorption loss against frequency. Salinity of the seawater is 3.4% at 15°C.

### 2.2.1 Attenuation and choice of frequency

The frequency of the acoustic energy transmitted in a sonar system is of prime importance. To achieve a narrow directive beam of energy, the radiating transducer is normally large in relation to the wavelength of the signal. Therefore, in order to produce a reasonably sized transducer emitting a narrow beam, a high transmission frequency needs to be used. The high frequency will also improve the signal-to-noise ratio in the system because ambient noise occurs at the lower end of the frequency spectrum. Unfortunately the higher the frequency used the greater will be the attenuation as shown in Figure 2.1.

The choice of transmission frequency is therefore a compromise between transducer size, freedom from noise, and minimal attenuation. Frequencies between 15 and 60 kHz are typical for depth sounders fitted in large vessels. A high power is transmitted from a large magnetostrictive transducer to indicate great depths with low attenuation. Small light craft use depth sounders that transmit in the band 200–400 kHz. This enables compact electrostrictive or ceramic transducers to be used on a boat where space is limited. Speed logs use frequencies in the range 300 kHz to 1 MHz depending upon their design and are not strictly sonar devices in the true definition of the sense.

#### *Beam spreading*

Transmission beam diverging or spreading is independent of fixed parameters, such as frequency, but depends upon distance between the transducer and the seabed. The greater the depth, the more the beam spreads, resulting in a drop in returned energy.

#### *Temperature*

Water temperature also affects absorption. As temperature decreases, attenuation decreases. The effect of temperature change is small and in most cases can be ignored, although modern sonar equipment is usually fitted with a temperature sensor to provide corrective data to the processor.

#### *Consistency of the seabed*

The reflective property of the seabed changes with its consistency. The main types of seabed and the attenuation which they cause are listed in Table 2.1. The measurements were made with an echo sounder transmitting 24 kHz from a magnetostrictive transducer.

**Table 2.1** Sea bed consistency and attenuation

| *Consistency* | *Attenuation (dB)* |
|---|---|
| Soft mud | 15 |
| Mud/sand | 9 |
| Sand/mud | 6 |
| Sand | 3 |
| Stone/rock | 1 |

These figures are typical and are quoted as a guideline only. In practice sufficient transmitted power will overcome these losses.

### 2.2.2 Salinity, pressure and the velocity of the acoustic wave

Since a depth sounder operates by precisely calculating the time taken for a pulse of energy to travel to the ocean floor and return, any variation in the velocity of the acoustic wave from the accepted calibrated speed of 1500 $ms^{-1}$ will produce an error in the indicated depth. The speed of acoustic waves in seawater varies with temperature, pressure and salinity. Figure 2.2 illustrates the speed variation caused by changes in the salinity of seawater.

Ocean water salinity is approximately 3.4% but it does vary extensively throughout the world. As salinity increases, sonar wave velocity increases producing a shallower depth indication, although in practice errors due to salinity changes would not be greater than 0.5%. The error can be ignored except when the vessel transfers from seawater to fresh water, when the indicated depth will be approximately 3% greater than the actual depth. The variation of speed with pressure or depth is indicated by the graph in Figure 2.3.

It can readily be seen that the change is slight, and is normally only compensated for in apparatus fitted on survey vessels. Seasonal changes affect the level of the thermocline and thus there is a small annual velocity variation. However, this can usually be ignored.

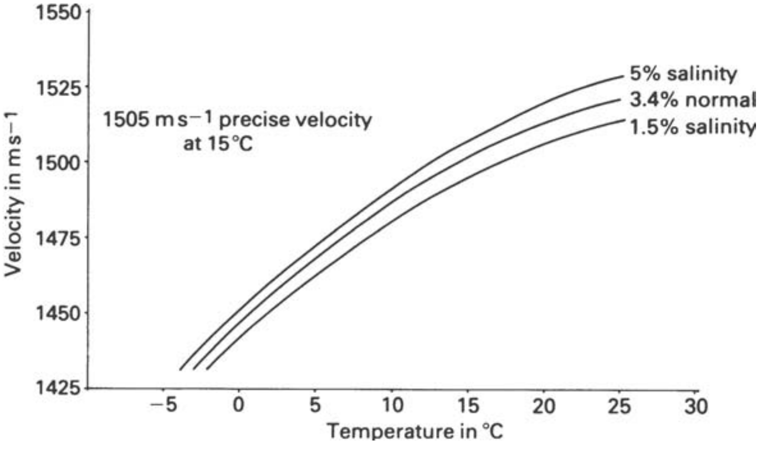


**Figure 2.2** Graph showing that the velocity of acoustic energy is affected by both the temperature and the salinity of seawater.

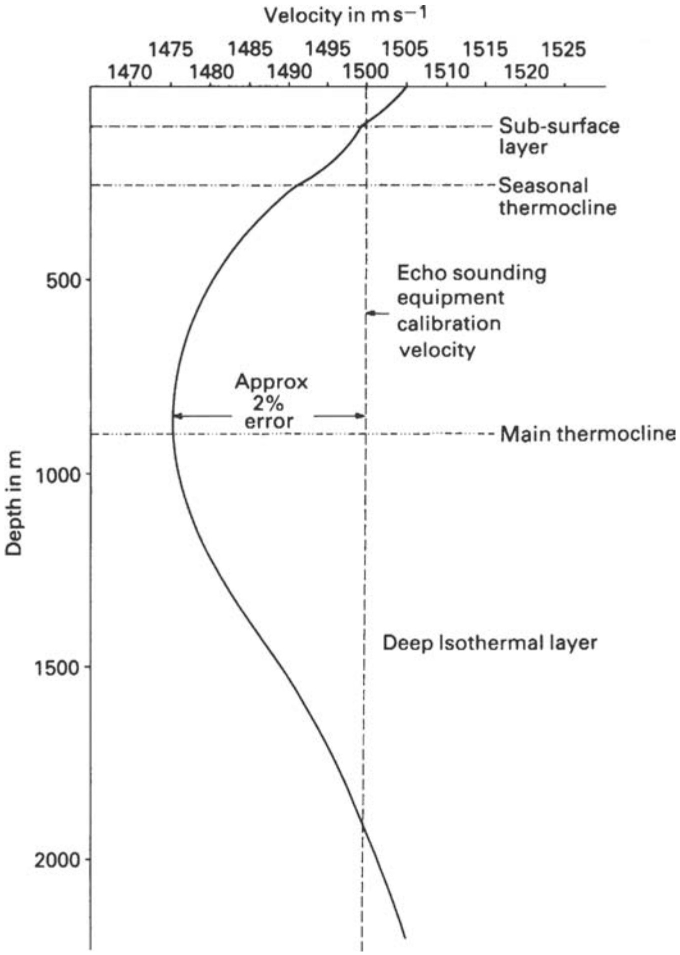


**Figure 2.3** Variation of the velocity of acoustic waves with pressure.

### 2.2.3 Noise

Noise present in the ocean adversely affects the performance of sonar equipment. Water noise has two main causes.

- The steady ambient noise caused by natural phenomena.
- Variable noise caused by the movement of shipping and the scattering of one's own transmitted signal (reverberation).

*Ambient noise*

Figure 2.4 shows that the amplitude of the ambient noise remains constant as range increases, whereas both the echo amplitude and the level of reverberation noise decrease linearly with range. Because of beam spreading, scattering of the signal increases and reverberation noise amplitude falls more slowly than the echo signal amplitude.

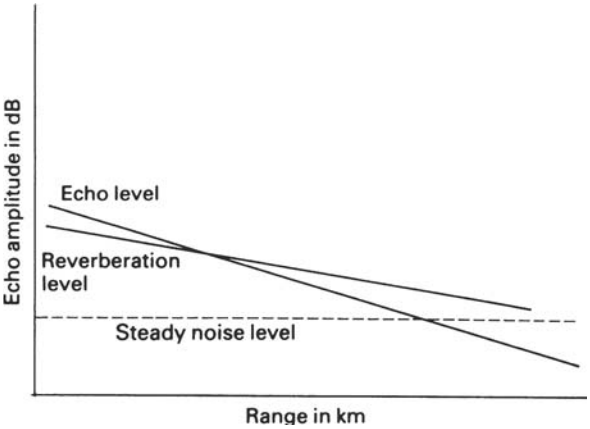


**Figure 2.4** Comparison of steady-state noise, reverberation noise and signal amplitude.

Ambient noise possesses different characteristics at different frequencies and varies with natural conditions such as rainstorms. Rain hitting the surface of the sea can cause a 10-fold increase in the noise level at the low frequency (approx. 10 kHz) end of the spectrum. Low frequency noise is also increased, particularly in shallow water, by storms or heavy surf. Biological sounds produced by some forms of aquatic life are also detectable, but only by the more sensitive types of equipment.

The steady amplitude of ambient noise produced by these and other factors affects the signal-to-noise ratio of the received signal and can in some cases lead to a loss of the returned echo. Signal-to-noise ratio can be improved by transmitting more power. This may be done by increasing the pulse repetition rate or increasing the amplitude or duration of the pulse. Unfortunately such an increase, which improves signal-to-noise ratio, leads to an increase in the amplitude of reverberation noise. Ambient noise is produced in the lower end of the frequency spectrum. By using a slightly higher transmitter frequency and a limited bandwidth receiver it is possible to reduce significantly the effects of ambient noise.

### *Reverberation noise*

Reverberation noise is the term used to describe noise created and affected by one's own transmission. The noise is caused by a 'back scattering' of the transmitted signal. It differs from ambient noise in the following ways.

- Its amplitude is directly proportional to the transmitted signal.
- Its amplitude is inversely proportional to the distance from the target.
- Its frequency is the same as that of the transmitted signal.

The signal-to-noise ratio cannot be improved by increasing transmitter power because reverberation noise is directly proportional to the power in the transmitted wave. Also it cannot be attenuated by improving receiver selectivity because the noise is at the same frequency as the transmitted wave. Furthermore reverberation noise increases with range because of increasing beamwidth. The area covered by the wavefront progressively increases, causing a larger area from which back scattering will occur. This means that reverberation noise does not decrease in amplitude as rapidly as the transmitted signal. Ultimately, therefore, reverberation noise amplitude will exceed the signal noise amplitude, as shown in Figure 2.4, and the echo will be lost. The amplitude of both the echo and reverberation noise decreases linearly with range. However, because of beam spreading, back scattering increases and reverberation noise amplitude falls more slowly than the echo signal amplitude. Three totally different ‘scattering’ sources produce reverberation noise.

- Surface reverberation. As the name suggests, this is caused by the surface of the ocean and is particularly troublesome during rough weather conditions when the surface is turbulent.
- Volume reverberation. This is the interference caused by beam scattering due to suspended matter in the ocean. Marine life, prevalent at depths between 200 and 750 m, is the main cause of this type of interference.
- Bottom reverberation. This depends upon the nature of the seabed. Solid seabeds, such as hard rock, will produce greater scattering of the beam than silt or sandy seabeds. Beam scattering caused by a solid seabed is particularly troublesome in fish finding systems because targets close to the seabed can be lost in the scatter.

## 2.3 Transducers

A transducer is a converter of energy. RF energy, when applied to a transducer assembly, will cause the unit to oscillate at its natural resonant frequency. If the transmitting face of the unit is placed in contact with, or close to, seawater the oscillations will cause acoustic waves to be transmitted in the water. Any reflected acoustic energy will cause a reciprocal action at the transducer. If the reflected energy comes into contact with the transducer face natural resonant oscillations will again be produced. These oscillations will in turn cause a minute electromotive force (e.m.f.) to be created which is then processed by the receiver to produce the necessary data for display.

Three types of transducer construction are available; electrostrictive, piezoelectric resonator, and magnetostrictive. Both the electrostrictive and the piezoelectric resonator types are constructed from piezoelectric ceramic materials and the two should not be confused.

### 2.3.1 Electrostrictive transducers

Certain materials, such as Rochelle salt and quartz, exhibit pressure electric effects when they are subjected to mechanical stress. This phenomenon is particularly outstanding in the element lead zirconate titanate, a material widely used for the construction of the sensitive element in modern electrostrictive transducers. Such a material is termed ferro-electric because of its similarity to ferro-magnetic materials.

The ceramic material contains random electric domains which when subjected to mechanical stress will line up to produce a potential difference (p.d.) across the two plate ends of the material section. Alternatively, if a voltage is applied across the plate ends of the ceramic crystal section its length will be varied. Figure 2.5 illustrates these phenomena.

The natural resonant frequency of the crystal slice is inversely proportional to its thickness. At high frequencies therefore the crystal slice becomes brittle, making its use in areas subjected to great stress forces impossible. This is a problem if the transducer is to be mounted in the forward section of a large merchant vessel where pressure stress can be intolerable. The fragility of the crystal also imposes limits on the transmitter power that may be applied because mechanical stress is directly related to power. The power restraints thus established make the electrostrictive transducer unsuitable for use in depth sounding apparatus where great depths need to be indicated. In addition, the low transmission frequency requirement of an echo sounder means that such a transducer crystal slice would be

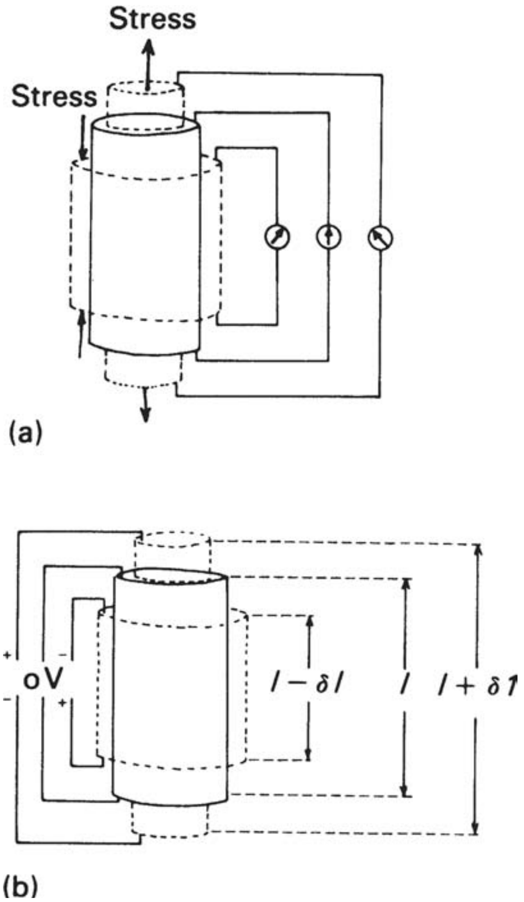


**Figure 2.5** (a) An output is produced when a piezoelectric ceramic cylinder is subjected to stress. (b) A change of length occurs if a voltage is applied across the ends of a piezoelectric ceramic cylinder.

excessively thick and require massive transmitter peak power to cause it to oscillate. The crystal slice is stressed by a voltage applied across its ends, thus the thicker the crystal slice, the greater is the power needed to stress it.

The electrostrictive transducer is only fitted on large merchant vessels when the power transmitted is low and the frequency is high, a combination of factors present in Doppler speed logging systems. Such a transducer is manufactured by mounting two crystal slices in a sandwich of two stainless steel cylinders. The whole unit is pre-stressed by inserting a stainless steel bolt through the centre of the active unit as shown in Figure 2.6.

If a voltage is applied across the ends of the unit, it will be made to vary in length. The bolt is insulated from the crystal slices by means of a PVC collar and the whole cylindrical section is made waterproof by means of a flexible seal. The bolt tightens against a compression spring permitting the crystal slices to vary in length, under the influence of the RF energy, whilst still remaining mechanically stressed. This method of construction is widely found on the electrostrictive transducers used in the Merchant Navy. For smaller vessels, where the external stresses are not so severe, the simpler piezoelectric resonator is used.

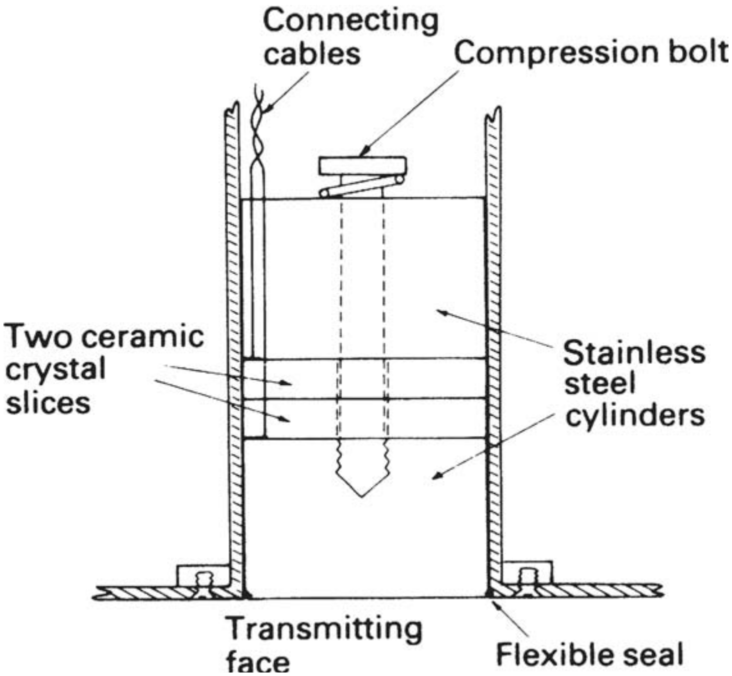


**Figure 2.6** Construction details of a ceramic electrostrictive transducer.

### 2.3.2 Piezoelectric resonator

This type of transducer makes use of the flexible qualities of a crystal slice. If the ceramic crystal slice is mounted so that it is able to flex at its natural resonant frequency, acoustic oscillations can be produced. The action is again reciprocal. If the ceramic crystal slice is mounted at its corners only, and is caused to flex by an external force, a small p.d. will be developed across the ends of the element. This phenomenon is widely used in industry for producing such things as electronic cigarette lighters and fundamental crystal oscillator units for digital watches. However, a ceramic crystal slice used in this way is subject to the same mechanical laws as have previously been stated. The higher the frequency of oscillation, the thinner the slice needs to be and the greater the risk of fracture due to external stress or overdriving. For these reasons, piezoelectric resonators are rarely used at sea.

### 2.3.3 Magnetostrictive transducers

Figure 2.7 shows a bar of ferromagnetic material around which is wound a coil. If the bar is held rigid and a large current is passed through the coil, the resulting magnetic field produced will cause the bar to change in length. This slight change may be an increase or a decrease depending upon the material used for construction. For maximum change of length for a given input signal, annealed nickel has been found to be the optimum material and consequently this is used extensively in the construction of marine transducers.

As the a.c. through the coil increases to a maximum in one direction, the annealed nickel bar will reach its maximum construction length $(l+\delta l)$. With the a.c. at zero the bar returns to normal $(l)$. The current now increases in the opposite direction and the bar once again constricts $(l-\delta l)$. The frequency

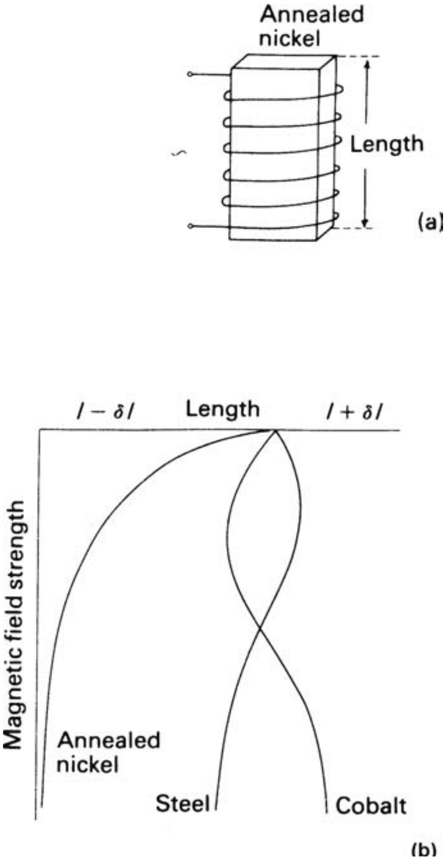


**Figure 2.7** (a) A bar of ferromagnetic material around which is wound a coil. (b) Relationship between magnetic field strength and change of length.

of resonance is therefore twice that of the applied a.c. This frequency doubling action is counteracted by applying a permanent magnet bias field produced by an in-built permanent magnet.

The phenomenon that causes the bar to change in length under the influence of a magnetic field is called ‘magnetostriction’, and in common with most mechanical laws possesses the reciprocal quality. When acoustic vibrations cause the bar to constrict, at its natural resonant frequency, an alternating magnetic field is produced around the coil. A minute alternating current is caused to flow in the coil and a small e.m.f. is generated. This is then amplified and processed by the receiver as the returned echo.

To limit the effects of magnetic hysteresis and eddy current losses common in low frequency transformer construction, the annealed nickel bar is made of laminated strips bonded together with an insulating material. Figure 2.8 illustrates the construction of a typical magnetostrictive transducer unit. The transmitting face is at the base of the diagram.

Magnetostrictive transducers are extremely robust which makes them ideal for use in large vessels where heavy sea pounding could destroy an unprotected electrostrictive type. They are extensively used with depth sounding apparatus because at the low frequencies used they can be constructed to an acceptable size and will handle the large power requirement of a deep sounding system. However,

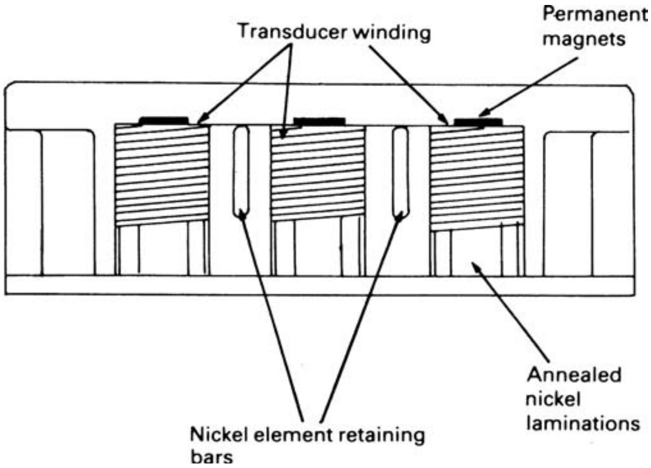


**Figure 2.8** Cross-section of a magnetostrictive transducer. (Reproduced courtesy of Marconi Marine.)

magnetic losses increase with frequency, and above 100 kHz the efficiency of magnetostrictive transducers falls to below the normal 40%. Above this frequency electrostrictive transducers are normally used.

### 2.3.4 Transducer siting

The decision of where to mount the transducer must not be made in haste. It is vital that the active face of the transducer is in contact with the water. The unit should also be mounted well away from areas close to turbulence that will cause noise. Areas close to propellers or water outlets must be avoided.

Aeration is undoubtedly the biggest problem encountered when transducers are wrongly installed. Air bubbles in the water, for whatever reason, will pass close to the transducer face and act as a reflector of the acoustic energy.

As a vessel cuts through the water, severe turbulence is created. Water containing huge quantities of air bubbles is forced under and along the hull. The bow wave is aerated as it is forced above the surface of the sea, along the hull. The wave falls back into the sea at approximately one-third the distance along the length of the vessel from the bow. A transducer mounted aft of the position where the bow wave re-enters the sea, would suffer badly from the problems of aeration. Mounting the transducer ahead of this point, even in the bulbous bow, would be ideal. It should be remembered, however, that at some stage maintenance may be required and a position in the bulbous bow may be inaccessible.

A second source of aeration is that of cavitation. The hull of a vessel is seldom smooth and any indentations or irregularities in it will cause air bubbles to be produced leading to aeration of the transducer face. Hull irregularities are impossible to predict as they are not a feature of the vessel's design.

## 2.4 Depth sounding principles

In its simplest form, the depth sounder is purely a timing and display system that makes use of a transmitter and a receiver to measure the depth of water beneath a vessel. Acoustic energy is transmitted perpendicularly from the transducer to the seabed. Some of the transmitted energy is reflected and will be received by the transducer as an echo. It has been previously stated that the velocity of sound waves in seawater is accepted to be 1500 $\text{ms}^{-1}$. Knowledge of this fact and the ability to measure precisely the time delay between transmission and reception, provides an accurate indication of the water depth.

$$\text{Distance travelled} = \frac{\text{velocity} \times \text{time}}{2}$$

where velocity = 1500 $\text{ms}^{-1}$ in salt water; time = time taken for the return journey in seconds; and distance = depth beneath the transducer in metres. Thus if the time taken for the return journey is 1 s, the depth of water beneath the transducer is 750 m. If the time is 0.1 s the depth is 75 m, and so on.

The transmitter and transducer, must be capable of delivering sufficient power and the receiver must possess adequate sensitivity to overcome all of the losses in the transmission medium (seawater and seabed). It is the likely attenuation of the signal, due to the losses described in the first part of this chapter, which determines the specifications of the equipment to be fitted on a merchant vessel.

### 2.4.1 Continuous wave/pulse system

The transmission of acoustic energy for depth sounding, may take one of two forms.

- A continuous wave system, where the acoustic energy is continuously transmitted from one transducer. The returned echo signal is received by a second transducer and a phase difference between the two is used to calculate the depth.
- The pulse system, in which rapid short, high intensity pulses are transmitted and received by a single transducer. The depth is calculated by measuring the time delay between transmission and reception.

The latter system is preferred in the majority of applications. Both the pulse length (duration) and the pulse repetition frequency (PRF) are important when considering the function of the echo sounding apparatus.

#### *Continuous wave system*

This system is rarely used in commercial echo sounding applications. Because it requires independent transmitters and receivers, and two transducer assemblies it is expensive. Also because the transmitter is firing continually, noise is a particular problem. Civilian maritime echo sounders therefore use a pulsed system.

#### *Pulsed system*

In this system the transmitter fires for a defined period of time and is then switched off. The pulse travels to the ocean floor and is reflected back to be received by the same transducer which is now switched to a receive mode. The duration of the transmitter pulse and the pulse repetition frequency (PRF) are particularly important parameters in this system

The pulse duration effectively determines the resolution quality of the equipment. This, along with the display method used, enables objects close together in the water, or close to the seabed, to be recorded separately. It is called target or echo discrimination. This factor is particularly important in fish finding apparatus where very short duration pulses (typically 0.25 or 0.5 ms) are used.

Echo discrimination *(D)* is:

$$D = V \times l \text{ (in metres)}$$

where $V$ = the velocity of acoustic waves, and $l$ = pulse length.

For a 0.5 ms pulse length:

$$D = 1500 \times 0.5 \times 10^{-3} = 0.75\,\text{m}$$

For a 2 ms pulse length:

$$D = 1500 \times 2 \times 10^{-3} = 3\,\text{m}$$

Obviously a short pulse length is superior where objects to be displayed are close together in the water. Short pulse lengths tend to be used in fish finding systems.

A short pulse length also improves the quality of the returned echo because reverberation noise will be less. Reverberation noise is directly proportional to the signal strength, therefore reducing the pulse length reduces signal strength which in turn reduces noise. Unfortunately, reducing the signal strength in this way reduces the total energy transmitted, thereby limiting the maximum depth from which satisfactory echoes can be received. Obviously, a compromise has to be made. Most depth sounders are fitted with a means whereby the pulse length can be varied with range. For shallow ranges, and for better definition, a short pulse length is used. On those occasions where great depths are to be recorded a longer pulse is transmitted.

For a given pulse length, the PRF effectively determines the maximum range that can be indicated. It is a measure of the time interval between pulses when transmission has ceased and the receiver is awaiting the returned echo.

The maximum indicated range may be determined by using the following formula:

$$\text{Maximum range indication } (r) = \frac{v \times t}{2} \text{ (in metres)}$$

where $v$ = velocity of sound in seawater ($1500\,\text{ms}^{-1}$) and $t$ = time between pulses in seconds. If the PRF is one per second (PRF = 60), the maximum depth recorded is 750 m. If the PRF is two per second (PRF = 120) the maximum depth recorded is 375 m.

The maximum display range should not be confused with the maximum depth. For instance, if the PRF is one per second the maximum display range is 750 m. If the water depth is 850 m, an echo will be returned after a second pulse has been transmitted and the range display has been returned to zero. The indicated depth would now be 100 m. A system of 'phased' ranges, where the display initiation is delayed for a pre-determined period after transmission overcomes the problem of over-range indication.

### 2.4.2 Transmission beamwidth

Acoustic energy is radiated vertically downwards from the transducer in the form of a beam of energy. As Figure 2.9 shows the main beam is central to the transducer face and shorter sidelobes are also produced. The beamwidth must not be excessively narrow otherwise echoes may be missed, particularly in heavy weather when the vessel is rolling. A low PRF combined with a fast ship speed

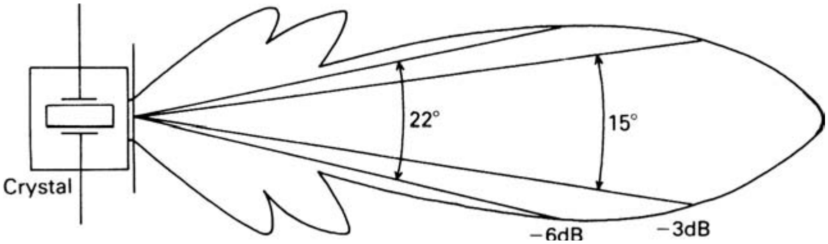


**Figure 2.9** Transmission beam showing the sidelobes.

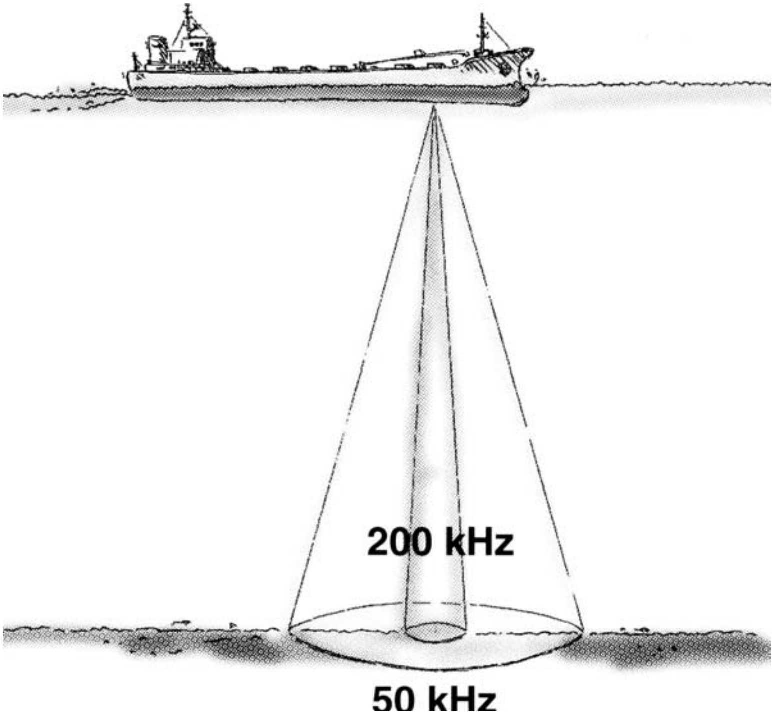


**Figure 2.10** Typical beamwidths for echo sounders transmitting low and high frequencies. (Reproduced courtesy Furuno Electric Co. Ltd.)

can in some cases lead to the vessel 'running away' from an echo that could well be missed. In general, beamwidths measured at the half-power points (–3 dB), used for depth sounding apparatus are between 15° and 25°. To obtain this relatively narrow beamwidth, the transducer needs to be constructed with a size equal to many wavelengths of the frequency in use. This fact dictates that the transducer will be physically large for the lower acoustic frequencies used in depth sounding.

In order to reduce the transducer size, and keep a narrow beamwidth, it is possible to increase the transmission frequency. However, the resulting signal attenuation negates this change and in practice a compromise must once again be reached between frequency, transducer size and beamwidth. Figure 2.10 shows typical beamwidths for a low frequency (50 kHz) sounder and that of a frequency four times greater.

## 2.5 A generic echo sounding system

Compared with other systems, echo sounder circuitry is relatively simple. Most manufacturers of deep sounding systems now opt for microprocessor control and digital displays, but it was not always so. Many mariners preferred the paper-recording echo sounder because the display was clear, easy to read and provided a history of soundings.

Marconi Marine's 'Seahorse' echo sounder (Figure 2.11) was typical of the standard paper-recording echo sounder. Built in the period before microprocessor control, it is used here to describe the relatively simply circuitry needed to produce an accurate read-out of depth beneath the keel. From the description it is easy to see that an echo sounding system is simply a timing device.

The system used a transmission frequency of 24 kHz and two ranges, either manually or automatically selected, to allow depths down to 1000 m to be recorded. The shallow range was 100 m and operated with a short pulse length of 200 µs, whereas the 1000 m range uses a pulse length of 2 ms. Display accuracy for the chart recorder is typically 0.5% producing indications with an accuracy of ±0.5 m on the 100 m range and ±5 m on the deepest range.

### 2.5.1 Description

*Receiver and chart recorder*

When chart recording has been selected, transmission is initiated by a pulse from a proximity detector which triggers the chart pulse generator circuit introducing a slight delay, pre-set on each range, to ensure that transmission occurs at the instant the stylus marks zero on the recording paper. This system trigger pulse or that from the trigger pulse generator circuit when the chart is switched off, has three functions:

- to initiate the pulse timing circuit
- to operate the blanking pulse generator
- to synchronize the digital and processing circuits.

The transmit timing circuit sets the pulse length to trigger the 24 kHz oscillator (transmission frequency). Pulse length is increased, when the deep range is changed, by a range switch (not shown). Power contained in the transmitted signal is produced by the power amplifier stage, the output of which is coupled to the magnetostrictive transducer with the neon indicating transmission.

When the transmitter fires, the receiver input is blanked to prevent the high-energy pulse from causing damage to the input tuned circuits. The blanking pulse generator also initiates the swept gain circuit and inhibits the data pulse generator. During transmission, the swept gain control circuit holds the gain of the input tuned amplifier low. At cessation of transmission, the hold is removed permitting the receiver gain to gradually increase at a rate governed by an inverse fourth power law. This type of inverse gain control is necessary because echoes that are returned soon after transmission ceases are of large amplitude and are likely to overload the receiver.

The echo amplitude gradually decreases as the returned echo delay period increases. Thus the swept gain control circuit causes the average amplitude of the echoes displayed to be the same over the whole period between transmission pulses. However, high intensity echoes returned from large reflective objects will produce a rapid change in signal amplitude and will cause a larger signal to be coupled to the logarithmic amplifier causing a more substantial indication to be made on the paper. The logarithmic amplifier and detector stages produce a d.c. output, the amplitude of which is logarithmically proportional to the strength of the echo signal.

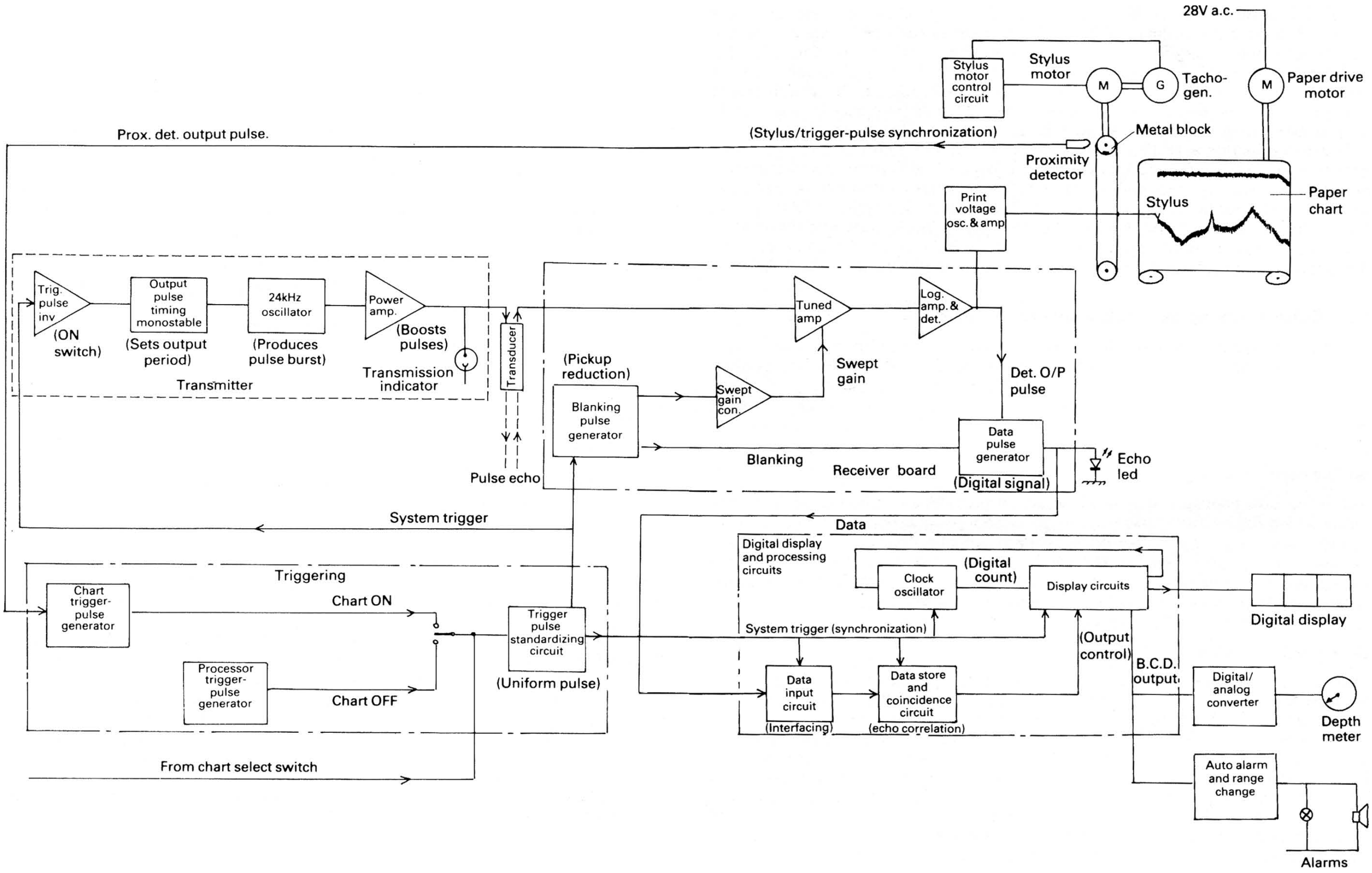


**Figure 2.11** A block schematic diagram of the Seahorse echo sounder. (Reproduced courtesy of Marconi Marine.)

In the chart recorder display, electrosensitive paper is drawn horizontally beneath a sharp stylus. The paper is tightly drawn over the grounded roller guides by a constant speed paper-drive motor. Paper marking is achieved by applying a high voltage a.c. signal to the stylus which is drawn at 90° to the paper movement, across the surface of the paper on top of the left-hand roller. The paper is marked by burning the surface with a high voltage charge produced through the paper between the stylus and ground. Depending upon the size of the returned echo, the marking voltage is between 440 and 1100 V and is produced from a print voltage oscillator running at 2 kHz. Oscillator amplifier output is a constant amplitude signal, the threshold level of which is raised by the d.c. produced by a detected echo signal. Thus a high-intensity echo signal causes the marking voltage to be raised above the threshold level by a greater amount than would be caused by a detected small echo signal.

For accurate depth marking it is essential that the stylus tracking speed is absolutely precise. The stylus is moved along the paper by a belt controlled by the stylus d.c. motor. Speed accuracy is maintained by a complex feedback loop and tacho-generator circuit.

### *Digital circuits*

The digital display section contains the necessary logic to drive the integral three-digit depth display, the alarm circuit, and the remote indicators. Pulse repetition frequency (PRF) of the clock oscillator is pre-set so that the time taken for the three-digit counter to count from 000 to 999 is exactly the same as that taken by the paper stylus to travel from zero to the maximum reading for the range in use. The counter output is therefore directly related to depth.

When the chart recorder is switched off, the digital processing section and the transmitter are triggered from the processor trigger pulse generator circuit. Both the transmit and receive sections work in the same way as previously described. A low logic pulse from the trigger pulse standardizing circuit synchronizes the logic functions. The d.c. output from the receiver detector is coupled via a data pulse generator circuit to the interface system. Unfortunately in any echo sounder it is likely that unwanted echoes will be received due to ship noise, aeration or other factors.

False echoes would be displayed as false depth indications on the chart and would be easily recognized. However, such echoes would produce instantaneous erroneous readings on the digital counter display that would not be so easily recognized. To prevent this happening echoes are stored in a data store on the processing board and only valid echoes will produce a reading on the display. Valid echoes are those that have indicated the same depth for two consecutive sounding cycles. The data store, therefore, consists of a two-stage counter which holds each echo for one sounding cycle and compares it with the next echo before the depth is displayed on the digital display.

The display circuit consists of three digital counters that are clocked from the clock oscillator circuit. Oscillator clock pulses are initiated by the system trigger at the instant of transmission. The first nine pulses are counted by the lowest order decade counter which registers 1–9 on the display least significant figure (LSF) element. The next clock pulse produces a 0 on the LSF display and clocks the second decade counter by one, producing a 1 in the centre of the display. This action continues, and if no echo is received, the full count of 999 is recorded when an output pulse from the counting circuit is fed back to stop the clock.

Each time transmission takes place the counters are reset to zero before being enabled. This is not evident on the display because the data output from the counters is taken via a latch that has to be enabled before data transfer can take place. Thus the counters are continually changing but the display data will only change when the latches have been enabled (when the depth changes). If an echo is received during the counting process, the output is stopped, and the output latches enabled by a pulse from the data store. The new depth is now displayed on the indicator and the counters are reset at the start of the next transmission pulse.

With any echo sounder, it is necessary that the clock pulse rate be directly related to depth. When the shallow (100 m) range is selected a high frequency is used which is reduced by a factor of 10 when the deep range (1000 m) is selected.

Modern echo sounders rely for their operation on the ubiquitous microprocessor and digital circuitry, but the system principles remain the same. It is the display of information that is the outward sign of the advance in technology.

## 2.6 A digitized echo sounding system

The Furuno Electric Co. Ltd, one of the world's big manufacturers of marine equipment, produces an echo sounder, the FE606, in which many of the functions have been digitized. Transmission frequency is either 50 or 200 kHz depending upon navigation requirements. A choice of 50 kHz provides greater depth indication and a wider beamwidth reducing the chance that the vessel may 'run away' from an echo (see Figure 2.10).

The pulse length increases with depth range from 0.4 ms, on the shallow ranges, to 2.0 ms on the maximum range. This enables better target discrimination on the lower ranges and ensures that sufficient pulse power is available on the higher ranges. Pulse repetition rate (sounding rate) is reduced as range increases to ensure adequate time between pulses for echoes to be returned from greater depths.

The system shown in Figure 2.12 is essentially a paper recorder and two LCD displays showing start depth and seabed depth. As before, transmission is initiated at the instant the stylus marks the zero line on the sensitive paper by a trigger sensor coupled to the control integrated circuits. Depending upon the range selected, the pulse length modulates the output from the transmit oscillator, which is power amplified and then coupled via a transmit/receive switch to the transducer.

A returned echo is processed in the receiver and applied to the logic circuitry. Here it is processed to determine that it is a valid echo and then it is latched through to a digital-to-analogue converter to produce the analogue voltage to drive the print oscillator. Thus the depth is marked on the sensitive paper at some point determined by the time delay between transmission and reception, and the distance the stylus has travelled over the paper.

## 2.7 A microcomputer echo sounding system

As you would expect, the use of computing technology has eliminated much of the basic circuitry and in most cases the mechanical paper display system of modern echo sounders. Current systems are much more versatile than their predecessors. The use of a computer enables precise control and processing of the echo sounding signal. Circuitry has now reached the point where it is virtually all contained on a few chips. However, the most obvious changes that users will be aware of in modern systems are the display and user interface.

Once again there are many manufacturers and suppliers of echo sounders or, as they are often now called, fish finders. The Furuno navigational echo sounder FE-700 is typical of many. Depending upon requirements the system is able to operate with a 200 kHz transmission frequency giving high-resolution shallow depth performance, or 50 kHz for deep-water sounding.

Seabed and echo data is displayed on a 6.5 inch high-brightness TFT colour LCD display which provides the navigator with a history of soundings over a period of 15 min, much as the older paper recording systems did (see Figure 2.13).

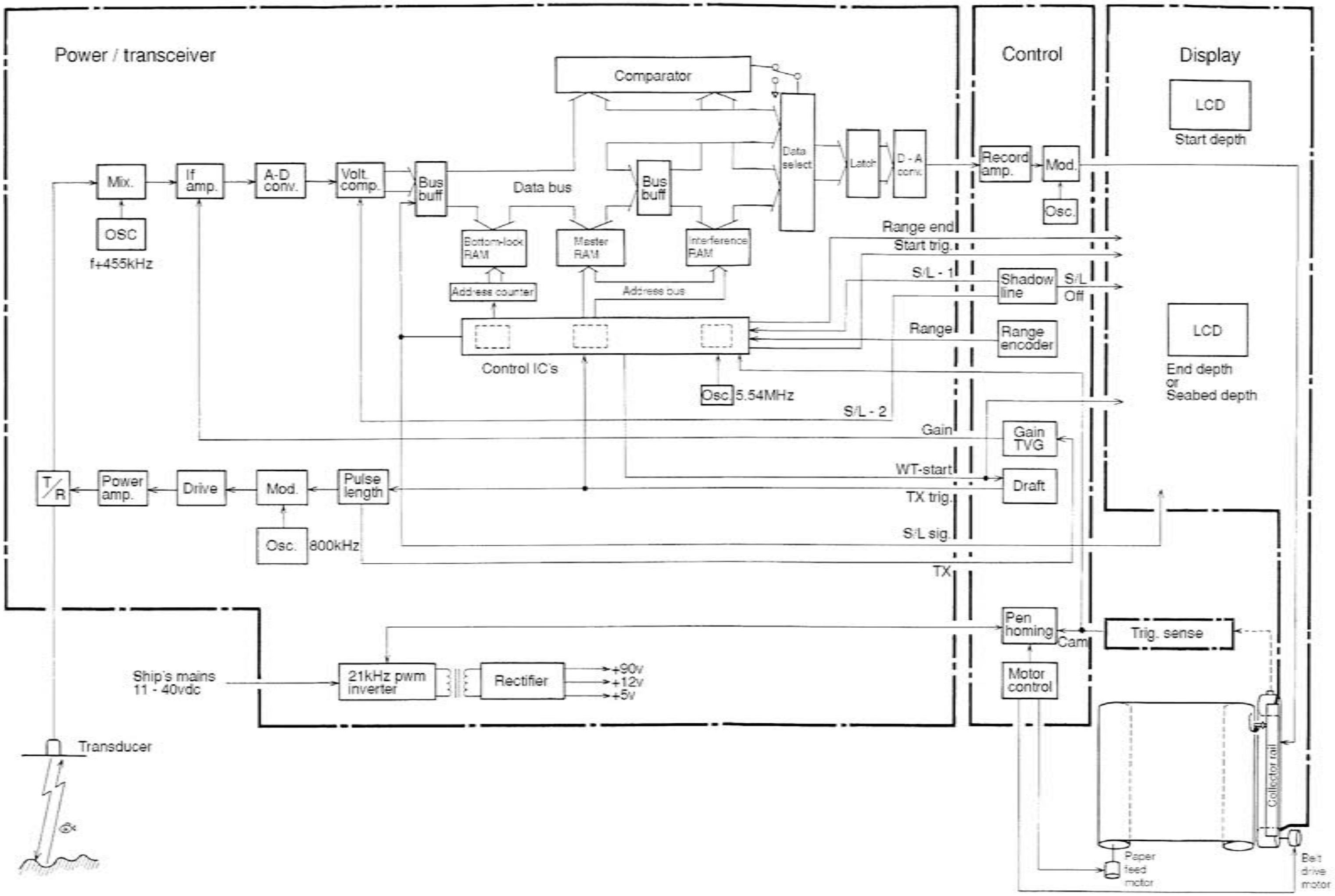


**Figure 2.12** Furuno FE-606 echo sounding system. (Reproduced courtesy of Furuno Electric Co.)

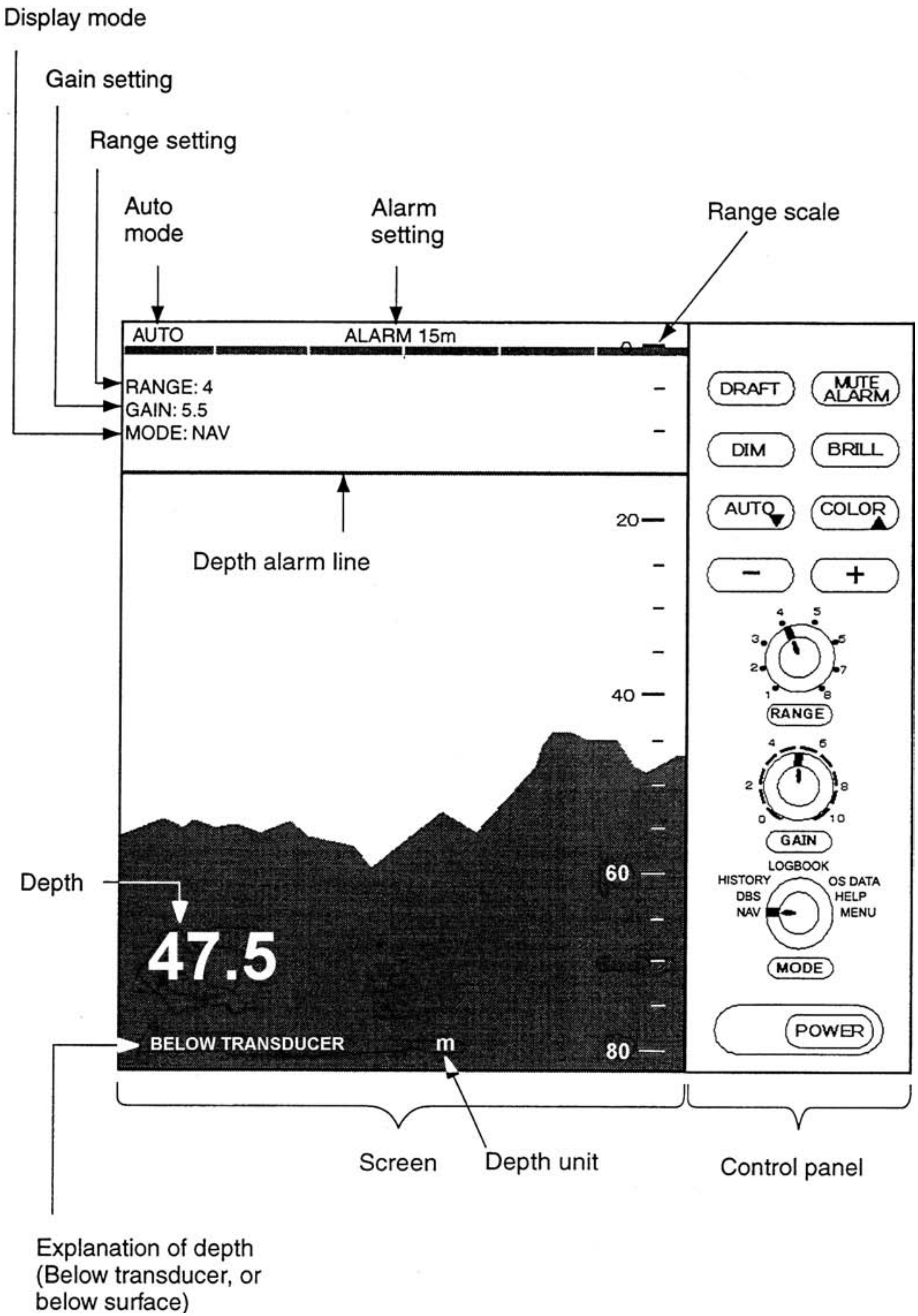


**Figure 2.13** Furuno FE-700 LCD TFT data display (Navigation Mode.) (Reproduced courtesy of Furuno Electric Co.)

Depths, associated time, and position are all stored in 24-h memory and can be played back at any time. This is a useful function if there is any dispute following an accident.

The main depth display emulates a cross-sectional profile of the ocean over the past 15 min. At the top of the display in Figure 2.13, the solid zero line marks the ocean surface or transducer level whichever is selected. At 15 m down, a second line marks the depth at which the alarm has been set. The undulating line showing the ocean floor depth is shown varying over 15 min from 58 to 44 m and the instantaneous depth, also shown as a large numerical display, is 47.5 m. Other operation detail is as shown in the diagram. What is not indicated on the display is the change of pulse length and period as selected by range.

As shown in Table 2.2, the pulse length is increased with the depth range to effectively allow more power to be contained in the transmitted pulse, whilst the pulse period frequency is reduced to permit longer gaps in the transmission period allowing greater depths to be indicated

**Table 2.2** Echo sounder range vs pulse length vs PRF

| *Depth (metres)* | *Pulse length (ms)* | *PRF (pulses per minute)* |
|---|---|---|
| 5, 10 and 20 | 0.25 | 750 |
| 40 | 0.38 | 375 |
| 100 | 1.00 | 150 |
| 200 | 2.00 | 75 |
| 400 and 800 | 3.60 | 42 |

In addition to the standard navigation mode, Furuno FE-700 users are provided with a number of options adequately demonstrating the capability of a modern echo sounder using a TFT LCD display (see Figure 2.14). All the selected modes display data as a window insert on top of the echo sounder NAV mode display.

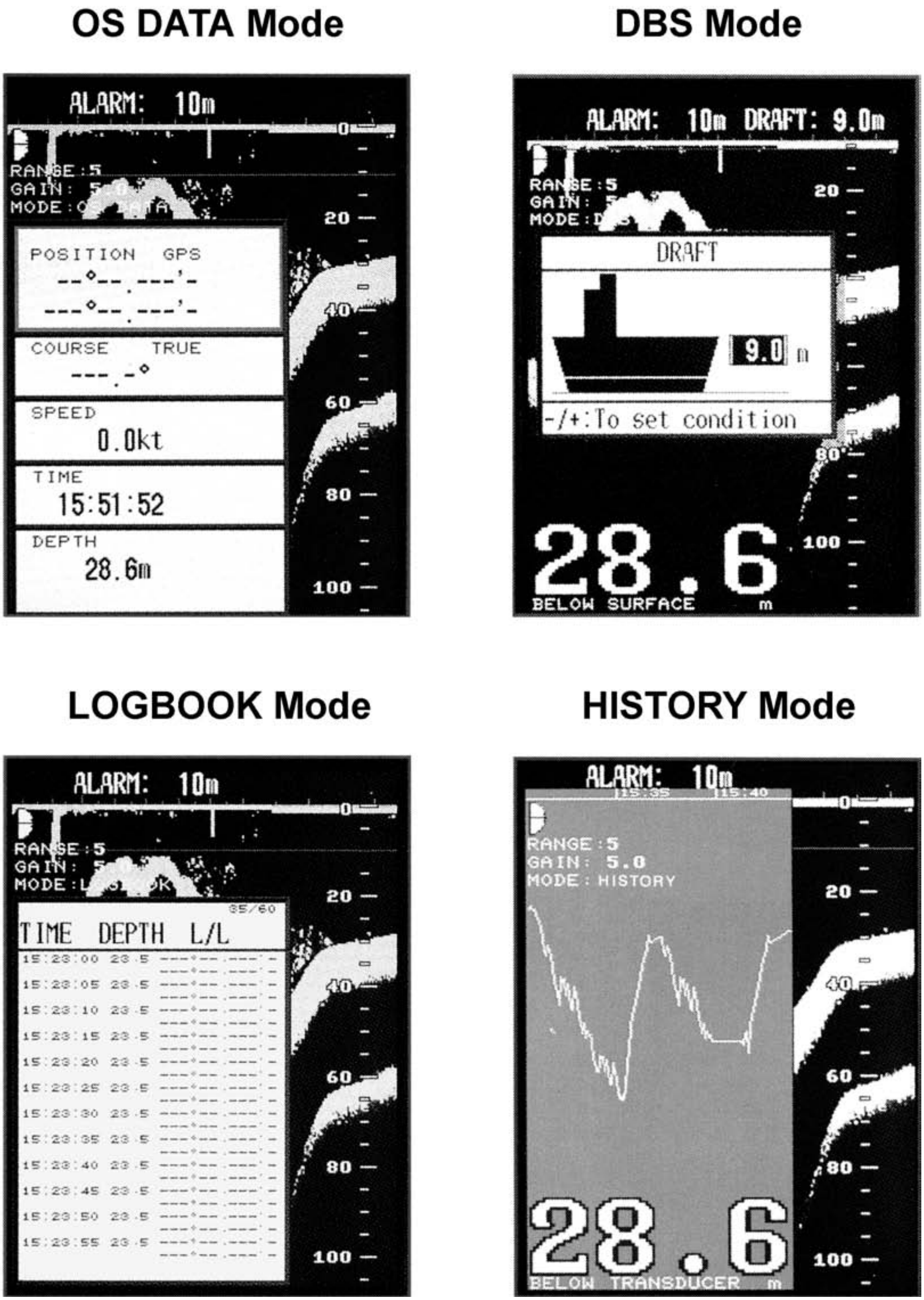
There are four display-mode areas.

- OS DATA mode. Indicates own ship position, GPS derived course, time and a digital display of water depth.
- DBS mode. Provides a draft-adjusted depth mode for referencing with maritime charts.
- LOGBOOK mode. As the name suggests, provides a facility for manually logging depths over a given period.
- HISTORY mode. Provides a mixture of contour and strata displays. The contour display can be shifted back over the past 24 h whilst the strata display (right-hand side of display) shows sounding data over the last 5 min.

## 2.8 Glossary

The following lists abbreviations, acronyms and definitions of specific terms used in this chapter.

| | |
|---|---|
| **Aeration** | Aerated water bubbles clinging to the transducer face cause errors in the system. |
| **Ambient noise** | Noise that remains constant as range increases. |



OS DATA Mode

ALARM: 10m

RANGE : 5
GAIN : 5.0
MODE : OS DATA

POSITION GPS
--°--.---'-
---°--.---'-

COURSE TRUE
---.-°

SPEED
0.0kt

TIME
15:51:52

DEPTH
28.6m

DBS Mode

ALARM: 10m DRAFT: 9.0m

RANGE : 5
GAIN : 5
MODE : DBS

DRAFT
9.0 m
-/+:To set condition

28.6
BELOW SURFACE m

LOGBOOK Mode

ALARM: 10m

RANGE : 5
GAIN : 5
MODE : LOGBOOK

35/60

| TIME | DEPTH | L/L |
|---|---|---|
| 15:23:00 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:05 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:10 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:15 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:20 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:25 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:30 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:35 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:40 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:45 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:50 | 23.5 | ---°--.---'-<br>---°--.---'- |
| 15:23:55 | 23.5 | ---°--.---'-<br>---°--.---'- |

HISTORY Mode

ALARM: 10m

15:35 15:40

RANGE : 5
GAIN : 5.0
MODE : HISTORY

28.6
BELOW TRANSDUCER m

**Figure 2.14** Different display modes demonstrating the flexibility of a microcomputer-controlled echo sounder. (Reproduced courtesy of Furuno Electric Co.)

| | |
|---|---|
| **Beam spreading** | The transmitted pulse of energy spreads as it travels away from the transducer. The use of a wide beam will cause noise problems in the receiver and a narrow beam may lead to an echo being missed as the vessel steams away from the area. |
| **Chart recorder** | A sensitive paper recording system which, when the surface is scratched by a stylus, marks the contour of the ocean floor. |
| **Continuous wave system** | An echo sounding system that uses two transducers and transmits and receives energy at the same time. |
| **Electrostrictive transducer** | A transducer design based on piezoelectric technology. It is used when a higher transmission frequency is needed such as in speed logging equipment or fish-finding sounders. |
| **Magnetostrictive transducer** | A design based on magnetic induction. A large heavy transducer capable of transmitting high power. Used in deep sounding systems. |
| **Pulse duration (length)** | The period of the transmitted pulse when the transmitter is active. |
| **Pulse repetition frequency (PRF)** | The number of pulses transmitted per minute by the system. Similar to RADAR |
| **Pulse wave system** | A system that, like RADAR, transmits pulses of energy from a transducer which is then switched off. The received energy returns to the same transducer. |
| **Reverberation noise** | Noise that decreases as range increases. |
| **Sonar** | *So*und *n*avigation *a*nd *r*anging. |
| **Velocity** | Speed of acoustic waves in seawater; 1505 $ms^{-1}$ or approximated to1500 $ms^{-1}$. |

## 2.9 Summary

- Sonar stands for *so*und *n*avigation *a*nd *r*anging.
- Sound travels relatively slowly in seawater at 1505 $ms^{-1}$. This is approximated to 1500 $ms^{-1}$ for convenience.
- The velocity is not a constant, it varies with the salinity of seawater. Ocean salinity is approximately 3.4%.
- Transmitted signal amplitude is attenuated by saltwater and the ocean floor from which it is reflected.
- Noise caused by sea creatures and ocean activity is a major problem affecting sonar equipment.
- The temperature of the seawater affects the velocity of the acoustic wave and consequently affects the accuracy of the displayed data. Temperature sensors are contained in the transducer housing to produce corrective data.
- Transducers are effectively the antennas of sonar systems. They transmit and receive the acoustic energy.
- There are two main types of transducer in use; magnetostrictive and electrostrictive. Magnetostrictive transducers are large and heavy and tend to be used only on large vessels. Electrostrictive transducers are lighter and often used in speed logging systems and on smaller craft.
- Low frequencies are often used in deep sounding systems typically in the range 10–100 kHz.
- The depth below the keel is related to the time taken for the acoustic wave to travel to the ocean floor and return. Put simply if the delay is 1 s and the wave travels at 1500 $ms^{-1}$ then the depth is $0.5 \times 1500 = 750$ m.

- Pulsed systems, like those used in maritime RADAR, are used in an echo sounder. The pulse length or duration determines the resolution of the equipment. A short pulse length will identify objects close together in the water. If all other parameters remain constant, the pulse repetition frequency (PRF), the number of pulses per minute, determines the maximum range that can be indicated.
- The width of the transmitted beam becomes wider as it travels away from the transducer. It should not be excessively narrow or the vessel may 'run away' from, or miss, the returned echo.
- Modern echo sounding equipment is computer controlled and therefore is able to produce a host of other data besides a depth indication.

## 2.10 Revision questions

1. Why do deep sounding echo sounders operate with a low transmission frequency?
2. For a given ocean depth, how is it possible for returned echoes to vary in strength?
3. If a vessel sails from salt water into fresh water the depth indicated by an echo sounder will be in error. Why is this and what is the magnitude of the error?
4. Noise can degrade an echo sounder display. How does narrowing the transmitted beamwidth reduce system noise and at what cost?
5. Why are electrostrictive transducers used in maritime applications in preference to piezoelectric resonators?
6. Why do marine echo sounding systems use pulsed transmission and not a continuous wave mode of operation?
7. Many echo sounders offer the ability to vary the transmission pulse duration. Why is this?
8. How are the pulse repetition frequency (PRF) and the maximum depth, indicated by an echo sounding system, related?
9. Why is the siting of an echo sounder transducer important?
10. What do you understand by the term target discrimination?
11. What effect may a narrow transmission beamwidth have on returned echoes if a ship is rolling in heavy seas?

# Chapter 3

# Speed measurement

## 3.1 Introduction

Speed measurement has always been of the utmost importance to the navigator. The accuracy of a dead reckoning position plotted after a long passage without star sights being taken, is dependent upon a sound knowledge of the vessel's heading and speed.

To be of value, the speed of any object must be measured relative to some other point. At sea, speed may be measured relative to either the seabed (ground reference speed) or to the water flowing past the hull (water reference speed). Both of these types of speed measurement are possible and both have their place in modern navigation systems.

This chapter deals with the methods of speed logging that are in general use on board modern vessels. One of these, the pressure tube log, is old but it still gives a satisfactory performance. Another, the electromagnetic log, is often used on smaller vessels and the popular Doppler speed log is to be found everywhere.

## 3.2 Speed measurement using water pressure

When a tube, with an opening at its base, is vertically submerged in water, a pressure, proportional to the depth to which the tube is submerged, will be developed in the tube. If the tube is held stationary the pressure remains constant and is termed 'static' pressure. If the tube is now moved through the water, whilst keeping the depth to which it is submerged constant, a second pressure called 'dynamic' pressure is developed. The total pressure in the tube, called a Pitot tube, is therefore the sum of both the static and dynamic pressures.

To ensure that the dynamic pressure reading, and thus speed, is accurate, the effect of static pressure must be eliminated. This is achieved by installing a second tube close to the first in such a way that the static pressure produced in it is identical to that created in the Pitot tube but without the pressure increase due to movement through the water (see Figure 3.1).

In a practical installation, tube B, the Pitot tube, extends below the vessel's hull to a depth *d*, whereas tube A, the static pressure intake tube, is flush with the hull. With the vessel stationary, the static pressures from tube A to the top of the diaphragm and tube B to its underside almost cancel. The unequal pressures, which cause a small indication of speed to be displayed when the vessel is stationary, are compensated for in the log electromechanical system and the erroneous indication is cancelled. As the vessel moves through the water, in the direction shown, water is forced into tube B producing a combined pressure in the lower half of the chamber equal to both the static and dynamic pressures. The difference in pressure, between upper and lower chambers, now forces the diaphragm upwards thus operating the mechanical linkage. Obviously the greater the speed of the vessel through the water, the more the diaphragm will move and the greater will be the speed indicated.

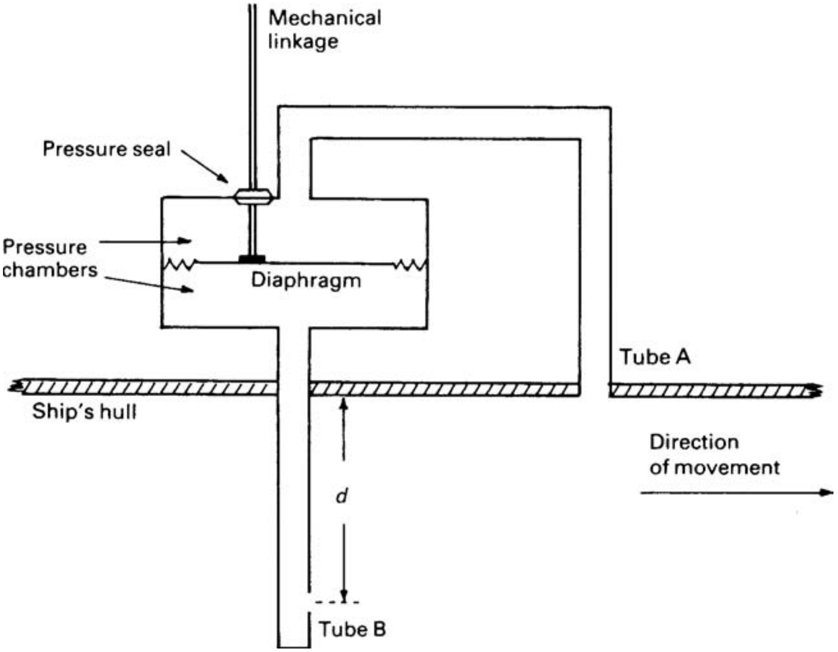


**Figure 3.1** The pressure tank and tube intakes of a pressure tube speed logging system.

Unfortunately, the dynamic pressure developed in tube B, by the relative movement through the water, is proportional to the square of the vessel's speed. Pitot's Law states that this pressure $p$ is proportional to the square of the ship's speed $v$ multiplied by the coefficient $K$.

$$p = K \times v^2$$

where the constant $K$ is derived from the vessel's tonnage, shape of hull, speed of the ship, and the length of the protruding part of the Pitot tube (distance $d$).

As shown in Figure 3.2, the speed indication produced is not linear. It is necessary therefore to eliminate the non-linear characteristics of the system and produce a linear speed indication. This is achieved mechanically, by the use of precisely engineered cones or electronically using CR (capacitive/resistive) time constant circuitry.

### 3.2.1 A pressure tube speed logging system

Figure 3.3 shows a typical installation of the Pitot system on board a vessel with a double bottom. The Pitot tube is encased in a sea-cock arrangement with valve control, to enable the tube to be withdrawn, without shipping water, when the vessel goes alongside. The static pressure opening is controlled by the use of a valve. Both dynamic and static pressures are transferred via air collectors and strainer valves to the pressure chamber. The strainer valves are designed to prevent water oscillations in the interconnecting pipes during operation. Such oscillations would cause the diaphragm to oscillate producing an erratic speed indication.

Figure 3.4 shows the basic speed and distance translating system of a Pitot tube log. The diagram includes two repeating systems for speed and distance data transmission to remote indicators on the

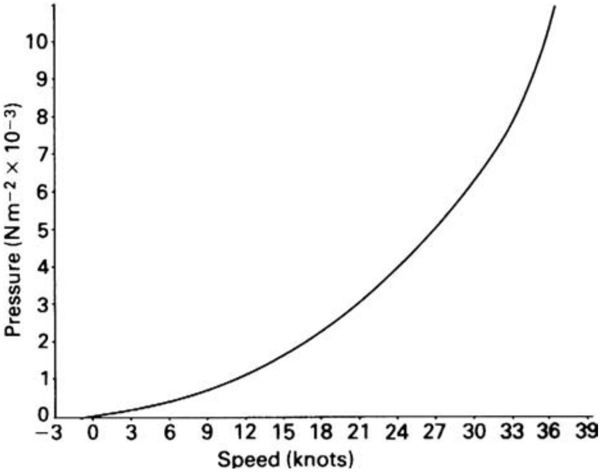


**Figure 3.2** Graph indicating the non-linear increase in pressure due to speed.

ship's bridge. This system was superseded by the SAL24E which replaced some of the mechanical apparatus with electronics. The original log has been included here because it is still in use on many vessels and is a fine example of a pressure type speed logging system.

### *Description of operation*

An increase in the vessel's speed will cause an increase in the dynamic pressure beneath the diaphragm in the pressure chamber (1). This causes the diaphragm to move upwards, pushing the pressure rod (2) and moving the lever (3) to the right on pivot (4). The upper end of the lever (3) moves the electric start contact (5) to the right to connect power to a reversible motor (6). The motor now turns causing the main shaft (7) to move a spiral cam (8) clockwise. This action tilts the lever (9), also pivoted on (4), to the left. The deflection stretches the main spring, producing a downward pressure on the diaphragm, via lever (3), causing it to cease rising at an intermediate position. This is achieved when equilibrium has been established between the dynamic pressure, acting on the lower side of the diaphragm, and the counter pressure from the spring on the upper side. At this point the motor (6) stops and thus holds the spiral cam (8) in a fixed position indicating speed.

This method of pressure compensation provides accurate indications of speed independent of alterations of the diaphragm caused by ageing. The shape of the spiral cam (8) has been carefully calculated to produce a linear indication of speed from the non-linear characteristics of the system. Also attached to the spiral cam is a second gearing mechanism (19) that transfers the movement of the speed indicator to the three-phase speed transmission system (20). An identical servo-receiver (22) is fitted in the remote speed repeater unit fitted on the ship's bridge and thus remote speed indication has been achieved.

Distance recording is achieved by using a constant speed motor (10) which drives the distance counter (11), via friction gearing. The constant speed motor has been used in order that a distance indication may be produced that is independent of the non-linear characteristic of the system. The motor is started by contact (5) as previously described. The main shaft (7), whose angle of rotation is directly proportional to the speed of the ship, is fitted with a screw spindle (12). The rotation of the shaft causes a lateral displacement of the friction wheel (13). At zero speed, the friction wheel rests against the apex of the distance cone (14), whilst at maximum speed the wheel has been displaced

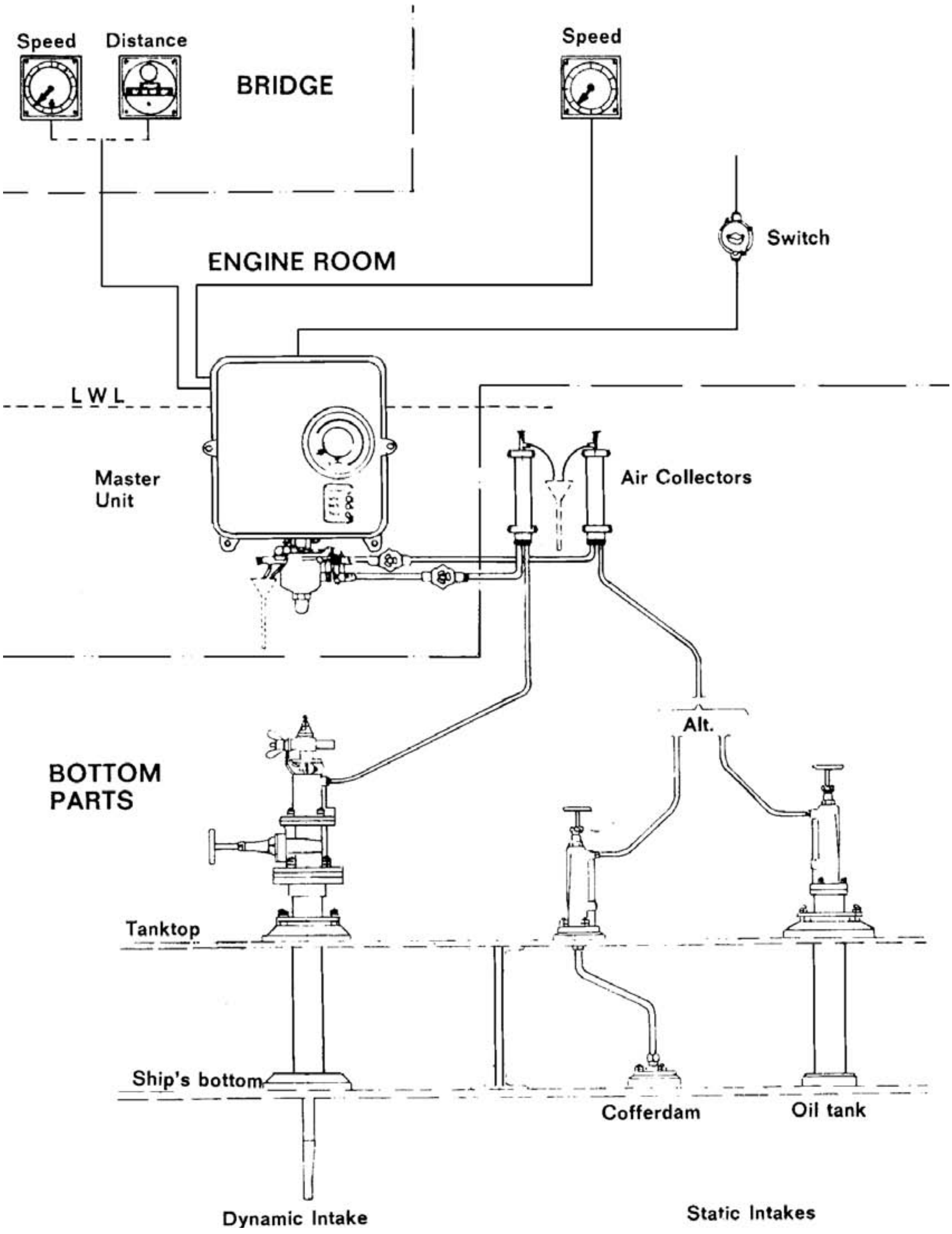
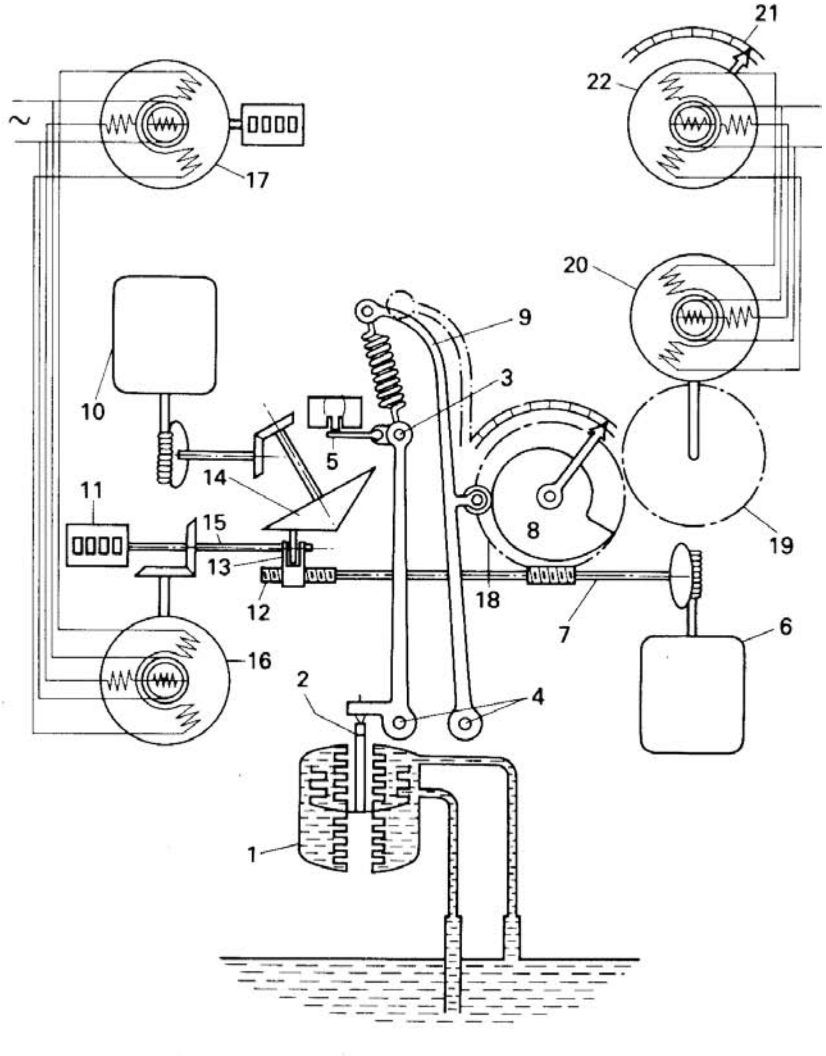


**Figure 3.3** A shipboard installation. (Reproduced courtesy of SAL Jungner Marine.)



1. pressure chamber
2. pressure rod
3. lever
4. pivot
5. electric start contact
6. reversible motor
7. main shaft
8. spiral cam
9. lever
10. constant speed motor
11. distance counter
12. screw spindle
13. friction wheel
14. distance cone
15. distance shaft
16. servo transmission system
17. servo transmission system
18. gear wheels
19. gear wheels
20. speed servo transmitter
21. remote speed indicator
22. servo receiver

**Figure 3.4** The mechanical speed translating system of the SAL 24 pressure tube log. (Reproduced courtesy of SAL Jungner Marine.)

along the cone to the rim. The distance indicator (11) is driven from the constant speed motor (10) via the cone. The nearer to the rim of the cone the friction wheel rides, the greater will be the distance indication. Revolutions of the distance shaft (15) are transmitted to the remote distance indicator via the servo transmission system (16 and 17).

### *Operation of the SAL 24E*

The SAL 24E utilizes the same system of tubes, pressure tank and diaphragm to convert pressure variations due to speed, to electrical pulses suitable to drive the electronic circuits that replace much of the mechanical arrangement of the SAL 24 log. The distance integration mechanism with servo, cone and counter has been fully replaced with electronic circuitry.

As previously described, when the vessel moves forwards, the dynamic pressure acting on the underside of the diaphragm causes it to move upwards forcing the pushrod upwards. As shown in Figure 3.5, this causes the pushrod arm assembly to move to the right on the pivot, increasing the

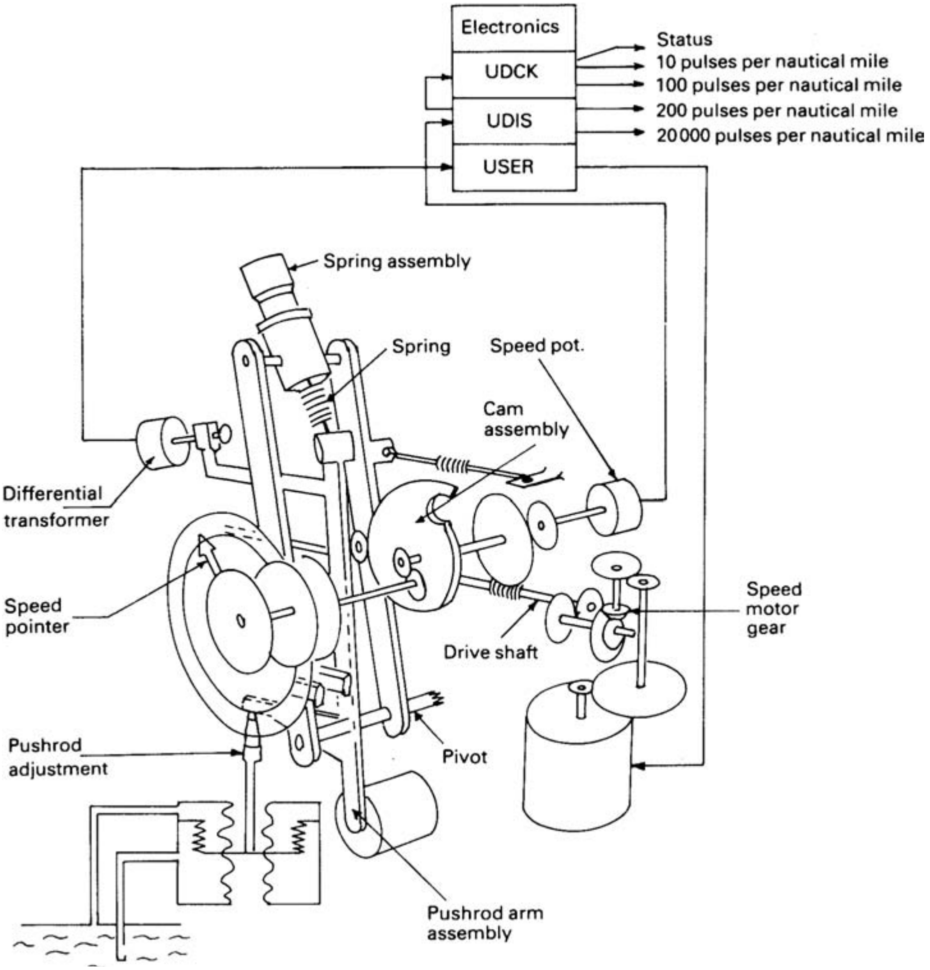


**Figure 3.5** Pressure/mechanical assembly of the SAL 24E electronic pressure speed log. (Reproduced courtesy of SAL Jungner Marine.)

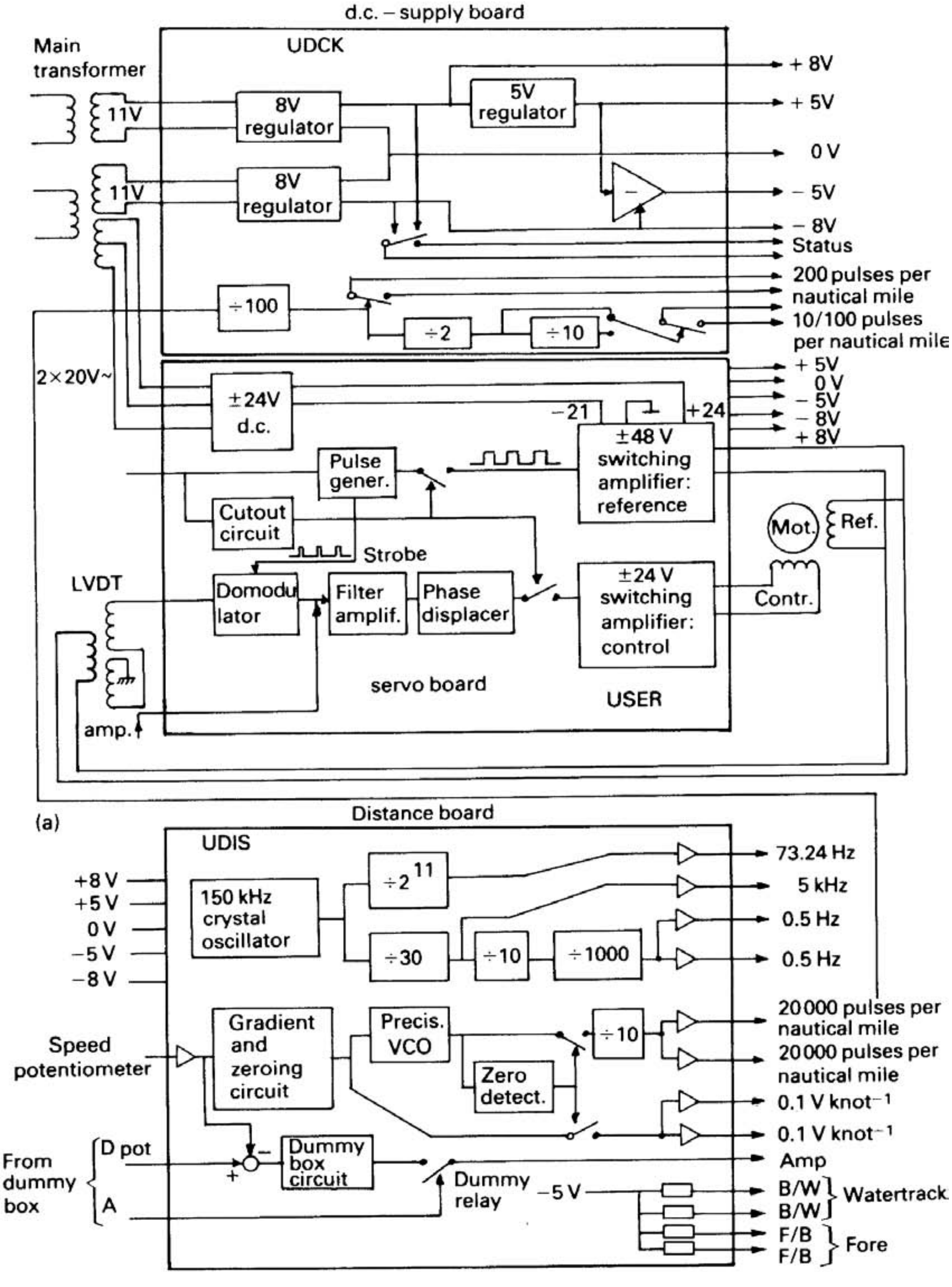


**Figure 3.6** The electronics unit. (Reproduced courtesy of SAL Jungner Marine.)

tension on the spring assembly and producing an output from the differential transformer. This output is applied to the USER board, shown in Figure 3.6, where it is processed to provide the drive for the speed servo-control winding via a ± 24 V switching amplifier. The servo now turns and rotates the cam assembly via gearing and the drive shaft. An increase in speed is now shown on the speed pointer. As the cam rotates it forces the balance arm to the left and tightens the spring until the pushrod arm and the diaphragm bellows are balanced. The cam is carefully designed so that the spring force is proportional to the square of the rotation angle and thus the non-linearity of the pressure system is counteracted. The speed potentiometer turns together with the speed pointer to provide an input to the UDIS board. This input produces a variety of outputs enabling the system to be interfaced with other electronic equipment.

The accuracy of the Pitot type speed log when correctly installed and calibrated is typically better than 0.75% of the range in use.

## 3.3 Speed measurement using electromagnetic induction

Electromagnetic speed logs continue to be popular for measuring the movement of a vessel through water. This type of log uses Michael Faraday's well-documented principle of measuring the flow of a fluid past a sensor by means of electromagnetic induction.

The operation relies upon the principle that any conductor which is moved across a magnetic field will have induced into it a small electromotive force (e.m.f.). Alternatively, the e.m.f. will also be induced if the conductor remains stationary and the magnetic field is moved with respect to it. Assuming that the magnetic field remains constant, the amplitude of the induced e.m.f. will be directly proportional to the speed of movement.

In a practical installation, a constant e.m.f. is developed in a conductor (seawater flowing past the sensor) and a minute current, proportional to the relative velocity, is induced in a collector. The magnetic field created in the seawater is produced by a solenoid which may extend into the water or be fitted flush with the hull. As the vessel moves, the seawater (the conductor) flowing through the magnetic field has a small e.m.f. induced into it. This minute e.m.f., the amplitude of which is dependent upon the rate of cutting the magnetic lines of force, is detected by two small electrodes set into the outer casing of the sensor.

Figure 3.7 shows a solenoid generating a magnetic field and a conductor connected in the form of a loop able to move at right angles to the field. If the conductor is moved in the direction shown, a tiny current will be induced in the wire and a small e.m.f. is produced across it. In the case of an electromagnetic speed log, the conductor is seawater passing through the magnetic field. Fleming's right-hand rule shows that the generated e.m.f. is at right angles to the magnetic field ($H$). Induced current flowing in the conductor produces an indication of the e.m.f. on the meter. If we assume that the energizing current for the solenoid is d.c. the induced e.m.f. is $\beta lv$, where $\beta$ = the induced magnetic field, $l$ = the length of the conductor, and $v$ = the velocity of the conductor.

$\beta$ is approximately equal to $H$, the magnetic field strength. Therefore, e.m.f. = $Hlv$ assuming no circuit losses.

To reduce the effects of electrolysis and make amplification of the induced e.m.f. simpler, a.c. is used to generate the magnetic field. The magnetic field strength $H$ now becomes $H\mathrm{m}\sin\omega t$ and the induced e.m.f. is: $H\mathrm{m}lv\sin\omega t$. If the strength of the magnetic field and the length of the conductor both remain constant then, e.m.f. $\simeq$ velocity.

Figure 3.8 illustrates that the changes of e.m.f., brought about by changes in velocity, produce a linear graph and thus a linear indication of the vessel's speed. The e.m.f. thus produced is very small

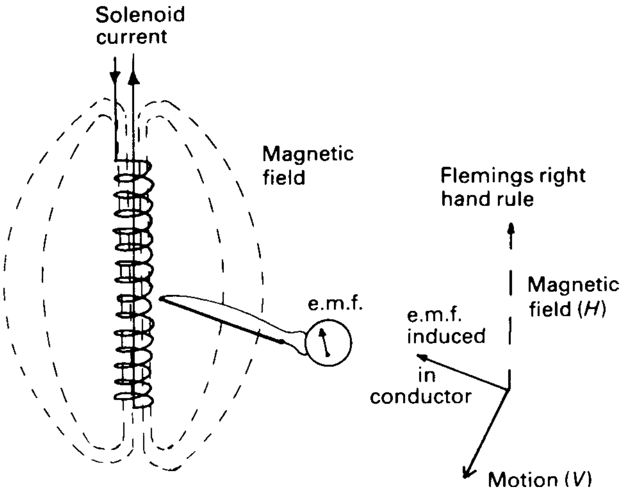


**Figure 3.7** Effect of moving a conductor through a magnetic field.

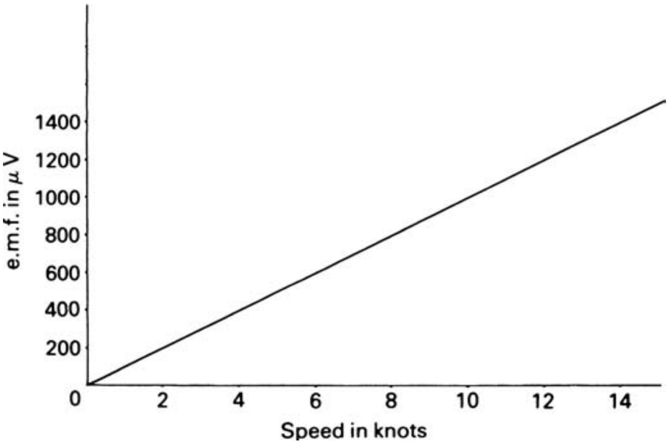


**Figure 3.8** Relationship between the vessel's speed and the output from the sensors.

but, if required, may be made larger by increasing the energizing current, or the number of turns of wire on the solenoid.

The following points should be noted.

- The a.c. supply to the solenoid produces inductive pick-up between the coil and the wires that carry the signal. This in turn produces a 'zero' error that must be compensated for by 'backing off' the zero setting of the indicator on calibration.
- The induced e.m.f. is very small (for reasonable amplitudes of energizing current), typically 100 μV per knot.
- The induced e.m.f. and hence the speed indication will vary with the conductivity of the water.
- The device measures the speed of the water flowing past the hull of the ship. This flow can vary due to the non-linearity of a hull design.

- Ocean currents may introduce errors.
- Pitching and rolling will affect the relationship between the water speed and the hull. Error due to this effect may be compensated for by reducing the sensitivity of the receiver. This is achieved using a CR timing circuit with a long time constant to damp out the oscillatory effect.
- Accuracy is typically 0.1% of the range in use, in a fore and aft direction, and approximately 2% thwartships.

Figure 3.9 shows a typical sensor cutaway revealing the solenoid and the pick-up electrodes. A speed translating system is illustrated in Figure 3.10.

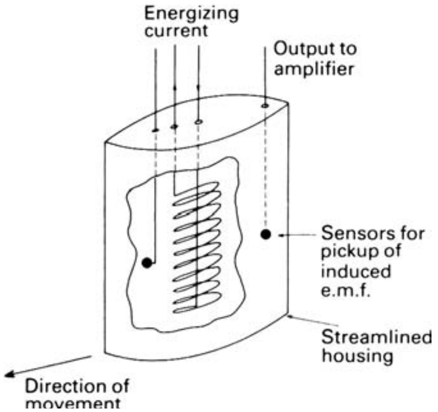


**Figure 3.9** Constructional details of an electromagnetic log sensor.

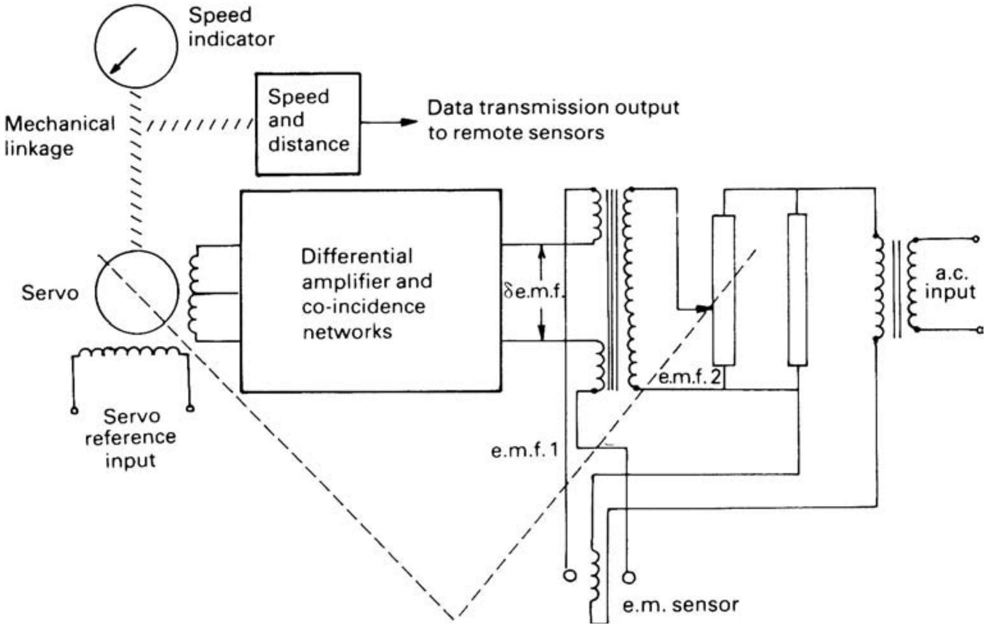


**Figure 3.10** An e.m. speed log translating system.

### *Description of the speed translating system*

The small signal speed voltage from the sensor, e.m.f.1, is applied to a differential transformer where it is compared to a reference voltage, e.m.f.2, produced from a potentiometer across the input a.c. supply. The potential difference produced across the reference resistor provides the energizing current for the solenoid in the sensor.

If the signal voltage e.m.f.1. differs from the reference voltage e.m.f.2. an error signal voltage δ e.m.f. is produced. This error voltage is applied to the speed signal amplifier where it is amplified to produce sufficient power to drive the servo motor. The servo will in turn produce a speed reading, via a mechanical linkage, on the indicator. Also coupled to the servo shaft is the slider of the speed potentiometer that turns in the direction to reduce the error voltage δ e.m.f. When this error voltage drops to zero the servo ceases to turn. The speed indicator is stationary until the next error voltage δ e.m.f. is produced. Each time an error voltage is created the servo turns to cancel the error and thus balances the system.

## 3.3.1 A practical electromagnetic speed logging system

The potential developed across the transducer electrodes is proportional to magnetic field strength (and consequently the energizing current) and the flow velocity in the volume of water influenced by the field. The magnetic field strength is in no way stabilized against any changes in the ship's main voltage, temperature, etc, but by effectively comparing the energizing current with the voltage at the electrodes, their ratio provides a measure of the ship's speed.

The input transformer T1 (shown in Figure 3.11) possesses a very high inductance and a step-down ratio of 5:1. This results in an input impedance, as seen by the pick-up electrodes, approaching 20 MΩ which when compared with the impedance presented by salt water can be considered an open circuit. Hence changes in salinity have no effect on the measured voltage and the resulting speed indication. A switched resistor chain (R1/R5) sets the gain of the overall amplifier in conjunction with resistor chain (R6/R10) which controls the amplitude of the feedback signal.

The output of IC1 is coupled, via IC2, which because of capacitive feedback (not shown), ensures that the circuit has a zero phase shift from T1 through T2, to the demodulator. Demodulation is carried out by TR1/TR2 that are switched in turn from an a.c. reference voltage derived from a toroidal transformer monitoring the energizing current of the transducer. By driving TR1/TR2 synchronously, the phase relationship of the voltage detected by the electrodes determines the polarity of the demodulated signal. 0° and 180° phasing produce a positive or negative component; 90° and 270° produce no output and hence a complete rejection of such phase-quadrature signals. The demodulated signal is applied to the Miller Integrator IC3 which in turn drives the current generator. Speed repeaters are current-driven from this source.

### *Operation of the loop*

With no vessel movement, there will be a zero signal at the input to IC1 and consequently there will be no signal at the multiplier chip input. No feedback signal is developed at the input to IC1. As the vessel moves ahead, the small signal applied to IC1 is processed in the electronic unit to produce a current flow through the speed repeaters and the multiplier. There now exists an output from the multiplier, proportional to the speed repeater current and the reference voltage produced by the toroidal transformer monitoring the transducer energizing current. The a.c. from the multiplier is fed back to IC1 in series with, and 180° out of phase with, the small signal secondary of T1. This a.c. signal rises slowly and eventually, with the time constant of the demodulator, is equal to the signal p.d.

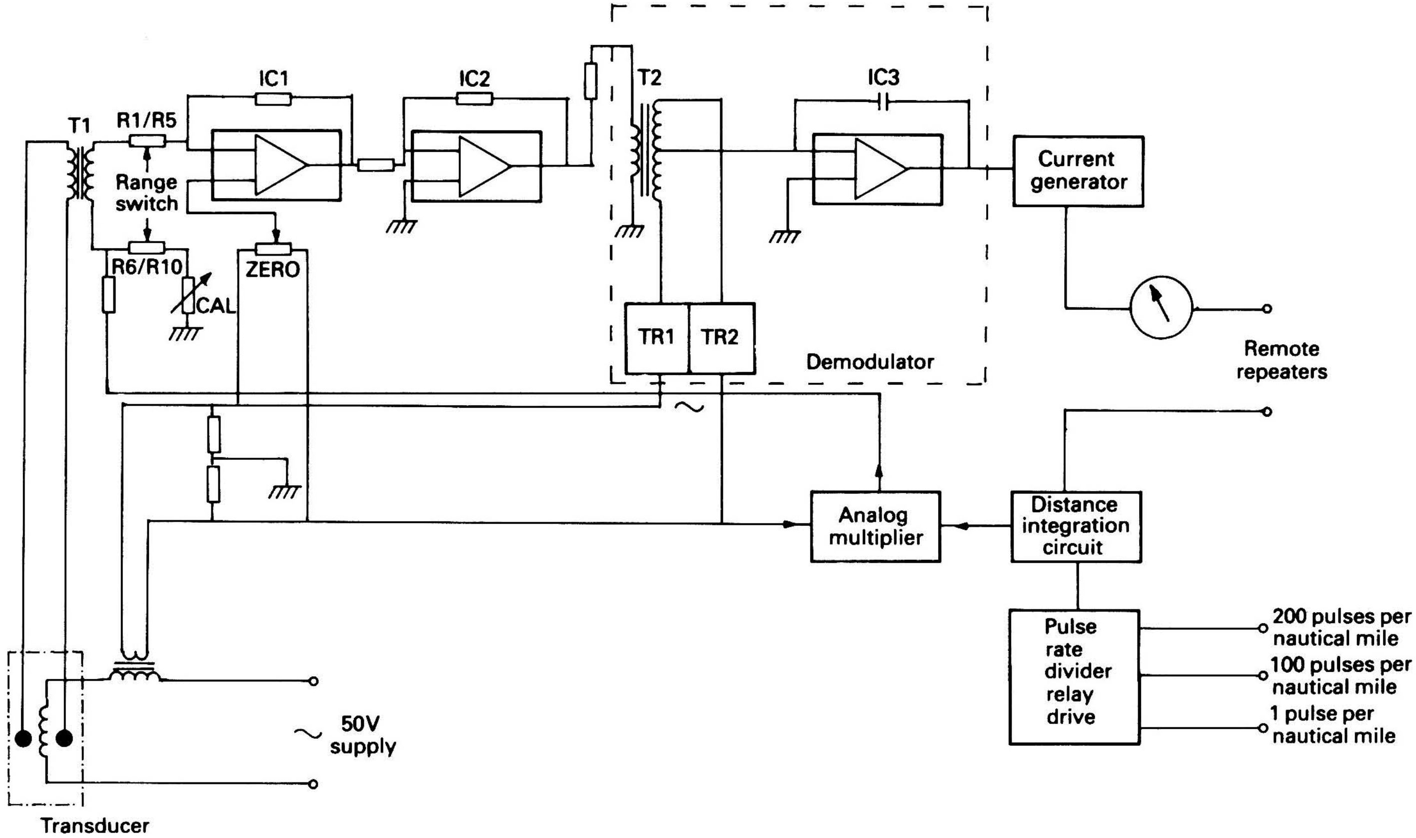


**Figure 3.11** Simplified diagram of an e.m. log. (Reproduced courtesy of Thomas Walker and Son Ltd.)

developed across T1. At this time the resultant signal applied to IC1 falls to zero and therefore the demodulator output remains at a constant figure. Any further change in speed results in an imbalance in the secondary of T1 producing a resultant a.c. signal to IC1. As a result, the demodulator output increases or decreases (faster or slower ship's speed) until the balance condition is restored. The speed repeaters will indicate the appropriate change of speed.

### *Distance integration*

The speed current is passed through a resistive network on the distance integration board, in order that a proportional voltage may be produced for integration. The output of this board is a pulse train, the rate of which is proportional to the indicated speed. The 10 ms pulses are coupled to the relay drive board which holds the necessary logic to give the following outputs: 200 pulses per nautical mile, 100 pulses per nautical mile, and 1 pulse per nautical mile.

## 3.4 Speed measurement using acoustic correlation techniques

Unlike the previously described speed log, which measure the vessel's speed with respect to water only, the SAL-ICCOR log measures the speed with respect to the seabed or to a suspended water mass. The log derives the vessel's speed by the use of signal acoustic correlation. Simply, this is a way of combining the properties of sonic waves in seawater with a correlation technique. Speed measurement is achieved by bottom-tracking to a maximum depth of 200 m. If the bottom echo becomes weak or

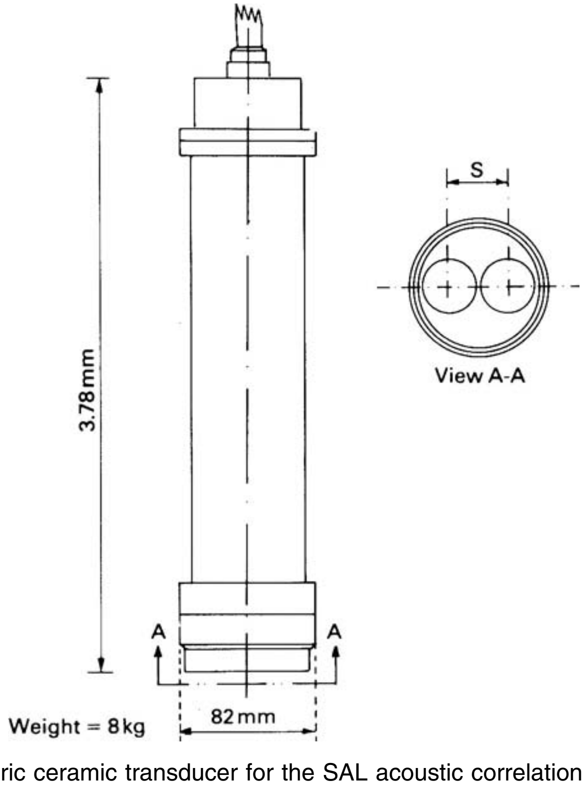


**Figure 3.12** Piezoelectric ceramic transducer for the SAL acoustic correlation speed log.

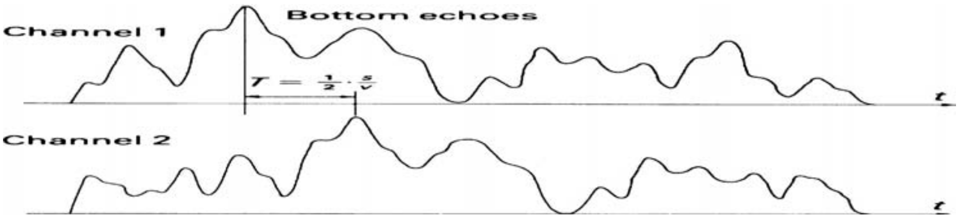


**Figure 3.13** Illustration of the time delay (*T*) between each channel echo signal.

the depth exceeds 200 m, the system automatically switches to water-mass tracking and will record the vessel's speed with respect to a water mass approximately 12 m below the keel.

The transducer transmits pulses of energy at a frequency of 150 kHz from two active piezoceramic elements that are arranged in the fore and aft line of the vessel (see Figure 3.12). Each element transmits in a wide lobe perpendicular to the seabed. As with an echo sounder, the transducer elements are switched to the receive mode after transmission has taken place.

The seabed, or water mass, reflected signals possess a time delay (*T*) dependent upon the contour of the seabed, as shown in Figure 3.13. Thus the received echo is, uniquely, a function of the instantaneous position of each sensor element plus the ship's speed. The echo signal, therefore, in one channel will be identical to that in the other channel, but will possess a time delay as shown.

The time delay (*T*), in seconds, can be presented as:

$$T = 0.5 \times s\ v$$

where $s$ = the distance between the receiving elements and $v$ = the ship's velocity.

In the SAL-ACCOR log (see Figure 3.14), the speed is accurately estimated by a correlation technique. The distance between the transducer elements ($s$) is precisely fixed, therefore when the time ($T$) has been determined, the speed of the vessel ($v$) can be accurately calculated.

It should be noted that the calculated time delay ($T$) is that between the two transducer echoes and not that between transmission and reception. Temperature and salinity, the variables of sound velocity in seawater, will not affect the calculation. Each variable has the same influence on each received echo channel. Consequently the variables will cancel.

It is also possible to use the time delay ($T$) between transmission and reception to calculate depth. In this case the depth ($d$),in metres, is:

$$d = \frac{T}{2} \times C$$

where $C$ = the velocity of sonic energy in seawater (1500 $\mathrm{ms}^{-1}$).

Dimensions of the transducer active elements are kept to a minimum by the use of a high frequency and a wide lobe angle. A wide lobe angle (beamwidth) is used because echo target discrimination is not important in the speed log operation and has the advantage that the vessel is unlikely to 'run away' from the returned echo.

### 3.4.1 System description

Initiating the sequence, the power amplifier produces the transmitted power, at the carrier frequency of 150 kHz, under the command of a pulse chain from the clock unit. Returned echoes are received by two

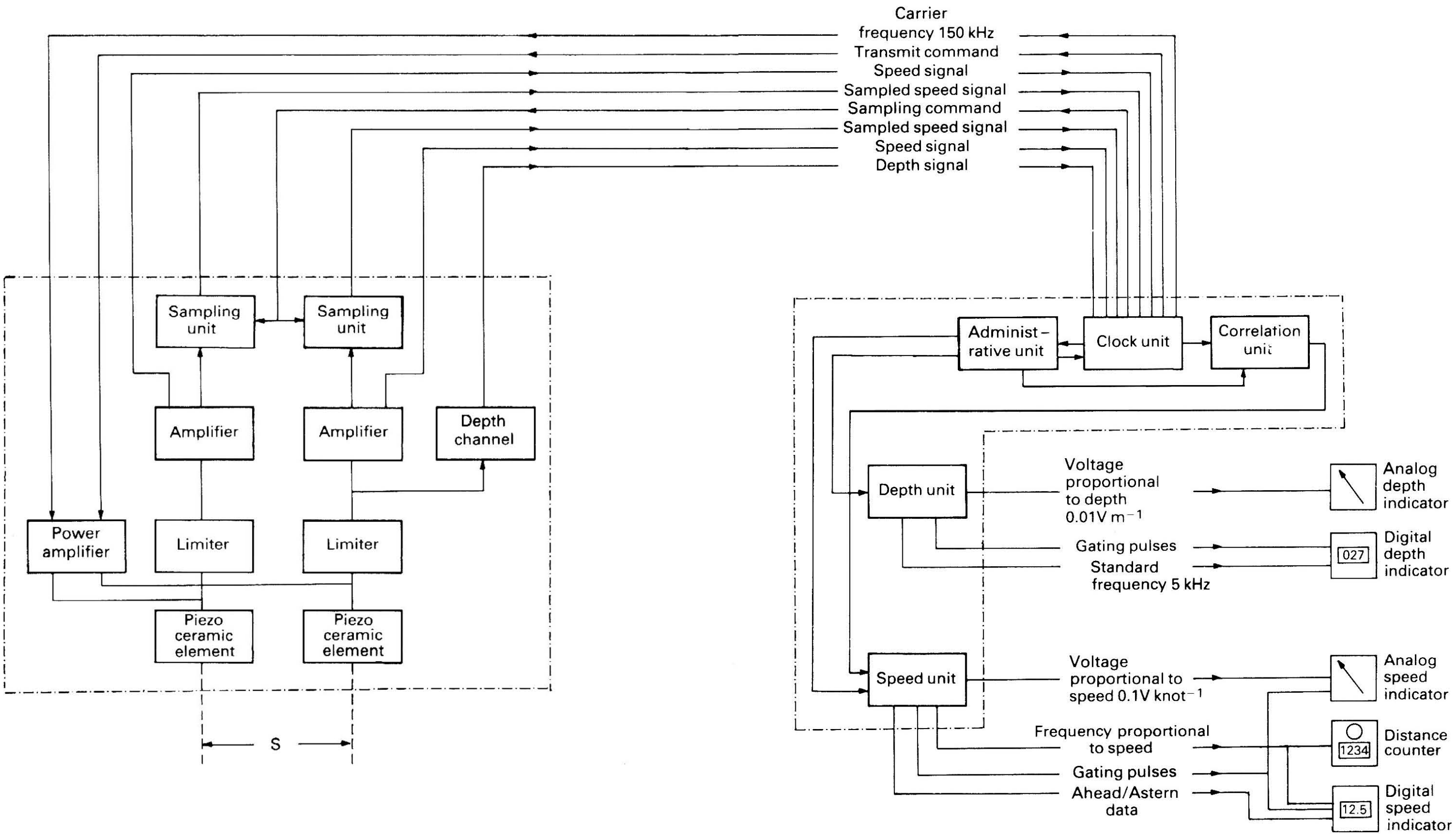


**Figure 3.14** System diagram of the SAL-ACCOR acoustic correlation speed log. (Reproduced courtesy of SAL Junger Marine.)

independent identical channels and are pre-amplified before being applied to sampling units. Each sampling unit effectively simplifies the echo signal to enable interconnection to be made between transducer and main unit without the risk of signal deterioration. As with other functions, sampling is commanded by a clock unit, which also provides a highly stable 150 kHz for the carrier frequency. This frequency is also used as a standard frequency for the other functions on the electronics board, where it is divided to produce the 5 kHz needed to operate some of the speed indicators.

As the name suggests, the administration block controls most of the electronic functions. This block initiates the transmit/receive cycle, determines whether the system selects B-track or W-track operation and supervises the speed and depth calculations. The unit is effectively a microprocessor operating to a pre-determined program. Actual speed calculation takes place in the correlation block. The process extracts the time delay by correlating the sampled output of each channel.

The speed unit provides the following outputs to drive both speed and distance counters.

- An analogue voltage, the gradient of which is 0.1 V/knot, to drive the potentiometer servo-type speed indicators.
- A pulse frequency proportional to speed. The frequency is 200/36 pulses/s/knot. Pulses are gated into the digital counter by a 1.8-s gate pulse.
- A positive/negative voltage level to set the ahead/astern indication or the B track/W track indication.
- 2000 pulses per nautical mile to drive the stepping motor in the digital distance indicator.

The depth unit provides the following outputs to drive the depth indicators when the echo sounding facility is used.

- An analogue voltage with a gradient of 0.01 $\text{Vm}^{-1}$, to drive the analogue depth indicator.
- Pulses of 2 $\text{ms m}^{-1}$, which are used to gate a 5 kHz standard frequency into the digital depth indicator.
- A positive/negative voltage level to cause the indicator to display 'normal operation' or 'over-range'.

When correctly installed and calibrated, a speed accuracy of ±0.1 knot is to be expected. Distance accuracy is quoted as 0.2%. The SAL-ICCOR speed log can be made to measure the vessel's transverse speed with the addition of a second transducer set at 90° to the first.

## 3.5 The Doppler Principle

In the early 19th century, Christian Doppler observed that the colour emitted by a star in relative movement across the sky appeared to change. Because light waves form part of the frequency spectrum, it was later concluded that the received wavelength must be changing and therefore the apparent received frequency must also change. This phenomenon is widely used in electronics for measuring velocity.

Figure 3.15(b) shows that the wavelength ($\lambda$) is compressed in time when received from a transmitter moving towards a receiver ($\lambda_1$) and expanded (Figure 3.15c) in time from a transmitter moving away ($\lambda_2$). Consider a transmitter radiating a frequency ($f_1$). The velocity of propagation of radiowaves in free space ($c$) is $300 \times 10^6\,\text{ms}^{-1}$ and in seawater it is much slower at approximately $1500\,\text{ms}^{-1}$. After a period of 1 s, one cycle of the transmitted acoustic wave in seawater will occupy a distance of 1500 m.

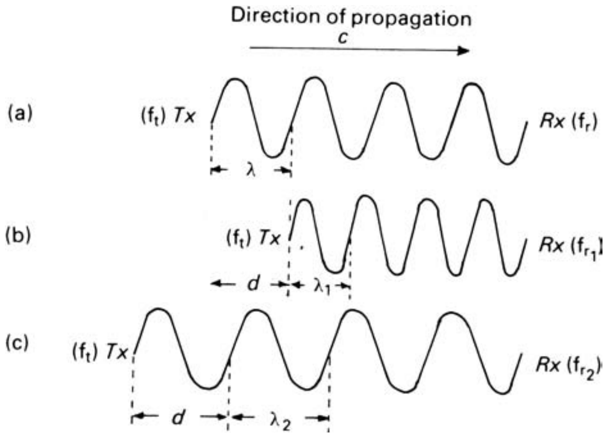


**Figure 3.15** Expansion and compression of wavelength.

If the transmitter moves towards an observer at speed ($v$) it will, at the end of 1 s, have travelled a distance ($d$) towards the receiver. Each transmitter wave has now been shortened because of the distance travelled by the transmitter towards the observer. By definition, a shorter wavelength defines a higher frequency ($f_r$). The shortened wavelength, or higher frequency, received is directly proportional to the speed of movement of the transmitter.

In Figure 3.15(b), the transmitter has moved towards an observer by a distance ($d$). This is the distance travelled during the time of generating one cycle ($T$).

$$T = \frac{1}{ft} \quad \text{and} \quad d = v \times T = \frac{v}{ft}$$

Therefore the apparent wavelength is

$$\lambda_1 = \lambda - \frac{v}{ft}$$

and the frequency is

$$fr_1 = \frac{c}{\lambda_1} = \frac{c}{\lambda - v/ft} = \frac{cft}{\lambda ft - v} = f_t \frac{c}{c - v}$$

For a moving transmitter that is approaching a receiver, the received frequency is apparently increased. The reverse is true of a transmission from a transmitter moving away from an observer, when the wavelength will be stretched and the frequency decreased.

$$\lambda_2 = \lambda + v/ft$$

$$f_r = f_t \frac{c}{c + v}$$

If an observer moves at velocity ($v$) towards a stationary sound source, the number of cycles reaching the receiver per second is increased, thus the apparent received frequency is increased. The received frequency is

$$f_r = f_r + v/\lambda$$

and

$$1/\lambda = f/c$$

therefore

$$f_t + f_v/c = f_t(1 + v/c) = f_t \frac{c + v}{c}$$

If the observer now moves away from the stationary transmitter the apparent received frequency is;

$$f_r = f_t \frac{c - v}{c}$$

If, as in the Doppler speed log, both the observer and the sound source (transmitter and receiver) are moving towards a reflecting surface, the received frequency is;

$$f_r = f_t \frac{c}{c - v} x \frac{c + v}{c} = f_t \frac{c + v}{c - v}$$

The Doppler frequency shift is

$$f_d = f_r - f_t \text{ (or } f_t - f_r)$$

$$f_d = f_t \, x \frac{c + v}{c - v} - f_t$$

$$= \frac{cf_t + vf_t - cf_t + vf_t}{c - v}$$

$$= \frac{2vf_t}{c - v}$$

The velocity of radio waves ($c$) is always far in excess of $v$ and therefore the expression above can be simplified to;

$$f_d = \frac{2vf_t}{c}$$

where $f_d$ = Doppler frequency shift in cycles per second, $v$ = relative speed in the direction of the transmitted wave, $f_t$ = transmitted frequency, and $c$ = velocity of propagation of the radio wave.

## 3.6 Principles of speed measurement using the Doppler effect

The phenomenon of Doppler frequency shift is often used to measure the speed of a moving object carrying a transmitter. Modern speed logs use this principle to measure the vessel's speed, with respect to the seabed, with an accuracy approaching 0.1%.

If a sonar beam is transmitted ahead of a vessel, the reflected energy wave will have suffered a frequency shift (see Figure 3.16), the amount of which depends upon:

- the transmitted frequency
- the velocity of the sonar energy wave
- the velocity of the transmitter (the ship).

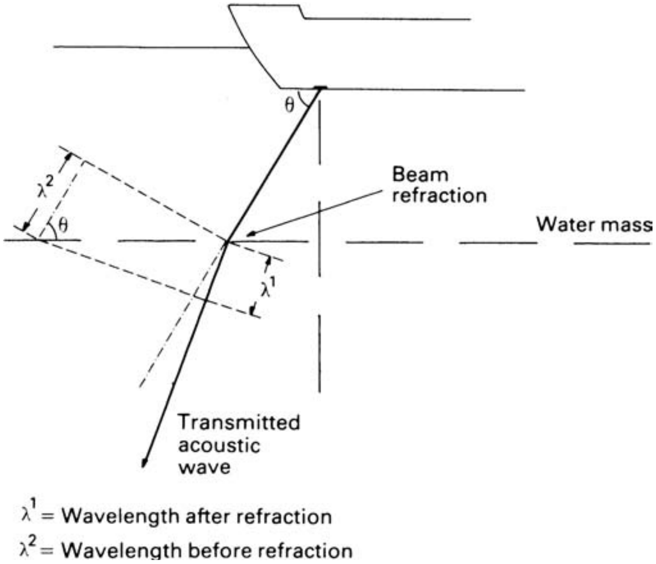


**Figure 3.16** Illustration of the change of wavelength that occurs when an acoustic wave crosses a water mass.

The frequency shift, in hertz, of the returned wave is:

$$fd = ft - fr$$

where $ft$ = the transmitted wave frequency, and $fr$ = the received wave frequency.

The Doppler shift formula, for a reflected wave, is given as:

$$fd = \frac{2vft}{c}$$

where $v$ = the velocity of the ship, and $c$ = the velocity of the sonar wave (1500 $\mathrm{ms}^{-1}$ in seawater).

Obviously there can be no objects directly ahead of a vessel from which the acoustic wave may be reflected. The wave is therefore transmitted towards the seabed, not vertically as with echo sounding, but ahead at an angle of 60° to the horizontal. This angle has been found to be the optimum angle of incidence with the seabed, which will reflect a signal of sufficient strength to be received by the transducer. The shape of the seabed has no effect on the frequency shift. Provided that the seabed is not perfectly smooth, some energy will be reflected.

The angle between the horizontal plane and the transmission must now be applied to the basic Doppler formula:

$$fd = \frac{2vft\cos\theta}{C} \text{ (in hertz)}$$

Figure 3.17(a) shows this angle. Using trigonometry, $\cos\theta$ = Adjacent/Hypotenuse. Therefore, Adjacent = $C\cos\theta$.

Given a propagation angle of 60°, $\cos\theta = 0.5$

$$fd = \frac{2vft\cos\theta}{C} = \frac{vft}{C}$$

It follows that if the angle changes, the speed calculated will be in error because the angle of propagation has been applied to the speed calculation formula in this way. If the vessel is not in correct trim (or pitching in heavy weather) the longitudinal parameters will change and the speed indicated will be in error. To counteract this effect to some extent, two acoustic beams are transmitted, one ahead and one astern. The transducer assembly used for this type of transmission is called a ‘Janus’ configuration after the Roman god who reputedly possessed two faces and was able to see into both the future and the past. Figure 3.17(b) shows the Janus assembly.

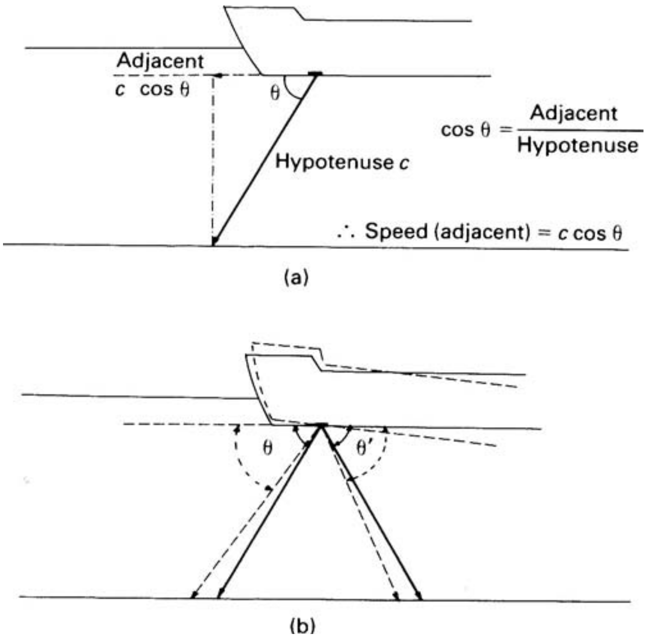


**Figure 3.17** (a) Derivation of longitudinal speed using trigonometry. (b) The effect of pitching on a Janus transducer configuration.

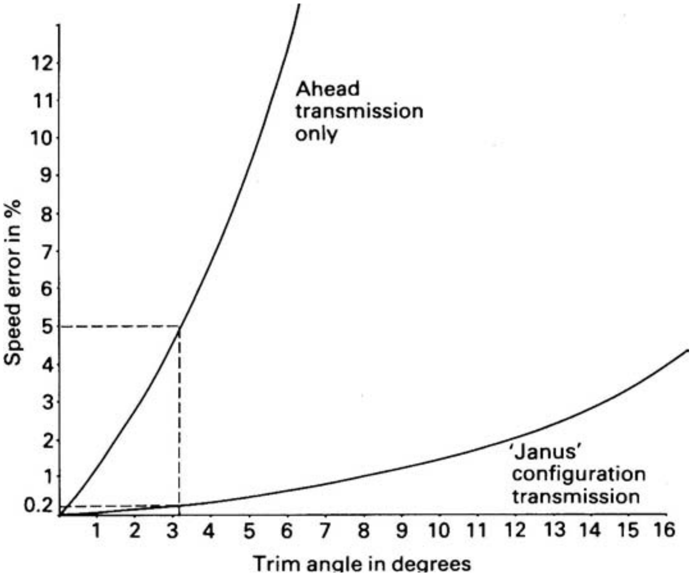


**Figure 3.18** Graphs of speed error caused by variations of the vessel's trim.

The Doppler frequency shift formula now becomes:

$$fd = \frac{2vft}{C}(+\cos\theta + \cos\theta')$$

(+ cos 60° + cos 60°′ = 1) therefore the transmission angle can effectively be ignored.

As Figure 3.17(b) shows, in heavy weather one angle increases as the other decreases effectively cancelling the effects of pitching on the speed indication.

Figure 3.18 shows the advantage of having a Janus configuration over a single transducer arrangement. It can be seen that a 3° change of trim on a vessel in a forward pointing Doppler system will produce a 5% velocity error. With a Janus configuration transducer system, the error is reduced to 0.2% but is not fully eliminated.

The addition of a second transducer assembly set at right angles to the first one, enables dual axis speed to be indicated (Figure 3.19).

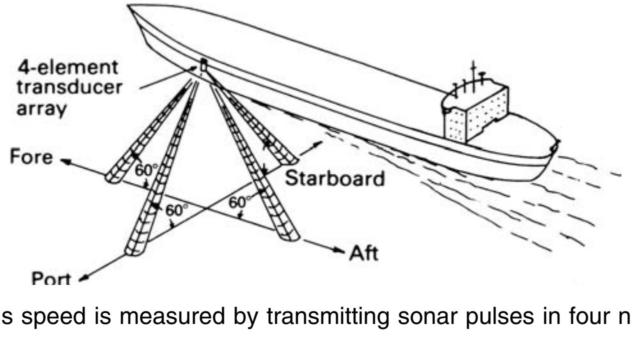


**Figure 3.19** Dual axis speed is measured by transmitting sonar pulses in four narrow beams towards the sea bed.

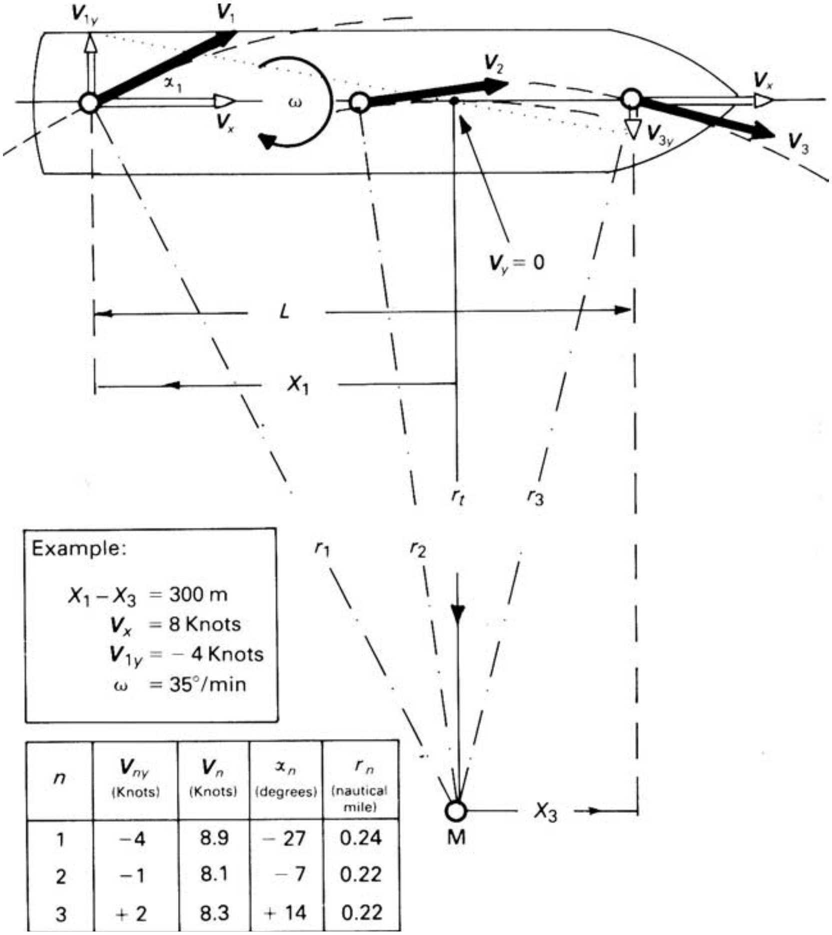
### 3.6.1 Vessel motion during turn manoeuvres

A precise indication of athwartships speed is particularly important on large vessels where the bow and stern sections may be drifting at different rates during docking or turning manoeuvres.

#### *Speed vectors during a starboard turn*

A dual axis Doppler speed log measures longitudinal and transverse speed, at the location of the transducers. If transducers are mounted in the bow and stern of a vessel, the rate of turn can be computed and displayed. This facility is obviously invaluable to the navigator during difficult manoeuvres.

Figure 3.20 shows the speed vectors plotted from bow and stern transducer data when a ship is turning to starboard without the effect of water current. When the rudder is put hard over, the transverse speed indication vector (**Vy**) can point either to the side to which the rudder has been moved or to the other side. This will depend upon the longitudinal speed, the angular speed (rate of turn) and

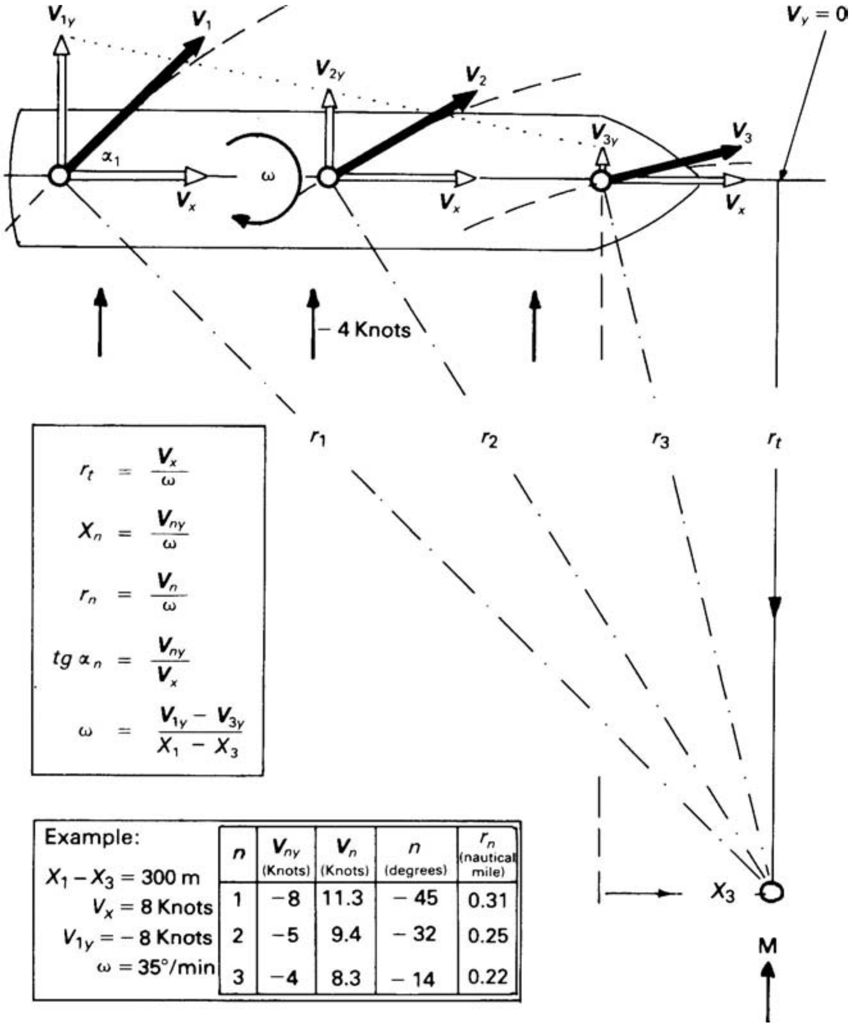


Example:

$X_1 - X_3 = 300$ m
$V_x = 8$ Knots
$V_{1y} = -4$ Knots
$\omega = 35°/\text{min}$

| $n$ | $V_{ny}$ (Knots) | $V_n$ (Knots) | $\alpha_n$ (degrees) | $r_n$ (nautical mile) |
|---|---|---|---|---|
| 1 | −4 | 8.9 | −27 | 0.24 |
| 2 | −1 | 8.1 | −7 | 0.22 |
| 3 | +2 | 8.3 | +14 | 0.22 |

**Figure 3.20** Speed vectors during a starboard turn with no current. (Reproduced courtesy of Krupp Atlas Elektronik.)



$r_t = \frac{V_x}{\omega}$

$X_n = \frac{V_{ny}}{\omega}$

$r_n = \frac{V_n}{\omega}$

$tg\,\alpha_n = \frac{V_{ny}}{V_x}$

$\omega = \frac{V_{1y} - V_{3y}}{X_1 - X_3}$

Example:

$X_1 - X_3 = 300$ m

$V_x = 8$ Knots

$V_{1y} = -8$ Knots

$\omega = 35°/\text{min}$

| $n$ | $V_{ny}$ (Knots) | $V_n$ (Knots) | $n$ (degrees) | $r_n$ (nautical mile) |
|---|---|---|---|---|
| 1 | −8 | 11.3 | − 45 | 0.31 |
| 2 | −5 | 9.4 | − 32 | 0.25 |
| 3 | −4 | 8.3 | − 14 | 0.22 |

**Figure 3.21** Speed vectors for a starboard turn under the influence of a four knot current . (Reproduced courtesy of Krupp Atlas Elektronik.)

weather/tide conditions. If the longitudinal speed and transverse speeds at two points of the vessel are known, the ship's movement is completely determinable. The bow transverse speed vector ($\mathbf{V}_{3y}$) points to starboard, the direction of the ship's turning circle.

Under the influence of the 4-knot current, shown in Figure 3.21, however, $\mathbf{V}_{3y}$ points to port. The transverse speed development along the ship's length is represented by a dotted line (between $\mathbf{V}_{1y}$ and $\mathbf{V}_{3y}$). The intersection of this line with the longitudinal axis produces a point at which the ship has longitudinal speed but no transverse speed. This point ($\mathbf{V}_y = 0$) is normally positioned, approximately, in the fore third of the vessel (see Figure 3.20) if the ship is to turn along a circle about point M (the instantaneous centre of rotation). The effect of current from the starboard side causes point $V_y = 0$ to be ahead of the vessel and the ship to turn around point M in Figure 3.21, which is shifted forward relative to that shown in Figure 3.20. It is obvious therefore that an accurate indication of transverse speeds at various points along the vessel enables the navigator to predict the movement of his ship.

### *Speed components with the rudder amidships*

Dual axis Doppler logs are able to measure accurately the ship's speed in a longitudinal direction ($\mathbf{V}_x$) and a transverse direction ($\mathbf{V}_y$). The data derived from these measurements enables the navigator to predict the course to steer in order to optimize the performance of the vessel. By measuring both speed components (i.e. the velocity vector) it is possible to optimize the vessel's course by computing the drift angle:

$$\propto = \text{arc tan}\,\frac{V_y}{V_x}$$

In the water-tracking mode this is the leeward angle (caused by wind) which is the angle between the true course (heading) and the course-made-good (CMG) through the water. In the bottom-tracking mode, it is the angle due to wind and tidestream between the heading and the CMG over the ground. With the help of a two-component log the ship can be navigated so that heading steered plus drift angle measured by the log, results exactly in the intended chart course (see Figure 3.22).

The transverse speed at the stern is computed from the transverse speed of the bow, the ship's rate of turn and the ship's length as follows:

$$V_{q2} = V_{q1} - \omega L$$

where $V_{q2}$ = stern transverse speed, $V_{q1}$ = bow transverse speed, $\omega$ = rate of turn (angular velocity), and $L$ = distance between bow and stern points of measurement.

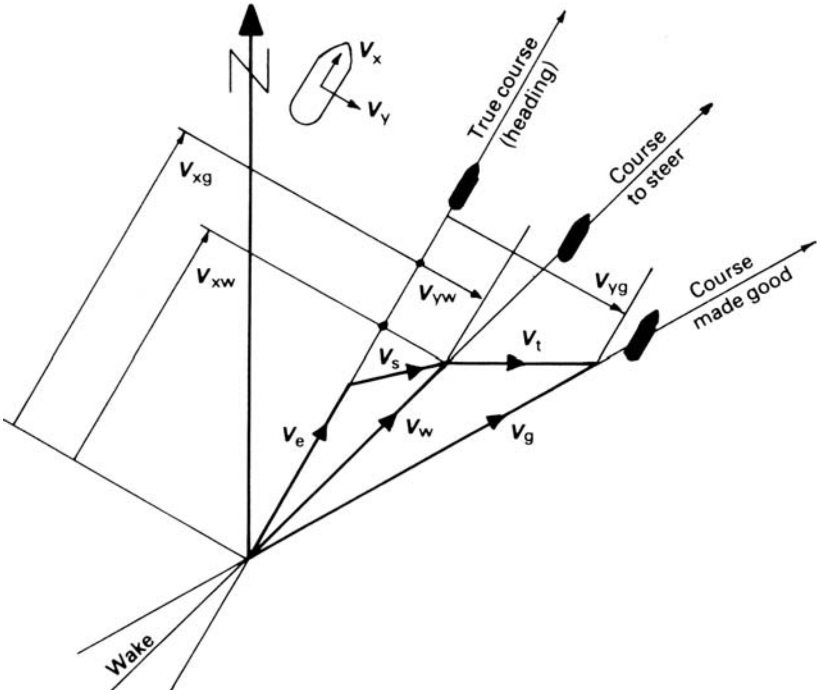


**Figure 3.22** External environmental effects of a vessel's track. (Reproduced courtesy of Krupp Atlas Elektronik.)

### 3.6.2 Choice of frequency/transducer

As with depth sounding, the size of the transducer can be kept within reasonable limits by using a high frequency. This is particularly important in the situation where many elements are to be mounted in the same assembly. Unfortunately, as has already been discussed, attenuation losses increase exponentially with the transmission frequency. The choice of frequency is therefore a compromise between acceptable transducer size and the power requirements of the acoustic wave in order to overcome the signal losses due to the transmission media. Frequencies used in speed logging systems vary widely and are usually in the range 100 kHz to 1 MHz.

The factor with the greatest effect on speed accuracy is the velocity of the acoustic wave in seawater. Propagation velocity is affected by both the salinity and the temperature of the seawater through which the wave travels. However, velocity error due to these two factors can be virtually eliminated by mounting salinity and temperature sensors in the transducer array. Data from both sensors are processed to provide corrective information for the system. Alternatively, the Krupp Atlas Alpha transducer system effectively counteracts the effects of salinity and temperature by the use of a phased beam.

#### *ALPHA transducer array*

The necessity of a tilted beam normally dictates that the transducer protrudes below the keel and therefore may suffer damage. It is possible to produce the required angle of propagation by the use of

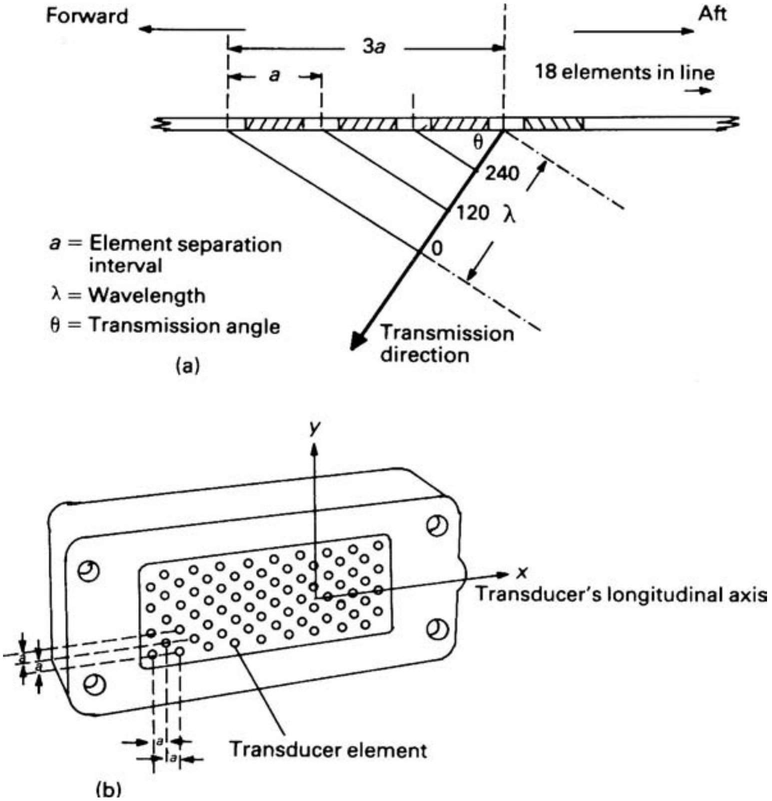


**Figure 3.23** (a) Principle of the alpha transducer array. (b) A 72-element alpha transducer array.

a number of flush fitting transducers. The Krupp Atlas Alpha (Atlas Low Frequency Phased Array) multiple transducer 'Janus' assembly uses (4 × 18 = 72) flush fitting elements in each of the fore and aft positions. In theory any number of elements may be used, but the spacing of the elements must not exceed certain limits in order to keep unwanted side lobes down to an acceptable level.

Figure 3.23(a) is a cut-away bow section of a vessel fitted with an Alpha transducer array. For clarity, only a three-element assembly is shown. If the three elements are fed with in-phase signal voltages the beam formed would be perpendicular. However, if the signal voltages to each element are phase delayed, in this case by 120°, the main lobe is propagated at an angle (which under these conditions is about 50°). In this case the elements are fed with three sine waves each shifted clockwise by 120°. For the Janus configuration the same elements are fed alternately clockwise and counter clockwise. The Alpha system also overcomes the external factors that influence the velocity of acoustic waves in salt water and is thus able to counteract the unwanted effects of salinity and temperature change.

The standard Doppler formula, from which velocity is calculated, comprises a number of parameters, two of which are variable. Ideally the vessel's speed ($v$) should be the only unknown factor in the formula, but unfortunately the velocity of acoustic waves ($C$) is also a variable. Since speed accuracy depends upon the accuracy of acoustic wave velocity in salt water it is advantageous to eliminate ($C$) from the formula.

$$fd = \frac{2vft}{C}\cos\theta$$

With the Alpha system, the angle of propagation (θ) is a function of the velocity of acoustic waves because of the geometry and mode of activating the multiple elements (see Figure 3.23(a)). The angle of propagation is:

$$\cos\theta = \frac{\lambda}{3a} = \frac{\mathrm{C}}{3a\,ft}$$

where $a$ = the transducer element spacing and is therefore a fixed parameter.

$$\lambda = \frac{\mathrm{C}}{ft} = \text{one acoustic wavelength in salt water}$$

If the two earlier equations in this section are now combined, the Doppler frequency shift is:

$$fd = \frac{2v}{3a}$$

$3a$ is a fixed parameter and therefore $v$ is now the only variable. Two modes of operation are possible.

### 3.6.3 Choice of transmission mode

#### *Continuous wave mode (CW) transmission*

Two transducers are used in each of the Janus positions. A continuous wave of acoustic energy is transmitted by one element and received by the second element. Received energy will have been reflected either from the seabed, or, if the depth exceeds a predetermined figure (20 m is typical), from a water mass below the keel. Problems can arise with CW operation particularly in deep water when the transmitted beam is caused to scatter by an increasing number of particles in the water. Energy due to scattering will be returned to the transducer in addition to the energy returned from the water mass. The receiver is likely to become confused as the returned energy from the water mass becomes weaker due to the increasing effects of scattering. The speed indication is now very erratic and may fall to zero. CW systems are rarely used for this reason.

### *Pulse mode operation*

To overcome the problems of the CW system, a pulse mode operation is used. This is virtually identical to that described previously for depth sounding where a high energy pulse is transmitted with the receiver off. The returned acoustic energy is received by the same transducer element that has been switched to the receive mode. In addition to overcoming the signal loss problem, caused by scattering in the CW system, the pulse mode system has the big advantage that only half the number of transducers is required.

### *Comparison of the pulse and the CW systems*

- Pulse systems are able to operate in the ground reference mode at depths up to 300 m (depending upon the carrier frequency used) and in the water track mode in any depth of water, whereas the CW systems are limited to depths of less than 60 m. However, CW systems are superior in very shallow water, where the pulse system is limited by the pulse repetition frequency (PRF) of the operating cycle.
- The pulse system requires only one transducer (two for the Janus configuration) whereas separate elements are needed for CW operation.
- CW systems are limited by noise due to air bubbles from the vessel's own propeller, particularly when going astern.
- Pulse system accuracy, although slightly inferior to the CW system, is constant for all operating depths of water, whereas the accuracy of the CW system is better in shallow water but rapidly reduces as depth increases.

## 3.6.4 Environmental factors affecting the accuracy of speed logs

Unfortunately environmental factors can introduce errors and/or produce sporadic indications in any system that relies for its operation on the transmission and reception of acoustic waves in salt water.

- *Water clarity.* In exceptional cases the purity of the seawater may lead to insufficient scattering of the acoustic energy and prevent an adequate signal return. It is not likely to be a significant factor because most seawater holds the suspended particles and micro-organisms that adequately scatter an acoustic beam.
- *Aeration.* Aerated water bubbles beneath the transducer face may reflect acoustic energy of sufficient strength to be interpreted erroneously as sea bottom returns producing inaccurate depth indications and reduced speed accuracy. Proper siting of the transducer, away from bow thrusters, for instance, will reduce this error factor.
- *Vessel trim and list.* A change in the vessel's trim from the calibrated normal will affect fore/aft speed indication and an excessive list will affect athwartship speed. A Janus configuration transducer reduces this error.

- *Ocean current profile.* This effect is prevalent in areas with strong tides or ocean currents. In the water track mode, a speed log measures velocity relative to multiple thermocline layers several feet down in the water. If these layers are moving in opposite directions to the surface water, an error may be introduced.
- *Ocean eddy currents.* Whilst most ocean currents produce eddies their effect is minimal. This problem is more likely to be found in restricted waters with big tidal changes or in river mouths.
- *Sea state.* Following seas may result in a change in the speed indication in the fore/aft and/or port/starboard line depending upon the vector sum of the approaching sea relative to the ship's axis.
- *Temperature profile.* The temperature of the seawater affects the velocity of the propagated acoustic wave (see Figure 2.2 in Chapter 2). Temperature sensors are included in the transducer to produce corrective data that is interfaced with the electronics unit.

## 3.7 Doppler speed logging systems

There are many maritime Doppler speed logging systems available, ranging from simple and inexpensive units designed for the leisure market to the complex rugged units fitted on modern merchant vessels, and they all rely for the operation on the first principles described in this chapter. The difference between the cheaper leisure Doppler logs and those designed for a more demanding environment, lies in their construction, their reliability under pressure, the facilities they offer, and the fact that they are type-approved for use on ocean trading vessels.

### 3.7.1 The Sperry SRD-500 Dual Axis Doppler Speed Log System

One of the traditional manufacturers Sperry Marine Inc., now a part of Litton Marine Systems, produces and markets a dual axis log, the SRD-500 (see Figure 3.24), which is a fine example of how microtechnology may be used in signal processing and data presentation in a modern Doppler speed log.

The SRD 500 is a dual axis speed log capable of indicating speed along the fore and aft axis, in the range -20 to 50 knots, and along the port and starboard axis, range ± 10 knots, using a four-element electrostrictive transducer assembly. Additional electronic signal processing circuitry enables an echo sounding function, when bottom tracking speed mode is available.

The transmission frequency is 307.2 kHz and the radiated power is 15–20 W.

#### *Display unit*

The SRD-500 main navigation display is shown in Figure 3.25. Whilst many of the unit controls will be familiar to navigators, a few are listed below to show how the company has used electronic processing to get the most out of the system. For ease of viewing, the equipment uses illuminated liquid crystal displays.

- *Speed display.* This shows the vessel's fore/aft speed in knots, m/s (metres per second) or ft/s (feet per second). The speed range is 0–20 knots astern and 0–50 knots ahead. Bottom-tracking speed accuracy is ± 0.1 knots below 2 knots speed and ± 0.05 knots above 2 knots. Accuracy when water tracking is ± 0.1 knots below 10 knots, ± 0.025 knots in the range 10–25 knots, and ± 0.1 knot above 25 knots.

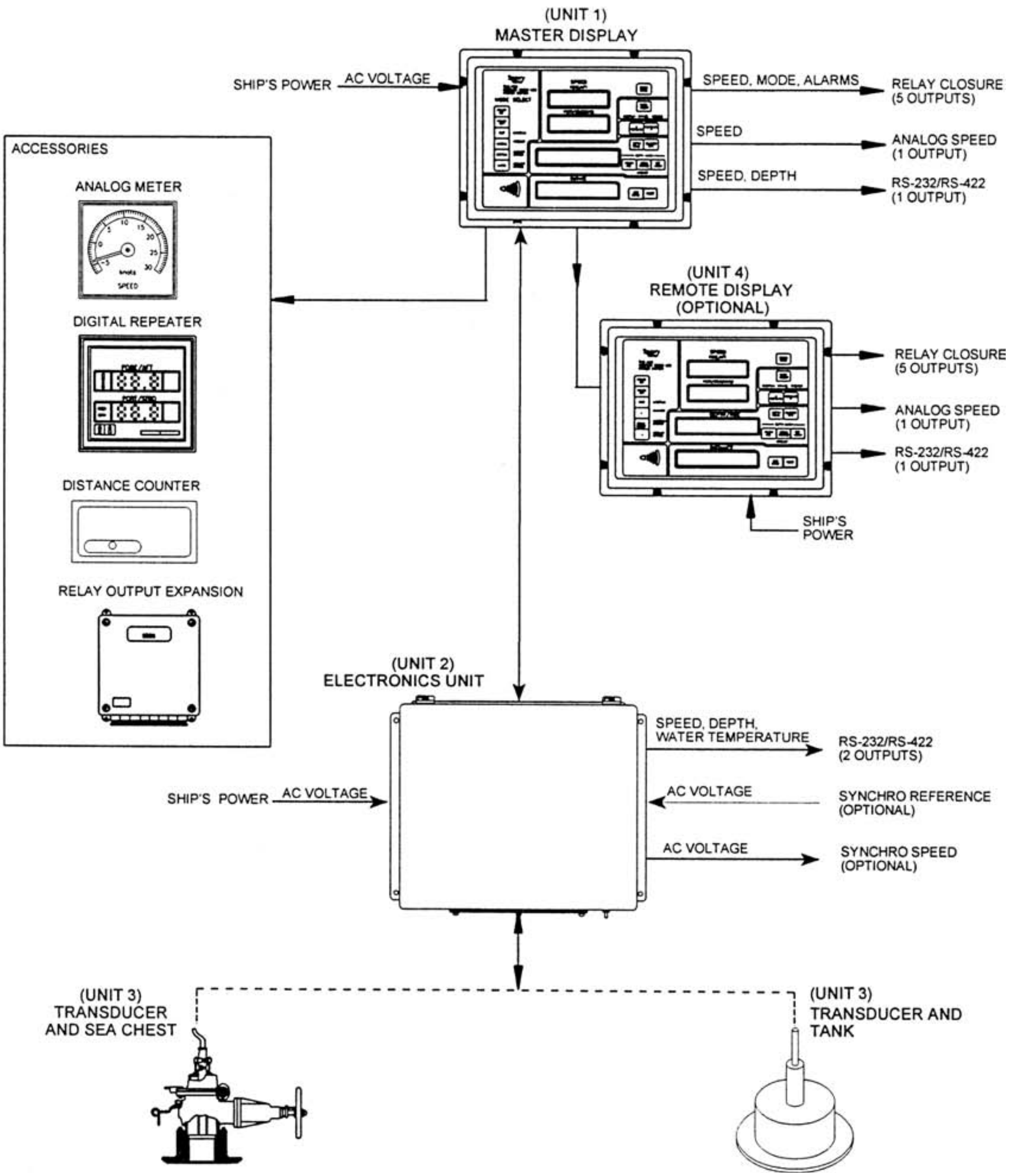


**Figure 3.24** Sperry SRD-500 dual axis Doppler speed log system. (Reproduced courtesy Litton Marine Systems.)

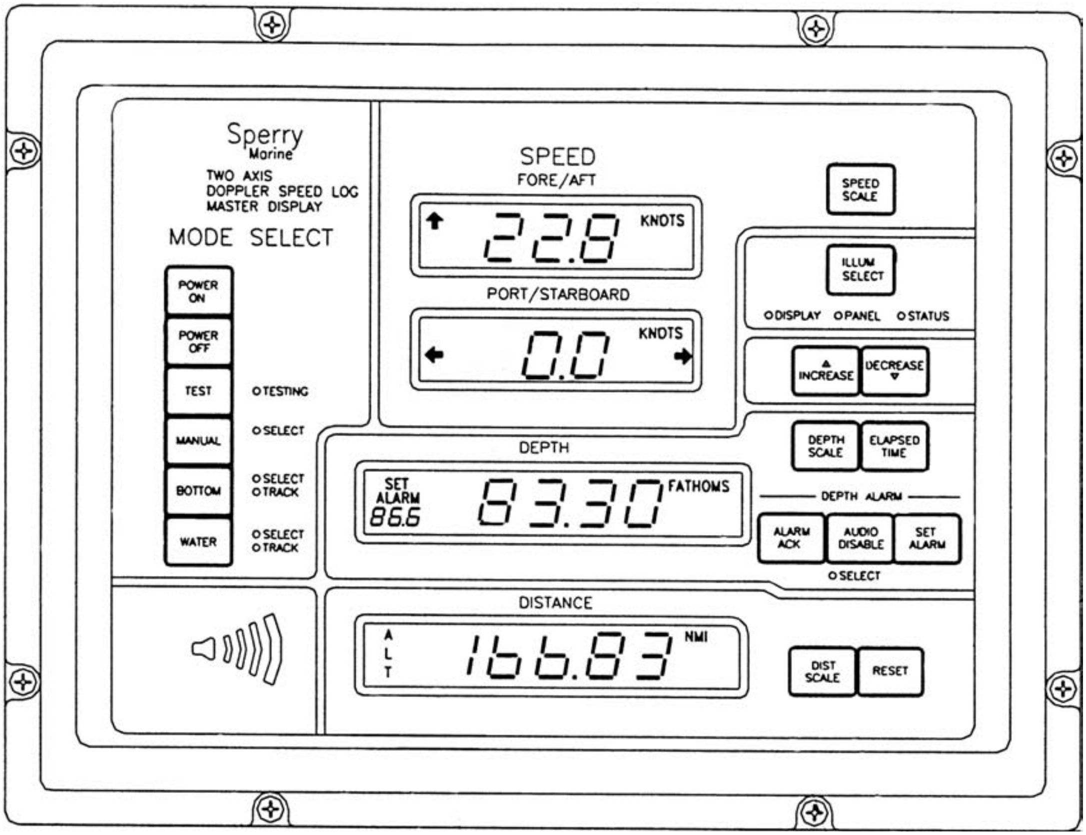


**Figure 3.25** Master display unit, controls and indicators. (Reproduced courtesy of Litton Marine Systems.)

- *Port/starboard display.* This indicates athwartship speed in knots, m/s or ft/s. The range is 0–10 knots.
- *Depth/time display.* This indicates water depth to the seabed, in fathoms, metres or feet, when in either water or bottom-tracking mode, providing the depth is within 200 m. The depth indication circuitry also includes a depth alarm.
- *Distance display.* This shows the distance run in nautical miles or km. Depending upon the selected mode and depth, the display indicates over-the-bottom distance or, when the unit is water tracking, the distance travelled through the water. If the ALT characters are showing, the system tracks both bottom and water simultaneously and provides both outputs to external devices. This display also provides a numerical indication which, when used in conjunction with the system manual, provides clues to any system malfunction.

### *System description*

The Sperry SRD-500 uses a four-element piezoelectric and prism transducer assembly with the four heads aligned as shown in Figure 3.26.

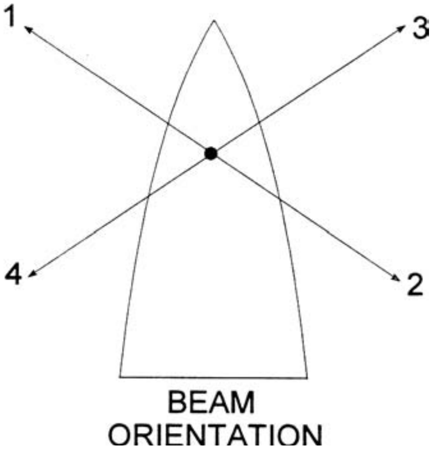


**Figure 3.26** Transducer beam orientation. (Reproduced courtesy of Litton Marine Systems.)

Because the beams are aligned at 45° to the ship's fore/aft and port/starboard axes, each returned signal will hold data about the fore/aft and port/starboard speed of the vessel. Each beam is inclined at 20° to the vertical axis. When the ship is moving ahead, beams 1 and 3 will contain a positive Doppler shift (a compression of the transmitted wavelength) whereas beams 2 and 4 will hold corresponding data.

The data line XMIT from the Processor Board initiates transmission (see Figure 3.27). Four independent transmit amplifiers are pulsed to generate 15–20 W of power causing the four piezoelectric elements to oscillate at 307.2 kHz and transmit acoustic energy pulses towards the seabed.

Sensors in the transducer assembly monitor the seawater temperature and provide, via an A to D converter, corrective data for the central processor.

Returned echoes from either the seabed or scattered from water particles, are received by the transducer elements and coupled to the Receive Amplifier and Frequency Comparison Board. The signals are amplified, gain controlled and then fed to a limiter circuit where the received signal amplitude is monitored and a receive signal strength indicator (RSSI) level is produced. The four levels (RSSI1–4) are coupled back to the processor, which uses the data to determine bottom detection, signal quality and to set the receiver gain. The four-channel signal data is then coupled to frequency comparators that are clocked by the frequency correlation and processing circuitry. From here the data is coupled with address information to the parallel interface and serial I/O boards for distribution.

The RSSI signals generated by the receive circuitry are used to determine the range to the seabed. Individual depths are computed for each beam. The minimum (shallowest) is used to determine if the water depth is sufficient for the log to operate in Water Track mode. The ability of the log to operate in Water Track mode requires a minimum of 3 m depth below the transducer. In deep water the ability to operate in Bottom Track mode is based on obtaining bottom-returned pulses on at least three beams. The displayed depth is an average of the individual depths indicated by the beams. System operation is commanded at start up by the Navigation Display Unit (see Figure 3.28)

These routines are performed when the system is switched on or the manual button is pressed. This includes initializing all program variables, the setting-up of peripheral components and communication controllers and the recall of display configuration data stored in NVRAM (non-volatile random access memory).

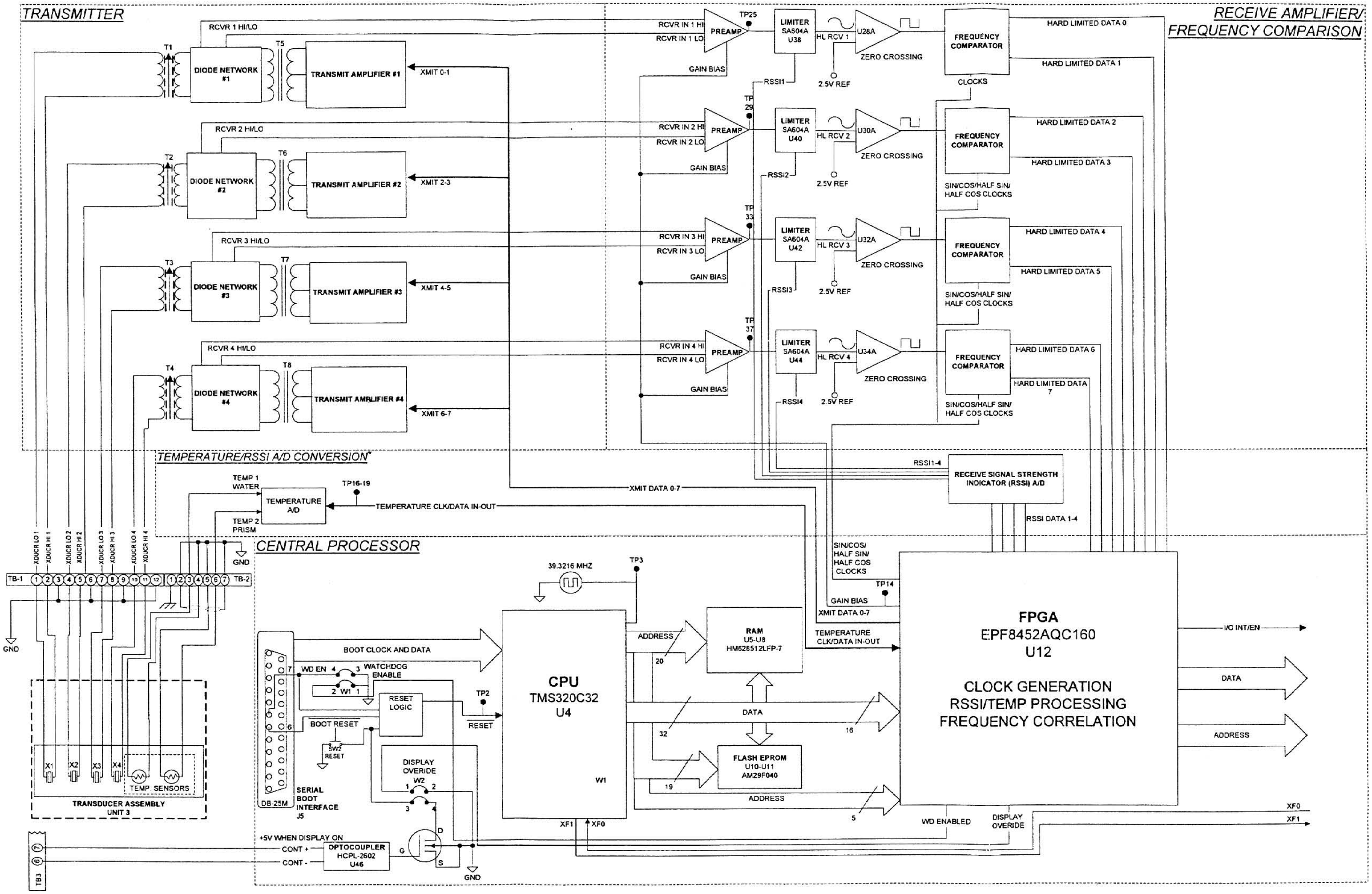


**Figure 3.27** Electronics unit functional block diagram. (Reproduced courtesy of Litton Marine Systems.)

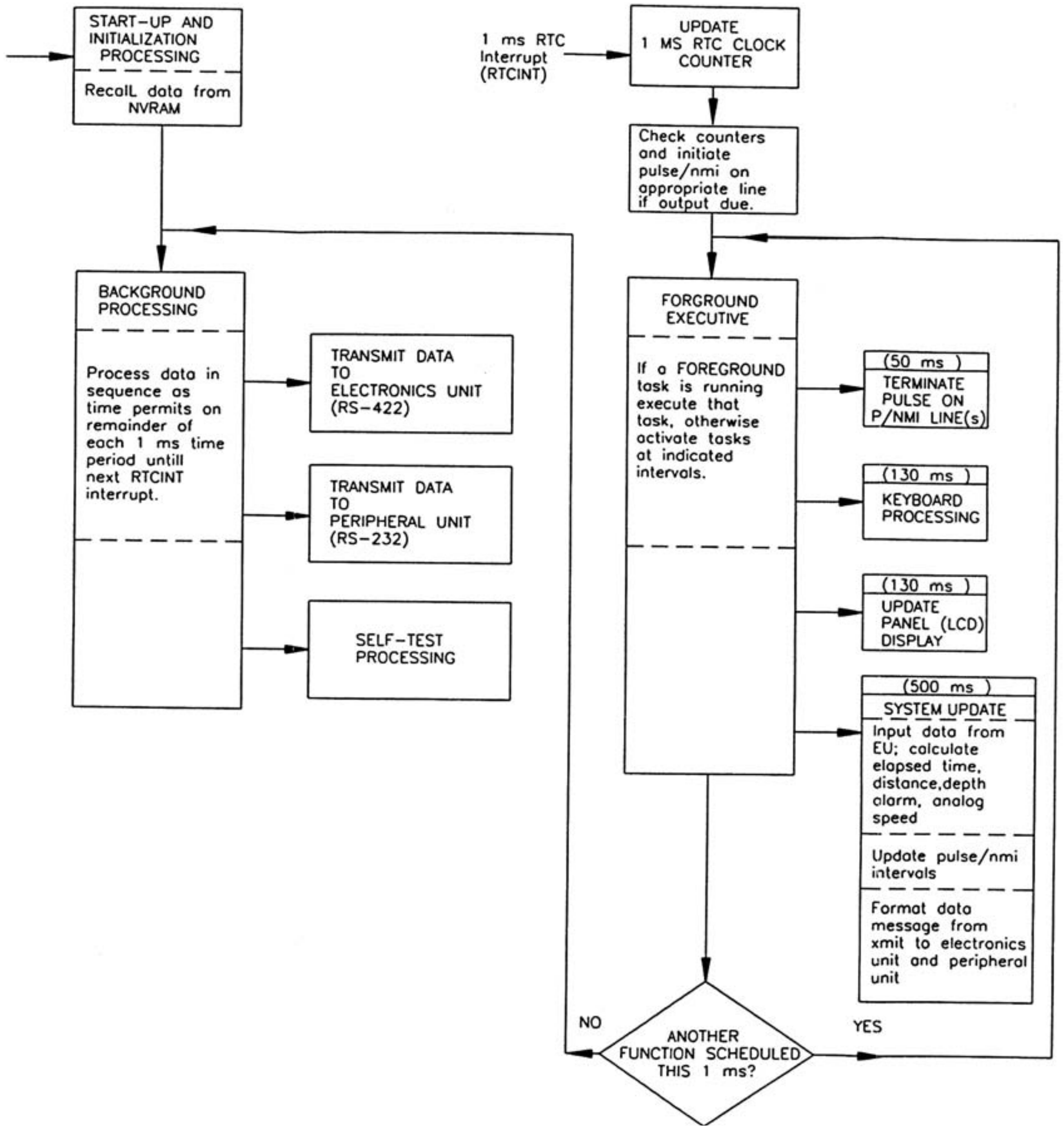


**Figure 3.28** Display unit start-up flow chart. (Reproduced courtesy of Litton Marine Systems.)

The foreground executive processing routines are initiated by a 1-ms real-time interrupt line (RTCINT), which instructs the circuitry to perform the following.

- Keyboard Data Processing: reads and processes any input from the keypad.
- Panel Processing: updates the LCD displays with the latest stored variables.
- System Update: incorporates the latest data from the Electronics Unit into the system variables for processing. Also formats data messages for transmission to the Electronics Unit and to peripheral equipment.

When the Foreground Executive processing routine has completed its tasks, the Background Processing Routine is initiated. This runs until the next RTCINT pulse arrives to retrigger the sequence. Background Processing enables the following actions.

- Data Message Output: enables the outputting of formatted data to both the Electronic Unit and Peripheral Units.
- Self-Test Processing: initializes the self-test program.

The system control flow chart, shown in Figure 3.29, illustrates the main functions. During operation, the foreground processing unit commands all system functions: determining, amongst other things, whether the log is in water or bottom-track mode, processing returned signal information, and operating the self-testing procedure. At 10-ms intervals the background processing subroutine commands data outputs to external navigation equipment and reads input from the master display.

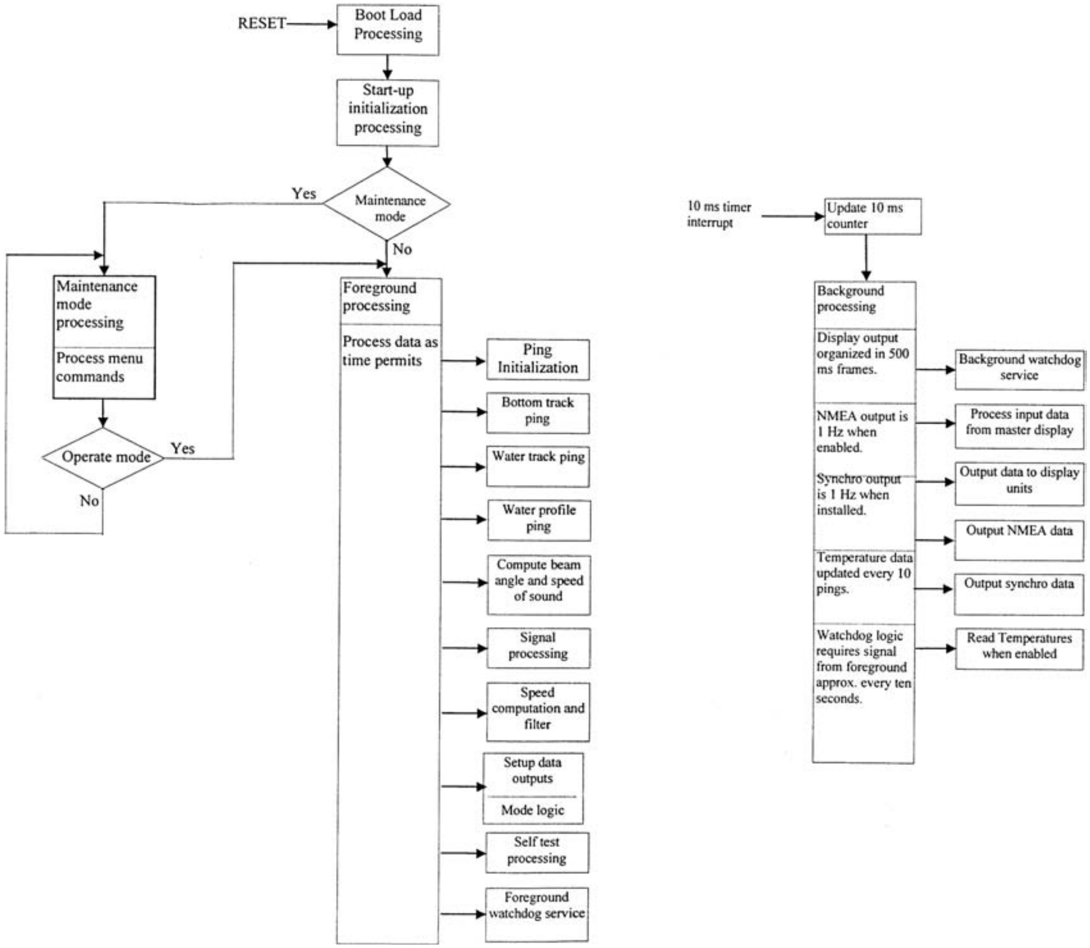


**Figure 3.29** System control flow chart. (Reproduced courtesy of Litton Marine Systems.)

### Built-in test circuitry

In common with most computer-controlled equipment, the system operates a self-test procedure every time it is switched on and performs a fault detection routine at regular intervals during operation. Fault diagnosis routine and testing can be performed manually via the Master Display Unit keypad. During a manual test sequence all LEDs are illuminated and the LCD digits are sequentially displayed. If the test is successful, PASS is displayed in the Distance display. If not, a fault code number is displayed.

As an example of this, the number 402 indicates that the Water Temperature reading is faulty and that the probable faulty component is the temperature sensor in the transducer or the wiring between it and the processing card 2A1. As a further indication of the depth to which the system is able to diagnose faults, the other codes are listed below.

- Codes 101 or 102: keypad faults in the Master Display Unit.
- Codes 201–208: communication faults between the Display Unit and the Main Electronics Unit.
- Codes 255 and 265: RS-232/422 outputs faulty from the Display Unit.
- Codes 301–308 and 355–365: refer to faults in the Main Electronics Unit.
- Codes 401–403: temperature measurement faults.
- Codes 490 and 491: memory test faults.
- Codes 520–524: transmit/receive ping faults.
- Codes 600–604 and 610–614: noise level and sensitivity faults.
- Codes 620–630: receive/transmit signal faults.

### Output data formats

Output data sent to remote navigation systems is formatted in the standard protocol for Interfacing Marine Electronics Navigation Devices developed by the National Marine Electronics Association NMEA 0183 (see Appendix 3 for full details). A second output, in Sperry Marine Format, is intended primarily for the direct printout of speed data.

### Display unit – serial data output format

The serial data output port of each display is configured so that the data can be communicated to peripheral processing devices. Data can be interfaced using RS-232 or RS-422 protocols.

Descriptions of the Sperry SRD-500 communications data format are given in Table 3.1. Examples of NMEA 0183 format messages sent by the RS 422 interface at 4800 bauds are as follows.

Speed message format: $VDVBW,sww.ww,sxx.xx,A,syy.yy,szz.zz,A*cc<CR><LF>
Depth message format: $VDDRU,ddd.dd,A,,V,*cc<CR><LF>
Water Track format: $VDVBW,16.24,-0.62,A,,,V*25
$VDDRU,,V,,V,*7D

The Sperry Marine Format, intended primarily for the direct printout of speed data, is shown in Table 3.2.

Data output is transmitted in ASCII coded format and is structured to be displayed or printed in six-headed columns on a standard page with a width of 80 characters and a length of 56 lines. The serial data interface is set up with 8 data and 2 stop bits and no parity. No handshaking lines are used. Messages are never repeated. A new set of data is formatted for transmission every 0.5 s.

**Table 3.1** Sperry SRD-500 communications data format – NMEA 0183. (Reproduced courtesy of Litton Marine Systems)

| *Data field* | *Description* |
|---|---|
| $ | Message header character |
| VD | Talker ID |
| VBW | Message type (speed bottom/water) |
| DPT | Message type (depth with keel offset) |
| DRU | Message type (depth) |
| MTW | Message type (water temperature) |
| XDR | Message type (transducer measurements) |
| G | Generic |
| eee.ee | Percentage good; first, second, and last 'e' omitted if not used |
| PCB1 | Beam one ID for bottom speed |
| PCW1 | Beam one ID for water speed |
| S | Sign – for aft/port speeds, omitted for fore/stbd speeds |
| ww.ww | Fore/Aft water speed (knots); first and last 'w' omitted if not used |
| xx.xx | Port/Stbd water speed (knots); first and last 'y' omitted if not used |
| zz.zz | Port/Stbd bottom speed (knots); first and last 'z' omitted if not used |
| ddd.dd | Depth (meters); first, second and last 'd' omitted if not used |
| oo.oo | Keel offset (decimeters); first and last 'o' omitted if not used |
| ttt.tt | Temperature (C°); first, second and last 't' omitted if not used |
| A | Data status (A = valid, V = invalid) |
| * | Message data trailer |
| cc | Checksum; 8 bit running XOR of character between $ and* |
| <CR> | Carriage return |
| <LF> | Line feed |

### *Electronics Unit – serial data output*

There are two bi-directional auxiliary ports (Aux 1 and Aux 2) in the Electronics Unit, each of which can be selected to output NMEA 0183 format data directly to peripheral devices.

The baud rate can be selected between 1200 and 115200 and defaults to 4800. Message words are 8 data bits long with selectable parity, a single Start bit and selectable Stop bits (one or two). The default communication setting for Aux 2 complies with NMEA 0183 version 2.1 recommendation: one Start bit, eight data bits, one Stop bit, no parity and a 4800 baud rate.

Both output serial ports send NMEA messages at a 1-s (1 Hz) data rate for speed, depth and water temperature; and at a 10-s rate (0.1 Hz) data rate for 'percent good pings for Bottom Speed' and 'percent good pings for Water Speed'.

### *Examples of output data formats*

Speed message format:
$VDVBW,sww.ww,sxx.xx,A,syy.yy,szz.zz,A*cc<CR><LF>

Depth message format:
$VDDPT,ddd.dd,oo.oo(keel offset)*cc<CR><LF>

Water temperature message format:
$VDMTW.ttt.tt,C*cc<CR><LF>

**Table 3.2** Sperry SRD-500 display unit – serial output data format (Sperry ASCII) (Reproduced courtesy of Litton Marine Systems)

| *Data* | *Format* | *Comments* |
|---|---|---|
| Operating mode | EUTESTâ<br>OPTESTâ<br>BOTTOMâ<br>WATERâ<br>WATBOTâ<br>MANUALâ | 7 character field |
| Fault code | FFFâââ<br>â*ââââ | 6 character field<br>FFF = 3 digit fault code<br>â*â = no fault |
| F/A bottom speed | svv.vvâm/sâââ | 13 character field<br>s = sign bit (-blank)<br>vv.vv = speed value, zero fill if necessary<br>m/s = unit indicator |
|  | âââ*****âââââ | speed undefined |
| F/A water speed | same as F/A bottom |  |
| P/S bottom speed | same as F/A bottom |  |
| P/S water speed | same as F/A bottom |  |
| Depth (altitude) | âddd.dâmâmâcle | 13 character field<br>âââd.d if altitude < 10 m<br>ââdd.d if altitude < 100 m<br>âddd.d if altitude > = 100 m<br>m = unit indicator<br>c = carriage return<br>l = line feed<br>e = end of text character |
|  | ââ*****âââcle | Depth defined |

Note: a 'â' character represents a blank character

Percent good pings for Bottom Speed message format:
$VDXDR,G,eee.ee, ,PCB1,G,eee.ee, ,PCB2,G, . . . ,PCB3,G, . . . ,PCB4*cc<CR><LF>

Percent good pings for Water Speed message format:
$VDXDR,G,eee.ee, ,PCW1,G,eee.ee, ,PCW2,G, . . . ,PCW3,G, . . . ,PCW4*cc<CR><LF>

To decode the above symbols, see Table 3.1.

### *Message example for Water Track Speed*

```
$VDVBW, 2.0,-0.25,A,,,V*
$VDDPT,2.5,-1.0,*79
$VDMTW,18.4,C*0C
$VDXDR,G,000,PCB1,G,000,PCB2,G,000,,PCB3,G,000,,PCB4*58
$VDXDR,G,100,,PCW1,G,100,PCW2,G,100,,PCW3,G,100,,PCW4*58
```

As an example, the first line of this message may be simply decoded as follows.

$ (header) VD (talker ID) VBW (speed bottom/water) s (aft/port speeds) 2.0 (aft speed in kts) s (aft/port speeds) 0.25 (port speed in kts) A (data status).

The above description is only a simple outline of how the NMEA 0183 protocol is used to interface data from this speed log with other electronic systems. Refer to Appendix 3 for a more detailed description of the protocol.

### 3.7.2 The Furono Doppler Sonar DS-50 System

Another respected manufacturer of marine equipment, Furuno, produces a Doppler sonar system, the DS-30, based on the principles of Doppler speed measurement. Whilst the system principles are the same as with other speed logs in this category, Furuno have made good use of the data processing circuitry and a full colour 10-inch wide LCD display to present a considerable amount of information to a navigator. The display modes or shown in Figure 3.30.

The system uses a triple beam, 440 kHz pulsed transmission and from the received Doppler shifted signal calculates longitudinal, thwartship speeds and depth beneath the keel at the bow.

In addition, a Laser Gyro may be fitted on the stern to provide a further data input of transverse speed and rate of turn information (see Figures 3.21 and 3.31). Position data from a GPS receiver may also be input to the CPU.

There are three principle modes of data display.

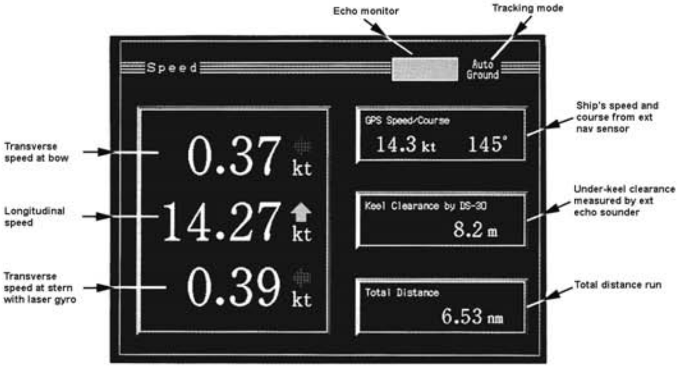
- The Speed Mode showing all the normal speed/depth/distance indications.
- The Berthing Mode which, with the additional inputs from a laser gyro at the stern, shows a vessel's movements during low speed manoeuvres (see Figure 3.31).
- The Nav Data Mode with a display reminiscent of an integrated navigation system.

#### *Berthing Mode display*

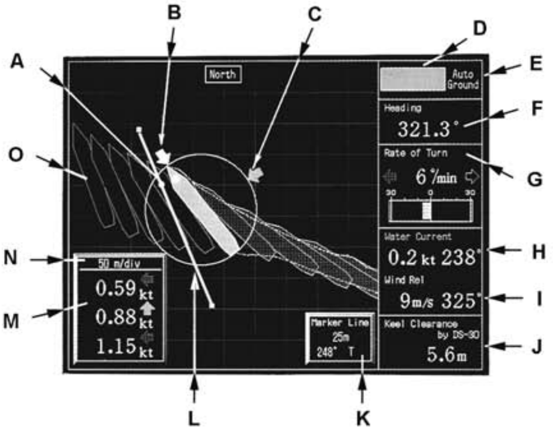
The display diagram key shows the following.

| | |
|---|---|
| A | Intersection of perpendicular from ship's ref. point to marker line. |
| B | Yellow arrowhead showing wind direction. |
| C | Blue arrowhead showing current direction. |
| D | Echo monitor. |
| E | Tracking mode. |
| F | Heading (input from gyro). |
| G | Rate of turn (measured by laser gyro). |
| H | Readout of speed and direction of water current. |
| I | Readout of wind speed and direction (input from wind sensors). |
| J | Under-keel clearance measured by an external echo sounder. |
| K | Range and bearing (true) to marker line. |
| L | Marker line. |
| M | Ship's speed: transverse, longitudinal and transverse at stern with laser gyro. |
| N | Grid scale and presentation mode. |
| O | Ship's predicted motion. |

**SPEED MODE**



**BERTHING MODE**



**NAV DATA MODE**

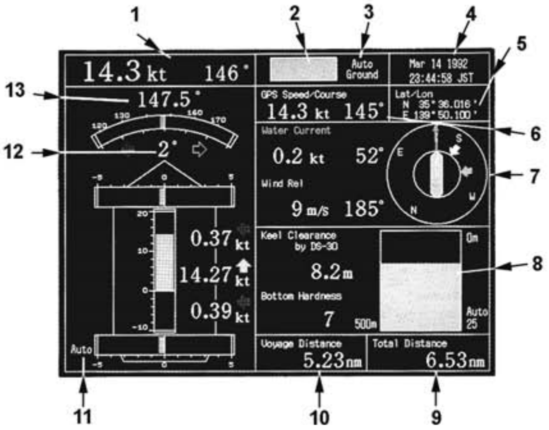


**Figure 3.30** Furuno Doppler Sonar DS-30 display modes. (Reproduced courtesy of Furuno Electric Co.)

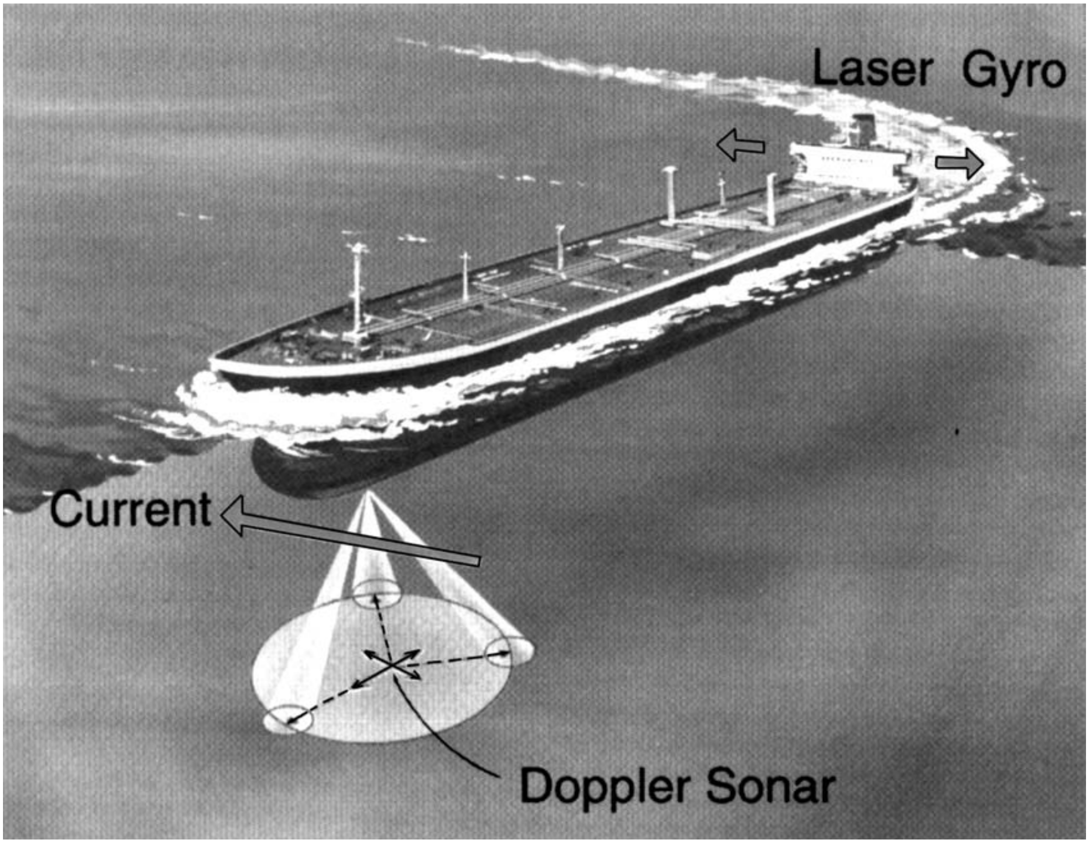


**Figure 3.31** Triple beam transducer configuration of the Furuno Doppler Sonar Log. Note the forces acting on the vessel during a starboard turn under the influence of a cross-current from the port side. (Reproduced courtesy of Furuno Electric Co.)

### *Nav Data Mode display*

The display diagram key for this mode shows the following.

1. Ship's speed and course.
2. Echo monitor.
3. Tracking mode and echo level indicator.
4. Date and time.
5. Position (input from external sensors).
6. Ship's speed and course (input from external sensors).
7. Current speed and direction (app.088°) and wind speed and direction (app. 038°).
8. Graphic presentation of under-keel clearance.
9. Total distance run.
10. Voyage distance from reset.
11. Ship's transverse speed at bow, longitudinal speed and transverse speed at stern with laser gyro.
12. Drift angle (deviation of course over ground from ship's course).
13. Course heading.

## 3.8 Glossary

| | |
|---|---|
| **Aeration** | The formation of bubbles on the transducer face causing errors in the system. |
| **ALPHA (Atlas Low Frequency Phased Array) transducer** | A flush fitting transducer using multiple elements to create the transmitted beam. |
| **Beamwidth** | The width of the transmitted acoustic pulsed wave. The beam spreads the further it travels away from a transducer. |
| **BITE** | Built-in test circuitry. A self-test or manually operated diagnostic system. |
| **CW mode** | Continuous wave transmission. Both the transmitter and receiver are active the whole time. Requires two transducers. |
| **Distance integrator** | The section of a speed log that produces an indication of distance travelled from speed and time data. |
| **Doppler principle** | A well-documented natural phenomenon enabling velocity to be calculated from a frequency shift detected between transmission and reception of a radio signal. |
| **E.M. log** | An electronic logging system relying on the induction of electromagnetic energy in seawater to produce an indication of velocity. |
| **G/T** | Ground-tracking or ground referenced speed. |
| **NMEA** | National Marine Electronic Association. Interfacing standards. |
| **Pitot log** | An electromechanical speed logging system using changing water pressure to indicate velocity. |
| **Pulse mode** | Acoustic energy is transmitted in the form of pulses similar to an echo sounding device or RADAR |
| **Transducer** | The transmitter/receiver part of a logging system that is in contact with the water. Similar to an antenna in a communications system. |
| **Translating system** | The electronic section of a logging system that produces the speed indication from a variety of data. |
| **W/T** | Water-tracking or water referenced speed. |

## 3.9 Summary

- To be accurate, speed must be calculated with reference to a known datum.
- At sea, speed is measured with reference to the ocean floor (ground-tracking (G/T)) or water flowing past the hull (water-tracking (W/T)).
- Traditionally, maritime speed logging devices use water pressure, electromagnetic induction, or the transmission of low frequency radio waves as mediums for indicating velocity.
- A water pressure speed log, occasionally called a Pitot log:
  - (a) measures W/T speed only;
  - (b) requires a complex arrangement of pressure tubes and chambers mounted in the engine room of a ship and a Pitot tube protruding through the hull;
  - (c) produces a non-linear indication of speed which must be converted to a linear indication to be of any value. This is achieved either mechanically or electrically in the system;
  - (d) speed indication is affected by the non-linear characteristics of the vessel's hull and by the vessel pitching and rolling;
  - (e) possesses mechanical sections that require regular maintenance.

- An electromagnetic speed log:
  - (a) measures W/T speed only;
  - (b) produces a linear speed indication;
  - (c) operates by inducing a magnetic field in the salt water flowing past the hull and detecting a minute change in the field;
  - (d) produces a varying speed indication as the conductivity of the seawater changes.
  - (e) Indication may be affected by the vessel pitching and rolling in heavy weather.
- Speed logs that use a frequency or phase shift between a transmitted and the received radio wave generally use a frequency in the range 100–500 kHz. They also use a pulsed transmission format.
- A log using the acoustic correlation technique for speed calculation:
  - (a) can operate in either W/T or G/T mode. G/T speed is also measured with respect to a water mass;
  - (b) measures a time delay between transmitted and received pulses;
  - (c) produces a speed indication, the accuracy of which is subject to all the environmental problems affecting the propagation of an acoustic wave into salt water. See Chapter 2.
- Doppler frequency shift is a natural phenomenon that has been used for many years to measure velocity. If a transmitter (TX) and receiver (RX) are both stationary, the received signal will be the same frequency as that transmitted. However, if either the TX or the RX move during transmission, then the received frequency will be shifted. If the TX and/or RX move to reduce the distance between them, the wavelength is compressed and the received frequency is increased. The opposite effect occurs if the TX and/or RX move apart.
- A Doppler speed logging system:
  - (a) transmits a frequency (typically 100 kHz) towards the ocean floor and calculates the vessel's speed from the frequency shift detected;
  - (b) measures both W/T and G/T speed;
  - (c) produces a speed indication, the accuracy of which is subject to all the environmental problems affecting the propagation of an acoustic wave in salt water;
  - (d) uses a Janus transducer arrangement to virtually eliminate the effects of the vessel pitching in heavy weather;
  - (e) may use more than one transducer arrangement. One at the bow and another at the stern to show vessel movement during turn manoeuvres.

## 3.10 Revision questions

1. A speed indication is only of value if measured against another parameter. What is the speed indication, produced by a pressure tube speed log, referenced to?
2. What is the approximate velocity of propagated acoustic energy in seawater?
3. In a pressure tube speed logging system, why is the complex system of cones required in the mechanical linkage?
4. What is the speed indication produced by an electromagnetic log referenced to?
5. How does the non-linearity of a ship's hull affect the speed indication produced by an electromagnetic speed log?
6. Does the amount of salinity in the water affect the speed indication produced by an acoustic correlation speed log?
7. Why do all Doppler speed logs use a Janus configuration transducer assembly?
8. How does aeration cause errors in the speed indicated by a Doppler log?

9 Using the $\mathbf{V_x}$ and $\mathbf{V_y}$ speed components produced by a Doppler speed log, how is it possible to predict a vessel's drift rate?

10 Why are pulsed transmission systems used in preference to a continuous wave mode of operation?

11 Why are water temperature sensors included in the transducer assembly of a Doppler speed logging system?

12 How may the distance run be calculated in a speed logging system?

# Chapter 4

# Loran-C

## 4.1 Introduction

Loran is an acronym for *lo*ng *ra*nge *n*avigation. It is an electronic system of land-based transmitters broadcasting low frequency pulsed signals that enable ships and aircraft to determine their position. A system that used this concept was first proposed in the 1930s and implemented as the British Gee system early in the Second World War. The Gee system used master and slave transmitters sited approximately 100 miles apart and used frequencies between 30 and 80 MHz. The use of frequencies in the VHF band constrained the system to 'line-of-sight' distance for coverage but this was not a problem at the time since the system was designed to aid bomber navigation on raids over Germany.

The system was further developed at the Radiation Laboratory of the Massachusetts Institute of Technology and the speed of development was such that by 1943 a chain of transmitters was in operation under the control of the United States Coastguard (USCG). This early system was later known as standard loran or Loran-A. This system operated in the frequency range 1850–1950 kHz with master and slave stations separated by up to 600 nmiles. Coverage of the system used groundwaves at ranges from 600 to 900 nmiles over seawater by day, and between 1250 and 1500 nmiles via sky wave reception at night, using the first-hop E layer mode of propagation. Loran-A has a typical accuracy of about 1 nmile for ground wave reception and 6 nmiles for sky wave reception.

Loran-A chains operate by measuring the difference in time arrival of the pulses from the master and the slave stations. Every time difference produces a line of position (LOP) for a master–slave pair and a positional fix is obtained by the intersection of two such LOPs using two suitable master–slave pairs. Two adjacent chains usually have a common master transmitter station. For each chain the slave station transmission is retarded in time compared to that of the master station. Such retardation is known as the coding delay and has a value such that within the coverage area of the chain the master pulse is always received at a receiver before the slave pulse. Known unreliable signals can be indicated by the master or slave signals, or both, being made to blink. Loran-A chains are identified by an alphanumeric which specifies the transmission frequency and the pulse repetition rate (determined by the number of pulses transmitted per second). The pulse repetition rate differs between station pairs in the same chain.

Loran-A was finally phased out in the United States in 1980 and replaced by Loran-C. The use of Loran-A continued in other parts of the world for a time before a change was made to the more universal Loran-C. The last operational Loran-A chains were based along the coast of China. The Loran-C system evolved from Loran-A and the basic principles of both systems are the same.

## 4.2 System principles

The loran transmitter stations send out a stream of pulses at a specified rate known as the pulse repetition frequency (PRF) or the pulse repetition rate (PRR). The pulse repetition period is the reciprocal of the PRF. Assume the PRF is 25, i.e. 25 pulses are transmitted every second, then the period of the pulse is 1/25 s or 40 000 μs. The pulse width is 40 μs for Loran-A and 250 μs for Loran-C.

Assuming that the velocity of radio waves in free space is $3 \times 10^8\,\mathrm{ms}^{-1}$, then the distance travelled by a pulse may be measured in terms of the time taken to travel that distance, i.e. if a pulse took 1000 μs to travel a certain distance then the distance is given by:

$$d = v \times t \tag{4.1}$$

where $d$ = distance in metres, $v$ = velocity of radio waves in $\mathrm{ms}^{-1}$, and $t$ = time in seconds taken for pulse to travel $d$ metres. Then $d = (3 \times 10^8 \times 1000 \times 10^{-6})$ metres or $d$ = 300 km. The velocity of light has been taken as $3 \times 10^8\ \mathrm{ms}^{-1}$ in the above calculation whereas in free space the actual value is $2.99792458 \times 108\,\mathrm{ms}^{-1}$. Also the time taken to travel a certain distance via a ground wave will be affected by the conductivity of the terrain over which it travels. The approximations made here and in the next section are for indicative purposes only.

### 4.2.1 Loran lines of position (LOPs)

Consider two transmitters A and B simultaneously transmitting the same pulse stream (Figure 4.1). If we assume that the distance between the transmitters is 972 nmiles or 1800 km (since 1 nmile = 1.85 km, approximately), then the time taken to cover the distance between the transmitters can be found from equation (4.1) to be:

$$t = d/v$$

$$\text{or } t = (1800 \times 10^3)\mathrm{m}/(3 \times 10^8)\mathrm{ms}^{-1} = 6000\ \mu\mathrm{s}$$

A receiver situated along the baseline joining the two transmitters would receive both pulse streams with the time of reception of each pulse stream determined by its position along the baseline. If the receiver was positioned 600 km from station A and 1200 km from station B then the pulse stream from

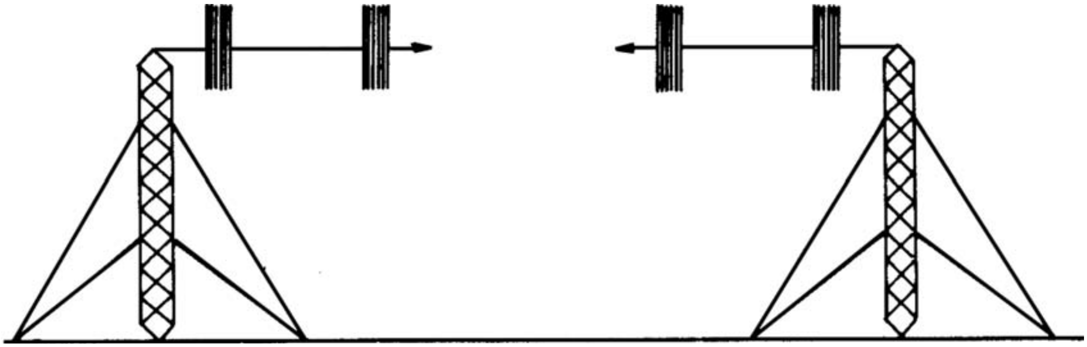


Figure 4.1 The loran system: two transmitters each radiating short pulses of specified length at a specified repetition interval.

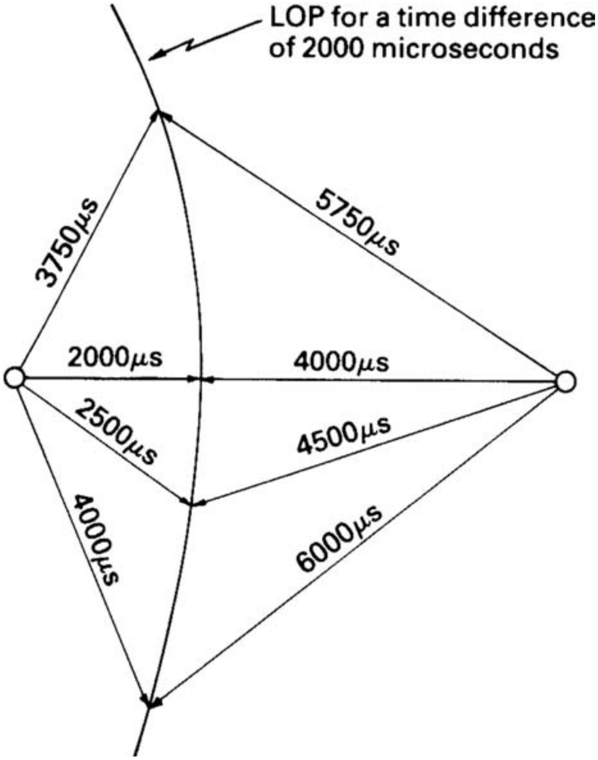


**Figure 4.2** Line of constant time difference (LOP) produced from two transmitter stations emitting pulses simultaneously.

station A would arrive after 2000 μs, while that from station B would arrive after 4000 μs. This means that there is a difference in arrival time of 2000 μs. There would be other receiver positions in the region between the transmitters, not necessarily on the baseline, where the difference in arrival time was 2000 μs. It follows that by connecting all possible points where there is a difference in arrival time of 2000 μs, a line of position (LOP) may be plotted. Figure 4.2 shows a plot of all possible positions where the time difference in pulse reception is 2000 μs.

The LOP shown in Figure 4.2 is a plot of a hyperbola with the transmitter stations as the foci. For this reason loran, and other similar systems, are known as hyperbolic systems. It follows that other hyperbolae may be plotted for other time differences and this has been done in Figure 4.3 for time differences in steps of 1000 μs.

Note that from this diagram the time difference LOPs are symmetrically disposed about the centre line, i.e. there are two 2000-μs LOPs. Hence if the only information at the receiver is the time difference value then an ambiguity can occur. The ambiguity may be avoided by causing the second station, say station B, to be triggered by the pulse received from station A. The hyperbolic LOPs for this arrangement are no different from the original arrangement but the values of time difference are different for each LOP, as shown in Figure 4.4.

Station A in this case is known as the ‘master’ station while station B is known as the ‘secondary’ station. This arrangement, although apparently solving the ambiguity problem, has in fact created another problem. As shown in Figure 4.4, in the region of the baseline extension for the secondary

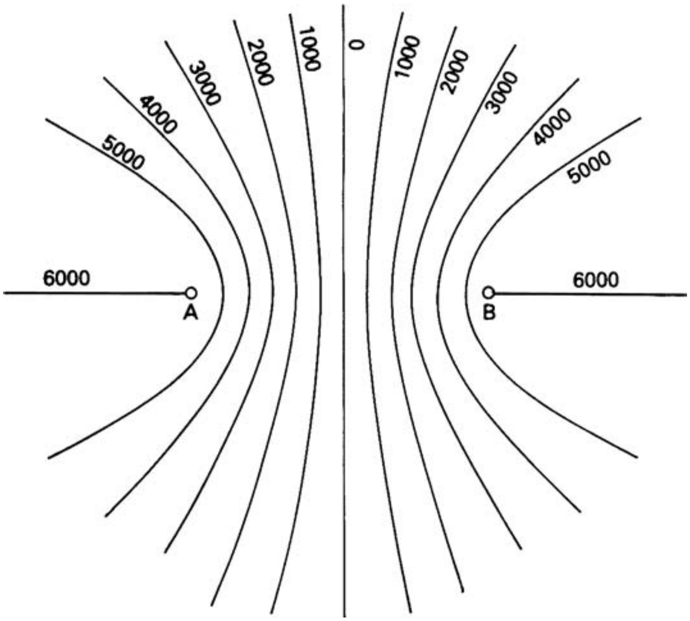


**Figure 4.3** Lines of constant time difference (LOP) produced from two transmitter stations emitting pulses simultaneously.

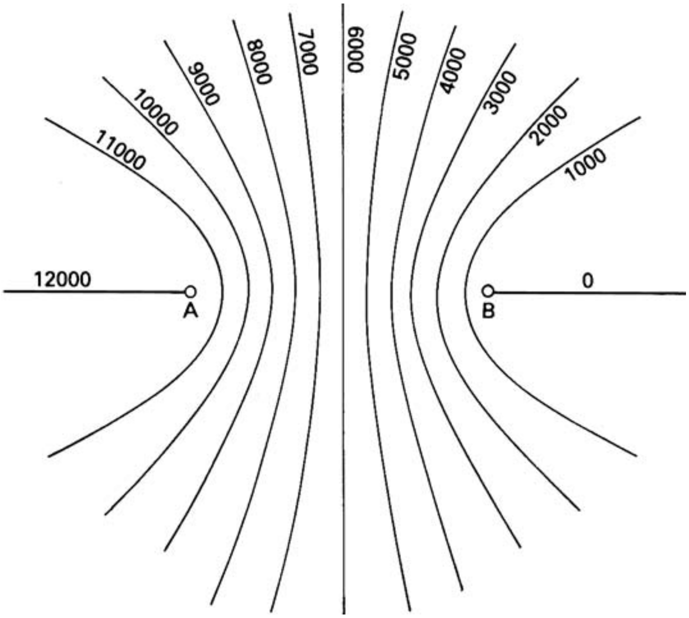


**Figure 4.4** Modification of the LOPs of Figure 4.3. Station B is not allowed to transmit until triggered by a pulse from Station A.

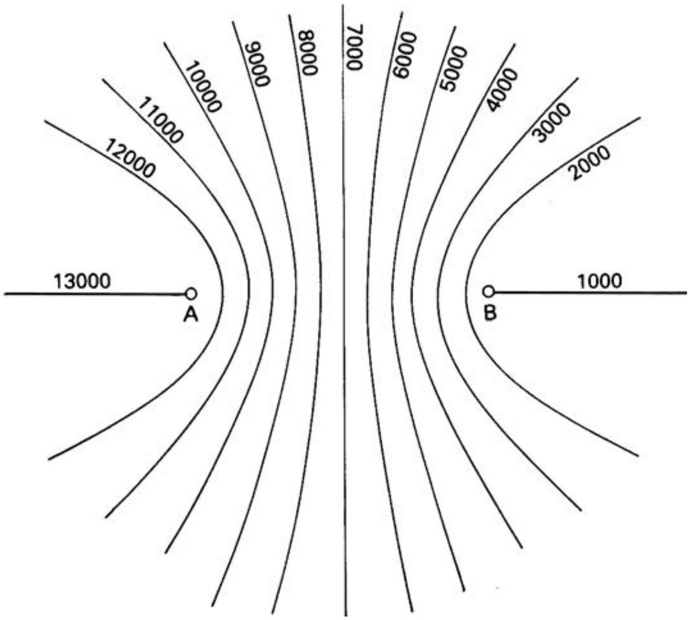


**Figure 4.5** A further modification to the LOPs of Figure 4.3. Not only must Station B wait for a pulse from Station A but there is also a coding delay (1000 μs in this example) which alters the time difference value of each LOP.

station B, the difference in arrival time of the two sets of pulses is smaller than the width of the actual pulse and is in fact zero on the baseline extension. Hence in these regions it would be impossible to separate the two pulses to measure the difference in arrival times.

This drawback is solved by delaying the transmission of the pulse from the secondary for a certain period of time after the pulse from the master has arrived. As mentioned in Section 4.1, this delay period is known as a coding delay. Figure 4.5 has been drawn indicating a coding delay of 1000 μs. The total elapsed time from the master transmission until secondary transmission occurs is known as the emission delay. This is equal to the sum of the time taken for the master signal to travel to the secondary (baseline travel time) and the coding delay. Details of coding delay and emission delay values for Loran-C transmitters may be found in Table 4.9.

Again no two LOPs have the same time difference, eliminating possible ambiguity, and the coding delay ensures that no area is unable to receive two distinctly separate pulses. It is important to ensure that the coding delay is kept accurately constant, since any variation in this value would cause errors in received time differences giving erroneous positioning of the vessel containing the receiver.

The LOPs are overprinted on charts showing the value of time difference for each LOP. Thus using an on-board receiver which is capable of comparing the delay in reception of the pulses from the master and secondary stations, it is possible to plot the position of the vessel along a particular LOP (or, by interpolation between two adjacent LOPs, if the time difference obtained is not the exact value printed on the chart). All that is necessary to establish a position fix for the vessel is to establish the position along a second, intersecting LOP (whether actual or interpolated) using another pair of transmitting stations, i.e. the master, common to all station pairs, and a second secondary station (see Figure 4.6).

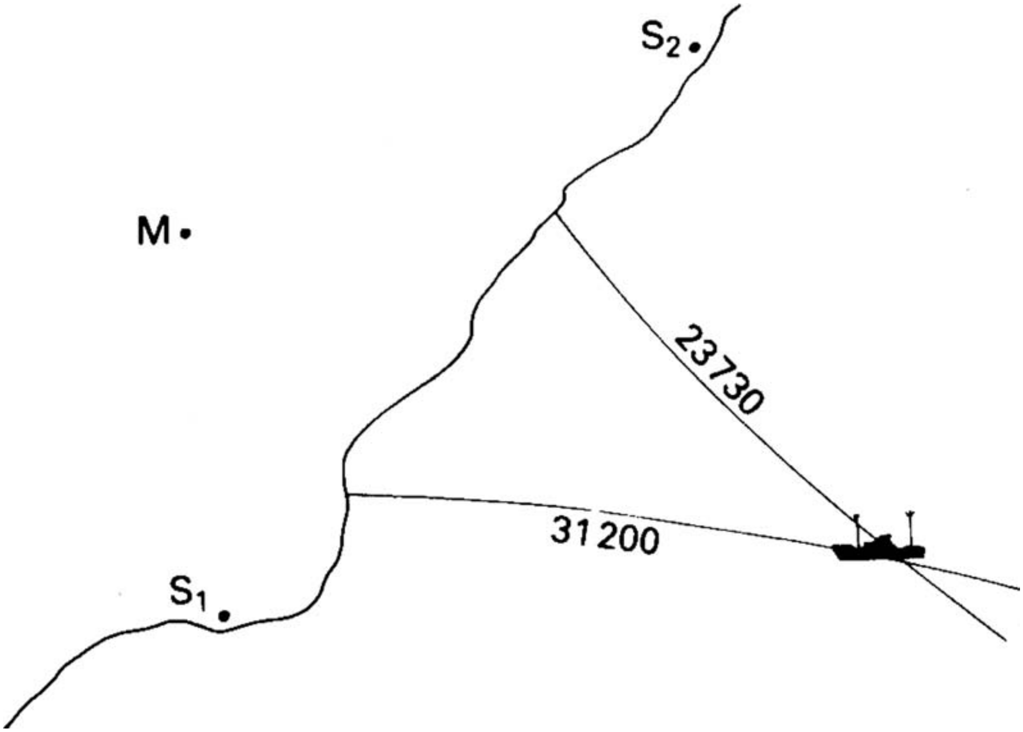


**Figure 4.6** Position fixing using LOPs from two pairs of master/secondary stations.

## 4.3 Basics of the Loran-C System

In the early 1970s the US Department of Transportation which, through the US Coastguard, was responsible for the loran stations, decided that the existing coverage and accuracy provided by the Loran-A stations was below standard and the system of Loran-C, already extant in some regions of the US, was adopted to replace it.

The Loran-C system usually comprises a chain of from three to five land-based transmitting stations, although one chain (see Table 4.9) actually has six transmitting stations, i.e. 9610 South Central US has Victor (V) based at Gillette. One station is always designated as the master (M), while the others are known as secondary stations, whisky (W), x-ray (X), yankee (Y) and zulu (Z) (see Figure 4.7).

All transmitters are synchronized so that signals from the secondaries have precise time-interval relationships with transmissions from the master. This is achieved by the use of atomic oscillators at the stations. Radiated power from Loran-C transmitters varies from a few kW to several hundred kW. The power radiated will affect the range at which usable signals are received and hence define the coverage area of a chain.

Loran-C uses a transmission frequency of 100 kHz and this lower frequency compared with Loran-A gives greater range of reception. The pulse width is 250 µs compared to 40 µs for Loran-A. The actual pulse shape is different for both systems as Figure 4.8 shows.

Since Loran-C achieves its greater accuracy by a process of 'cycle-matching', i.e. matching specified cycles of the received master and secondary pulses rather than the envelope as in Loran-A, the Loran-C pulse is subject to stringent specification requirements.

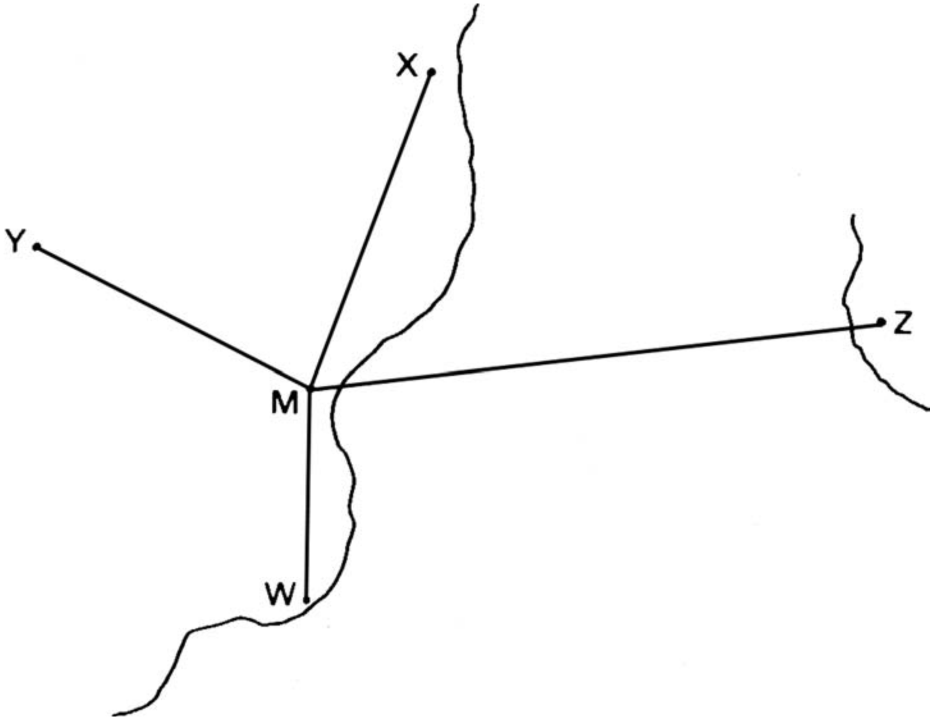


**Figure 4.7** A chain may be configured from a master and up to four secondaries.

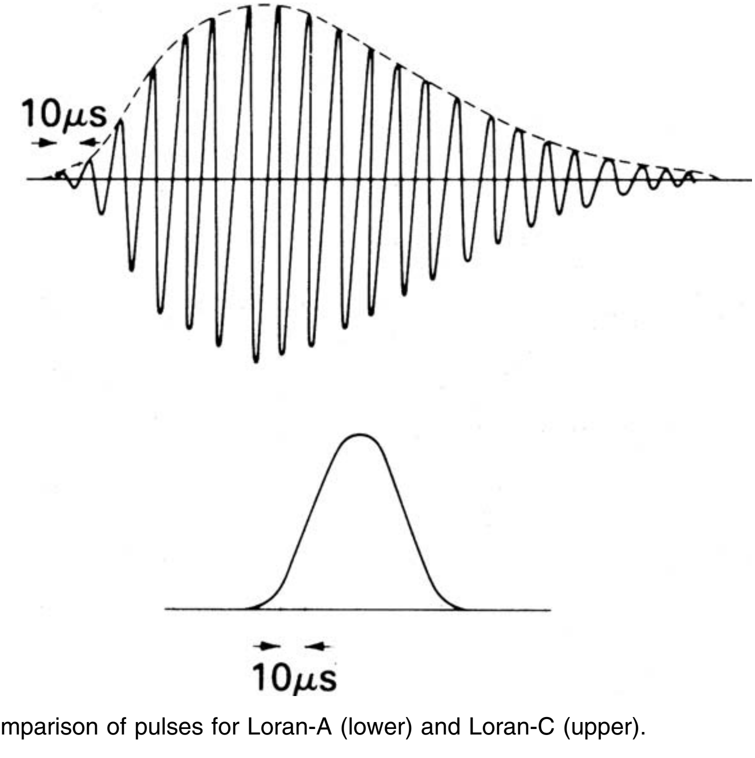


**Figure 4.8** Comparison of pulses for Loran-A (lower) and Loran-C (upper).

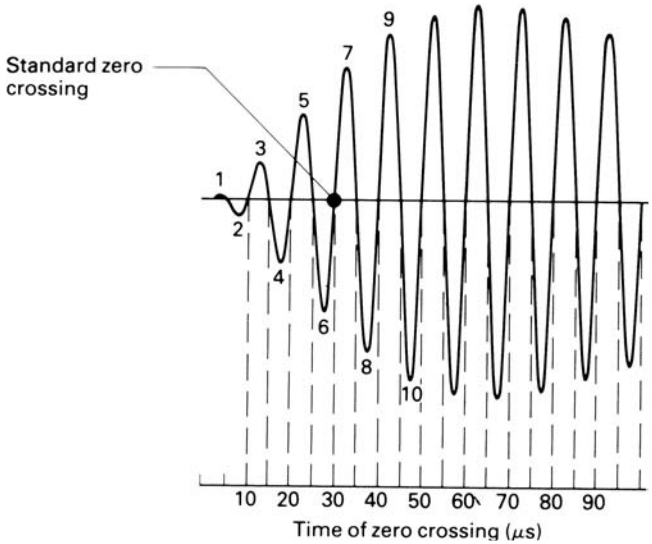


**Figure 4.9** Zero crossing times and labels for half-cycles. This figure shows a 'positive' pulse. For a 'negative' pulse the polarity changes but the labels remain the same.

Each station transmits a pulse which increases rapidly in amplitude and decays at a rate depending on the particular transmitter (Figure 4.9).

The standard pulse leading edge requirement is defined as

$$i(t) = 0; \text{ for } t <_t$$
$$i(t) = A(t - \tau)^2 exp[(- 2(t - \tau))/65]sin(0.2\pi\ t + PC); \text{ for } \tau < t < (65 + \tau)$$

where $i(t)$ is the Loran-C antenna waveform, $A$ is the normalization constant related to the peak antenna current magnitude in amperes, $t$ is the time in μs, $\tau$ is the envelope to cycle difference (ECD) in μs, $PC$ is the phase-code parameter, in radians, which is 0 for positive phase code and $\pi$ for negative phase code.

The ECD is determined as the difference in time between the actual waveform, sampled at the first eight half-cycle peaks, and the standard leading edge as defined above. This deviation is minimized in a root-mean-square sense over ECD and the first 40 μs. The ECD of the pulse is that value which minimizes this deviation. The best nominal ECD for a transmitting station over an all-seawater path is determined from the empirical formula:

$$\text{ECD} = 2.5 + \text{NECD} - 0.00025d$$

where NECD is the nominal ECD of a transmitting station, and $d$ is the distance in nautical miles from the transmitting station.

The pulse trailing edge (that portion of the pulse following the peak of the pulse, or 65 μs, whichever occurs first) is controlled in order to maintain spectrum requirements. At different transmitting sites, or with different transmitting equipments, the pulse trailing edge may differ significantly in appearance and characteristics. Regardless of these differences, for each pulse and for all $t > 500$ µs, $i(t)$ satisfies the pulse trailing-edge tolerances based upon peak amplitude ($A$).

$$\text{Category 1: } i(t) \leq 0.0014\,A$$

$$\text{Category 2: } i(t) \leq 0.016\,A$$

There is a tolerance placed on the amplitude of half-cycles both individually and as a group (considering only the first eight half-cycles). Zero crossing times and tolerances of the first group of pulses are shown in Table 4.1 for the first pulse. The zero crossing times are measured with respect to the standard zero crossing which gives a positive-going zero crossing at 30 µs for a positively coded pulse. ECDs in the range –2.5 to +2.5 µs are assumed.

**Table 4.1** Zero crossing times (with respect to the standard zero crossing) and tolerances

| | | *Tolerance (ns)* | |
|---|---|---|---|
| *Zero crossing (µs)* | *Time (µs)* | *Category 1* | *Category 2* |
| 5 | –25 | ±1000 | ±2000 |
| 10 | –20 | ±100 | ±1500 |
| 15 | –15 | ±75 | ±1000 |
| 20 | –10 | ±50 | ±500 |
| 25 | –5 | ±50 | ±250 |
| 30 | Standard zero crossing | (Time reference) | |
| 35 | 5 | ±50 | ±100 |
| 40 | 10 | ±50 | ±100 |
| 45 | 15 | ±50 | ±100 |
| 50 | 20 | ±50 | ±100 |
| 55 | 25 | ±50 | ±100 |
| 60 | 30 | ±50 | ±100 |

Beyond 60 µs the zero crossings conform to 100 kHz ± 1 kHz

### 4.3.1 Pulse groups

Each Loran-C station operates with a specified group repetition interval (GRI) which are multiples of 10 µs, from 40 000µs up to 99 990µs. The particular GRI is recognized by its GRI value divided by 10, i.e. 7980 would define a GRI of 79 800 µs.

Secondary pulse groups are transmitted with the same GRI as the master pulse group and are linked in time to the master. The delays in transmissions from secondary stations with respect to the master are selected to ensure that the following criteria are met wherever signals can be received for any particular chain.

- Minimum time difference between any secondary and master is 10 900 µs.
- Minimum difference of any two time differences is 9900 µs.
- Maximum time difference is the Group Repetition Interval minus 9900 µs.

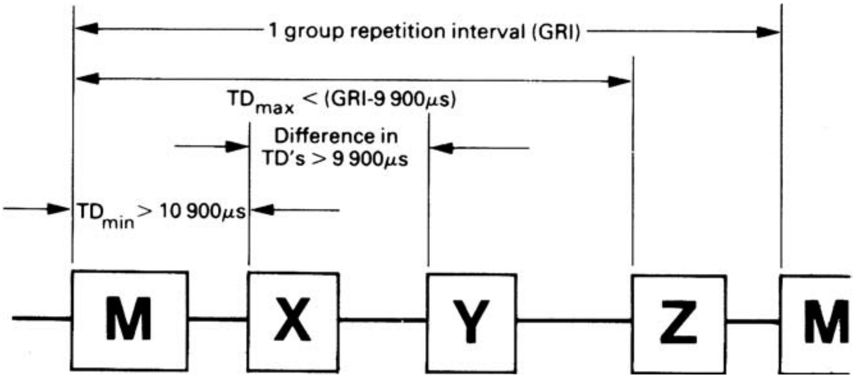


**Figure 4.10** Constraints for assignment of emission delay.

- Minimum spacing between corresponding points of the last pulse of any stations group and the first pulse of the next group in the same chain is 2900 μs. The minimum spacing between the master's ninth pulse and the next secondary pulse (of the same chain), however, may be as little as 1900 μs. This is a direct result of applying the first three criteria.

Figure 4.10 gives an indication of the constraints for emission delay.

### *Uniformity of pulses within a pulse group*

The uniformity of pulses within a pulse group depends not only on the equipment used but whether the station is single-rated (SR) or dual-rated (DR). Dual-rated means that the master station is common to two chains and transmits on two different GRIs. The amplitude of the smallest pulse in the group compared with the amplitude of the largest pulse in the same group should not differ by more than the limits specified in Table 4.2.

Percentage droop is given by:

$$D = \frac{I_{\text{pk.max}} - I_{\text{pk.min}}}{I_{\text{pk.max}}} \times 100$$

where $I_{\text{pk.max}}$ is the value of $i(t)$ at the peak of the largest pulse and $I_{\text{pk.min}}$ is the value of $i(t)$ at the peak of the smallest pulse.

**Table 4.2** Pulse-to-pulse amplitude tolerance, or percentage droop ($D$)

| | *Category 1 (%)* | *Category 2 (%)* |
|---|---|---|
| Single rate | 5 | 10 |
| Dual rate | 10 | 20 |

### 4.3.2 Pulse-to-pulse ECD tolerances

The pulse-to-pulse ECD tolerances account for the pulse-to-pulse leading edge differences and the pulse-to-pulse zero crossing differences. The ECD of any single antenna current pulse does not differ from the average ECD of all pulses by more than the values given in Table 4.3.

**Table 4.3** Pulse-to-pulse ECD tolerances

| | *Category 1 (μs)* | *Category 2 (μs)* |
|---|---|---|
| Single rate | 0.5 | 1.0 |
| Dual rate | 0.7 | 1.5 |

### 4.3.3 Transmission of Loran-C pulses

Whereas Loran-A transmitted one pulse from the master and another from the secondary, with the two pulses compared in the receiver to obtain a time delay, the Loran-C transmitter emits a series of pulses from the master and secondary stations. A typical transmission sequence for a Loran-C chain is shown in Figure 4.11.

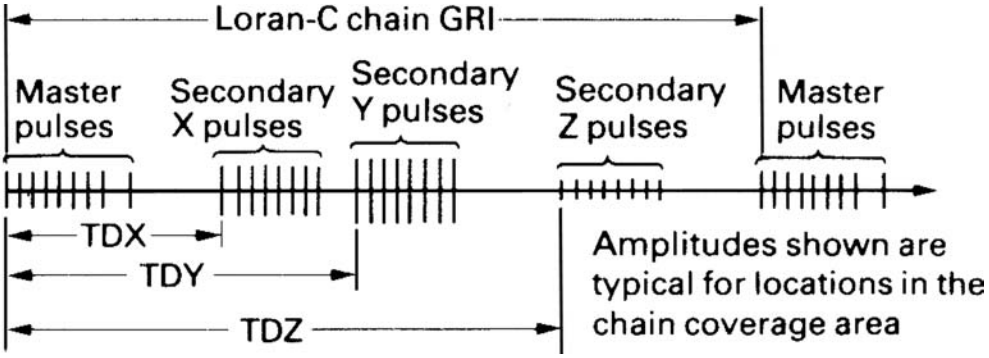


**Figure 4.11** Loran-C chain group repetition interval (GRI) showing the receipt of master and X, Y and Z secondaries.

The GRI is defined as the time interval between successive pulse groups measured from the third cycle (or zero crossover) of the first pulse of any one station in the group, to the third cycle of the first pulse of the same station in the following pulse group. All stations in the chain have the same GRI, and the GRI expressed in tens of microseconds is the identifier for that chain and is called the chain 'rate'. The master transmitter sends out a series of nine pulses while the secondaries transmit only eight pulses.

### 4.3.4 Phase coding

Each Loran-C station phase-codes the series of pulses in accordance with Table 4.4. For identification, the first group of pulses in the sequence is labelled Group A and the second group, one GRI later, is labelled Group B. A transmission sequence (the phase-code interval or PCI) comprises both Group A and Group B, and the PCI sequence is thereafter repeated. The minus sign in Table 4.4 stands for a pulse that is 180° out of phase with the 'normal' pulse, i.e. the phase of the pulse is inverted.

**Table 4.4** Loran-C phase modes

| *Group* | *Station* | |
|---|---|---|
| | *Master* | *Secondary* |
| A | + + − − + − + − + | + + + + + − − + |
| B | + − − + + + + + − | + − + − + + − − |

### 4.3.5 Pulse-to-pulse timing tolerances

Pulses two to eight of a group are referenced in time to the first pulse of each group. The timing relationship and tolerances of the standard zero crossings of pulses two to eight with respect to pulse one standard zero crossing are shown in Table 4.5. The ninth pulse of the master transmission is spaced 2000 μs from the eighth pulse of the group. This pulse is used primarily as a visual aid to master group identification and not as an aid to navigation.

**Table 4.5** Pulse-to-pulse timing tolerances. N is the pulse number (2–8) of the pulses which follow the first pulse within each group. C is 0 for positively phase-coded pulses; |C| ≤ 150 μs for negatively phase-coded pulses. The standard zero crossing of pulse one is the time reference within each group

| | *Category 1* | *Category 2* |
|---|---|---|
| Single rate | (N–1) 1000 μs ± 25 ns | (N–1) 1000 μs ± 50 ns + C |
| Dual rate | (N–1) 1000 μs ± 50 ns | (N–1) 1000 μs ± 100 ns + C |

The use of phase coding allows automatic Loran-C receivers to distinguish between master and secondary transmissions and also assists the receiver to operate when the loran signals are weak in the presence of noise.

### 4.3.6 Blink

Blink is a repetitive on–off pattern (approximately 0.25 s 'on' and 3.75 s 'off') of the first two pulses of the secondary signal which indicates that the baseline is unusable for one of the following reasons:

(a) TD out of tolerance;
(b) ECD out of tolerance;
(c) improper phase code or GRI;
(d) master or secondary station operating at less than one-half of specified power output or master station off the air.

Blink continues until the out of tolerance condition is eliminated. The ninth pulse of the master may also be blinked simultaneously, but by itself master blink is not an indication of an out of tolerance condition.

Master blink is normally only used for internal Loran-C system communication. If used, the master's ninth pulse will be blinked in accordance with the code shown in Figure 4.12.

A selection of Loran-C stations together with their category status, with regard to pulse generation, is shown in Table 4.6. It should be noted from this table that many stations are dual rated. For example, Dana is the zulu secondary transmitter for the Northeast US Chain (9960) and also acts as the master transmitter for the Great Lakes chain (8970).

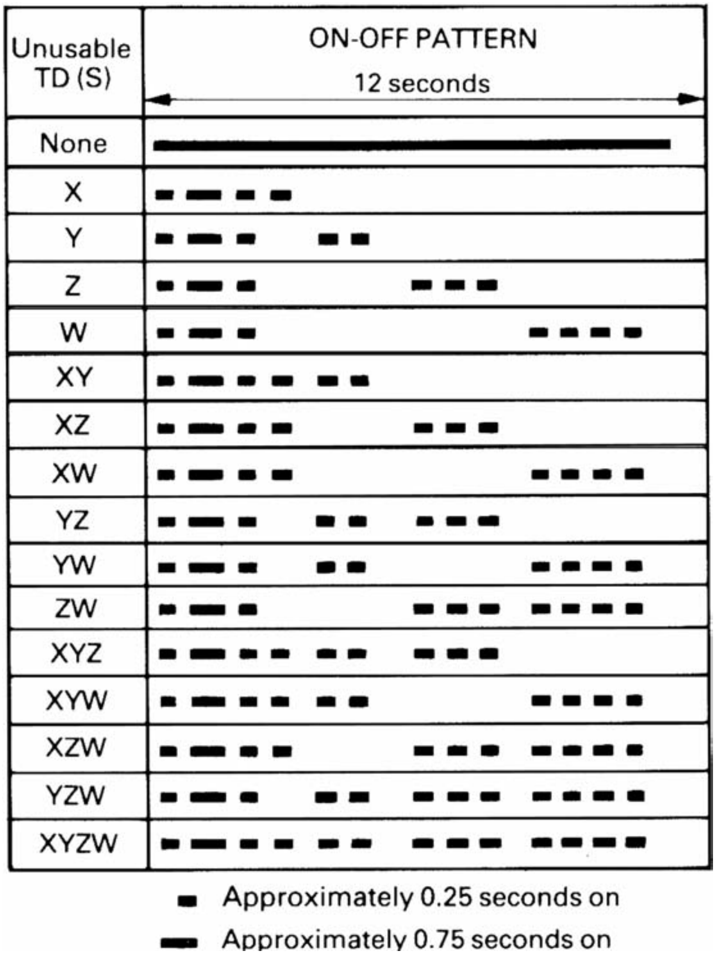


**Figure 4.12** Master ninth pulse blink codes.

**Table 4.6** Compliance requirements for selected Loran-C transmitter stations

| *Station* | *Rate* | *Compliance requirements categories* | | | | |
|---|---|---|---|---|---|---|
| | | *Pulse trailing edge* | *Pulse zero crossings* | *Amplitude* | *ECD* | *Timing* |
| St Paul, AK | 9990 | 2 | 2/1 | 2/1 | 2/1 | 2/1 |
| Attu, AK | 9990/5980 | 1 | 1 | 1 | 1 | 1 |
| Port Clarence, AK | 9990/7960 | 1 | 2/1 | 2/1 | 2/1 | 2/1 |
| Kodiak, AK | 9990/7960 | 1 | 2 | 1 | 1 | 1 |
| Tok, AK | 7960 | 1 | 1 | 1 | 1 | 1 |
| Shoal Cove, AK | 7960/5990 | 1 | 1 | 1 | 1 | 1 |
| Williams Lake, BC | 5990/8290 | 2 | 1 | 1 | 1 | 1 |
| George, WA | 5990/9940 | 1 | 1 | 1 | 2 | 1 |
| Port Hardy, BC | 5990 | 1 | 1 | 1 | 1 | 1 |
| Fallon, NV | 9940 | 2 | 1 | 1 | 1 | 1 |
| Middletown, CA | 9940 | 2 | 1 | 1 | 1 | 1 |
| Searchlight, NV | 9940/7980 | 1 | 1 | 1 | 1 | 1 |
| Malone, FL | 8970/7980 | 1 | 1 | 1 | 1 | 1 |
| Grangeville, LA | 7980/9610 | 1 | 1 | 1 | 1 | 1 |
| Raymondville, TX | 7980/9610 | 1 | 1 | 1 | 1 | 1 |
| Jupiter, FL | 7980 | 1 | 1 | 1 | 1 | 1 |
| Carolina Beach, NC | 9960/7980 | 1 | 1 | 1 | 1 | 1 |
| Seneca, NY | 9960/8970 | 1 | 1 | 1 | 1 | 1 |
| Caribou, ME | 9960/5930 | 1 | 1 | 1 | 1 | 1 |
| Nantucket, MA | 9960/5930 | 1 | 1 | 1 | 1 | 1 |
| Dana, IN | 9960/8970 | 2 | 2 | 1 | 1 | 1 |
| Cape Race NFLND | 5930/7270 | 1 | 1 | 1 | 1 | 1 |
| Fox Harbour, LABR | 8970/7270 | 2 | 1 | 1 | 1 | 1 |
| Boise City, OK | 8970/9610 | 1 | 1 | 1 | 1 | 1 |
| Gillette, WY | 8290/9610 | 1 | 1 | 1 | 1 | 1 |
| Havre, MT | 8290 | 1 | 1 | 1 | 1 | 1 |
| Comfort Cove, NFLND | 7270 | 1 | 1 | 1 | 1 | 1 |

### 4.3.7 Elimination of sky wave reception

Normal operation of Loran-C assumes reception by ground wave for high accuracy of position fixing. Sky waves always arrive later than ground waves although this difference in arrival time becomes less as the distance from the transmitter increases. However, the time difference is never less than 30 μs anywhere in the Loran-C coverage area. If, therefore, only the first 30 μs of the Loran-C pulse is used then sky wave contamination cannot occur.

At distances greater than 1000 nautical miles (1852 km), the ground wave is likely to be unusable because it suffers more attenuation than the sky wave. Thus the sky wave may be used beyond this range but reception of sky wave signals gives lower accuracy and corrections must be applied to compensate for the difference in path travelled compared to the ground wave.

### 4.3.8 Cycle matching

The technique of matching the pulse envelope, as used in Loran-A, is also used in Loran-C. However, this is only used to give coarse position fixing. Greater accuracy is obtained with cycle matching. With this technique the receiver has a flywheel oscillator which acquires the frequency and phase of the incoming 100 kHz master pulses. Thus the receiver has a reference frequency which is continually updated by the master pulses and has the same phase as the master signals.

The difference in phase between the flywheel oscillator and the secondary station pulses as received, is measured in the receiver and displayed as a time difference down to 0.1 µs. This is possible since the period of one cycle at 100 kHz is 10 µs and the phase difference can be measured up to approximately 1/100 of a cycle. For example, suppose a phase difference of 0.63 cycle is measured, then the phase difference in microseconds is given by (0.63 × 10)µs = 6.3 µs.

The envelope matching method gives the phase difference in tens of thousands, thousands, hundreds, and tens of microseconds with a tolerance of ± 4 µs while the cycle-matching gives the units and tenths of a unit. Thus if envelope matching gives a time difference of, say, 52 700 µs and cycle matching gives 4.3 µs, then the accurate value of time difference is 52 704.3 µs.

One method of automatic pulse envelope matching is to compare the received pulses from the secondary station with the pulses from the receiver flywheel oscillator after the latter have been passed through a variable time delay circuit. The delay circuit is necessary because the master pulses will always arrive first anywhere in the system (this is because the secondaries are triggered after the master transmission and there is a coding delay). If the timing of the pulses does not coincide then an error voltage is produced which adjusts a time delay until the start of the two pulses is caused to coincide. When this happens the delay voltage is reduced to zero. The value of the delay must be the same as the delay between the received master and secondary pulses, and if displayed in digital form would give the coarse delay figure in microseconds. One method of fine matching is to use the technique as illustrated in Figure 4.13.

The received pulse is amplified by a specific amount and shifted in phase by 180°. This new wave is algebraically added to the original wave to produce a resultant wave with a well-defined minimum at a time before the value where sky wave contamination can occur. The difference in time between the sampling of the master pulse and the secondary pulse is determined in the same way as for envelope matching and the result is also presented on the digital display.

An automatic method of fine matching to allow coverage on an extended range uses cycle matching on the seventh cycle of the received pulse. A mode switch allows the operator to choose third or seventh cycle matching facility. Matching may be extended to ranges 500 nautical miles (925 km) greater than with normal matching because of the greater amplitude of the seventh cycle. The inherent inaccuracies due to possible sky wave contamination must be taken into account but may be acceptable at the longer ranges involved.

## 4.4 Loran-C charts

Nautical charts, overprinted with Loran-C LOPs, are available from several sources, including:

- The US National Ocean Survey, for charts principally around the US coast
- The Defense Mapping Agency, for world-wide charts
- The Canadian Hydrographic Service, for charts of Canadian waters
- The UK Hydrographic Office, for charts of British waters.

Catalogues of charts and the areas covered are available from the organizations mentioned. The charts are identified in terms of the area covered and the designations of the stations serving the area; for example, a chart serving the Southeast US may require Loran-C LOPs for a master and, say, four

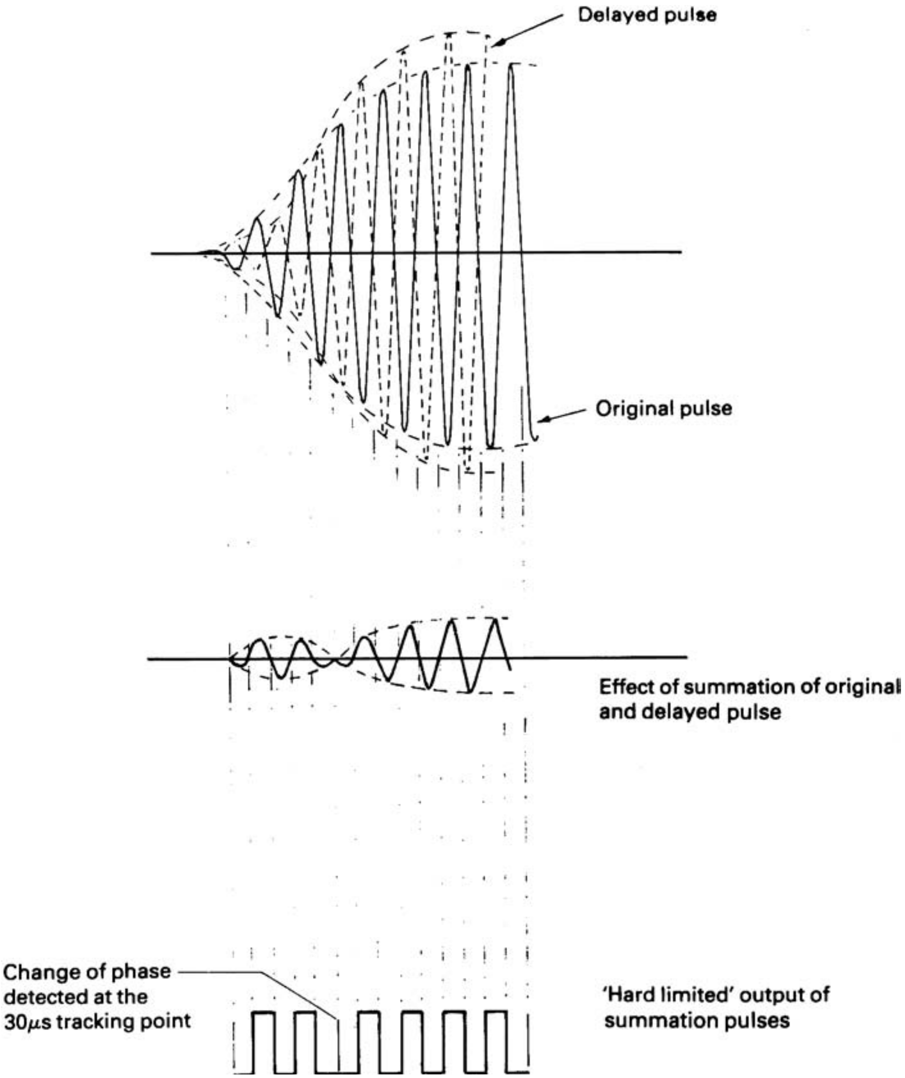


**Figure 4.13** Third cycle tracking by the use of the original pulse and a pulse delayed by 5 µs, amplified and summed with the original pulse, to give a change of phase at the standard zero crossing point.

secondary stations. In this case the stations are identified by the GRI number and the secondary designation, i.e. 7980-W, 7980-X, 7980-Y and 7980-Z.

Not all LOPs are printed on charts. Usually LOPs separated by time intervals of 10 µs are used to give lines spaced at reasonable intervals. For LOPs not specified on the chart the operator must interpolate between the lines.

The National Ocean Survey first edition charts are produced with the Loran-C LOPs based on predicted coverage rather than actual field measurements. Because there are factors which affect the propagation of the loran signal, any measured time difference may be slightly in error. Correction factors may need to be applied to ensure that the designed accuracy limits of the system are met. There are three correction factors that may be applied and these are called phase factors.

- Primary Phase Factor (PF). This allows for the fact that the speed of the propagated signal in the atmosphere is slightly slower than in a vacuum. This difference is due to the fact that the index of refraction of the atmosphere is slightly greater than unity.
- Secondary Phase Factor (SF). This allows for the fact that the speed of propagation of the signal is slowed when travelling over seawater because of the lower conductivity of seawater compared to land. This factor allows for the extra time needed for the propagated signal to travel over an all-seawater path compared to an all-land path.
- Additional Secondary Phase Factor (ASF). Because the Loran-C transmitters are land based the propagated signal will travel over land and sea. The ASF may be calculated by treating the signal path as separate segments, each with a uniform conductivity value depending on whether the segment path is over land or seawater. The matter is complicated by the fact that the ASFs at a fixed point in the coverage area may vary with time. Such variations are caused mainly by seasonal variations in temperature and by local weather activity.

The ASF corrections are incorporated into most Loran-C over-printed charts and many of the Loran-C receivers.

Since additional secondary phase factors can vary from one location to another, there will be points on the chart where there are differences, although usually small, between the actual additional phase factor and the average value that was used for making the chart. In such circumstances there would be a large difference between the Loran-C readings measured and the location on the chart where these readings would be plotted. It is expected that, when necessary, future chart editions will remedy this situation by using varying values for additional phase factors on a chart rather than just a single reading.

Loran-C correction tables are available for those charts that have not been corrected for ASF errors. These tables contain a complete chain and a table section is prepared for each master–secondary pair in that chain. Each page of the correction tables covers an area 3° of latitude and 1° of longitude. Examples are shown for the Northeast US (NEUS) chain master–whiskey pair (Table 4.7) and the master–yankee pair (Table 4.8). The ASF corrections can be either positive or negative; negative values are indicated by a negative sign preceding the number, the positive values have no sign. The ASF correction tables are intended primarily for the situation where the Loran-C time differences are converted electronically to geographic co-ordinates.

To use the tables the position of the vessel must first be determined to the nearest 5 minutes of arc in longitude and latitude and the relevant page of the table referred to, to find the value of the correction. The ASF correction is added algebraically to the time difference for the Loran-C pair.

Consider the following example. Loran-C receiver dial readings are 12 153.31 μs and 44451.83 μs for pairs 9960-W and 9960-Y, respectively. From these readings the computer determines a position of 44° 15.1′ N latitude and 67°25.4′ W longitude. Entering the page index of Section W with the latitude and longitude nearest to the computed position of the vessel, the page number containing the derived geographics is found to be 17W (see Table 4.7). Entering page 17W, the correction at 44°15′N and 67°25′W is +1.5 μs. On page 17Y (Table 4.8), at the same position the correction is +2.7 μs.

The ASF corrections would be applied to the dial readings as follows:

| | | | |
|---|---|---|---|
| W TD | 12153.31 | Y TD | 44451.83 |
| ASF correction | + 1.5 | ASF correction | + 2.7 |
| Corrected TD | 12154.81 | Corrected TD | 44454.53 |

**Table 4.7** Extract from Loran-C correction tables. (Reproduced courtesy of the Defense Mapping Agency Hydrographic/Topographic Center)

| LATITUDE NORTH | | | | | | | 9960-W | | | | | 17W | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *68°* | | | | | *Longitude west* | | | | | | *67°* |
| | | *0′* | *55* | *50* | *45* | *40* | *35* | *30* | *26* | *20* | *15* | *10* | *5* | *0′* |
| 45° | 0′ | | | | | | | | | | | | | |
| | 55 | | | | | | | | | | | | | |
| | 50 | | | LAND | | | | | | | | | | |
| | 45 | | | | | | | | | | | | | 2.0 |
| | 40 | | | | | | | | | | | | 1.7 | 1.6 |
| | 35 | | | | | | | | | 1.7 | 1.7 | 1.6 | 1.6 | 1.6 |
| | 30 | | | | | | | 1.7 | 1.6 | 1.6 | 1.6 | 1.5 | 1.5 | 1.5 |
| | 25 | | | | | 1.6 | 1.5 | 1.5 | 1.6 | 1.6 | 1.6 | 1.5 | 1.5 | 1.5 |
| | 20 | 1.6 | 1.5 | 1.6 | 1.6 | 1.5 | 1.4 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| | 15 | 1.6 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | $\boxed{1.5}$ | 1.5 | 1.5 | 1.5 | 1.4 | 1.4 |
| | 10 | 1.5 | 1.4 | 1.5 | 1.5 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.4 | 1.3 |
| | 5 | 1.4 | 1.4 | 1.5 | 1.4 | 1.4 | 1.4 | 1.4 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | |
| 44° | 0′ | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.2 | 1.3 | 1.3 | | | |
| | 55 | 1.3 | 1.3 | 1.2 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.2 | | | | |
| | 50 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | 1.3 | | | | | | |
| | 45 | 1.3 | 1.3 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | | |
| | 40 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | | | |
| | 35 | 1.3 | 1.2 | 1.3 | | | | | | | | | | |
| | 30 | 1.3 | 1.3 | 1.3 | | | | | | | | | | |
| | 25 | 1.2 | 1.3 | 1.3 | | | | | OUTSIDE CCZ | | | | | |
| | 20 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | | | | |
| | 15 | 1.2 | 1.3 | 1.3 | 1.3 | 1.3 | | | | | | | | |
| | 10 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | | | |
| | 5 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | | |
| 43° | 0′ | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | |
| | 55 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | | |
| | 50 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | | |
| | 45 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | | | | |
| | 40 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | 1.3 | | | |
| | 35 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | 1.3 | | | |
| | 30 | 1.3 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | | | |
| | 25 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.3 | 1.3 | 1.2 | 1.3 | 1.3 | 1.3 | 1.3 |
| | 20 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 |
| | 15 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.2 |
| | 10 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.3 | 1.2 | 1.2 | 1.2 | 1.3 | 1.2 |
| | 5 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |
| 42° | 0′ | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 | 1.2 |

**Table 4.8** Extract from Loran-C correction tables. (Reproduced courtesy of the Defense Mapping Agency Hydrographic/Topographic Center)

| | | *9960-Y* | | | | | | | | | | | *17Y* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *Longitude west* | | | | | | | | | | | | |
| | | *68°* | | | | | | | | | | | | *67°* |
| LATITUDE NORTH | | *0′* | *55* | *50* | *45* | *40* | *35* | *30* | *25* | *20* | *15* | *10* | *5* | *0′* |
| 45° | 0° | | | | | | | | | | | | | |
| | 55 | | | | | | | | | | | | | |
| | 50 | | | | | | | | | | | | | |
| | 45 | | | LAND | | | | | | | | | 3.4 | |
| | 40 | | | | | | | | | | | | 3.1 | 3.0 |
| | 35 | | | | | | | | | 3.0 | 3.1 | 3.0 | 2.9 | 2.9 |
| | 30 | | | | | | | 2.9 | 2.9 | 2.9 | 2.9 | 2.8 | 2.8 | 2.8 |
| | 25 | | | | | 2.7 | 2.8 | 2.8 | 2.8 | 2.8 | 2.8 | 2.8 | 2.7 | 2.7 |
| | 20 | 2.8 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 | 2.7 | *2.8* | 2.7 | 2.7 | 2.6 | 2.7 | 2.6 |
| | 15 | 2.7 | 2.6 | 2.6 | 2.7 | 2.8 | 2.7 | 2.7 | $\boxed{2.7}$ | 2.7 | 2.6 | 2.6 | 2.6 | 2.6 |
| | 10 | 2.7 | 2.6 | 2.6 | 2.7 | 2.7 | 2.6 | 2.6 | 2.5 | 2.5 | 2.6 | 2.6 | 2.6 | 2.6 |
| | 5 | 2.6 | 2.6 | 2.6 | 2.6 | 2.5 | 2.5 | 2.4 | 2.4 | 2.4 | 2.5 | 2.5 | 2.5 | |
| 44° | 0′ | 2.5 | 2.5 | 2.5 | 2.5 | 2.4 | 2.4 | 2.4 | 2.3 | 2.4 | 2.4 | | | |
| | 55 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | |
| | 50 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | |
| | 45 | 2.4 | 2.4 | 2.3 | 2.4 | 2.4 | 2.4 | | | | | | | |
| | 40 | 2.3 | 2.3 | 2.3 | 2.4 | 2.4 | | | | | | | | |
| | 35 | 2.4 | 2.4 | 2.4 | | | | | | | | | | |
| | 30 | 2.4 | 2.4 | 2.4 | | | | | | | | | | |
| | 25 | 2.4 | 2.4 | 2.4 | | | | | OUTSIDE CCZ | | | | | |
| | 20 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | | | | |
| | 15 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | | | |
| | 10 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | | | |
| | 5 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | | |
| 43° | 0′ | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | |
| | 55 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | | |
| | 50 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | | |
| | 45 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | | |
| | 40 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | |
| | 35 | 2.5 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | |
| | 30 | 2.5 | 2.5 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | | | |
| | 25 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.5 |
| | 20 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| | 15 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| | 10 | 2.4 | 2.3 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| | 5 | 2.4 | 2.3 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |
| 42° | 0′ | 2.4 | 2.3 | 2.3 | 2.4 | 2.3 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 | 2.4 |

The corrected dial readings are used to re-compute a new latitude and longitude for the Loran-C fix. The new position is: 44°15.4′N latitude and 67°26.4′W longitude.

The Loran-C correction tables for a particular chain may be obtained from the US Defense Mapping Agency in the LCPUB221 series.

## 4.5 Position fixing using the Loran-C System

For a particular location covered by more than one Loran-C chain, the operator should select the best chain available, and where possible, select a chain that can be used throughout the voyage so that the receiver can 'lock on' to the signal and 'track' throughout the trip.

Having selected a chain, it is necessary to select secondary stations which give the best fix. There may be a choice of more than two master–secondary station pairs and it is essential to choose those two pairs which give the most accurate fix. Consider Figure 4.14 which shows LOPs for two master–secondary stations, the lines shown for a particular pair being separated by 10 µs.

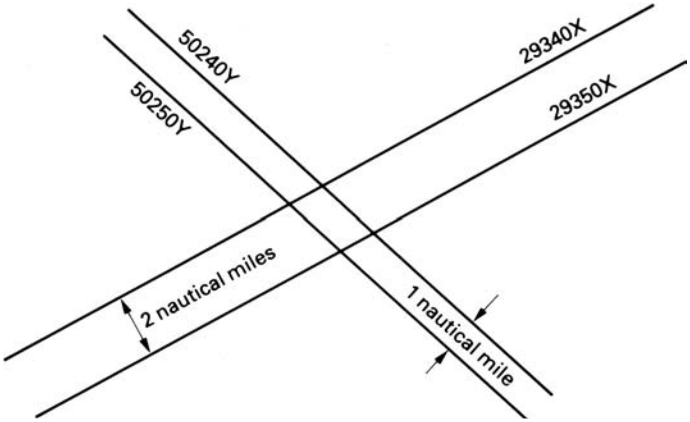


**Figure 4.14** Loran-C gradients and crossing angles.

The distance between the 'Y' lines is 1 nautical mile (1.85 km), while the distance between the 'X' lines is 2 nautical miles (3.7 km). Assuming an error in the loran reading of ±0.01 µs, then the error in terms of distance is ±0.01 nautical miles (18.5 m) for the 'Y' lines and ±0.02 nautical miles (37.5 m) for the 'X' lines.

Consider now Figure 4.15 with much larger gradients for both sets of lines. In this case the order of error, assuming an accuracy of ±0.01 µs as before, is ±166.5 m for the Z lines and ±222.0 m for the W lines. Given gradients as shown in these examples, the X and Y secondaries would be chosen in preference to the W and Z secondaries.

Ideally two LOPs that cross at right angles should always be used since this would give the greatest accuracy. Since this is not always possible to achieve, then LOPs that intersect as close as possible to 90°, such as shown in Figure 4.14, should be used subject of course to suitable values of gradient.

The area in the region of the baseline extension of a master–secondary pair should never be used since, as Figure 4.16 shows, the gradients near these lines become very large, giving rise to potentially very large errors. Baseline extensions are always indicated on the charts.

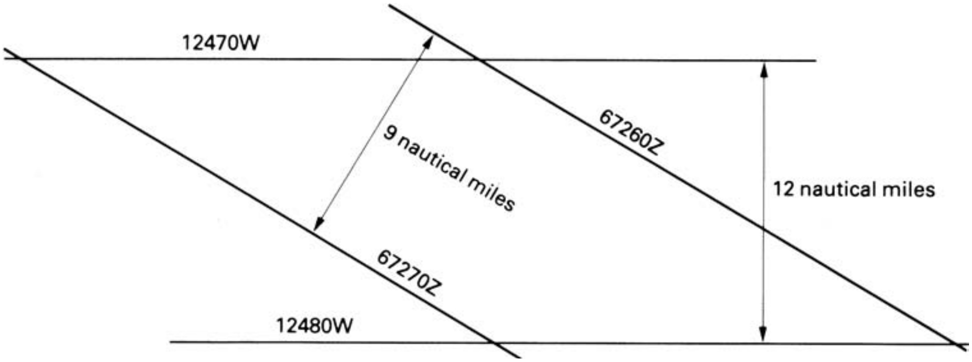


**Figure 4.15** Loran-C gradients and crossing angles.

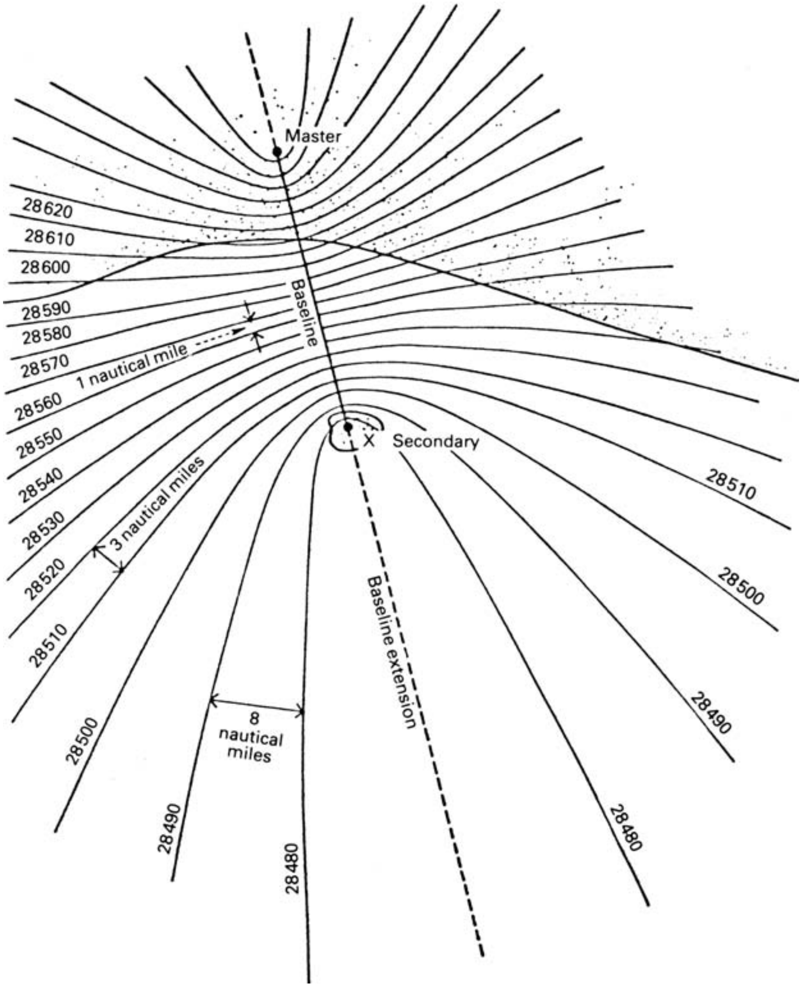


**Figure 4.16** Master/slave station pair LOPs illustrating the large gradients near the baseline extension.

**Table 4.9** Loran-C chain information in WGS 84 co-ordinates

| *Chain* | *Latitude* | *Longitude* | *Emission delay* | *Coding delay* | *Power (kW)* |
|---|---|---|---|---|---|
| *5543 Calcutta* | | | | | |
| M Balasore | 21°29′08.000″N | 86°55′18.000″E | | | 45 |
| W Diamond Harbour | 22°10′18.000″N | 88°12′25.000″E | 18510.68 | 18000 | 11 |
| X Patpur | 20°26′48.000″N | 85°49′47.000″E | 36542.75 | 36000 | 11 |
| *5930 Canadian East Coast* | | | | | |
| M Caribou | 46°48′27.305″N | 67°55′37.159″W | | | 800 |
| X Nantucket | 41°15′12.046″N | 69°58′38.536″W | 13131.88 | 11000 | 350 |
| Y Cape Race | 46°46′32.286″N | 53°10′27.606″W | 28755.02 | 25000 | 1000 |
| Z Fox Harbour | 52°22′35.252″N | 55°42′27.862″W | 41594.59 | 38000 | 900 |
| *5980 Russian-American* | | | | | |
| M Petropavlovsk | 53°07′47.584″N | 157°41′42.900″E | | | 700 |
| W Attu | 52°49′44.134″N | 173°10′49.528″E | 14467.56 | 11000 | 400 |
| X Alexandrovsk | 51°04′42.800″N | 142°42′04.950″E | 31506.50 | 28000 | 700 |
| *5990 Canadian West Coast* | | | | | |
| M Williams Lake | 51°57′58.876″N | 122°22′01.686″W | | | 400 |
| X Shoal Cove | 55°26′20.940″N | 131°15′19.094″W | 13343.60 | 11000 | 560 |
| Y George | 47°03′48.096″N | 119°44′38.976″W | 28927.36 | 27000 | 1400 |
| Z Port Hardy | 50°36′29.830″N | 127°21′28.489″W | 42266.63 | 41000 | 400 |
| *6042 Bombay* | | | | | |
| M Dhrangadhra | 23°00′14.000″N | 71°31′39.000″E | | | 11 |
| W Veraval | 20°57′07.000″N | 70°20′13.000″E | 13862.41 | 13000 | 11 |
| X Billamora | 20°45′40.000″N | 73°02′073.02″E | 40977.61 | 40000 | 11 |
| *6731 Lessay* | | | | | |
| M Lessay | 49°08′55.224″N | 01°30′17.029″W | | | 250 |
| X Soustons | 43°44′23.029″N | 01°22′49.584″W | 13000 | 10992.53 | 250 |
| Y Loop Head | 52°35′03.000″N | 09°49′06.000″W | 27300 | 24968.61 | 250 |
| Z Sylt | 54°48′29.975″N | 08°17′36.856″E | 42100 | 39027.54 | 250 |
| *6780 China South Sea* | | | | | |
| M Hexian | 23°58′03.847″N | 111°43′10.298″E | | | 1200 |
| X Raoping | 23°43′25.951″N | 116°53′44.826″E | 14464.69 | 12700 | 1200 |
| Y Chongzuo | 22°32′35.452″N | 107°13′21.665″E | 26925.76 | 25300 | 1200 |
| *7001 Bø* | | | | | |
| M Bø | 68°38′06.216″N | 14°27′47.350″E | | | 400 |
| X Jan Mayen | 70°54′51.478″N | 08°43′56.525″W | 14100 | 11014.42 | 250 |
| Y Berlevåg | 70°50′43.014″N | 29°12′15.980″E | 29100 | 27032.68 | 250 |
| *7030 Saudi Arabia South* | | | | | |
| M Al Khamasin | 20°28′02.025″N | 44°34′52.894″E | | | 1000 |
| W Salwa | 24°50′01.631″N | 50°34′12.574″E | 13620.00 | 11000 | 1000 |
| X Afif | 23°48′36.952″N | 42°51′18.184″E | 27265.00 | 26000 | 1000 |
| Y Ash Shaykh Humayd | 28°09′15.997″N | 34°45′40.544″E | 41414.00 | 40000 | 1000 |
| Z Al Muwassam | 16°25′56.028″N | 42°48′04.884″E | 57664.00 | 56000 | 1000 |
| *7270 Newfoundland East Coast* | | | | | |
| M Comfort Cove | 49°19′53.570″N | 54°51′42.570″W | | | 250 |
| W Cape Race | 46°46′32.286″N | 53°10′27.606″W | 12037.49 | 11000 | 500 |
| X Fox Harbour | 52°22′35.252″N | 55°42′27.862″W | 26148.01 | 25000 | 900 |

**Table 4.9** Continued

| *Chain* | *Latitude* | *Longitude* | *Emission delay* | *Coding delay* | *Power (kW)* |
|---|---|---|---|---|---|
| *7430 China North Sea* | | | | | |
| M Rongcheng | 37°03′51.765″N | 122°19′25.954″E | | | 1200 |
| X Xuancheng | 31°04′07.937″N | 118°53′09.625″E | 13459.70 | 11000 | 1200 |
| Y Helong | 42°43′11.562″N | 129°06′27.213″E | 30852.32 | 28000 | 1200 |
| *7499 Sylt* | | | | | |
| M Sylt | 54°48′29.975″N | 08°17′36.856″E | | | 250 |
| X Lessay | 49°08′55.224″N | 01°30′17.029″W | 14100 | 11027.54 | 250 |
| Y Værlandet | 61°17′49.435″N | 04°41′46.618″E | 29500 | 26986.19 | 250 |
| *7950 Eastern Russia ‘Chayka’* | | | | | |
| M Alexsandrovsk | 51°04′42.800″N | 142°42′04.950″E | | | 700 |
| W Petropavlovsk | 53°07′47.584″N | 157°41′42.900″E | 14506.50 | 11000 | 700 |
| X Ussuriisk | 44°31′59.702″N | 131°38′23.403″E | 33678.00 | 30000 | 700 |
| Y Tokachibuto | 42°44′37.214″N | 143°43′09.757″E | 49104.15 | 46000 | 600 |
| Z Okhotsk | 59°25′02.050″N | 143°05′22.916″E | 64102.05 | 61000 | 10 |
| *7960 Gulf of Alaska* | | | | | |
| M Tok | 63°19′42.884″N | 142°48′31.346″W | | | 550 |
| X Narrow Cape | 57°26′20.301″N | 152°22′10.708″W | 13804.45 | 11000 | 400 |
| Y Shoal Cove | 55°26′20.940″N | 131°15′19.094″W | 29651.14 | 26000 | 550 |
| Z Port Clarence | 65°14′40.372″N | 166°53′11.996″W | 47932.52 | 44000 | 1000 |
| *7980 Southeast U.S.* | | | | | |
| M Malone | 30°59′38.870″N | 85°10′08.751″W | | | 800 |
| W Grangeville | 30°43′33.149″N | 90°49′43.046″W | 12809.54 | 11000 | 800 |
| X Raymondsville | 26°31′55.141″N | 97°49′59.539″W | 27443.38 | 23000 | 400 |
| Y Jupiter | 27°01′58.528″N | 80°06′52.876″W | 45201.88 | 43000 | 400 |
| Z Carolina Beach | 34°03′46.208″N | 77°54′46.100″W | 61542.72 | 59000 | 800 |
| *7990 Mediterranean Sea* | | | | | |
| M Sellia Marina | 38°52′20.707″N | 16°43′06.713″E | | | 165 |
| X Lampedusa | 35°31′20.912″N | 12°31′30.799″E | 12755.98 | 11000 | 325 |
| Y Kargabarun | 40°58′21.066″N | 27°52′.02.074″E | 32273.29 | 29000 | 165 |
| Z Estartit | 42°03′36.629″N | 03°12′16.066″E | 50999.71 | 47000 | 165 |
| *8000 Western Russian* | | | | | |
| M Bryansk | 53°07′50.600″N | 34°54′44.800″E | | | 1150 |
| W Petrozavodsk | 61°45′32.400″N | 33°41′40.400″E | 13217.21 | 10000 | 1150 |
| X Slonim | 53°07′55.200″N | 25°23′46.000″E | 27125.00 | 25000 | 1150 |
| Y Simferopol | 44°53′20.600″N | 33°52′32.100″E | 53070.25 | 50000 | 1150 |
| Z Syzran (Karachev) | 53°17′17.600″N | 48°06′53.400″E | 67941.60 | 65000 | 1150 |
| *8290 North Central U.S.* | | | | | |
| M Havre | 48°44′38.589″N | 109°58′53.613″W | | | 400 |
| W Baudette | 48°36′49.947″N | 94°33′17.915″W | 14786.56 | 11000 | 800 |
| X Gillette | 44°00′11.305″N | 105°37′23.895″W | 29084.44 | 27000 | 400 |
| Y Williams Lake | 51°57′58.876″N | 122°22′01.686″W | 45171.62 | 42000 | 400 |
| *8390 China East Sea* | | | | | |
| M Xuancheng | 31°04′07.937″N | 118°53′09.625″E | | | 1200 |
| X Raoping | 23°43′25.951″N | 116°53′44.826″E | 13795.52 | 11000 | 1200 |
| Y Rongcheng | 37°03′51.765″N | 122°19′25.954″E | 31459.70 | 29000 | 1200 |
| *8830 Saudi Arabia North* | | | | | |
| M Afif | 23°48′36.952″N | 42°51′18.184″E | | | 1000 |
| W Salwa | 24°50′01.631″N | 50°34′12.574″E | 13645.00 | 11000 | 1000 |
| X Al Khamasin | 20°28′02.025″N | 44°34′52.894″E | 27265.00 | 25000 | 1000 |
| Y Ash Shaykh Humayd | 28°09′15.997″N | 34°45′40.544″E | 42645.00 | 40000 | 1000 |
| Z Al Muwassam | 16°25′56.028″N | 42°48′04.884″E | 58790.00 | 56000 | 1000 |

**Table 4.9** Continued

| *Chain* | *Latitude* | *Longitude* | *Emission delay* | *Coding delay* | *Power (kW)* |
|---|---|---|---|---|---|
| *8930 North West Pacific* | | | | | |
| M Niijima | 34°24′11.943″N | 139°16′19.473″E | | | 1000 |
| W Gesashi | 26°36′25.038″N | 128°08′56.920″E | 15580.86 | 11000 | 1000 |
| X Minamitorishima | 24°17′08.007″N | 153°58′53.779″E | 36051.53 | 30000 | 1100 |
| Y Tokachibuto | 42°44′37.214″N | 143°43′09.757″E | 53349.53 | 50000 | 600 |
| Z Pohang | 36°11′05.450″N | 129°20′27.440″E | 73085.64 | 70000 | 150 |
| *8970 Great Lakes* | | | | | |
| M Dana | 39°51′07.658″N | 87°29′11.586″W | | | 400 |
| W Malone | 30°59′38.870″N | 85°10′08.751″W | 14355.11 | 11000 | 800 |
| X Seneca | 42°42′50.716″N | 76°49′33.308″W | 31162.06 | 28000 | 800 |
| Y Baudette | 48°36′49.947″N | 94°33′17.915″W | 47753.74 | 44000 | 800 |
| Z Boise City | 36°30′20.783″N | 102°53′59.487″W | 63669.46 | 59000 | 800 |
| *9007 Ejde* | | | | | |
| M Ejde | 62°17′59.837″N | 07°04′26.079″W | | | 400 |
| W Jan Mayen | 70°54′51.478″N | 08°43′56.525″W | 14200 | 10983.83 | 250 |
| X Bø | 68°38′06.216″N | 14°27′47.350″E | 28000 | 23951.92 | 400 |
| Y Værlandet | 61°17′49.435″N | 04°41′46.618″E | 41100 | 38997.27 | 250 |
| Z Loop Head | 52°35′03.000″N | 09°49′06.000″W | 55700 | 52046.62 | 250 |
| *9610 South Central U.S.* | | | | | |
| M Boise City | 36°30′20.783″N | 102°53′59.487″W | | | 800 |
| V Gillette | 44°00′11.305″N | 105°37′23.895″W | 13884.48 | 11000 | 400 |
| W Searchlight | 35°19′18.305″N | 114°48′16.881″W | 28611.81 | 25000 | 550 |
| X Las Cruces | 32°04′18.130″N | 106°52′04.388″W | 42044.93 | 40000 | 400 |
| Y Raymondsville | 26°31′55.141″N | 97°49′59.539″W | 56024.80 | 52000 | 400 |
| Z Grangeville | 30°43′33.149″N | 90°49′43.046″W | 69304.00 | 65000 | 800 |
| *9930 East Asia* | | | | | |
| M Pohang | 36°11′05.450″N | 129°20′27.440″E | | | 150 |
| W Kwang Ju | 35°02′23.966″N | 126°32′27.295″E | 11946.97 | 11000 | 50 |
| X Gesashi | 26°36′25.038″N | 128°08′56.920″E | 25565.52 | 22000 | 1000 |
| Y Niijima | 34°24′11.943″N | 139°16′19.473″E | 40085.64 | 37000 | 1000 |
| Z Ussuriisk | 44°31′59.702″N | 131°38′23.403″E | 54162.44 | 51000 | 700 |
| *9940 U.S. West Coast* | | | | | |
| M Fallon | 39°33′06.740″N | 118°49′55.816″W | | | 400 |
| W George | 47°03′48.096″N | 119°44′38.976″W | 13796.90 | 11000 | 1600 |
| X Middletown | 38°46′57.110″N | 122°29′43.975″W | 28094.50 | 27000 | 400 |
| Y Searchlight | 35°19′18.305″N | 114°48′16.881″W | 41967.30 | 40000 | 550 |
| *9960 Northeast U.S.* | | | | | |
| M Seneca | 42°42′50.716″N | 76°49′33.308″W | | | 800 |
| W Caribou | 46°48′27.305″N | 67°55′37.159″W | 13797.20 | 11000 | 800 |
| X Nantucket | 41°15′12.046″N | 69°58′38.536″W | 26969.93 | 25000 | 400 |
| Y Carolina Beach | 34°03′46.208″N | 77°54′46.100″W | 42221.65 | 39000 | 800 |
| Z Dana | 39°51′07.658″N | 87°29′11.586″W | 57162.06 | 54000 | 400 |
| *9990 North Pacific* | | | | | |
| M Saint Paul | 57°09′12.350″N | 170°15′06.245″W | | | 325 |
| X Attu | 52°49′44.134″N | 173°10′49.528″E | 14875.25 | 11000 | 625 |
| Y Port Clarence | 65°14′40.372″N | 166°53′11.996″W | 32068.95 | 29000 | 1000 |
| Z Narrow Cape | 57°26′20.301″N | 152°22′10.708″W | 46590.45 | 43000 | 400 |

## 4.6 Loran-C coverage

Loran-C coverage is dependent on land-based transmitters grouped into chains. The current information relating to the chains, their group repetition interval (GRI), location, emission and coding delay and nominal radiated power is shown in Table 4.9.

Diagrams are available which show the predicted ground wave coverage for each chain. Briefly the coverage diagrams are generated as follows.

- **Geometric-fix accuracy limits**. Each of two LOPs in a chain is assigned a TD standard deviation of 0.1 μs. The geometric-fix accuracy is assigned a value of 1500 feet, $2d_{\text{RMS}}$ where $d_{\text{RMS}}$ is the radial or root mean square error. Using these constraints a contour is generated within the chain area representing the geometric-fix accuracy limits.
- **Range limits**. Predicted atmospheric noise and cross-rate Loran-C interference is compared with estimated Loran-C signal strength for each Loran-C transmitting station to obtain an expected 1:3 SNR (signal-to-noise ratio) range limits for each transmitted signal.
- **Predicted accuracy**. The predicted Loran-C coverage for each chain is the result of combining the geometric-fix accuracy limits and predicted SNR range limits. Where the geometric-fix accuracy limits extend beyond the range limits, the range limits are used on the coverage diagrams and vice versa.

Figure 4.17 shows the $2d_{\text{RMS}}$ coverage for various station pairs in the Northeast US (NEUS) chain. Diagram A, for example, shows the accuracy contours for the master–whiskey and the master–yankee station pairs. The solid line in the diagrams show the $2d_{\text{RMS}}$ contour of 1500 ft absolute accuracy, the dashed line 1000 ft and the dotted line 500 ft. Similar diagrams for other pair combinations are also shown in Figure 4.17.

A composite coverage diagram for the NEUS (9960) chain is shown in Figure 4.18.

Associated with each chain (not shown in Figure 4.18) are unmanned monitor sites (lormansites) which continuously check the loran signals received to detect any out-of-tolerance conditions so that corrections can be relayed back to the transmitting site for implementation of those corrections.

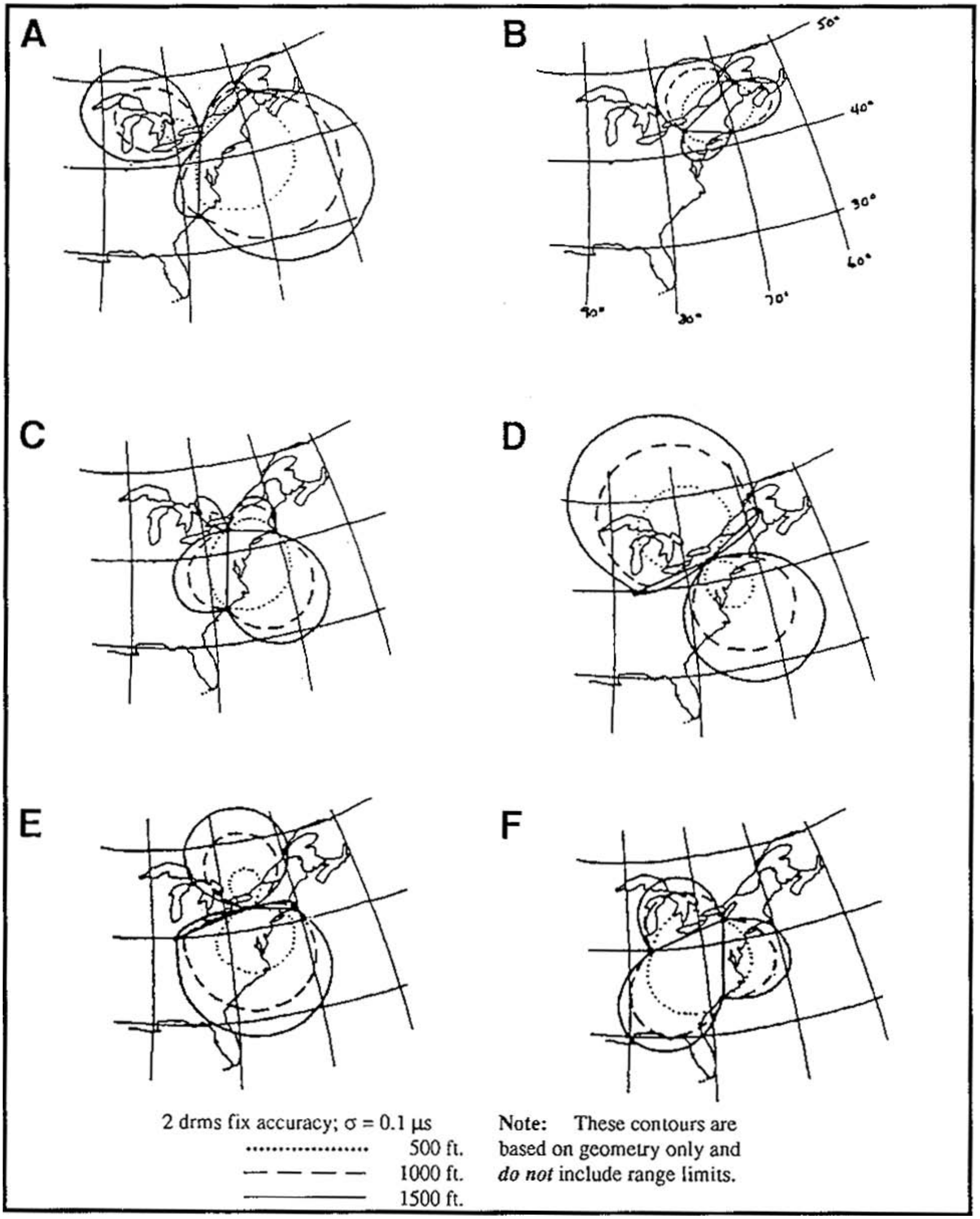
Clarinet Pilgrim (CP) and Clarinet Pilgrim with TTY2 is a system used, at specified stations, where certain pulses in each group are subject to pulse position modulation of ± 1 μs to provide back-up administrative and control signals.

Radial or root mean square error, $d_{\text{RMS}}$, is defined as the radius of the error circle produced from the square root of the sum of the square of the sigma error components along the major and minor axes of a probability ellipse (see Figure 4.19). The ellipse is produced by virtue of the deviation expected along each LOP as indicated by $\delta 1$ and $\delta 2$ in Figure 4.22, and varies according to the gradient and angle of cut of the LOPs at that point.

$1d_{\text{RMS}}$ is defined as the radius of a circle obtained when $\delta x = 1$, and $\delta y$ varies from 0 to 1. $2d_{\text{RMS}}$ is defined as the radius of a circle obtained when $\delta x = 2$ and $\delta y$ varies from 0 to 2. The relationship between $\delta 1$,$\delta 2$ and $\delta x$,$\delta y$ and the probability values associated with $1d_{\text{RMS}}$ or $2d_{\text{RMS}}$ values are beyond the scope of this book but may be obtained from standard reference books.

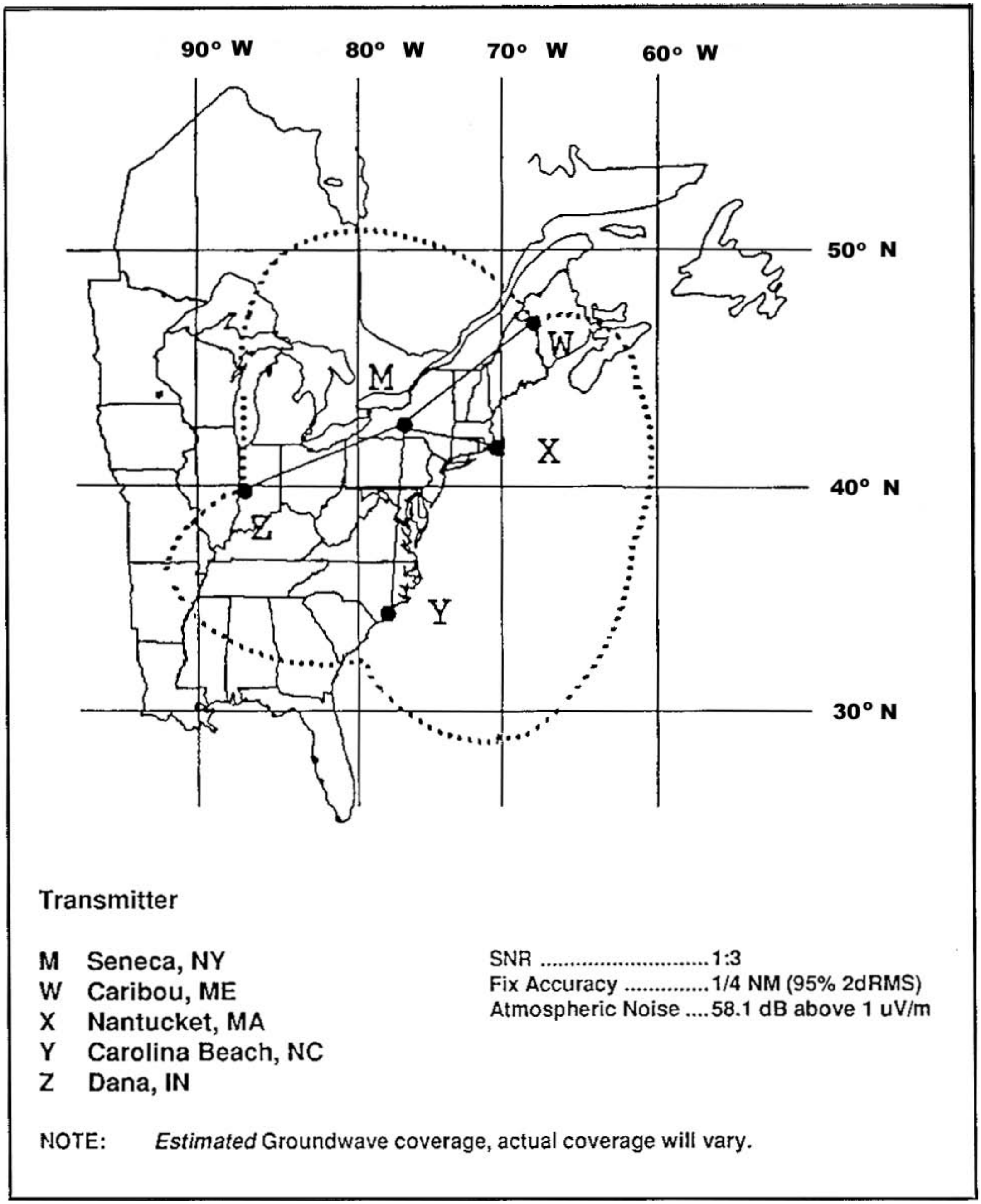
As far as the accuracy of Loran-C coverage is concerned the coverage diagram (Figure 4.18) shows that for ground wave reception areas, the fix probability is 95% ($2d_{\text{RMS}}$) at 1500 ft with a standard deviation of 0.1 μs and 1/3 SNR. Sky wave reception will extend the coverage area but accuracy cannot be guaranteed.

For the Loran-C system the absolute accuracy, i.e. the ability to determine the true geographic position (latitude and longitude), is claimed to be from 0.1 to 0.25 nautical mile (185–463 m) depending on the position of the receiver within the coverage area. Repeatable accuracy is the measure



**Note:** These contours are based on geometry only and *do not* include range limits.

**Figure 4.17** Contours of equal $2d_{\mathrm{rms}}$ for various triads in the 9960 Loran-C chain.



**Transmitter**

| | |
|---|---|
| **M** | **Seneca, NY** |
| **W** | **Caribou, ME** |
| **X** | **Nantucket, MA** |
| **Y** | **Carolina Beach, NC** |
| **Z** | **Dana, IN** |

SNR ............................ 1:3
Fix Accuracy ............. 1/4 NM (95% 2dRMS)
Atmospheric Noise .... 58.1 dB above 1 uV/m

NOTE: *Estimated* Groundwave coverage, actual coverage will vary.

**Figure 4.18** Loran-C GRI 9960 Northeast US (NEUS) chain.

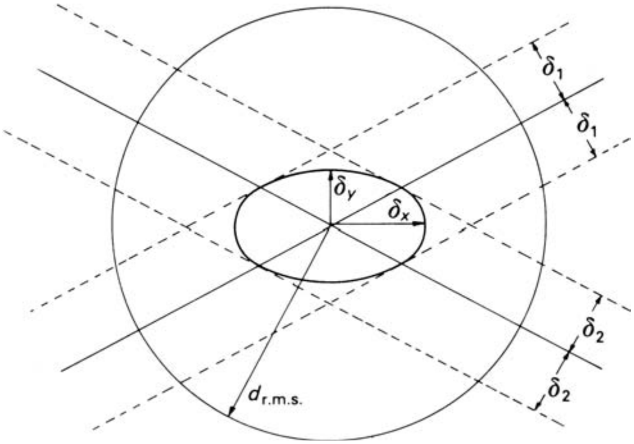


**Figure 4.19** The error ellipse.

of the ability to return to a previously plotted position, time and time again by using Loran-C readings for that position as a reference. For Loran-C the repeatable accuracy is claimed to be from 0.008 to 0.05 nautical mile (15–90 m). The global Loran-C coverage is shown in Figure 4.20.

Mariners should consult relevant local Notice to Mariners, whereby official notification of changes to the Loran-C system can be found.

## 4.7 Loran-C receivers

A Loran-C receiver which is capable of measuring position with the claimed accuracy for the system should possess the following characteristics.

- Acquire the Loran-C signals automatically.
- Identify master and secondary ground wave pulses automatically, and accomplish cycle matching on all eight pulses for each master–secondary pair used.
- Track the signals automatically once acquisition has been achieved.
- As a minimum requirement, display two time-difference readings, to a precision of at least 0.1 μs.
- Incorporate notch filters, adjusted by the manufacturer if required, to minimize the effects of radio frequency interference in the area in which the user expects to operate.

With some older Loran-C receivers it is necessary to select the chain and station pairs during the set-up process. Newer receivers possess an automatic initialization process whereby the operator enters the vessel's latitude/longitude and the receiver selects the best chain and station pairs for that position. This automatic selection process can be overridden if necessary. Having selected a suitable master and secondaries, the system should then acquire the signals with sufficient accuracy to permit settling and tracking to occur. Settling involves the detection of the leading edge of the signal pulse and the selection of the third cycle of the pulse for tracking purposes. Tracking involves the maintenance of the synchronization of the third cycle of the master and secondary signals. The time taken for the receiver to complete the 'acquire–settle–track' process will depend on the characteristics of the receiver and the S/N ratio of the received signals.

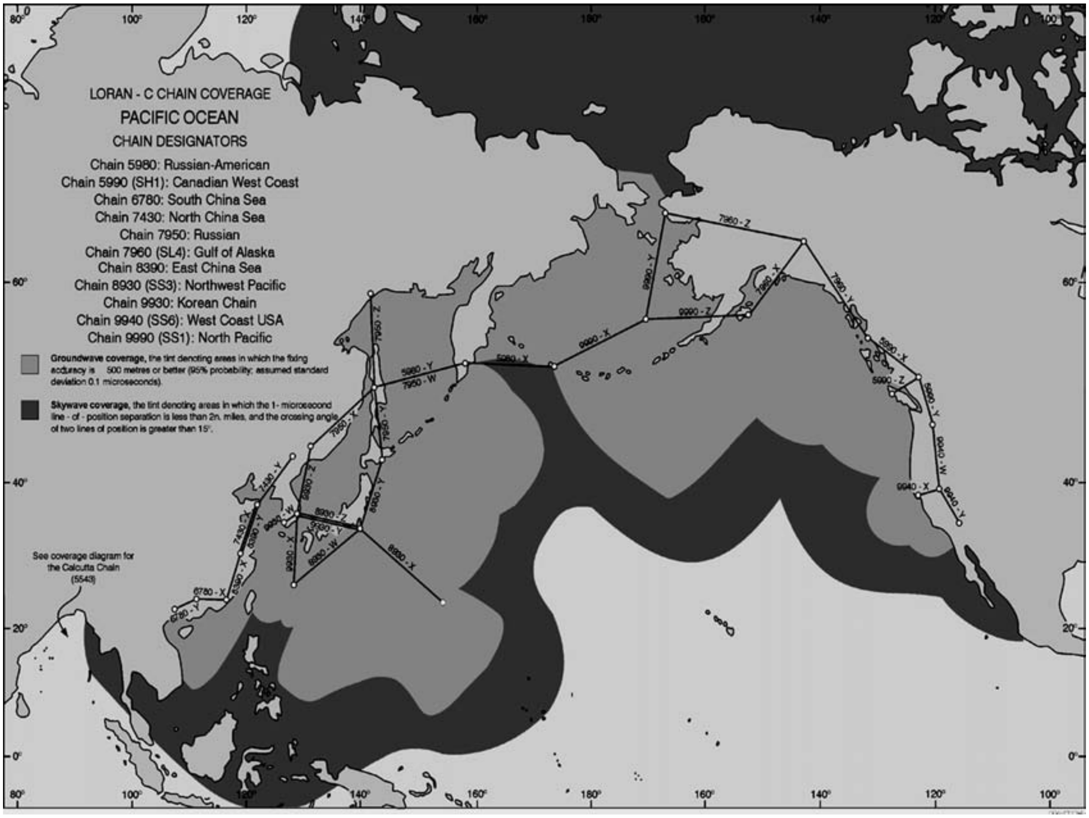


**Figure 4.20a** Loran-C global coverage. (Reproduced from Admiralty List of Radio Signals volume 2 by permission of the Controller of Her Majesty's Stationery Office and the UK Hydrographic Office.)

Signal reception may be impaired by interference from other signals which could act as a noise input and reduce the S/N ratio of the received loran signal and degrade positional accuracy. Notch filters within the receiver can assist in minimizing the effect of the interference. The notch filters may be either preset by the manufacturer or be adjustable on site.

Modern Loran-C receivers are designed with a front panel that contains a display element (usually a liquid crystal display (LCD) which is easily read under all lighting conditions and energy efficient) and a keypad with function keys and numeric keys to enter data and change the data displayed. Displays will indicate information such as: status and warning data; information on the GRI in use and the secondaries chosen; alarm settings; positional information in time differences (TDs) or as latitude/longitude and navigation information such as waypoint indicators; bearing and distance to waypoint; time to go (TTG); cross-track errors (XTE); speed and course etc. Some displays may use pages of information that can be selected as required by the operator. Time differences are measured by the receiver and may be converted to latitude/longitude by computer algorithms; such algorithms would most likely incorporate additional secondary factor (ASF) corrections, which are stored in the computer memory.

Modern receivers have the facility for the operator to monitor the progress of the voyage and allow for course corrections as necessary. The receiver gives a position (in TD or latitude/longitude) and has a precise clock so that it is possible to produce navigational information, such as vessel's speed and

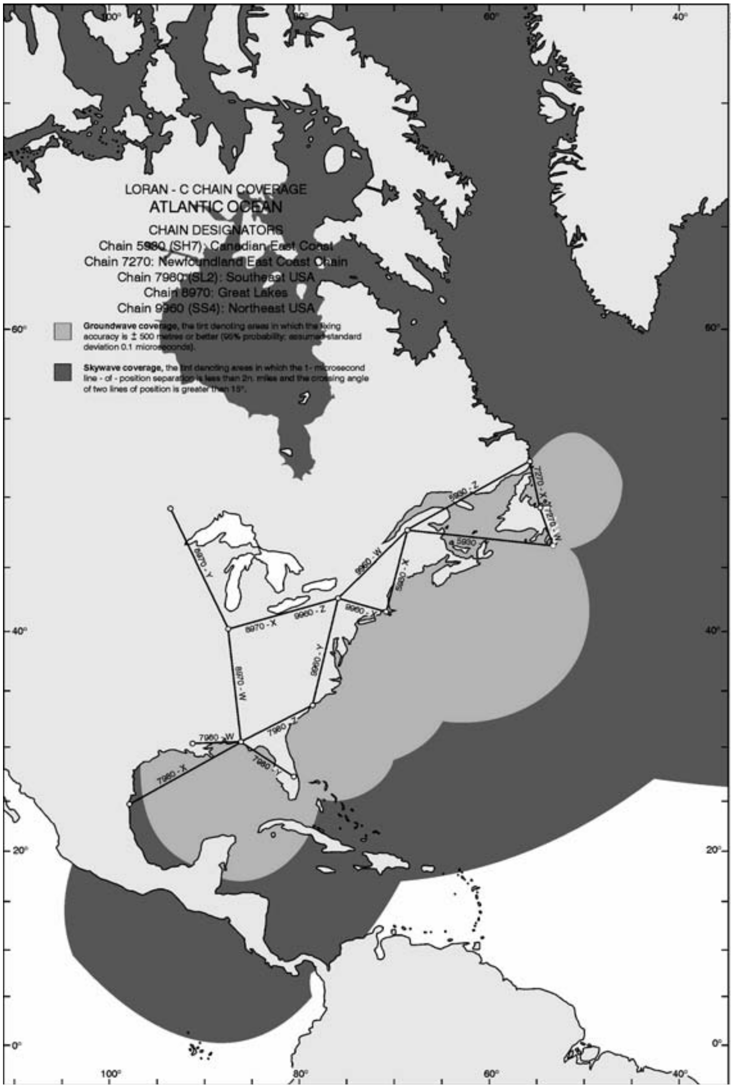


**Figure 4.20b** (continued).

course. A waypoint is a set of co-ordinates that indicate a location of interest to the navigator, such as wrecks, buoys, channel information, and previously productive fishing areas. Waypoints can usually be stored in the receiver memory by entering the waypoint co-ordinates or as a distance and bearing from another waypoint before pressing the appropriate control button. Waypoints may be used by the navigator as route indicators for a planned route. The receiver can track progress between waypoints allowing the operator to monitor data, such as bearing to the next waypoint, time-to-go (TTG) to reach the next waypoint, and cross-track error (XTE). The latter indicates a deviation from the planned course and shows the perpendicular distance from present position to the intended track between waypoints.

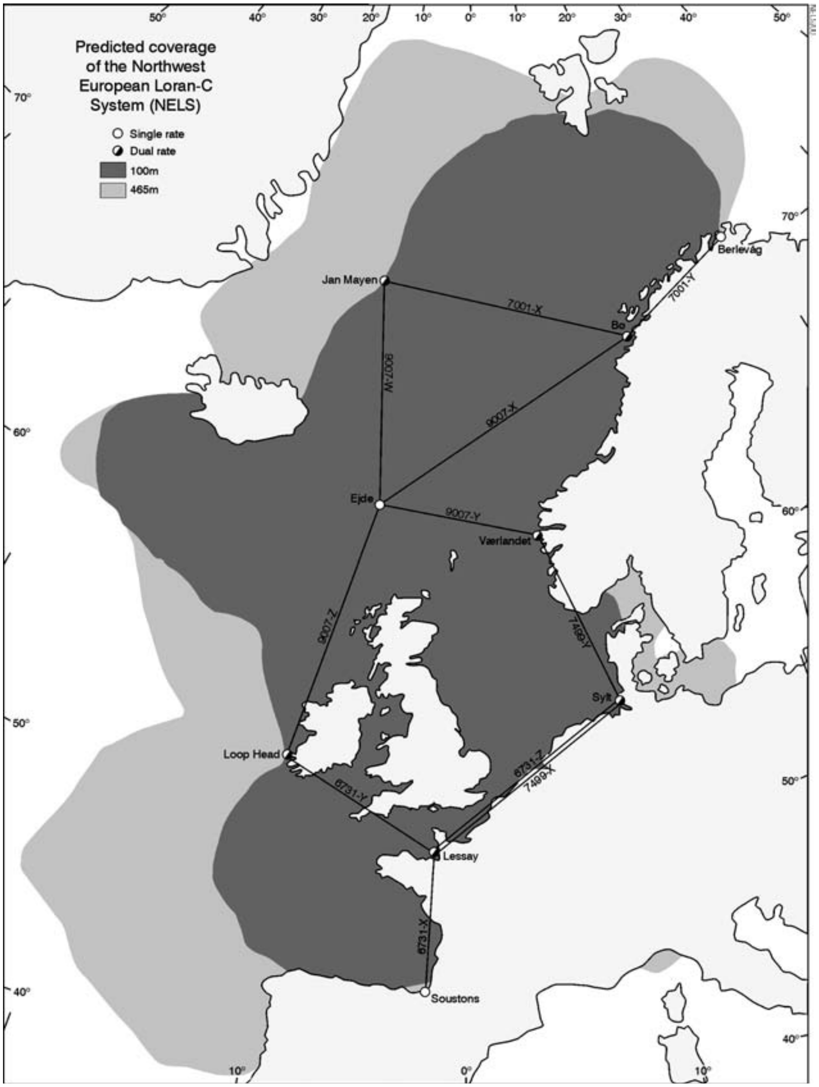


**Figure 4.20c** (continued).

In addition, magnetic variation data apposite to the loran coverage area may be stored in memory allowing the operator to navigate with reference to either true or magnetic north. The use of magnetic north would be indicated by some means on the display to inform the operator that directions are with reference to magnetic and not true north.

Loran receivers may stand alone or be integrated with other equipment, such as a plotter or GPS (Global Positioning System). In addition, modern receivers are able to provide outputs to other electronic equipment using protocols such as the NMEA (National Marine Electronics Association) 0180, 0182, 0183 and 2000 formats where applicable. Such outputs may thus be connected to autopilots, plotters, radars etc, while it is also possible to connect with a gyrocompass and speed log to enable the set and drift of the current to be determined.

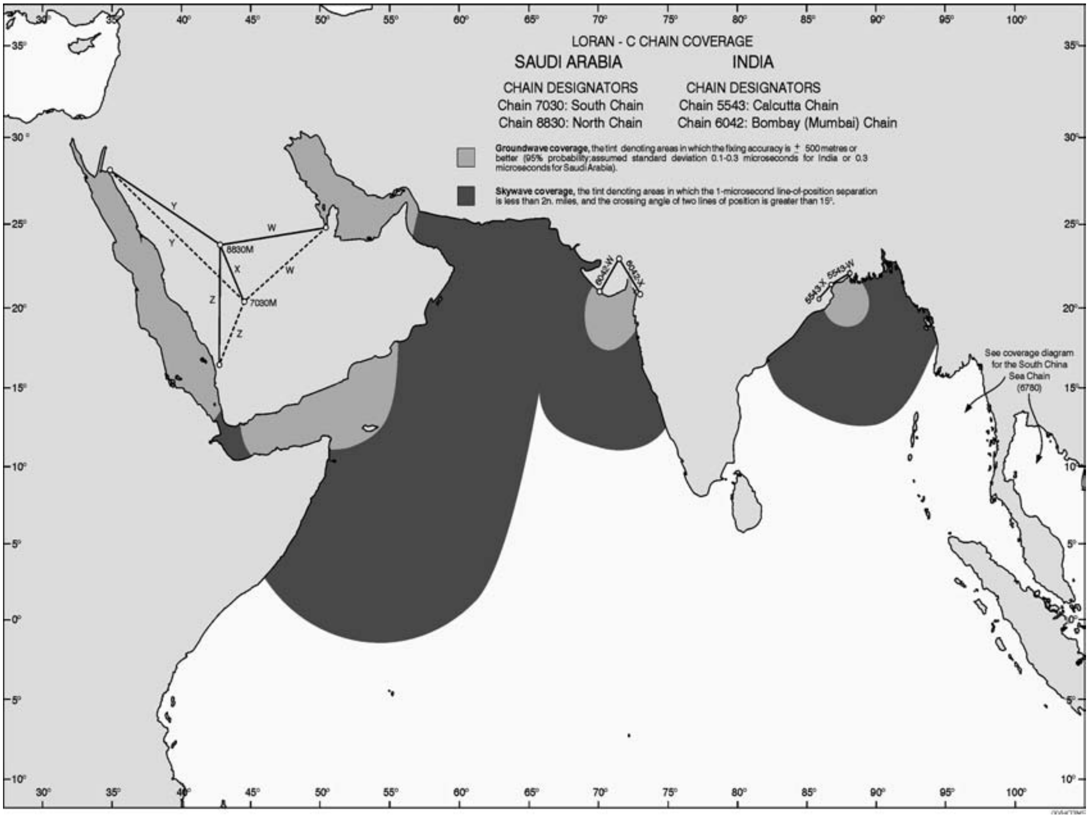


**Figure 4.20d** (continued).

A typical automatic receiver is illustrated in Figure 4.21. This is the Koden Electronics LR-707 receiver. Although this receiver is of an older generation of receivers, the way in which it operates is no different to its more modern counterparts. Details of the functions offered by this receiver and its operation are described in the following paragraphs.

### 4.7.1 Station selection

Switches S1 and S2 control the two time displays. When the receiver is first initialized (see p. 124) display 2 will be rolling, i.e. displaying various secondary time differences in an ascending sequence. The roll frequency is once every 3 s. When the required secondary time difference appears on display 2, pressing switch 2 will retain that output. If it is required to change the chosen secondary time difference, pressing S2 again will restore the roll action. S1 serves the same function as S2 except that it controls display 1.

An exception to the functions performed by the two switches is that if display 1 is adjusted for roll, display 1 will indicate all time differences including that being shown by display 2. With S1 adjusted for non-roll, S2 will indicate a time difference reading other than that indicated on display 1, i.e. it will skip that time difference. As a result of this feature, when only two secondaries are acquired (or available) and Sl is adjusted for non-roll, S2 will also appear to be adjusted for non-roll. The S1 display is also used to indicate certain alarm functions and to supply technical data.

**Figure 4.21** Koden Electronics LR-707 Loran-C receiver.

### *Function switch*

When initializing the receiver the function switch must be set to SEL. After the settling alarms have been extinguished, the function switch should be placed in NORM position to inhibit cycle selection of all stations and to enable certain functions of +/MEMO and –/RECALL (see under the appropriate heading for a description of these functions). Setting the function switch in either Sl, M or S2 position allows the time difference of the selected station to be manipulated using the +/MEMO and –/RECALL controls.

The cycle selection process is inhibited for all other stations except the selected station. With the function switch set to SEL, the cycle selection process is activated for all stations. In addition, the +/MEMO and –/RECALL buttons will jump all stations by 10 µs depending on the button chosen and the number of times it is pressed. If the control is left in this position, the readings should return to correct values provided propagation conditions are normal and the +/MEMO and –/RECALL buttons have not been pressed excessively, which would cause the tracking point to move off the pulse. Simultaneously pressing +/MEMO and –/RECALL will initialize the receiver.

+/MEMO

- With the function switch in TEST and +/MEMO pressed, the display will indicate the oscillator offset frequency. Pressing the button again will restore the normal technical information to the display.
- With the function switch set to SEL, the tracking points of all stations will shift by +10 µs each time the button is pressed.
- With the function switch in NORM, pressing the +/MEMO button will ‘freeze’ the display and place all acquired time differences into memory. Pressing +/MEMO again will restore the display to time difference readings.

- With the function switch in Sl, pressing the +/MEMO button will cause the tracking point of the station appearing on display 1 to move +10 μs.
- With the function switch set to M, pressing the +/MEMO button will cause the tracking point of the master station to move by +10 μs, causing S1 and S2 display to indicate 10 μs lower.

–/RECALL

- With the function switch set to NORM, pressing the –/RECALL button will recall and display all time differences previously entered into the memory. Pressing the button again will restore display to the normal tracking mode.

Notch Filters
These controls are used to eliminate interference that is sinusoidal. When not in use, two should be tuned fully clockwise and two tuned fully anticlockwise, or improper operation may result (see page 122).

Tune Control
Used in conjunction with the tune meter to locate interference.

Tune Noise Meter
Together with Tune Control this meter will locate interference. It does not indicate signal strength of the Loran-C signal.

Signal-to-noise Alarms
When lit, these indicate a possible problem with the associated station. When operating at great distances from the station or under adverse weather conditions, these alarms may light from time to time. Simultaneous flashing of all three alarms indicates that the RECALL control has been activated.

Settling Alarms
These indicate that the associated station is settled and is ready for tracking. Simultaneous flashing of all three alarms indicates that the +/MEMO control has been activated.

Dimmer Control
This controls the intensity of both displays and all six LED alarms.

Chain Selector
This must be used prior to initializing the receiver to select the required Loran-C chain. To determine the setting of the required GRI number, reference should be made to the appropriate Loran-C chart for the area of operation. Only the first three digits of the chain identification need be set since the last (fourth) digit is always zero.

### 4.7.2 Normal operation

The chain selector should be set for the chain of the area in which the vessel is operating. Next the function switch should be set to SEL and the notch filters detuned by setting two of them completely clockwise and two completely anticlockwise. The dimmer switch should also be set fully clockwise.

The power switch should then be turned ON and after about 4 s both displays should sequentially indicate all secondaries acquired. When the required time difference appears on the display, the wanted secondaries can be selected by pressing display control S1 and S2. When the settling alarms are no longer alight the function switch should be set to NORM. The unit will then have acquired the wanted signals and will track those signals.

### *Use of the notch filters*

Rotate Tune Control and check for signal interference. When Tune Control is in the '6-o'clock' position, it indicates the centre of the loran frequency and the tune meter should indicate a reasonably large deflection. When rotated either side of the central position, the tune meter should indicate a smaller deflection. Any 'bouncing' or increased deflection of the meter indicates the presence of noise.

Noise may be eliminated by using the notch filters which are highly tuned circuits and can sharply reduce the signal level of any frequency if the filters are tuned to that frequency. Thus, if Tune control finds any interfering signals in the frequency range of the loran signals, the notch filter controls may be adjusted to eliminate that interference. The technique to be used is described as follows.

(a) Turn all notch filter knobs fully clockwise.
(b) Set the Tune Control knob to its centre ('6 o'clock') position and note the deflection on the tune meter which is an indication of the loran signal.
(c) Turn the Tune Control knob slowly anticlockwise and note the abrupt deflection on the tune meter. A similar effect should be found if the knob is rotated slowly clockwise. See Figure 4.22(a).
(d) Set the Tune Control knob to the point where the meter deflection is greatest in the anticlockwise direction.
(e) Reset notch filter knob N1 to the centre position and slowly rotate it anticlockwise until the meter deflection is minimized.
(f) Check that the meter deflection for the interference signal is less than the loran signal and if not, repeat steps (d) and (e) above using the notch filter N2.
(g) Reset the Tune Control knob to its centre position and slowly rotate clockwise until the interference frequency below the loran centre frequency is found.
(h) Reset notch filter knob N3 to its centre position and slowly rotate clockwise until the meter deflection is minimized.
(i) Use the notch filter N4, by turning it clockwise from its centre position, if the use of notch filter N3 has not reduced the interference signal level below that of the wanted loran signal.
(j) Repeat step (c) and note that the levels of the interference signals are reduced below the level of the loran signal above and below the loran centre frequency. See Figure 4.22(b).

### *Receiver alarm indications*

The various alarms that are possible with this receiver as an aid to the operator are as follows.

(a) **Secondary blink**. This is indicated when the third and fourth digit of either display is flashing. During the blink alarm, only the time difference reading of the secondary station at fault will flash. This station should not therefore be used for position fixing. The blink alarm will not automatically reset itself. When two or more secondaries are flashing, it is usually an indication of problems with the master station and all time difference readings should be used with extreme caution.

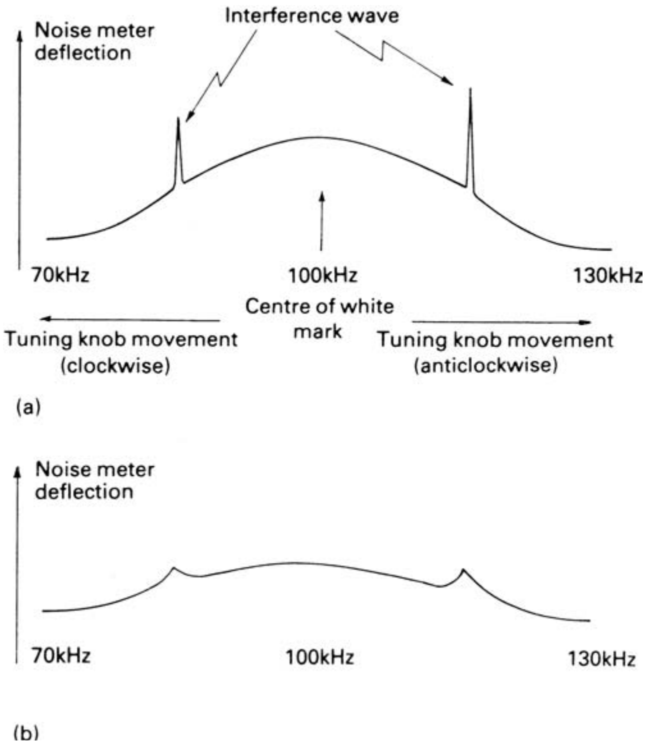


**Figure 4.22** (a) Possible interference levels prior to adjustment of the tuning controls.(b) Possible interference levels after adjustment of the tuning controls. The interference level should always be set to less than the level of the loran signal.

(b) **Test alarm**. When the function switch is set to TEST, the second digit of both displays will flash once every 3 s.
(c) **Memo alarm**. When the display is 'frozen' by activating the +/MEMO button, all three settling alarms will flash once every 3 s.
(d) **Recall alarm**. When the –/RECALL control has been activated, all three signal-to-noise alarms will flash once every 3 s.
(e) **Function switch alarm**. If the function switch was in any position other than SEL when power was applied to the receiver, the number 9 will appear in the window of each display. To correct, the receiver should be turned OFF and the function switch set to SEL before restoring the supply.
(f) **Signal-to-noise alarms**. When lit, the signal-to-noise alarms indicate a poor signal-to-noise ratio. If the alarm is lit for 50% or more of the time, the tracking capabilities on the problem station will be severely impaired or, in some cases, impossible to track. This alarm should be ignored during the settling process.
(g) **Settling alarm**. This alarm will light any time the cycle selection circuit is not satisfied with a decision. Since propagation conditions are variable, the settling alarm may light even though the displayed time difference reading is correct. With the function switch in the NORM position, no 10-μs jump will occur even though a jump is indicated by the settling alarm. If the function switch is in the SEL position, a jump will occur automatically.

To cancel any alarm function, first turn the function switch to TEST and then back to NORM. If the receiver detects that the alarm condition still exists, the alarm will, after a short delay, become active again.

### *Other functions of the display*

When the function switch is moved to TEST, both displays will automatically indicate the number 8 in each position, allowing the operator to check that all display segments are operating correctly. In addition the cycle selection alarms and the signal-to-noise alarms will light. This type of display will remain for 3 s, then the following information becomes available.

- The first two digits of each display indicate the station under observation. The first two digits correspond to the first two numbers of the respective time differences (99 indicates the master).
- The second digit will flash, indicating a non-tracking condition for the displays.
- The third digit of the time difference indicates the signal condition at the beginning of the pulse.
- The fourth digit indicates the signal condition at the tracking point of the pulse.
- The fifth digit indicates the signal condition at the crest of the pulse. In each of these cases ‘0’ is the lowest value and ‘9’ is the highest value.
- The sixth digit indicates the mode of the receiver, with 4 being the final and tracking mode.

When the +/MEMO button is pressed, the oscillator offset frequency is displayed on a scale of –20 to +20, with ‘0’ indicating no offset. Press again to reset. If zeros appear in the first to fourth digits, the frequency is low. The converse is true if no zeros appear.

### *Initializing procedure*

The receiver is initialized in four stages. The modes can be checked by the sixth digit when the function switch is set to TEST.

- **Mode 1 stage.** For those loran signals received from a chain, the GRI of which is preset on the front panel, the received pulse is compared with an internally generated pulse. The master pulse is detected first and then the secondary pulses. The time difference values are displayed on the display panels.
- **Mode 2 stage.** The detection and tracking of the zero cross point of the carrier is commenced and the tracking point is transferred to the start of the loran signal in 10-μs steps until the signal becomes zero in the signal-to-noise detection circuit; the noise indicator lamps will then light. This operation is performed for master and secondary stations independently.
- **Mode 3 stage.** The tracking point is now transferred in the signal direction in 10-μs steps until the signal is detected. The noise indicator lamps should now be extinguished. Once again the operation is performed independently for master and secondary stations. The function of modes 2 and 3 is to ensure that the pulses for master and secondary stations are overlapped correctly, i.e. pulse 1 of the master is overlapped with pulse 1 of the secondary station.
- **Mode 4 stage.** At the end of mode 3 stage, the 10-μs step operation switches from the signal-to-noise detection circuit to the third-wave detection circuit. The tracking point is now set to the correct tracking position, namely the point after the third wave as seen from the pulse leading edge. The set indicator lamps are then extinguished. The initialization operation is now complete. When the function switch is set to SEL, the check operation continues and if the circuit decides that the

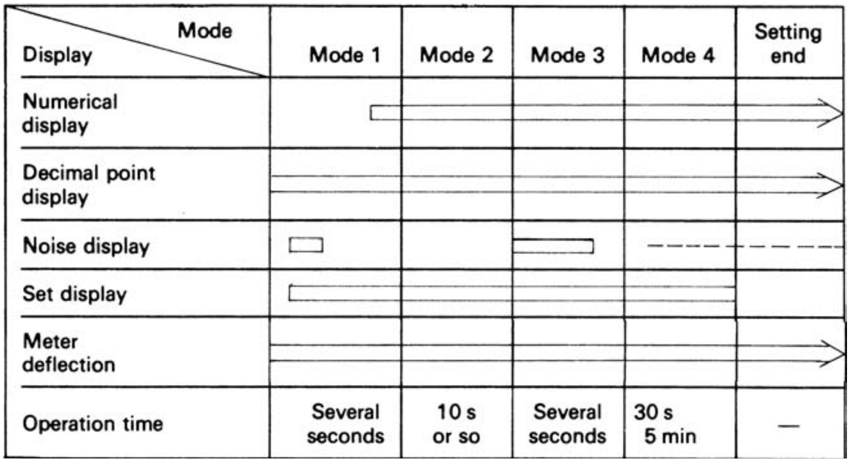


**Figure 4.23** Initialization and lighting/extinction of indicator lamps.
Notes: (1) It is possible that if the signal level is lower than -20dB or the S/N ratio (SNR) is very low, mode 1 will not proceed to the next stage, and no display appears. (2) The noise indicator lamp may light during mode 4 operation or after setting ends if the S/N ratio (SNR) is too low.

previously determined position is incorrect, the 10-μs step sequence is re-started and the set indicator lamps are then lit. The indicator lamps will only be extinguished after a second setting-up routine has been performed.

Refer to the diagram of lamp lighting sequence for the initialization routine (Figure 4.23).

### 4.7.3 Circuit description

This receiver uses a microprocessor and associated logic circuitry to detect and track the Loran-C 100-kHz pulse trains from master and secondary transmitter stations. The system also presents the time differences between the receipt of the master and secondary pulse trains as a direct visual display. Two such time differences can be indicated which give a position fix as the point of intersection of the two time difference lines (LOPs). The microprocessor used is a Motorola 6800 equivalent. The means of detection, sampling and tracking the signals is initiated by the use of interrupt signals IRQ (interrupt request) and NMI (non-maskable interrupt). A basic block diagram is shown in Figure 4.24.

The antenna coupler provides some filtering and gives some initial amplification prior to connection to the receiver block. The receiver block provides bandpass filtering and amplification. Separate circuits are provided for the CYCLE and ENVELOPE outputs, with the signals hard limited to give digitized values. The bandpass filter allows for a restriction on the received signal frequencies to a range of 70–130 kHz. The notch filter can be used to minimize the effects of noise signals within the pass band. The logic block is shown in more detail in Figure 4.25.

The incoming CYCLE signal to the logic block is fed to a sampling circuit consisting of 50-bit shift registers and a D-type flip-flop. The shift registers are integrated circuits 9C, 10C, 12C and 13C while the flip-flop is integrated circuit 8C.

The loran signal format is eight pulses of 100 kHz, each pulse lasting for 250 μs. The signal, after passing through the receiver and being hard limited, appears as digital pulses. The pulse train is

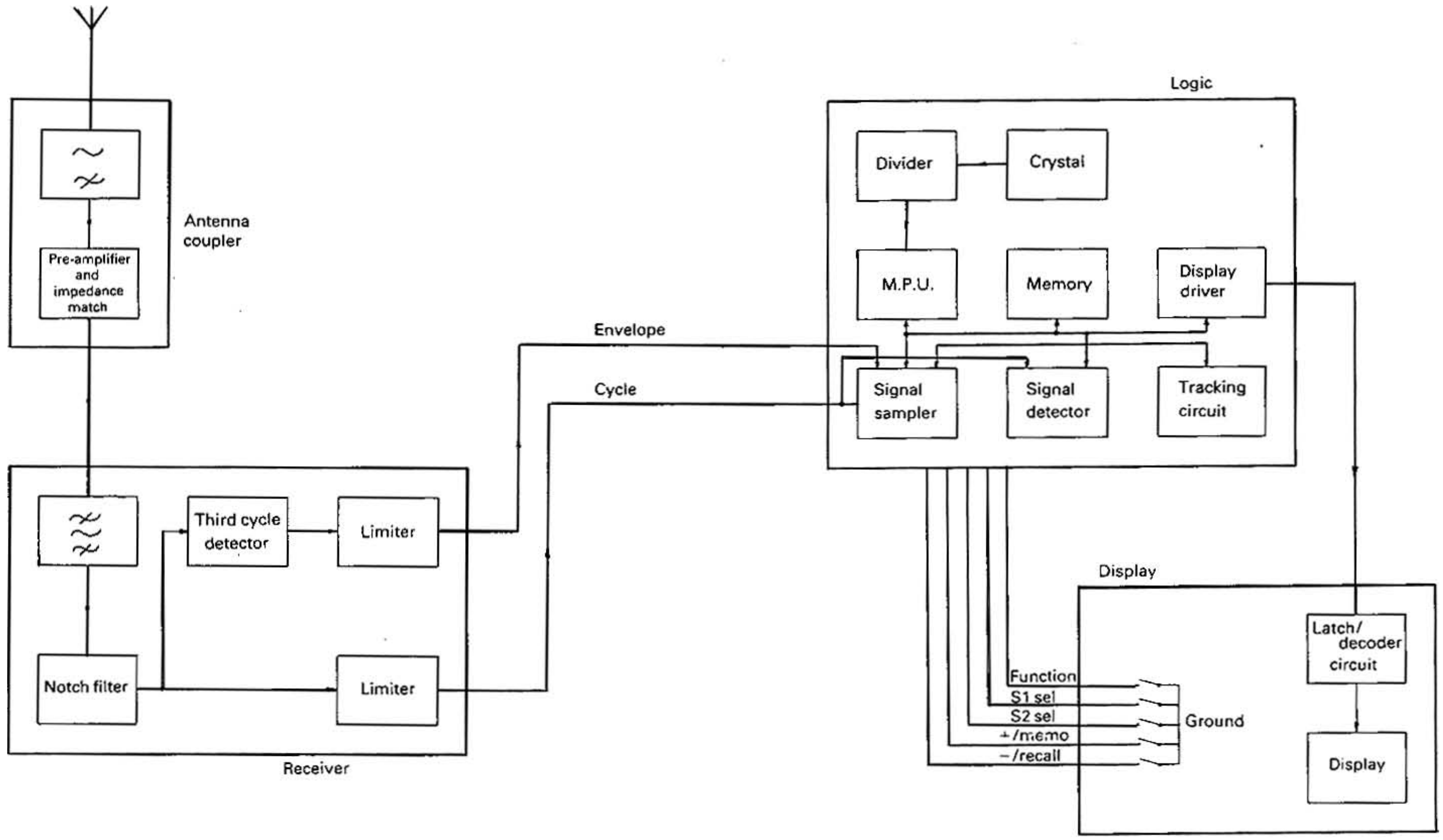


**Figure 4.24** Basic block diagram of Loran-C receiver.

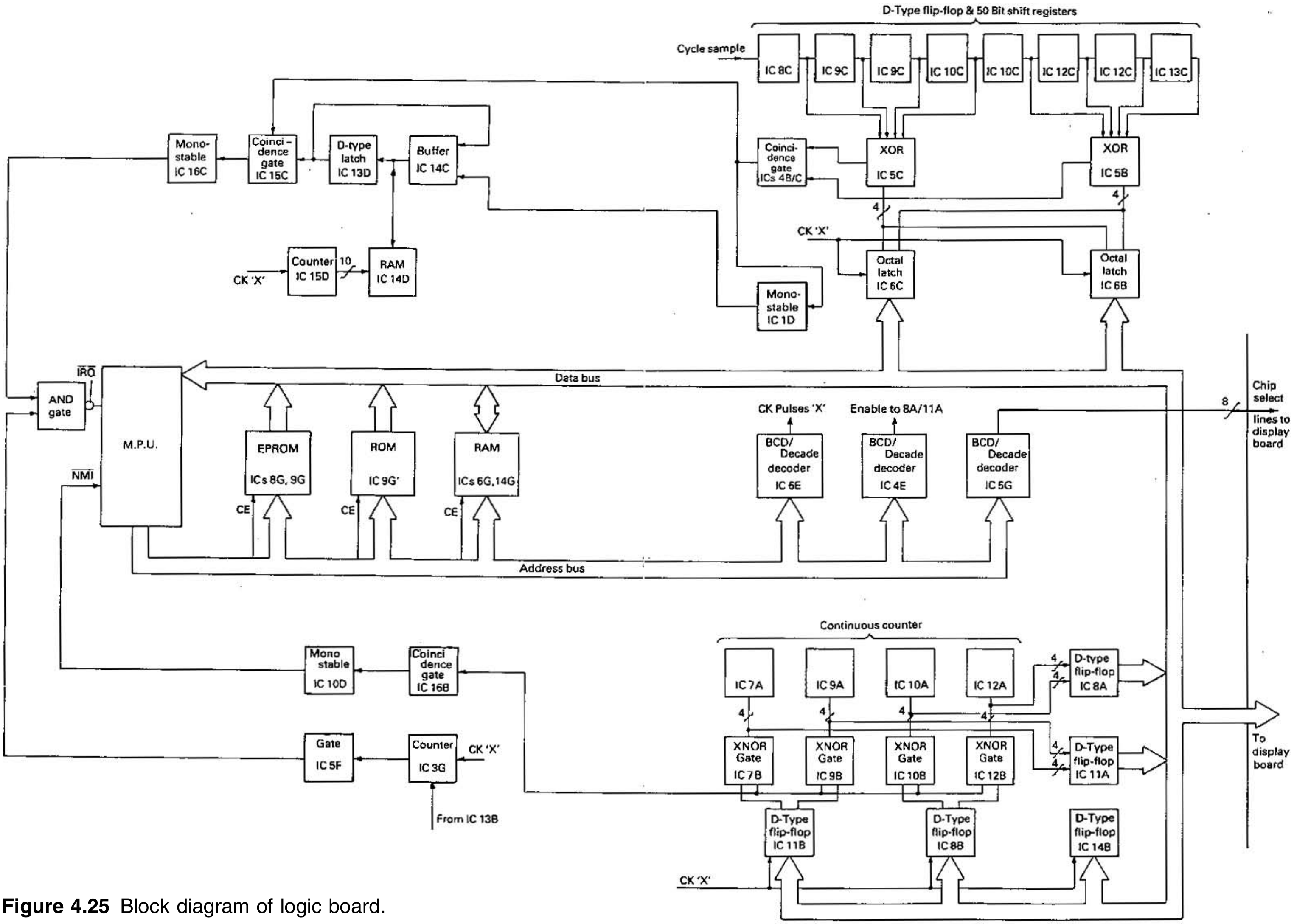


**Figure 4.25** Block diagram of logic board.

clocked through the shift registers using a clock-pulse duration of average value 20 µs, so that for each pulse period of the received pulse train of 1000 µs there are 50 bits. These bits are shifted continuously through the registers recording the presence, or absence, of a pulse as the case may be. As an example, considering master transmissions only and with a GRI of, say, 79600 µs, then after the reception of the eighth pulse (ignoring the ninth pulse for the moment) there is a time difference of (79600 – 7250) µs or 72350 µs, before the next pulse is received. Obviously the receipt of secondary station signals will occur during this period.

Considering only master signals for the moment, the phase coding of the eight pulses has the form:

Group A + + – – + – + –
Group B + – – + + + + +

and the phase code interval (PCI) has the form A, B, A, B etc., for successive master transmissions.

For the receipt of a master pulse train the CPU can cause the A and B code to be latched into a D-type octal latch. The required code could be output and compared with a sample from each of the outputs of the 50-bit registers (and D-type flip-flop) as shown in Figure 4.26.

The coincidence of a sampled signal with, say, the A code results in an output signal from the quad XOR gates that is logic '0' and this, through the gate circuit shown, results in a logic '1' output from the gate circuit. The coincidence output is shaped by a monostable circuit and fed via a buffer circuit to the input of a RAM 1024 × 4 bit memory circuit. The memory circuit address 000–3FF (0–1024) is selected by the output of a binary counter. The rate of data input to the memory address locations,

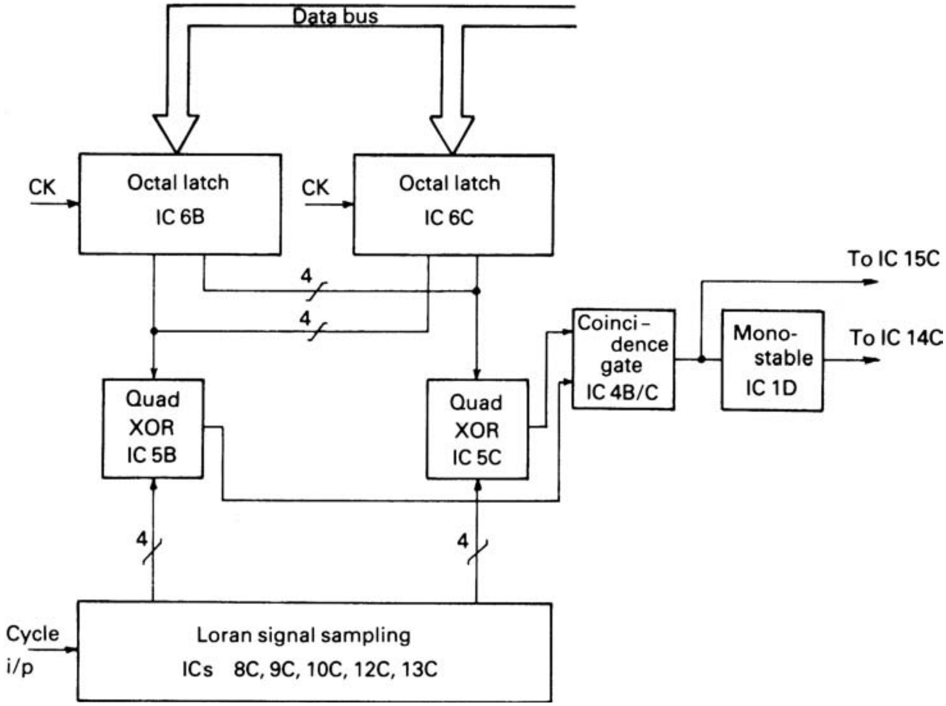


**Figure 4.26** Sampling and coincidence circuit.

governed by the counter, is 100 μs. Since the counter has 1024 locations to access before resetting there is a total of 102 400 μs to be represented by the 1024 bits of the memory. The bits corresponding to GR1 can be represented within the memory since no GRI exceeds 99 900 μs and the counter is reset every GRI.

After the memory has been loaded for 1GRI and the counter is reset, the procedure is repeated for 2GRI, 3GRI etc. The memory chip is configured as four rows of 1024 locations and for each address location data is latched from row 1 to row 2, row 2 to row 3, and row 3 to row 4 as new data are written into row 1. This means that row 1 is used for the latest GRI with the previous GRI latched into row 2; the GRI before that is in row 3 etc. Thus the results of four previous GRIs are held in the memory and these results are available on the output data lines as each memory is accessed. These four previous GRI outputs are checked, together with outputs of present GRI, in a coincidence gate (see Figure 4.27).

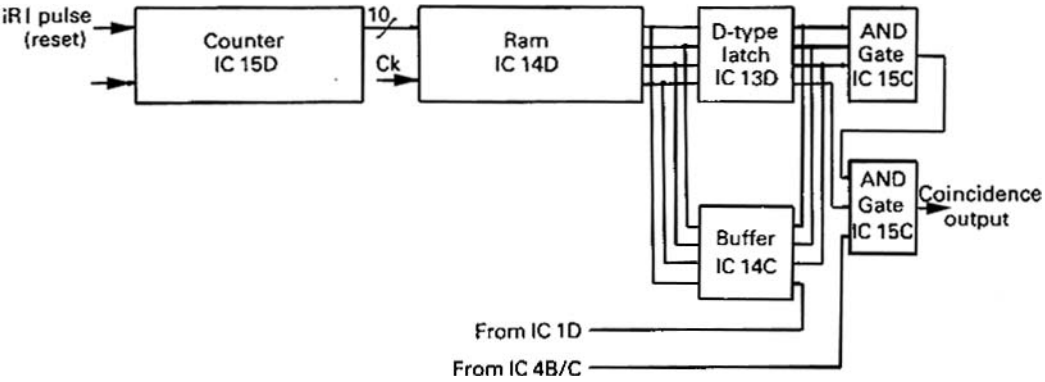


**Figure 4.27** Production of coincidence (IRQ) circuit.

The output of the coincidence AND-gate arrangement is only a logic '1' if all inputs are logic '1', i.e. the data output of each address location accessed for the four rows (each of the four previous GRI) and the data for the present GRI are all identical at logic '1'.

After pulse shaping, the coincidence output is used as an input to the interrupt request (IRQ) input of the MPU chip. The receipt of an IRQ input causes the microprocessor to finish any current instruction and to move to a high order address location where the starting address of a required subroutine is stored. In this case the routine causes the MPU to read the flip-flop (IC 8C) in the sampling circuit to determine whether the master signal was detected by an A or B code. This determines whether the phase coding of the secondaries should be A or B code. The code for the secondaries is as follows:

Group A + + + + + − − +

Group B + − + − + + − −

The interrupt subroutine causes the correct code to be latched into the phase code latch circuit (ICs 6B and 6C) ready to detect the received secondary signals, which are processed in exactly the same way as described above for the master signals.

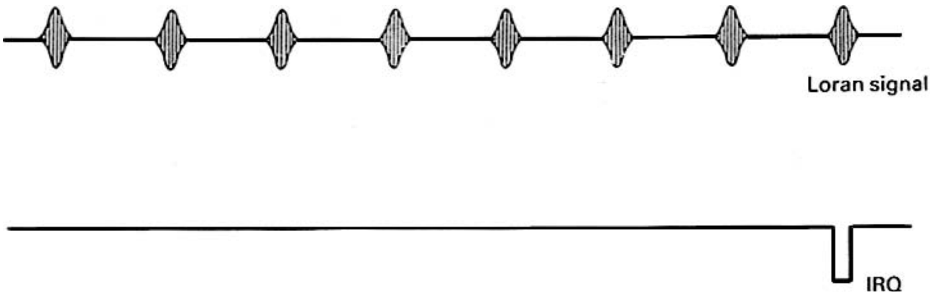


**Figure 4.28** Timing of IRQ pulse with received loran signal.

The MPU outputs the GRI pulse which resets the counter (IC 15D) ready for the next GRI input data sequence. The coincidence output (GRI signal) is also used to latch the outputs of a continuous counter to the data bus, using D-type flip-flops (ICs 8A and 11A), and to the MPU. The timing of the IRQ pulse in relation to the loran signal is shown in Figure 4.28.

The MPU calculates the time 2 ms before the first position of the loran signal in the next GRI cycle to set the values of octal D-type flip flops (ICs 8B, 11B and 14B). The MPU waits until that time for an interrupt. When the values set in the D-type flip-flops coincide with the values of the continuous counter then an NMI (non-maskable interrupt) signal is sent to the MPU (see Figure 4.29).

The NMI interrupt performs a similar function to that of the IRQ signal in that a jump to a subroutine is initiated. The difference is that the IRQ request will only be obeyed if the interrupt mask bit in the MPU flag register is not set. The NMI request will always be obeyed since the interrupt mask bit has no effect on NMI. On receipt of the NMI signal the MPU clears a counter (IC 3G) which masks

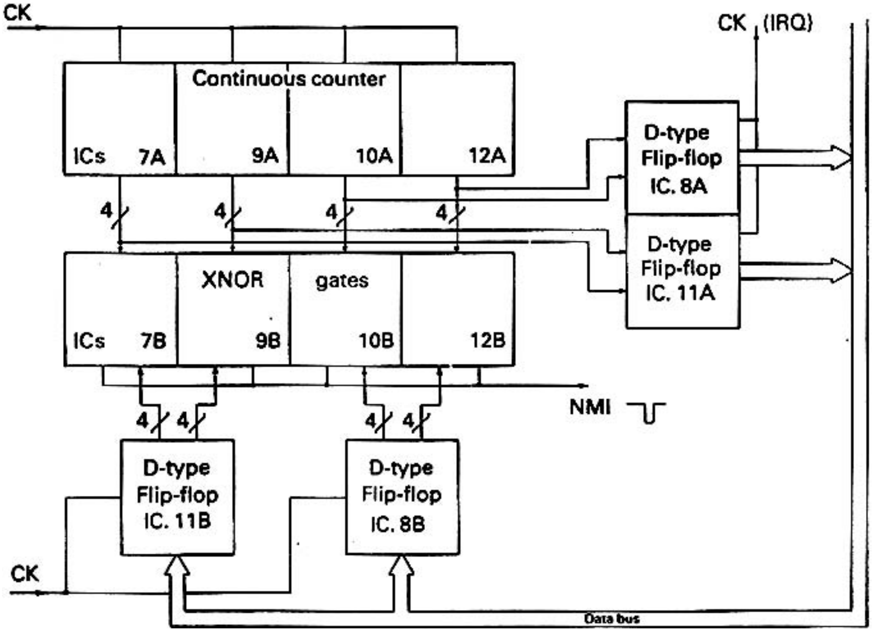


**Figure 4.29** Production of the NMI interrupt pulse.

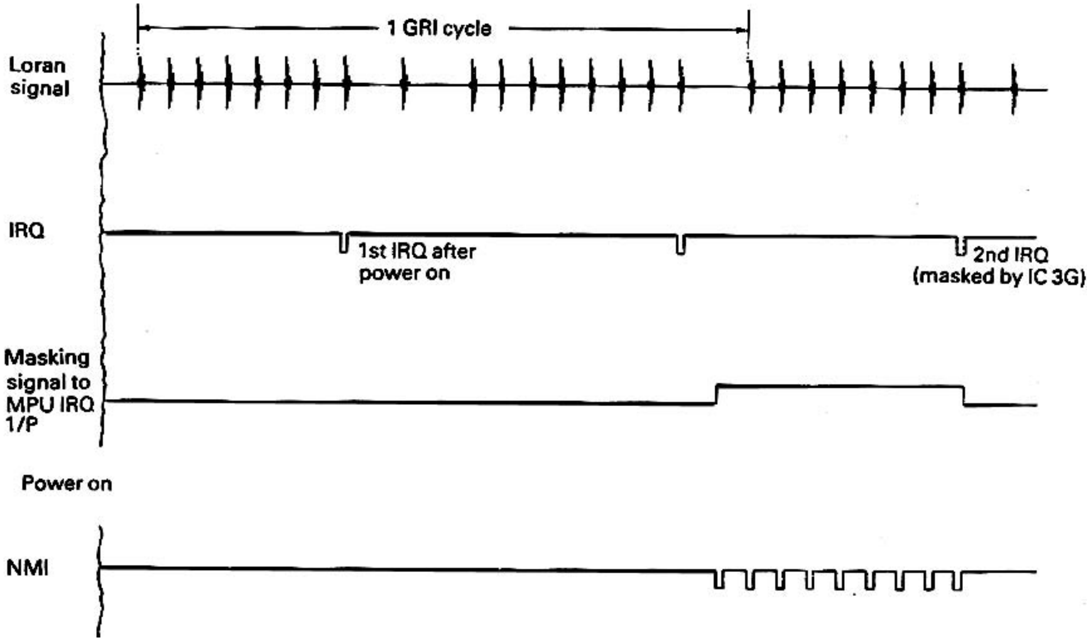


**Figure 4.30** Timing diagram showing the 'masking' of the IRQ input when NMI is generated.

the IRQ interrupt for about 9 ms [* (read data) + 1GRI – approximately 9 ms = time about 2 ms before the position of loran signal in the next cycle]. See Figure 4.30.

The loran signal tracking point is set by the MPU by adding 2 ms to the value obtained previously (see * above) and setting this value, via the data bus, to the octal D-type flip-flops. The MPU then waits for an interrupt, which recurs after 2 ms and coincides with the reception of the first loran pulse.

The first loran signal pulse is sampled when the MPU outputs a CYCLE ENV pulse (see Figure 4.31).

After sampling, the MPU adds 1 ms to the previously set value of the octal D-type flip-flops and waits for another interrupt pulse. This second loran pulse is sampled as for the first pulse. This

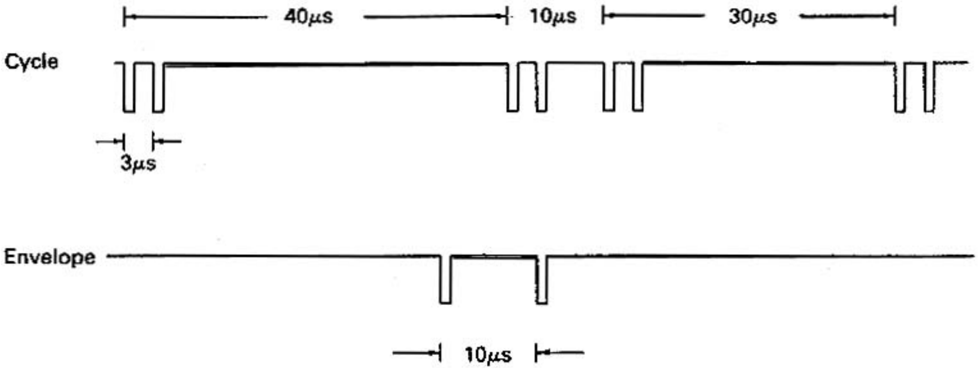


**Figure 4.31** Timing diagram for loran signal sampling.

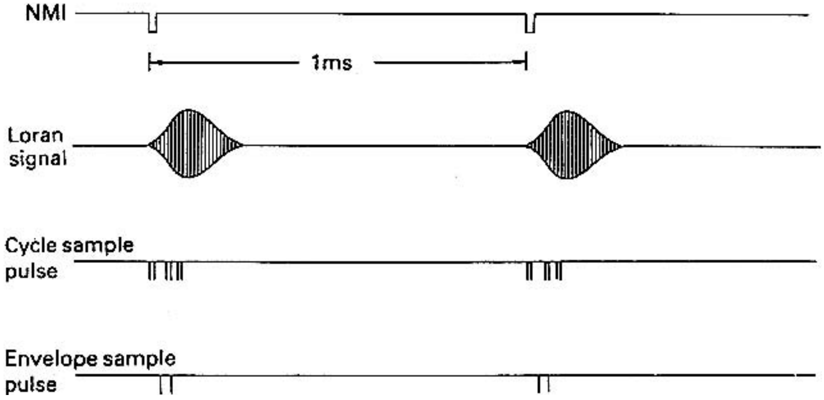


**Figure 4.32** Timing sequence for NMI pulses.

procedure is repeated for all the loran pulses and the complete sequence is repeated for any secondaries that need to be tracked. Figure 4.32 should make the sequence clear.

The display board contains two sets of six-filament displays, each of which is fed from a BCD to a seven-segment C-MOS decoder/driver with integral latch. The six-filament display elements are arranged to give the time differences between the reception of master and secondary station signals in tens of thousands, thousands, hundreds, tens, units and tenths of microseconds. Each decoder is fed from four of the eight data bus lines, so that time multiplexing is employed to give a full display. Figure 4.33 shows the arrangement.

Although only one set of display elements is shown, the other circuit is an exact duplicate. Each pair of decoders is enabled via a chip select line, which will go active low, to latch data into the decoder. The chip select line is in turn fed from a BCD to decimal decoder on the logic board, which operates under the control of the MPU.

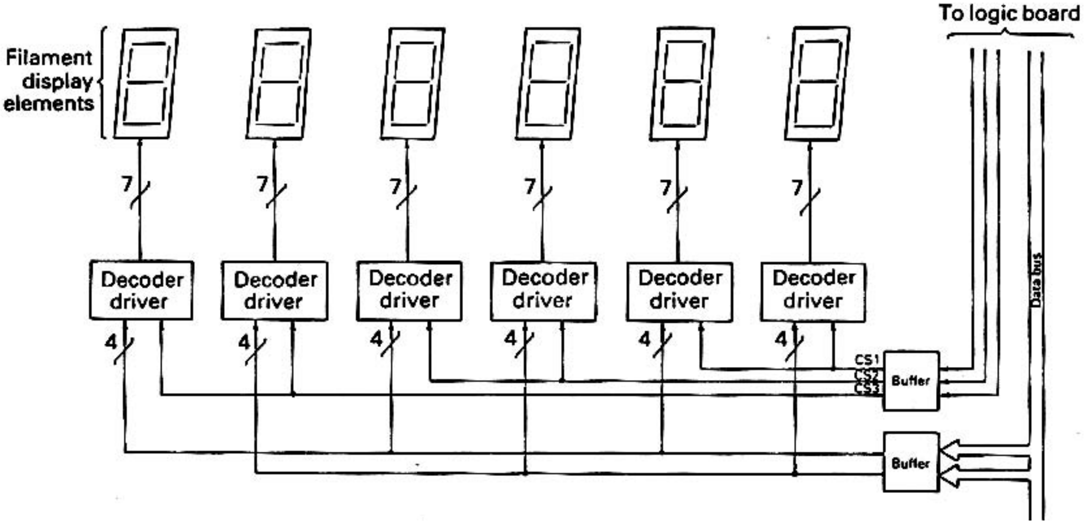


**Figure 4.33** Display elements, buffers and drivers.

In addition, the display board contains the function switch and the S1 SEL, S2 SEL, +/MEMO and –/RECALL switches. Connections when made will connect GROUND to that input on the logic board via a 44-way plug and socket arrangement which allows the data line inputs and chip select lines to be connected to the display board from the logic board. Connecting an input, or combination of inputs, to GROUND will, via the logic circuitry, fulfil the conditions as explained in pages 119–125 when describing receiver function.

The logic circuits on the logic board concerned with the function switch inputs have not been shown in Figure 4.25 in order to keep the diagram simple. As an example, however, of the circuit action, consider the case when the function switch is set to S1 and the +/MEMO switch is pressed. The logic circuit concerned would cause the value set in to the D-type flip-flops (ICs 8B, 11B and 14B) to change by +10 μs.

Also on the display board and connected via the data bus, when chip select allows, is the information regarding the settling alarm and signal-to-noise alarm indication using LEDs.

The use of microprocessors for Loran-C receivers has improved the reliability of positional information and its presentation for the operator's use; the Koden Electronics LR-707 receiver gives a good indication of this. Although Koden may no longer manufacture Loran-C receivers they still produce a range of marine electronic equipment (details may be obtained from their website at www.koden-electronics.co.jp).

An example of a modern Loran-C receiver which meets, or exceeds, the USCG standard for a Loran-C receiver and Automatic Co-ordinate Conversion System is the Furuno Model LC-90 Mark-II, the front panel of which is illustrated in Figure 4.34.

As can be seen from Figure 4.34, the front panel has a touchpad section for entering data and a five-line liquid crystal display that indicates system data. The top two lines provide positional information in either time difference (TD) format or latitude/longitude. The remaining three lines can provide different computed navigational data, as required by the operator. Additional lines at the top and bottom of the display give a constant readout of alarm and system status. The LC-90 Mark-II provides an automatic selection of optimum master and secondaries or, if preferred, it can allow the operator to override the automatic functions manually. The use of automatic selection will provide

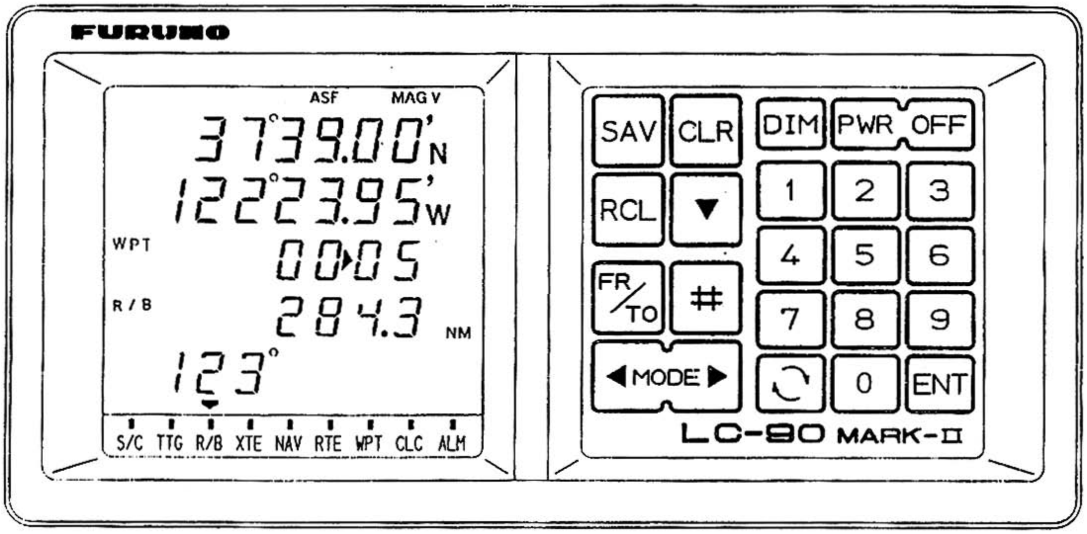


**Figure 4.34** LC-90 MkII front panel layout. (Reproduced courtesy of Furuno Electric Co. Ltd.)

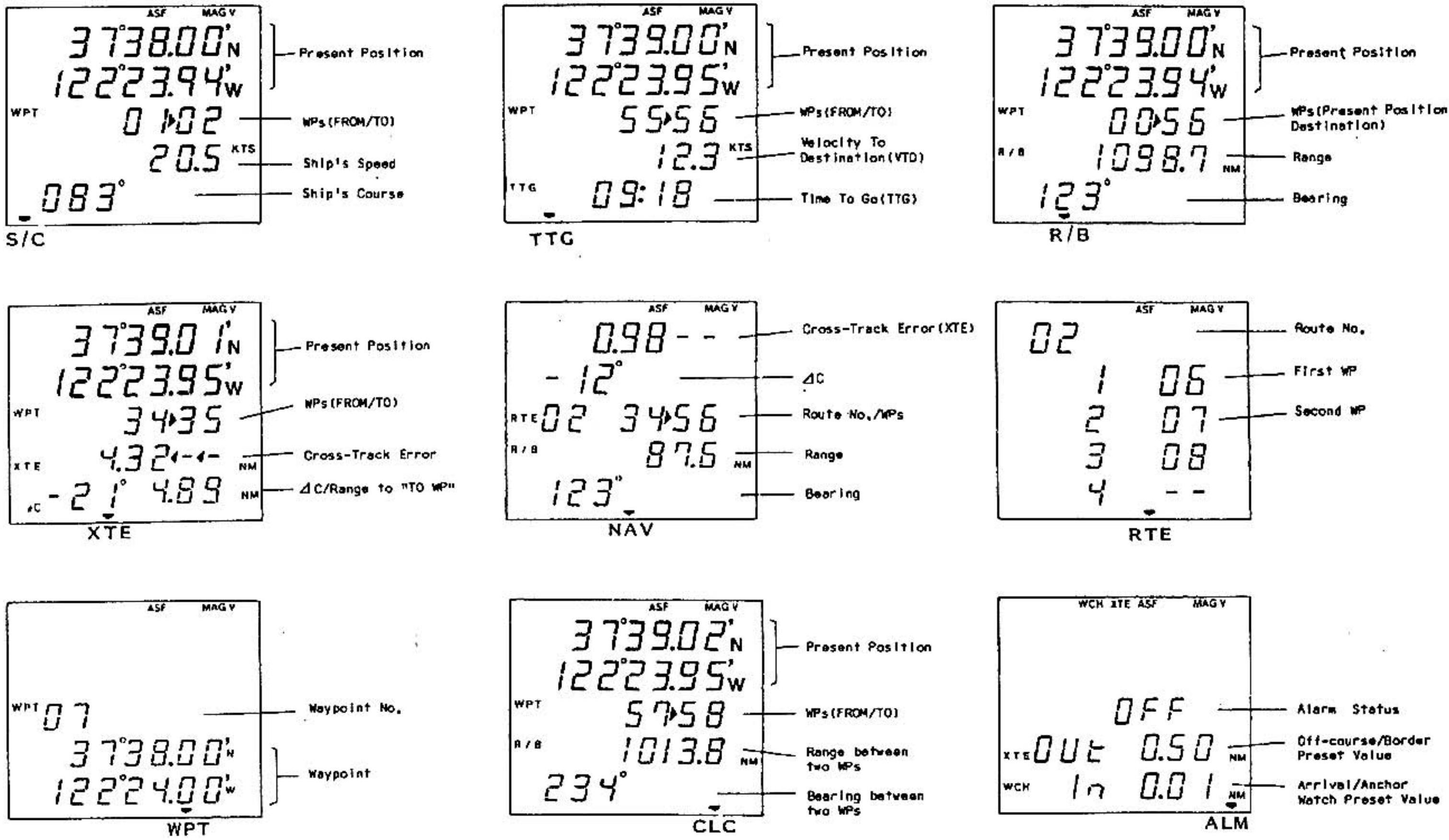


**Figure 4.35** LC-90 MkII typical mode screens. (Reproduced courtesy of Furuno Electric Co. Ltd.)

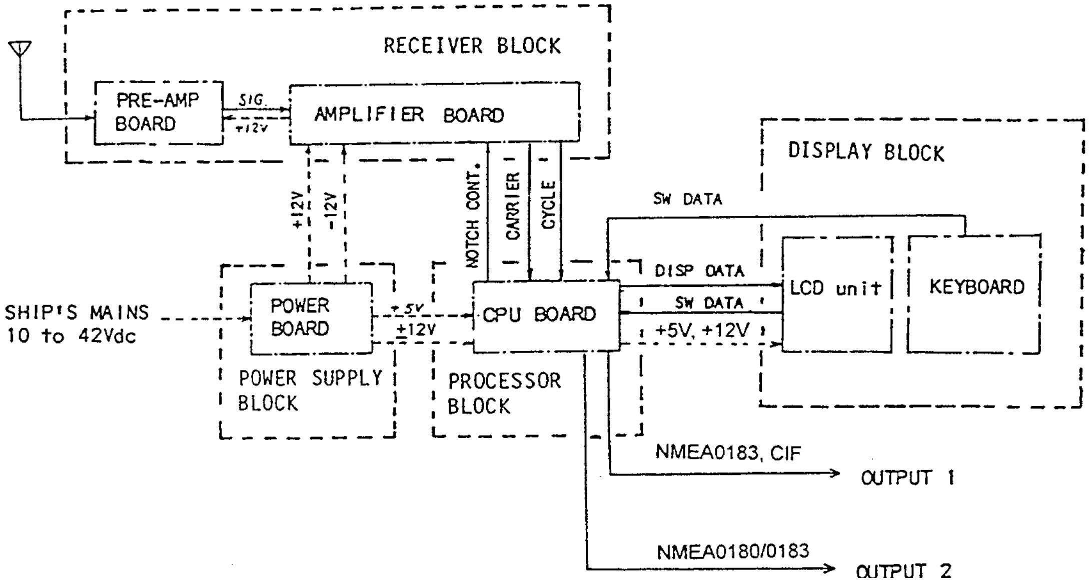


**Figure 4.36** Block diagram of the LC-90 MKII receiver. (Reproduced courtesy of Furuno Electric Co. Ltd.)

compensation for ASF and magnetic variations and set all four notch filters to eliminate interference. Entering an estimated local latitude and longitude will enable the LC-90 Mark-II receiver to determine the best available chain, and compute the correct latitude and longitude corresponding to the ship's position.

Up to 100 waypoints can be entered by either TD, latitude/longitude or range/bearing from any position. Also by use of the SAV (save) key up to 20 events can be stored in TD or latitude/longitude. The display will provide data on: range/bearing to a waypoint and between waypoints, speed and course made good, velocity and time to go, cross-track error, and course offset. Typical mode screens are illustrated in Figure 4.35. The information from the display is also indicated using values shown. Outputs include Furuno CIF, NMEA 0180 and NMEA 0183 outputs, which can be used as inputs to other equipment.

The LC-90 Mark-II receiver can be subdivided into four basic blocks as shown in Figure 4.36. The receiver block consists of a pre-amplifier and amplifier, the output of which is passed to the processor board, the other inputs to which come from the display block as SW data. Outputs from the processor board go to the liquid crystal display unit and are also available to feed other equipment via the output 1 and output 2 leads.

## 4.7.4 Specification of the LC-90 Mark-II Receiver

| | |
|---|---|
| Receiver sensitivity: | 1 μV/m |
| Differential dynamic range: | 80 dB |
| Interference rejection: | Six notch filters, four of which are auto and two are preset. |
| Tracking capacity: | Master and up to a maximum of five secondaries. Tracking speed 80 knots nominal. |
| Settling time: | Nominally 5 min, depending on signal conditions. |

Display resolution

- TD: 0.1 µs
- L/L: 0.01 min
- Range: 0.01 nautical miles

Display of signal status and alarms

- Status: S/N, CYC, tracking point, interference frequency and level, notch filter settings etc.
- Alarms: XTE, border, arrival, anchor watch.

Computation base:

- TD to L/L conversion: WGS-72
- Range/bearing: Great circle

Save function (entry of waypoint and event):

- Waypoint memory: 100 points (from no.00 (OS position) to no. 99)
- Event memory: 20 points (from no.100 to no. 119).

Other functions:

- Ground speed and true course display
- Range and bearing display to waypoint
- Velocity to destination and time-to-go display
- Range and bearing from waypoint to waypoint
- Route planning and automatic route following
- Manual compensation for TD and L/L
- Auto or manual compensation for magnetic variation
- Automatic selection of ASF or manual correction factors
- Memory back-up
- Cross-track error, course to steer to get back to an intended track.

Output (dual ports provided):

CIF

- Ship L/L, TD, Wp L/L, ship speed/course, event L/L, system time.

NMEA 0183

- Sort 1; $LCGLL/$LCAAM/$LCXTE/$LCBOD/$LCBWC/$LCVTG
- Sort 2; $LCBWW/$LCWNC/$LCWCV/$LCZTG/$LCWPL
- Sort 3; Sort 1 plus Sort 2
- Sort 4; $LCRMA/$LCRMB
- Sort 5; Sort 1 plus Sort 4

- Sort 6; Sort 2 plus Sort 4
- Sort 7; All data.

Power supply

- 10 to 42 VDC, universal, 9 W
- 110/220 VAC, 50–60 Hz CW/Rectifier Unit)

Details of the Furuno LC-90 Mk-II Loran-C receiver, and other marine electronic equipment, may be found on their website, www.furuno.com.

## 4.8 Glossary

| | |
|---|---|
| **Acquisition** | Reception and identification of Loran-C signals from a master and selected secondaries to allow a measurement of time differences (TDs) to be made. |
| **ASF** | Additional Secondary Phase Factor. Factors caused by variation in the conductivity of the surface of the earth depending on whether the loran signal path is over land or sea. The factor could cause errors in the measured Loran-C position. |
| **Attenuation** | A reduction in signal strength of a signal as it travels further from its source. The signal could be travelling in free space or in a transmission line. |
| **Baseline** | That segment of a great circle that defines the shortest distance between a master and secondary station in a loran chain. |
| **Baseline extension** | The extension of a baseline beyond the master and secondary stations in a loran chain. Measurements in the region of a baseline extension should be avoided because of possible large measurement errors in that area. |
| **Blink** | An indication that the master or secondary signals received from a loran chain are out of tolerance and would not produce reliable measurements. There are both master and secondary blink conditions. |
| **Co-ordinate conversion** | That process which changes co-ordinates produced using one system to co-ordinates in another system, i.e. when using Loran-C changing from time differences (TDs) to geodetic co-ordinates. This could be achieved by interpolation on Loran-C overprinted charts or automatically by the Loran-C receiver. |
| **Coverage area** | That coverage provided by loran signals where signal reception is of sufficient level to allow the determination of position to a specified level of accuracy and at a specified signal-to-noise ratio (SNR). |
| **Coverage diagram** | A diagram showing the coverage area for a particular master–secondary pair in a Loran-C chain. |
| **Cross track error** | See under XTE. |
| **Dual-rated (DR)** | A term used to indicate that a station in one Loran-C chain is also used in another Loran-C chain. The stations could be a master or a secondary station. |
| **Emission delay** | That time difference measured in microseconds between the emission of a signal from a master station and the emission of a signal from a specified secondary in that chain. The emission delay is the sum of the time taken for |

| | |
|---|---|
| | the master transmission to cover the baseline and the secondary coding delay. |
| **Envelope to cycle difference (ECD)** | A time difference between the phase of a Loran-C carrier. and the time origin of the pulse envelope waveform. ECD is zero when the 30-μs point of the Loran-C pulse envelope coincides in time with the third positive zero crossing of the 100-kHz carrier. |
| **Group repetition interval (GRI)** | That time interval between the start of a transmission from a master station in a Loran-C chain and the start of the next. Time is measured in microseconds and the chain is designated by its GRI value with the last zero term omitted, i.e. the North West Pacific chain has a GRI of 89 300 μs and the chain designation is 8930. |
| **IRQ** | Interrupt Request. A signal used in a microcomputer system to service an external device and which causes the current program to be interrupted to run a subroutine used to service the external device. Once the subroutine is finished the computer restores the original program and continues to execute it from the point where it was interrupted. The IRQ can be enabled/disabled according to the setting of an interrupt flag in the processor flag register. |
| **LCD** | Liquid Crystal Display. A form of display commonly used with Loran-C receivers. The display elements are typically dark coloured alphanumerics on a grey screen. The display is easily read, even in bright light conditions. |
| **LED** | Light Emitting Diode. A form of display with, typically, red alphanumerics on a dark background. Less popular than the LCD display and less easily read in strong light conditions. |
| **LOP** | Line of Position. In loran terms, a line where the time difference (TD) of signals received from the master and a specified secondary in a chain has a constant value. |
| **Loran** | *Lo*ng *ra*nge *n*avigation. |
| **Loran monitor site (lormonsite)** | A monitor site used to observe parameters of a transmitted signal as received in the coverage area. |
| **Master station** | The station in a Loran-C chain that transmits the signals identifying that chain (i.e. its GRI) and is the common base against which all time differences are established. |
| **MPU** | Microprocessor. That integrated circuit (IC) which forms the central processing unit (CPU) of a microcomputer. |
| **Nautical mile** | That unit of distance used at sea which is equivalent to 1852 metres. |
| **NECD** | Nominal ECD of a transmitting station. |
| **NEUS** | Northeast US chain. The Loran-C chain operating with a GRI designation of 9960. |
| **NMEA** | National Marine Electronics Association. An organisation comprising manufacturers and distributors. Responsible for agreeing standards for interfacing between various electronic systems on ships. NMEA 0183 version 2.3 is the current standard. |
| **NMI** | Non-maskable Interrupt. Unlike the IRQ interrupt which can be enabled or disabled by the setting of an interrupt flag, the NMI cannot be disabled and must execute the appropriate service subroutine when activated by an external device. |
| **PCI** | Phase Code Interval. That interval over which the phase code repeats. For Loran-C, phase codes repeat every two GRIs. |

| | |
|---|---|
| **PF** | Primary Phase Factor. A correction factor applied to a Loran-C signal reading made necessary by the difference in signal propagation through the atmosphere as opposed to propagation in free space. The speed of Loran-C signals through the atmosphere is equal to the speed through free space divided by the atmospheric index of refraction. The speed is taken as $2.996\,911\,62 \times 10^8\,\mathrm{ms}^{-1}$. |
| **PRF/PRR** | Pulse Repetition Frequency/Pulse Repetition Rate. The number of pulses transmitted in a specified time. For the Loran-C system the PRF/PRR is given by the reciprocal of the GRI. Hence a chain with a GRI of 80000 μs would have a PRF/PRR of 12.5 Hz. |
| **Root mean square (RMS)** | That value of a time varying signal which has the equivalent heating effect to that of a d.c. quantity. |
| **Secondary coding delay** | That time interval in microseconds between when a secondary station receives the master transmission and a transmission occurs from the secondary station. |
| **Secondary phase factor (SF)** | That amount of time, in microseconds, by which the predicted time differences (TDs) of a pair of Loran-C station signals travelling over an all-seawater path differ from those that travel through the atmosphere. |
| **Secondary station** | One of the possible maximum number of five stations that, together with the master station, comprise the Loran-C chain. |
| **Signal-to-noise ratio (SNR or S/N)** | The ratio of signal strength compared to the strength of electrical noise present with the signal in a given bandwidth. The coverage diagrams for Loran-C are calculated using an SNR of at least 1:3. SNR is often quoted in decibels (db) where the db value is given by $20\log_{10}(\mathrm{SNR})$ so that with an SNR of 1:3, the decibel value is –9.54, which is often approximated to –10db. |
| **Single-rated (SR)** | Those stations in a Loran-C chain which do not share transmissions with other chains. Compare with Dual-rated. |
| **Speed** | Rate of travel. For a vessel travelling relative to the water over a horizontal distance the speed of the vessel is measured in knots. |
| **Time difference (TD)** | In Loran-C, TD is the time difference in microseconds between the receipt of the master and secondary transmitted signals. |
| **Time to go (TTG)** | The time calculated to elapse before the next waypoint is reached. Time obtained by dividing distance to go by the groundspeed. |
| **Waypoint** | A point entered into a loran receiver and used as a reference point for navigational calculations. Planned voyages would have a series of waypoints indicating legs of the voyage. A modern Loran-C receiver is capable of storing multiple waypoints. |
| **XOR** | Exclusive-OR gate. A digital circuit that, for a two-input gate, only produces a logical 1 output when the two inputs are of opposite sign. |
| **XTE** | Cross-Track Error. That distance between the vessel's actual position and the direct course between two specified waypoints. |

## 4.9 Summary

- Loran-C is an electronic system of land-based transmitters broadcasting low-frequency pulsed signals capable of reception aboard a ship, or aircraft, and being used by the receiver to determine position in time difference or longitude/latitude.

- Loran-C uses a chain of typically three to five transmitters broadcasting at 100 kHz with a specially shaped pulse of 250 μs duration repeated at a particular rate.
- One transmitter of a Loran-C chain is designated the master (M) while the others are secondary stations known as whisky (W), x-ray (X), yankee (Y) and zulu (Z). The chain is formed of master–secondary pairs, i.e. M–W, M–X, M–Y and M–Z.
- The master station always transmits its signal first and this signal is used to trigger emissions from the secondary stations. An additional time delay is added at the secondary station. The total elapsed time between master transmission and secondary transmission is known as the emission delay.
- The emission delay ensures no ambiguity in reception within the coverage area for a chain. The unique time difference between reception of the master pulse and reception of a relevant secondary gives a specific line-of-position (LOP) for that pair. A unique LOP for a second master–secondary pair gives a point of intersection which determines the position of the receiver.
- Each Loran-C station operates with a specified group repetition interval (GRI) which are multiples of 10 μs from 40 000 up to 99 990 μs. A Loran-C chain is designated by its GRI value divided by 10, i.e. the Northeast US (NEUS) chain is designated 9960 which defines a GRI of 99 600 μs.
- Each Loran-C pulse is mathematically defined and transmissions are monitored to ensure compliance with the specified model.
- Normal operation of Loran-C assumes reception by ground waves. A ground wave signal will always arrive before a sky wave signal with a time difference of not less than 30 μs anywhere in the Loran-C coverage area, hence if only the first 30 μs of a pulse is used it will be a ground wave. Sky waves can be used at greater distances (>1000 nautical miles) where ground wave reception is unreliable but sky wave correction factors will need to be applied.
- There are possible corrections to be applied to data produced by received signals to allow for different conductivity of the surfaces over which the transmitted signal travels. The corrections, known as additional secondary phase factor (ASF) corrections, are incorporated with most Loran-C overprinted charts and many Loran-C receivers.
- Loran-C coverage is defined by geometric-fix accuracy and range limits to give what is known as the $2d_{\mathrm{RMS}}$ value with a 1:3 SNR.
- A Loran-C receiver should be able to acquire the signal automatically, identify the master and secondary pulses of a given chain pair and track the signal. As a minimum requirement it should display the time difference readings with a precision of at least one tenth of a microsecond. The receiver should also possess notch filters, used to eliminate unwanted interference, and alarms which can be used to inform the operator about signal status and receiver conditions.

## 4.10 Revision questions

1. Explain briefly the concept behind the use of low-frequency pulsed signals transmitted from land-based stations to determine the position of a ship, or aircraft, that carries a receiver suitable for the reception of such signals.
2. A transmitter emits a pulse which is intercepted by a second transmitter 150 km away. If the speed of transmission of the pulse is $3 \times 10^8\,\mathrm{ms}^{-1}$, how long does it take the pulse to travel between the stations?
   [Answer: 500 μs]

3 What would be the time taken in question 2 if the speed of transmission of the pulse was 2.997 924 58 × $10^8\,\mathrm{ms}^{-1}$?
[Answer: 500.1257 μs]

4 A transmitter emits a pulse which is intercepted by a second transmitter 1000 μs later. If the speed of transmission of the pulse is 3 × $10^8\,\mathrm{ms}^{-1}$, how far away is the second transmitter?
[Answer: 300 km]

5 How far away would the second transmitter be in question 4 if the speed of transmission of the pulse is taken as 2.997 924 58 × $10^8\,\mathrm{ms}^{-1}$?
[Answer: 299.792 458 km]

6 Explain what you understand by emission delay for a master–secondary pair in a loran system. A Loran-C master–secondary pair transmit with an emission delay of 12 000 μs of which 10 000 μs is coding delay. Sketch a typical series of LOPs, including baseline extensions, for such a master–secondary pair. What is the time difference value in microseconds of the LOP that bisects the line joining the master–secondary pair? What is the time difference value in microseconds of the baseline extensions?
[Answer: 12 000 μs; 14 000 μs (beyond master station); 10 000 μs (beyond secondary station)]

7 Loran-C stations operating in a chain have a particular GRI designation and secondary pulse groups are transmitted at the same GRI and linked in time to the master. Secondary transmission delays are selected to ensure certain criteria are met for signal reception. What are the values specified below?
(a) Minimum time difference between any secondary and master.
(b) Minimum time difference of any two time differences.
(c) Maximum time difference.
(d) Minimum spacing between corresponding points of the last pulse of any station group and the first pulse of the next group.

8 What is meant by the terms single-rated and dual-rated, as applied to a Loran-C station? Give an example of a dual-rated Loran-C station.

9 What do you understand by the term phase coding as applied to a Loran-C signal? What is the phase code for group A for both the master and secondary of a Loran-C pair? What is the phase code for group B for both the master and secondary of a Loran-C pair?

10 What is meant by the term ‘blink’ as applied to a Loran-C signal? Give an example of the use of blink.

11 Explain the technique, used in Loran-C receivers, known as ‘cycle matching’. What is the claimed advantage of such a technique?

12 Explain why it is preferable to use LOPs from two master–secondary pairs that cross at right angles to each other. Why should areas in the region of baseline extensions never be used?

13 What factors are taken into account to produce the predicted ground wave coverage for a chain? What do you understand by the term $2d_{\mathrm{RMS}}$? What is the specified SNR range limit for each transmitted signal?

14 What are the main features of a Loran-C receiver, which are necessary to measure position with the claimed accuracy for the system?

15 For the Koden Electronics LR-707 receiver shown in Figure 4.21 briefly explain the purpose of switches S1 and S2. What are the effects of moving the function switch to each of its different settings?

16 For the Koden Electronics LR-707 receiver shown in Figure 4.21 briefly explain the function of the +/MEMO and -/RECALL buttons.

17 For the Koden Electronics LR-707 receiver shown in Figure 4.21 briefly explain the use of the notch filters.

18 Using the basic block diagram of the Koden Electronics LR-707 receiver shown in Figure 4.24, describe the basic function of each block.

19 Using the logic board diagram and the sampling and coincidence circuit diagram of the Koden Electronics LR-707 receiver shown in Figures 4.25 and 4.26, respectively, describe how the incoming CYCLE signal is converted into a time difference reading fed to the display.

20. Using the information given in the text, make a comparison between an older type of receiver, such as the Koden Electronics LR-707, and a more modern receiver, such as the Furuno LC-90 Mk-II. Comment on any major differences.

# Chapter 5

# Satellite navigation

## 5.1 Introduction

It is surprising that the space technology that we rely on so heavily today had its origins over 50 years ago when, in the early 1950s, with the shock launching by the USSR of a man-made satellite into low orbit, the United States space programme was born. Although a tiny vehicle by present day standards, the USSR's 'Sputnik' had a radio transmitter on board, the frequency of which exhibited a pronounced Doppler shift when observed from any fixed point on the earth's surface. The Doppler phenomenon was well documented but this was the first time the effect had been produced by and received from a man-made orbiting satellite. Space engineers soon recovered from the initial shock and were quick to see that the effect could be exploited to create a truly accurate global positioning system, free from many of the constraints of the existing earth-bound hyperbolic navigation systems.

The first commercially available system to be developed, the Navy Navigation Satellite System (NNSS), made good use of the Doppler effect and provided the world's shipping with precise position fixing for decades. However, nothing lasts forever. The technology became old and the system was dropped on 31 December 1996 in favour of the vastly superior Global Positioning System (GPS). Although a number of NNSS Nova satellites are still in orbit, the system is no longer used for commercial navigation purposes.

## 5.2 Basic satellite theory

Whilst it is not essential to understand space technology, it is helpful to consider a few of the basic parameters relating to satellite orbits and the specific terminology used when describing them. A satellite is placed in a pre-determined orbit, either in the nose of an expendable launch vehicle or as part of the payload of a space shuttle flight. Either way, once the 'bird' has been delivered into the correct plane, called the 'inclination', that is the angle formed between the eastern end of the equatorial plane and the satellite orbit, it is subject to Kepler's laws of astrophysics.

Figure 5.1 shows orbits of zero inclination for the equatorial orbit, 45°, and for a polar orbit, 90°. The final desired inclination partly determines the launching site chosen. In practice it is difficult to achieve an inclination which is less than the latitude of the launching site's geographical location. A zero inclination orbit is most effectively produced from a launch pad situated on the equator, but this is not always possible and a compromise is often made. Launch normally takes place in an easterly direction because that way it is possible to save fuel, and thus weight, by using the earth's rotational speed to boost the velocity of the accelerating rocket. For an easterly launch from a site on the equator, the velocity needed to escape the pull of gravity, is 6.89 km $s^{-1}$, whereas for a westerly launch it is 7.82 km $s^{-1}$. Launch velocities also vary with latitude and the direction of the flight path.

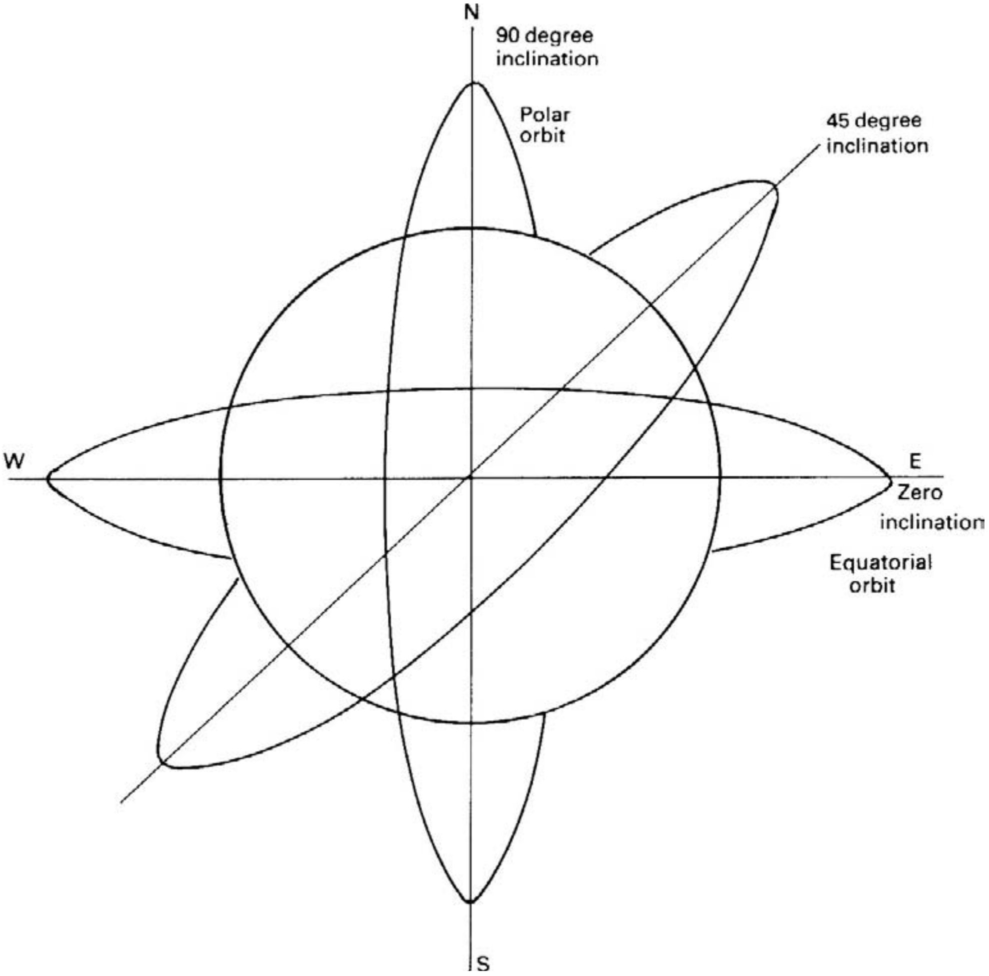


**Figure 5.1** Illustration of orbital inclination.

### 5.2.1 Kepler's Laws

Essentially, an artificial earth-orbiting satellite obeys three laws that were predicted in the late 16th century by Johannes Kepler (1571–1630) who also developed theories to explain the natural orbits of the planets in our solar system. When applied to artificial orbiting satellites, Kepler's laws may be summarized as follows.

- A satellite orbit, with respect to the earth, is an ellipse.
- Vectors drawn from the satellite orbit to the earth describe equal areas in equal times.
- The square of the period of the orbit is equal in ratio to the cube of its mean altitude above the earth's surface.

True to Kepler, artificial earth satellites follow elliptical orbits. In some cases the ellipse eccentricity is large and is a requirement of the first stage of a launch to the higher geostationary orbit, but in most

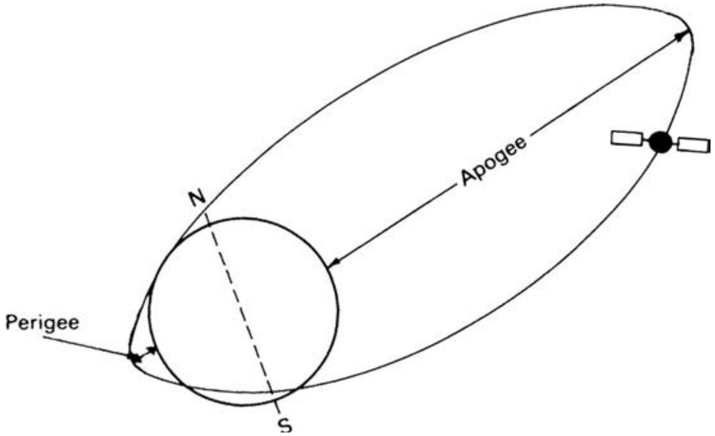


**Figure 5.2** Illustration of apogee and perigee.

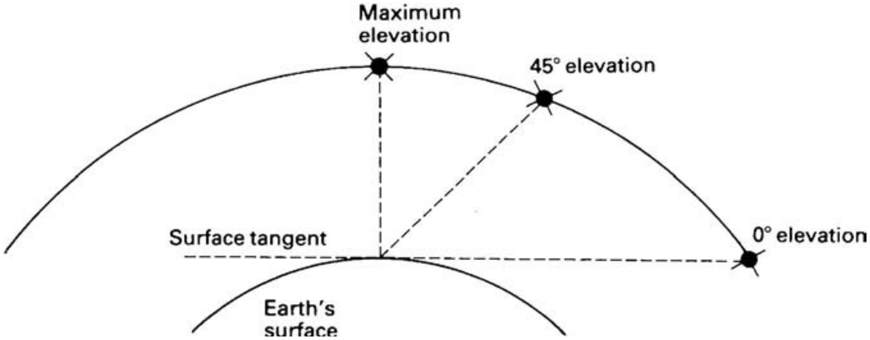


**Figure 5.3** Showing the changing angle of elevation during a satellite pass. The angle reaches a maximum at the closest point of approach to the earth bound observer.

cases it is created because the earth is not a perfect sphere. The closest point of approach to the earth of any elliptical orbit is called the 'perigee' and the furthest distance away is the 'apogee', as shown in Figure 5.2. The direction vector to the satellite from a fixed point on the earth is called the 'azimuth' and is quoted in degrees. The angle between the satellite, at any instant, and the earth's surface tangent is the 'elevation' and again is quoted in degrees (see Figure 5.3).

### 5.2.2 Orbital velocity

A satellite can only remain in orbit if its velocity, for a given altitude, is sufficient to defeat the pull of gravity ($9.81\,\mathrm{ms}^{-1}$) and less than that required to escape it. The velocity must be absolutely precise for the orbital altitude chosen. Eventually, drag will slow the satellite causing it to drop into a lower orbit and possibly causing it to re-enter the atmosphere and burn-up. The nominal velocity for a satellite at any altitude can be calculated by using the formula:

$$V = \frac{K}{(r + a)^{1/2}}\ \mathrm{kms}^{-1}$$

where $V$ = orbital velocity in $\text{kms}^{-1}$,
$a$ = altitude of the satellite above the earth's surface in km,
$r$ = the mean radius of the earth (approximately 6370 km), and
$K$ = 630 (a constant derived from a number of parameters).

The earth is not a perfect sphere and therefore its radius with respect to orbital altitude will vary. However, to derive an approximate figure for velocity, an earth radius figure of 6370 km is close enough. The velocity of a satellite with an altitude of 200 km would be:

$$V = \frac{630}{(6370 + 200)^{\frac{1}{2}}} = 7.77\,\text{kms}^{-1}$$

Orbital paths can be transferred to a Mercator projection chart as shown in Figure 5.4. The inclination will be the same in both northern and southern hemispheres and corresponds to latitude. The six orbits shown are for Navstar (GPS) satellites with an orbital inclination of 55°.

### 5.2.3 Orbital period

The time period for one complete orbit of a satellite can be readily calculated using the simple formula below:

$$P = K\left(\frac{r + a}{r}\right)^{3/2}$$

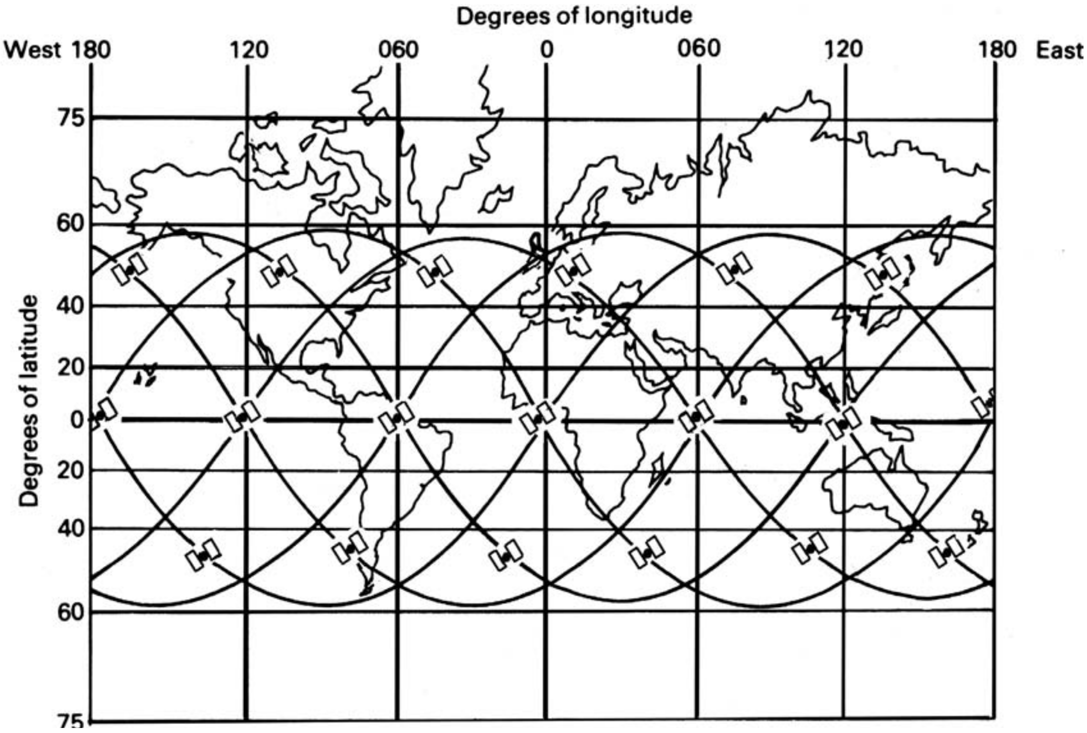


**Figure 5.4** Mercator presentation of the orbital inclination paths described by satellite orbits.

where $P$ = the period of one orbit in min,
$a$ = the altitude of the orbit above the earth's surface in km,
$r$ = the mean radius of the earth in km, and
$K$ = 84.49 (a constant derived from a number of parameters).

The orbital period for a satellite at an altitude of 200 km is:

$$P = 84.49 \left( \frac{6371 + 200}{6371} \right)^{3/2} = 88.45 \text{ min}$$

## 5.3 The Global Positioning System (GPS)

In 1973 a combined US Navy and US Air force task-force set out to develop a new global satellite navigation system to replace the ageing Navy Navigation Satellite System (NNSS).

The original test space vehicles (SVs) launched in the new programme were called Navigation Technology Satellites (NTS) and NTS1 went into orbit in 1974 to became the embryo of a system that has grown into the Global Positioning System (GPS). GPS was declared to be fully operational by the US Air Force Space Command (USAFSC) on 27April 1995, and brought about the demise of the NNSS which finally ceased to provide navigation fixes at midnight on 31 December 1996.

The GPS, occasionally called NAVSTAR, shares much commonality with the Russian Global Navigation System (GLONASS), although the two are in no way compatible. The GPS consists of three segments designated Space, Control and User.

### 5.3.1 The space segment

Satellite constellation calls for 24 operational SVs, four in each of six orbital planes, although more satellites are available to ensure the system remains continuously accessible (see Figure 5.5). SVs orbit the earth in near circular orbits at an altitude of 20 200 km (10 900 nautical miles) and possess an inclination angle of 55°.

Based on standard time, each SV has an approximate orbital period of 12 h, but when quoted in the more correct sidereal time, it is 11 h 58 min. Since the earth is turning beneath the SV orbits, all the satellites will appear over any fixed point on the earth every 23 h 56 min or, 4 min earlier each day. This, totally predictable, time shift is caused because a sidereal day is 4 min shorter than a solar day and all SVs complete two orbits in one day. To maintain further orbital accuracy, SVs are attitude-stabilized to within 1 m by the action of four reaction wheels, and on-board hydrazine thrusters enable precision re-alignment of the craft as required.

This orbital configuration, encompassing 24 SVs, ensures that at least six SVs, with an elevation greater than 9.5°, will be in view of a receiving antenna at any point on the earth's surface at any time. When one considers the problems of rapidly increasing range error caused by the troposphere at low SV elevations, 9.5° has been found to be the minimum elevation from which to receive data when using a simple antenna system.

The original satellites, numbered 1–11 and designated Block I, have ceased operation. Currently, the GPS constellation is based on the next generation of SVs, designated Block II. Block II (numbers 13–21) and block IIA (numbers 22–40) satellites, manufactured by Rockwell International, were launched from Cape Canaveral between February 1989 and November 1997. Each SV holds four atomic clocks, two rubidium and two caesium, and has selective availability (SA) and anti-spoofing (A-S) capabilities, although the US Government has now given an assurance that the system

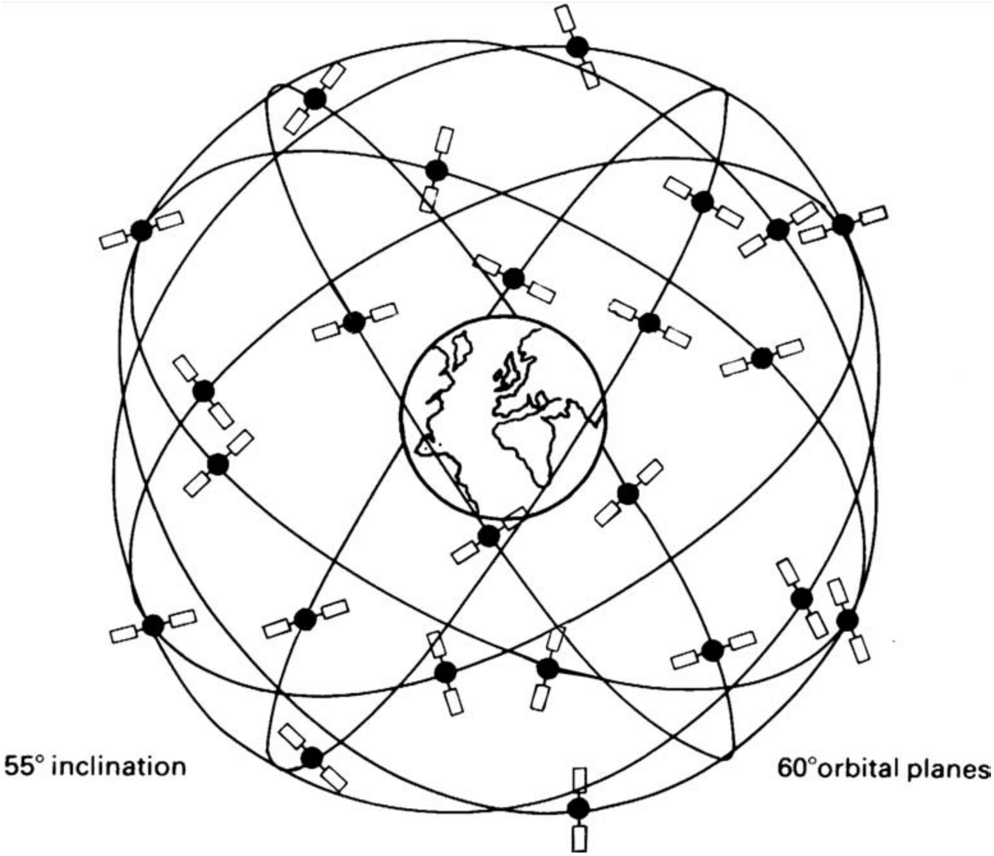


**Figure 5.5** GPS satellite coverage. Twenty-four satellites provide global coverage; four in each of six orbital planes.

downgrading functions, SA and A-S, will no longer be implemented in the GPS. Block IIR SVs (numbers 41–62) are replenishment satellites and have been designed for an operational life of 7.8 years.

All SVs transmit a navigation message comprising orbital data, clock timing characteristics, system time and a status message. They also send an extensive almanac giving the orbital and health data for every active SV, to enable a user to locate all SVs once one has been acquired and the data downloaded.

### 5.3.2 The control segment

The GPS is controlled from Schriever Air Force Base (formerly Falcon AFB) in Colorado. It is from there that the SV telemetry and upload functions are commanded. There are five monitor stations (see Figure 5.6), which are situated in the Hawaii Islands in the Pacific Ocean, on Ascension Island in the Atlantic, on Diego Garcia in the Indian Ocean, on Kwajalein Island, again in the Pacific, and at Colorado Springs on mainland US territory. SV orbital parameters are constantly monitored by one or more of the ground tracking stations, which then pass the measured data on to the Master Control Station (MCS) at Schriever. From these figures the MCS predicts the future orbital and operational

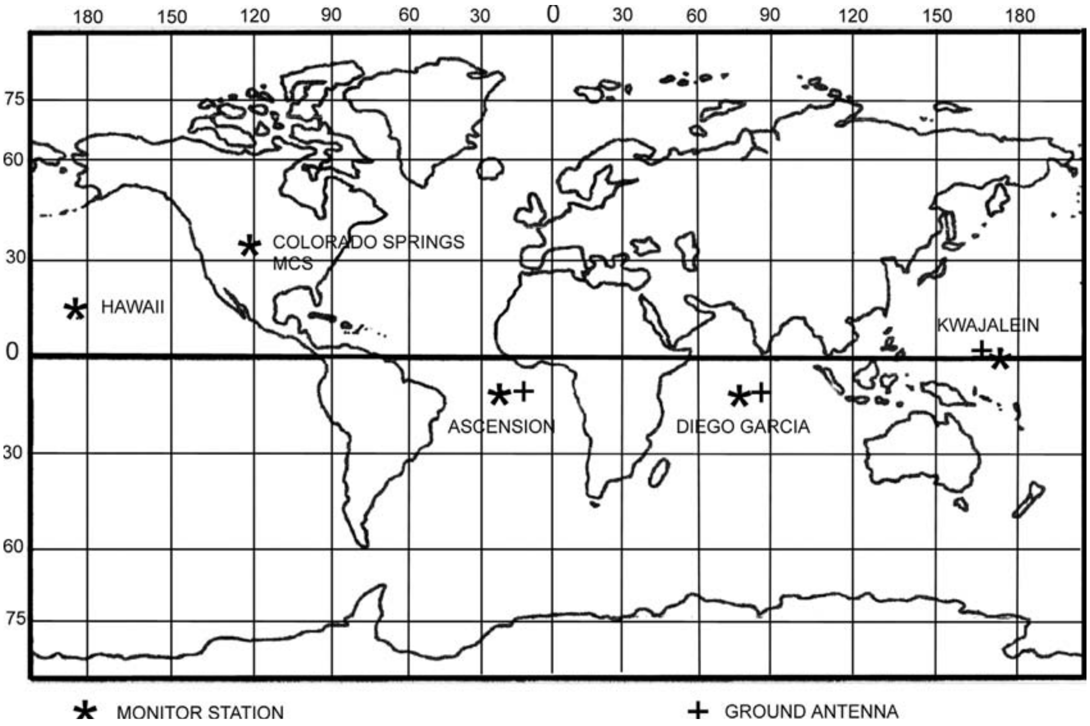


**Figure 5.6** GPS control segment stations.

parameters to be fed to the Upload Stations (ULS) on Ascension, Diego Garcia and Kwajalein Islands. All ground station locations have been precisely surveyed with respect to the World Geodetic System 1984 (WGS-84). Data are transmitted to each SV from a ULS, to be held in RAM and sequentially transmitted as a data frame to receiving stations.

### *Signal parameters*

Navigation data are transmitted from the SV on two frequencies in the L band (see Table 5.1). In practice the SV clock is slightly offset to a frequency of 10.229 999 995 45 MHz to allow for the effects of relativity. SV clock accuracy is maintained at better than one part in $10^{12}$ per day. Dual frequency transmission from the SV ensures that suitably equipped receivers are able to correct for signal delay (range error) caused by the ionosphere. Ionosopheric delays are proportional to $1/f^2$ hence the range error produced will be different on each frequency and can be compensated for in the receiver.

The C/A (Coarse and Acquire) code, see Figure 5.7, is a PRN (pseudo random noise) code stream operating at 1.023 megabits/s and is generated by a 10-bit register. C/A code epoch is achieved every 1 ms (1023 bits) and quadrature phase modulates the $L_1$ carrier only. This code has been designed to be easily and rapidly acquired by receivers to enable SPS fixing. Each SV transmits a unique C/A code that is matched to the locally generated C/A code in the receiver. A unique PRN is allocated to each SV and is selected from a code series called Gold codes. They are specifically designed to minimize the possibility that a receiver will mistake one code for another and unknowingly access a wrong satellite. Navigation data is modulated onto the $L_1$ C/A code at a bit rate of 50 Hz.

**Table 5.1** SV transmission frequencies

| *Band* | *Derivation (MHz)* | *Frequency (MHz)* | *Wavelength (cm)* | *Code* |
|---|---|---|---|---|
| $L_1$ | 154 × 10.23 | 1575.42 | 19 | C/A |
| $L_2$ | 120 × 10.23 | 1227.60 | 24.5 | C/A & P |

Both carriers are derived from the SV clock frequency 10.23 MHz

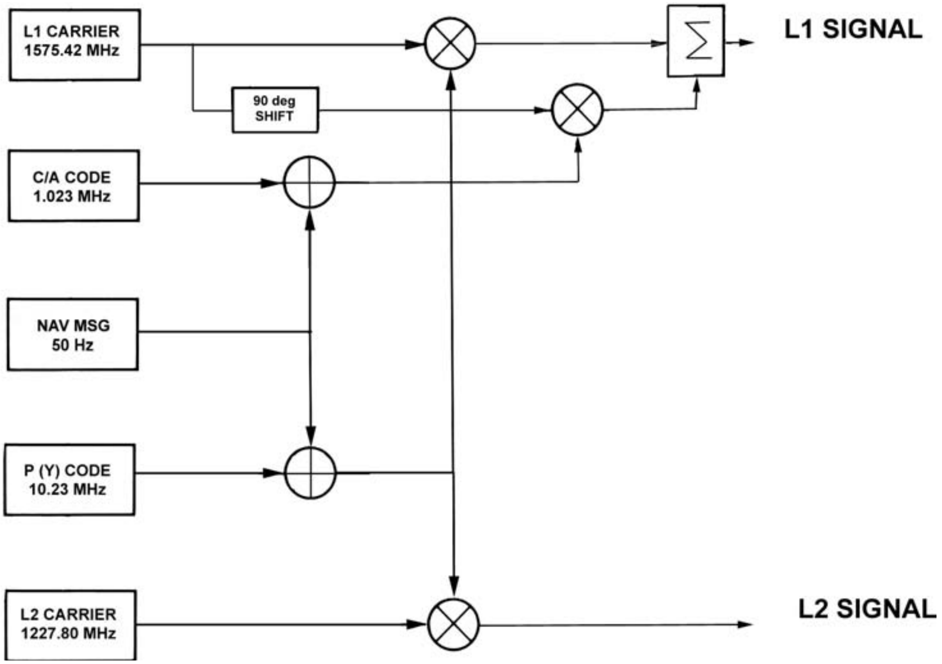


**Figure 5.7** Schematic diagram of a SV modulation circuit.

The P (Precise) code, operating at 10.23 MHz, is a PRN code produced as the modulo 2 sum of two 24-bit registers, in the SV, termed X1 and X2. This combination creates a PRN code of $2^{48-1}$ steps equating to a complete code cycle (before code repetition occurs) of approximately 267 days. Each SV employs a unique and exclusive 7-day long phase segment of this code. At midnight every Saturday, GPS time, the X1 and X2 code generators are reset to their initial state (epoch) to re-initiate the 7-day phase segment at another point along the 267-day PRN code cycle. Without prior knowledge of the code progression, it is not possible to lock into it.

### *The navigation data message*

A 50-Hz navigation message is modulated onto both the P code and C/A codes. One data frame is 1500 bits and takes 30 s to complete at the bit rate of 50 bit $s^{-1}$. Navigation data are contained in five subframes each of 6 s duration and containing 300 bits. Table 5.2 shows the data format structure.

**Table 5.2** Data format structure

| Five words 300 bits each with a total of 6 s | | | |
|---|---|---|---|
| | *30 bits* | *30 bits* | *240 bits* |
| 01 | TLM | HOW | Data block 1: Clock correction data. Accuracy and health of the signal. |
| 02 | TLM | HOW | Data block 2: Ephemeris data. Precise orbital parameters to enable a receiver to compute the position of an SV. |
| 03 | TLM | HOW | Data block 3: Ephemeris. Continued. |
| 04 | TLM | HOW | Data block 4: Almanac. Orbital data, low-precision clock data, simple health and configuration status for every SV, user messages, ionosopheric model data and UTC calculations. |
| 05 | TLM | HOW | Data block 5. Almanac. Continued. |

Subframes 4 and 5 hold low precision data, common to all SVs, and less critical for a satellite to acquire quickly.

As shown in Figure 5.8, each of the five subframes commences with a 14-bit TLM word (telemetry) containing SV status and diagnostic data. This is followed by a 17-bit handover word (HOW). HOW data enables a receiver, which has knowledge of the code encryption, to acquire the P code. Data subframe block 1 contains frequency standard corrective data enabling clock correction to be made in the receiver. Data blocks 2 and 3 hold SV orbit ephemeris data. The two blocks contain such data as orbit eccentricity variations and Keplerian parameters. Message block 4 passes alphanumeric data to the user and is only used when the ULS has a need to pass specific messages. Block 5 is an extensive almanac that includes data on SV health and identity codes.

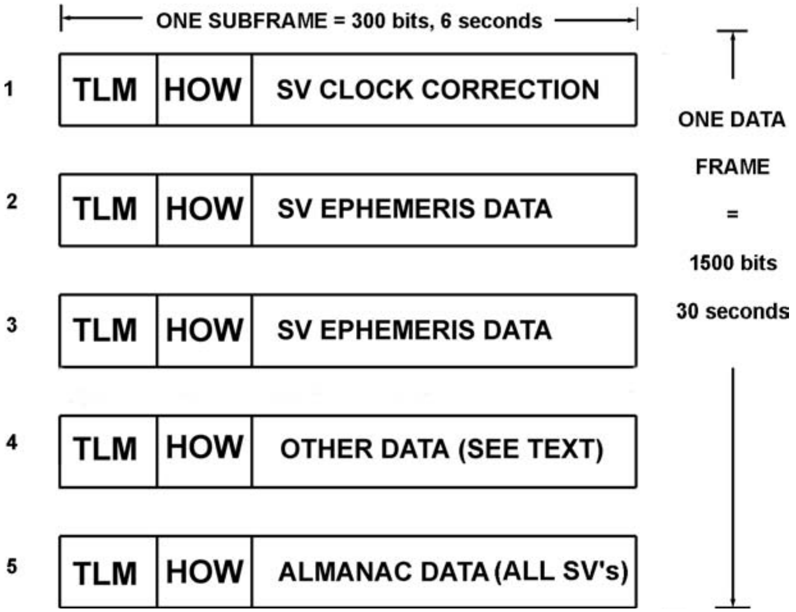


**Figure 5.8** Navigation data format.

**Table 5.3** Summary of data in a 30-s frame

| | |
|---|---|
| A | SV orbital parameters |
| B | SV clock error data |
| C | Sidereal correction figures |
| D | Almanac of all operational SVs |
| E | Polar wander data (Earth axis wander) |
| F | SV performance status |
| G | Time of last data inject |
| H | Data to enable P code acquisition (HOW) |
| I | Telemetry data (TLM) |
| J | SV number |
| K | Specific messages as required (i.e. an indication that an SV is off station) |
| L | Receiver clock correction data |

At the 50-Hz transmission rate, it takes 6 s to download a subframe, 30 s for one data frame (see Table 5.3) and a full 12.5 min to access all 25 frames.

The $L_1$ signal carrier is BPSK-modulated by both the P and C/A PRN codes and the navigation message. Modulation possesses both in-phase and quadrature components as shown in Figure 5.9.

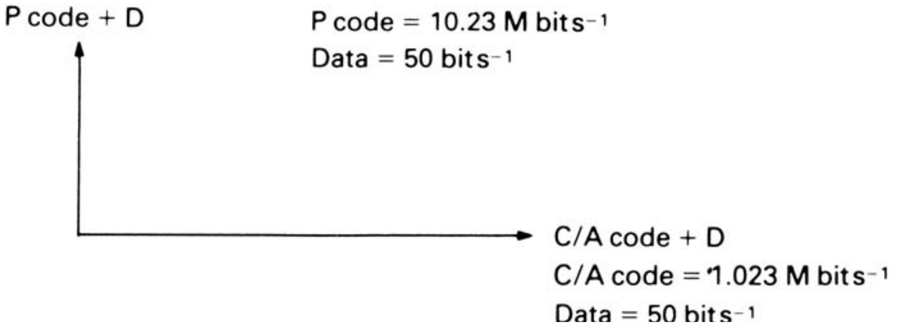


**Figure 5.9** Phase relationship between the P and C/A codes.

P code amplitude is –3dB down (half the power level) on the C/A code signal strength, thus the slower C/A code provides a better signal-to-noise ratio at the antenna. This makes the C/A code easier to access. The $L_2$ carrier is BPSK-modulated by the P code and the navigation message. The use of BPSK modulation causes a symmetrical spread of the code bandwidth around the carrier frequency. The frequency spectrum produced by both P and C/A codes on the $L_1$ carrier is shown in Figure 5.10. The bandwidth of the C/A code is 2.046 MHz and that of the P code is 20.46 MHz. The C/A code component of the $L_1$ signal possesses a power of –160 dBW (with respect to 1 watt), the $L_1$ P code a power of –163 dBW, and the $L_2$ P code signal has a power level of –166 dBW.

It should be noted that data modulation at 50 bit s$^{-1}$ produces a bandwidth of 100 Hz that is impossible to illustrate on this scale. Signal bandwidth, code matching and data stripping are further explained in the GPS receiver pages later in this chapter.

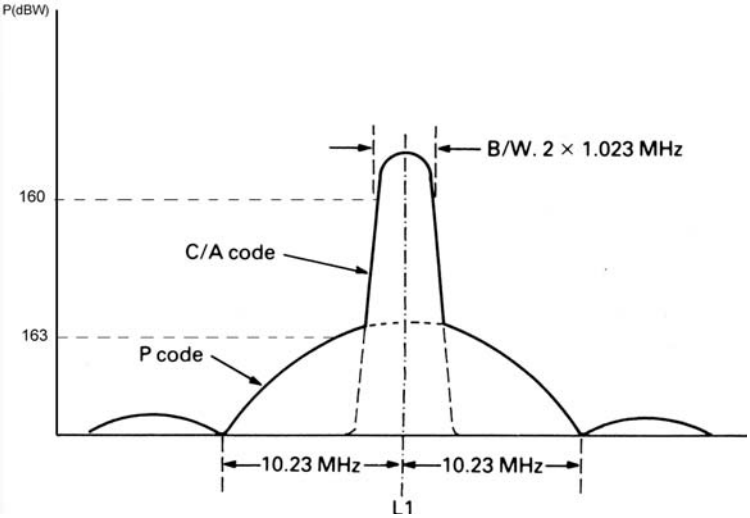


**Figure 5.10** Bandwidth power distribution curves for the P and C/A codes.

### *Frequency stability*

SV clock frequency stability is of major importance in any system that relies upon the accurate measurement of range for its operation. Stability is not easy to maintain in an electronic unit that is subjected to constantly varying ambient temperatures. The SV is travelling through a hostile environment where temperatures can vary by as much as 300°C. In addition, at the high altitudes of any SV, there is little protection from the sun's radiation. For these reasons the clock oscillators in SVs are under constant scrutiny

Since the early days of radiocommunication development, oscillator stability has been a major problem and it is one that has been compounded with the need to send clock oscillators into space. Older SVs, such as the Transit and Nova range on which the earlier NNSS sat-nav system was based, used quartz-controlled clock oscillators to give a short-term stability of $10^{-11}$ with a 24-h change less than $10^{-9}$. Timation SVs, the first to provide navigation capability by the calculation of the range between satellite and receiver, carried a quartz clock oscillator with a stability of 1 part in $10^{-11}$ per day. Timation SVs carried a new frequency standard unit formed by a quartz oscillator locked to an atomic resonance line of rubidium.

The technology used in rubidium and caesium clock oscillators is beyond the scope of this book. However, it should be noted that use of this type of oscillator in NTS1 produced the two transmission signals (UHF and L band) to an accuracy of 1 part in $10^{-12}$ per day. Caesium/quartz units offer even greater frequency stability and in 1975 the second generation of NTS vehicles was launched into orbit. NTS2 carried a caesium frequency standard unit from which were produced the carrier frequencies (SHF, $L_1$ and $L_2$) with an accuracy of 1 part in $10^{-13}$ per day. These oscillators are still in orbit and still being tested by the armed forces. Caesium clocks, however, require regular updating from the ground and in an effort to further improve and maintain stability for extended periods, clock units using hydrogen maser technology are being considered.

The clock oscillators used in current Navstar SVs are caesium/quartz with rubidium/quartz back-up units.

### *System time*

GPS system time is locked to the Master Clock (MC) at the USNO and further synchronized to UTC from which it will never deviate by more than 1 µs . Actual system time is given by its Composite Clock (CC) or, as it is often called a 'paper' clock, which had its epoch at 0000 UTC on 17 June 1990. Information about the GPS time difference and rate of system time against UTC (USNO) is contained in the navigation message transmitted to all users. Once a satellite has been accessed the user equipment clock is corrected.

## 5.4 The position fix

The GPS provides two levels of service known as Precise Positioning Service (PPS) and Standard Positioning Service (SPS), the accuracy of which were defined in the 1994 US Federal Radionavigation Plan. The PPS predictable accuracy is given in Table 5.4.

**Table 5.4** PPS predictable accuracy

| | |
|---|---|
| Horizontal accuracy | 21 m |
| Vertical accuracy | 27.7 m |
| Time transfer accuracy | 197 ns |

Based on a 95% Rayleigh distribution probability

PPS fixes are based on range measurement and the acquiring and integrating of the C/A code and the complex P code transmitted on both the $L_1$ and $L_2$ carrier frequencies. The method provides highly accurate positioning, timing and velocity figures for users authorized by the US Government. PPS users were generally the US military, government agencies and approved allied forces, but since 1 May 2000, when selective availability was ended, PPS fix accuracy is available to anyone with suitable equipment.

Selective availability (SA) was the name given to a process employed by the US Department of Defence to deny PPS accuracy to civilian users. SA was applied by offsetting SV clock frequency (dithering), and/or manipulating navigation orbit data (epsilon). To guard against the fake transmission of SV data, a system called anti-spoofing (A-S) was used whereby the P code was encrypted becoming the Y code. By Presidential order, on 1 May 2000, the US Government ceased to apply SA to the GPS and thus there is now little difference between SPS and PPS fix accuracy (see Table 5.5).

**Table 5.5** SPS predictable accuracy

| | *Prior to 1 May 2000* | *Subsequent to 1 May 2000* |
|---|---|---|
| Horizontal error | 100 m | 25 m |
| Vertical error | 156 m | 30 m |
| Time transfer error | 340 ns | 200 ns |

Based on a 95% Rayleigh distribution probability
Note: On 1 May 2000, Selective Availablity (S/A) was set to zero and SPS accuracy was thus improved by a factor of almost 10. The figures in column 3 are an approximation.

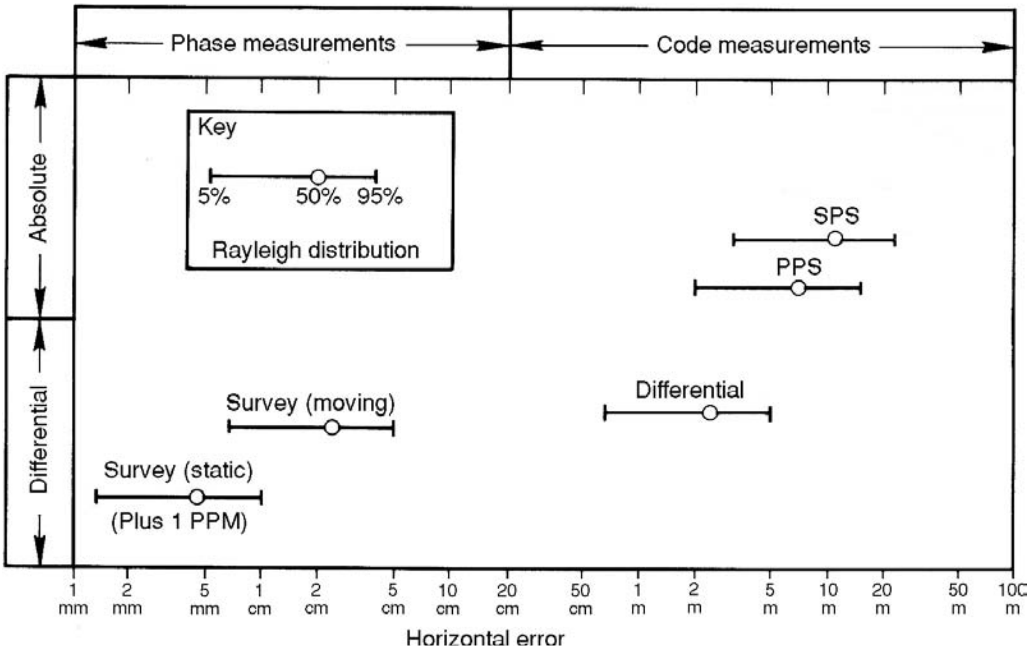


**Figure 5.11** Levels of GPS accuracy. (Reproduced courtesy of Magnavox.)

The decision to remove SA from the GPS was taken because it would have minimal impact on national security. Based on threat assessment analysis, it is possible for the US Government to selectively deny GPS signals on a regional basis if national security is threatened.

SPS fixes are based on acquiring and integrating the C/A code data transmitted on the $L_1$ carrier frequency, measuring ranges and decoding the navigation message. SPS fix accuracy can be extensively improved by using Differential GPS (see Figure 5.11). Data is received, at both a mobile and a ground station, from multiple SVs and, after the computation of correction figures at the fixed station, is retransmitted to the mobile receiver. The process is achieved in real time although because of the relatively short distances travelled by a ship between fixes it is possible to apply corrections to subsequent computations.

The upper part of Figure 5.11 shows the anticipated levels of accuracy of a standard position fix without the aid of differential techniques, whereas the lower half shows fix accuracy for receivers with a differential input. It also demonstrates that the use of phase measurement in addition to code measurement improves the fix still further. All fix lines are shown as Rayleigh distribution data.

GPS position fixes are achieved by the precise measurement of the distance between a number of SVs and a receiver at an instant in time and/or by phase measurement. It is possible for a receiver, with a precise clock and with a knowledge of altitude above the earth reference spheroid, to fix its position in three dimensions by interrogating a minimum of three SVs. But in practice, modern equipment provides for more precise position fixing using the data from four or more SVs. By interrogating multiple SVs it is possible to obtain accurate fixes in three dimensions (XYZ) plus time. All fixes computed by a receiver are known as earth-centred-earth-fixed (ECEF) locations and therefore navigation fixes are often quoted as ECEF XYZ positions.

To measure the precise distance between the transmitter and the receiver requires highly accurate time clocks in both vehicles. The satellite clock is monitored from the ground and is corrected by atomic standard time. During calculations, it is accepted therefore, that this clock, which is used to generate the transmission frequencies, is accurate and the receiver clock may be in error.

For this reason range measurements are termed false or 'pseudo-ranges', and must be corrected in the receiver. The pseudo-range measurement for a receiver with an imprecise clock is given as:

$$PsR = Rt + C\Delta td + C(\Delta tu - \Delta ts)$$

where range figures are in metres and time in seconds, $PsR$ = pseudo-range between satellite and receiver, $Rt$ = true range, $C$ = speed of light ($3 \times 10^8$ ms$^{-1}$), $\Delta ts$ = satellite clock error from GPS time, $\Delta tu$ = receiver clock error from GPS time, and $\Delta td$ = propagation delays due to both the ionosphere and the troposphere.

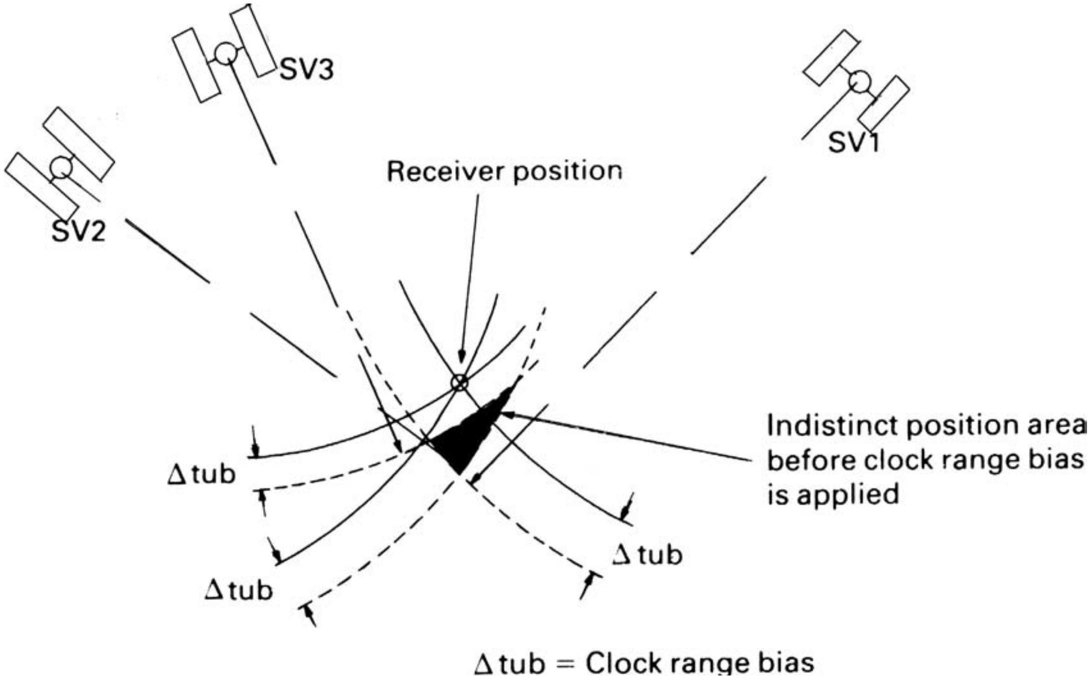


**Figure 5.12** Showing the indistinct position fix obtained from three SVs before clock range bias is applied.

The GPS receiver calculates the pseudo-range time taken for the transmission by measuring the phase shift of the P code and comparing it with a locally generated code in the receiver computer. Figure 5.12 illustrates that the pseudo-ranges calculated for three satellites will not converge at a specific point unless the receiver clock error is corrected.

The computed position in XYZ co-ordinates is converted as a function of the receiver algorithm to geodetic latitude, longitude and altitude above the reference ellipsoid. The ship's position is solved with reference to Cartesian co-ordinates as shown in Figure 5.13 with reference to a minimum of three celestial 'fixed' points (the SVs).

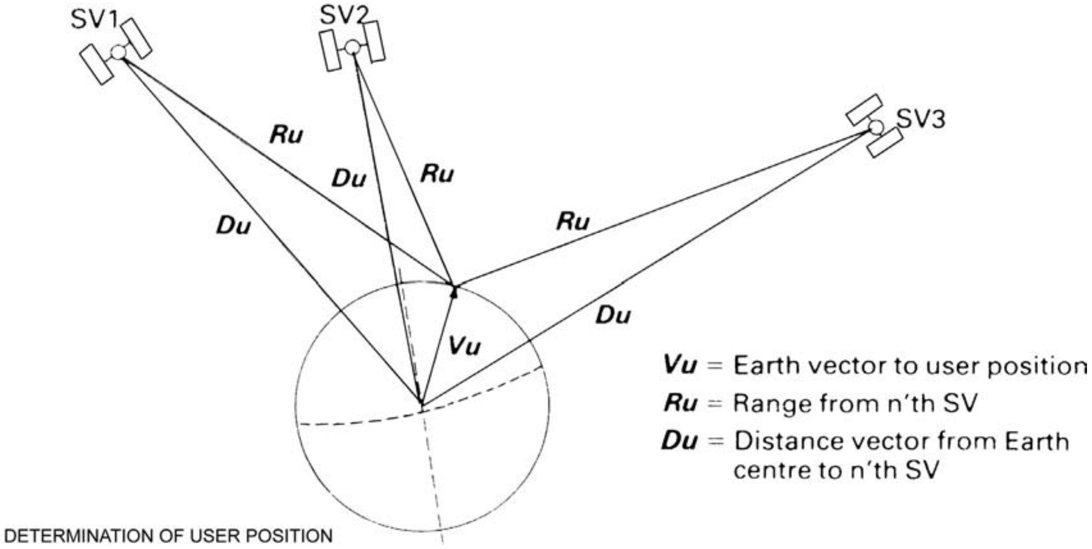


**Figure 5.13** Using Cartesian co-ordinates to determine an earth centred position fix.

## 5.5 Dilution of Precision (DOP)

Dilution of Precision (DOP) is a term used for expressing the mathematical quality of a solution. DOP can exist in one dimension only. Examples are; time DOP (TDOP); horizontal DOP; vertical DOP and geometric DOP, referring to SV geometry. But it is the position dilution of precision, PDOP, that is of most value to a navigator. PDOP in the GPS has an optimum value of unity. If the figure is higher the solution is degraded (diluted). The PDOP will approach unity when a solution is made with a satellite overhead and three other satellites evenly spaced at low elevation angles. Alternatively, if all satellites are in the same plane, PDOP would be near infinity and the navigation fix solution would be unsound. The PDOP figure has a direct bearing on user range error (URE). For example, for a URE of 50 m and a PDOP of unity, the best fix accuracy is 50 m. If the PDOP is 2, the accuracy drops to 100 m. Modern GPS receivers may be programmed to reject a position solution if the PDOP level is high.

The geometry of the satellite orbital cage can seriously affect the accuracy of a position fix. With 24 satellites in six orbits there is a better than average chance that as many as six will be in view of a receiver at any given time. When pseudo-ranges are measured from SVs that are close together in the sky (Figure 5.14(a)), the result is an enlarged area of improbability resulting in a bad GDOP, as shown above. Alternatively if the SVs are well spaced, the improbability area will be smaller. Modern GPS receivers pick the optimum SVs from those available before correcting timing errors.

## 5.6 Satellite pass predictions

The system is so well documented and controlled that it has become increasingly easy to predict satellite passes at a given location. Trimble Navigation Limited, one of the biggest manufacturers of GPS equipment, operates a world wide web site that will be of interest to students. It is called GPS Mission Planning and is accessed on http://www.trimble.com/cgi/satview.cgi. It is also interactive and

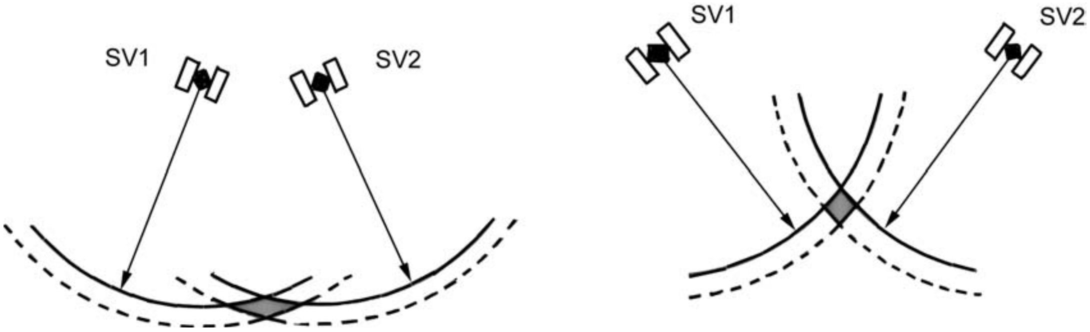


**Figure 5.14** Fix accuracy can be improved by selecting appropriate SVs. (a) Two SVs giving a poor GDOP and (b) two SVs providing a much better solution.

provides six different charts of predictions. User parameters for all the plots are input into boxes as shown below. Latitudes south of the equator and longitudes west of the Greenwich Meridian are identified with a minus sign. The time input in GMT is in two figures between 00 and 23.

Using this system it is easy to predict SV passes at a given location and consequently it is simple to select the appropriate SVs to give a good GDOP.

| | | | |
|---|---|---|---|
| Latitude: | 32.43 | Date: | 00–00–2000 |
| Longitude: | –117.10 | Starting hour GMT: | 00 hours |
| Mask: | 15.0 degs. | Duration: | 4 hours |

The six plots are as follows.

- *Azimuth Plot.* Use this plot to locate SVs with optimum azimuth angle for a given location.
- *DOP Plot.* A low DOP indicates a high probability of accuracy, whereas a high DOP shows a low probability. The plot shown in Figure 5.15 is the result of calculations evaluating the geometry of four available SVs that will provide the most accurate fix. The plot has three data lines corresponding to HDOP, VDOP and PDOP predictions.
- *Elevation Plot.* This plot (Figure 5.16) shows the paths of all the satellites in view for a specified time period at a specific location. An SV reaching an elevation of 90° will pass directly overhead.
- *Sky Plot.* This plot (Figure 5.17) is oriented so that the GPS receiver is in the centre of concentric rings spaced at 15° intervals. The outer ring represents the horizon. Using this plot it is easy to see if a SV could suffer signal block from buildings or trees because it is low on the horizon.
- *Total-in-View Plot.* This is a graph showing the total SVs in view over a specified elevation angle. It is particularly useful for checking if sufficient satellites will be in view to make a good fix.
- *Visibility Periods Plot.* Another form of presentation showing the time periods when satellites will be in view above the angle of elevation specified.

## 5.7 System errors

Errors in any system arise from a number of sources. They can be predictable or not and avoidable or not. The GPS is no exception. If suffers from error-inducing factors which will downgrade its

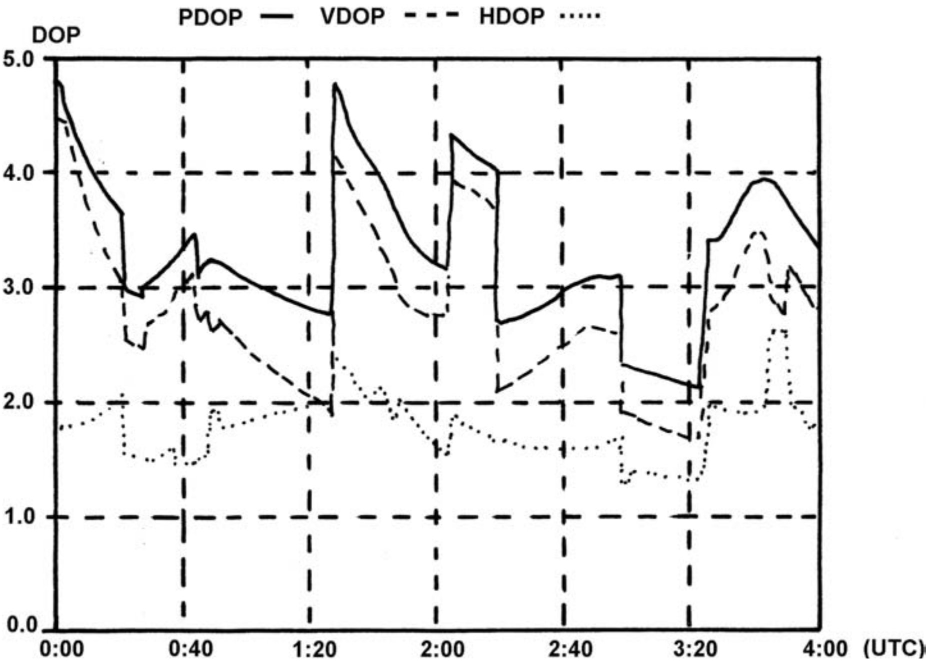


**Figure 5.15** Trimble mission planning DOP graph taken over 4 hours. A low DOP indicates a high level of accuracy.

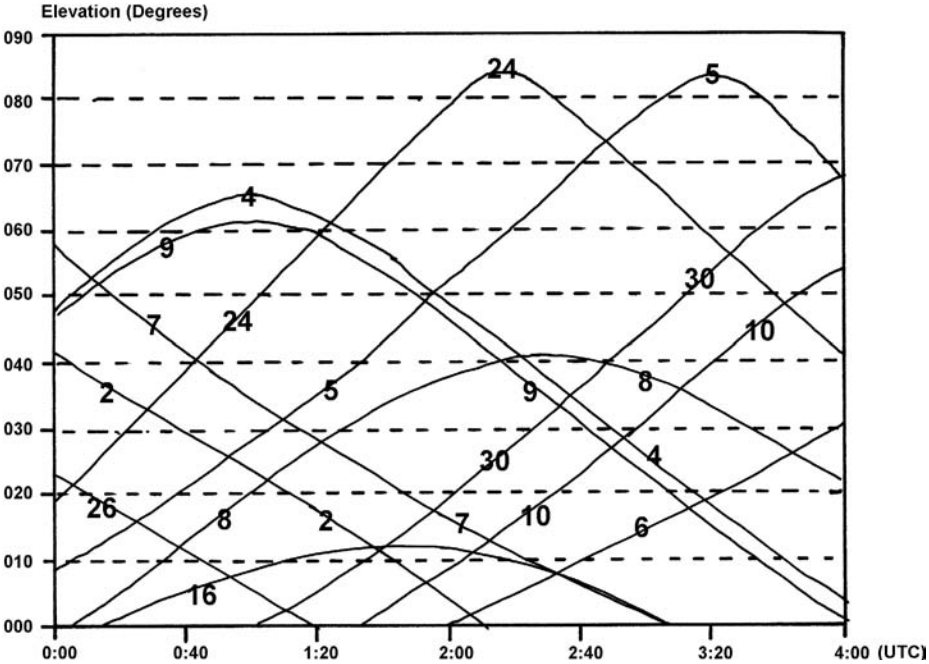


**Figure 5.16** Trimble SV elevation plot. A 4-h plot showing all SVs in view.

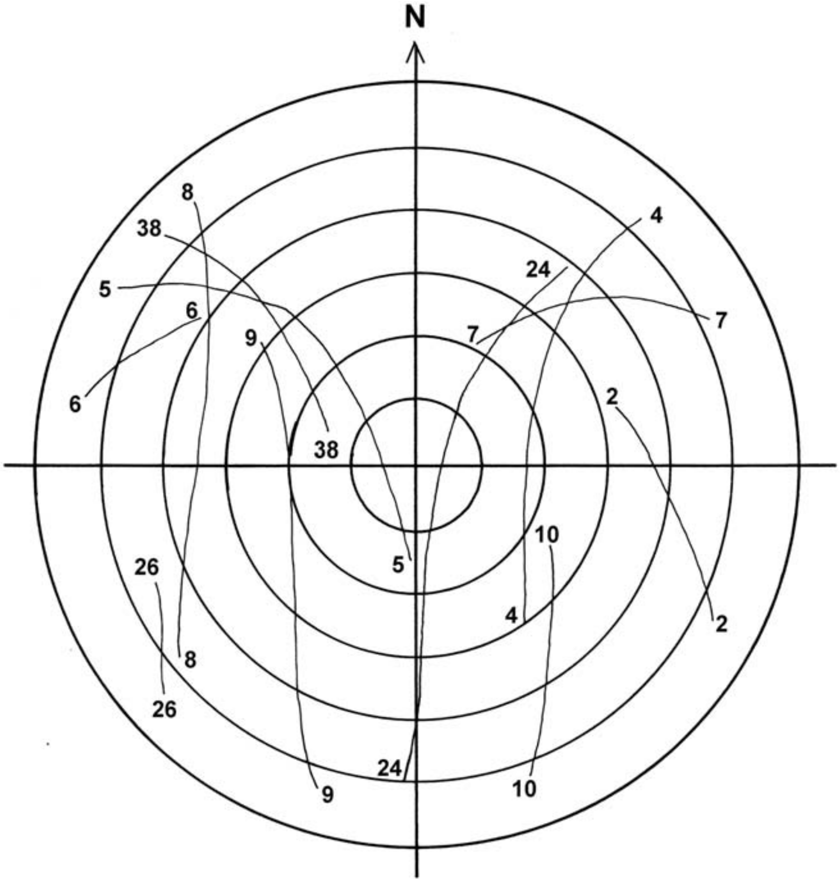


**Figure 5.17** Trimble SV sky plot presentation. A GPS receiver is in the centre of concentric circles. The outer ring represents the horizon or zero elevation.

performance as a position fixing system. However, the total error produced by a combination of all error-producing factors is very small. Assuming that the system is free from operator error (corrupt data inputting), the error most likely to downgrade system accuracy is an error in the SV clock, which in turn will cause range measurement error.

GPS accuracy is promulgated in a number of ways as indicated below.

- *Circular Error Probable (CEP).* This represents an accuracy figure achievable 50% of the time in two dimensions only. This is a fix error in latitude and longitude.
- *Spherical Error Probable (SEP).* An accuracy that is achievable 50% of the time in all three dimensions.
- *Root Mean Square Radial Distance error ($d_{RMS}$).* A circle around the true position containing 95% of the fix calculations.
- *User Equivalent Range Error (UERE).* This is determined by summing the squares of the individual range errors and then taking the square root of the total.

The following errors affect the accuracy of GPS position fixes.

### *Satellite clock error*

It has already been stated that a satellite clock oscillator is a precision instrument, but it is still necessary to re-adjust it from the ground support network. Error introduced by SV clock error is unlikely to exceed 1 m and regular uplinking of clock data reduces it to a minimum. Block IIA and Block IIR satellites, the latest SVs, carry better clock oscillators and will consequently provide higher accuracy fixes.

### *Ionospheric delay error*

As the two transmitted carriers must pass through the ionosphere, a speed reduction caused by refraction of the radio wave occurs. The extent of the delay, and consequently the error introduced into the pseudo-range measurement calculation, is dependent upon the electron density the radio wave encounters along the signal path. Electron density is itself dependent upon three main factors:

- the time of day
- the SV elevation
- the latitude of the receiver.

Fortunately, ionospheric error is inversely proportional to the square of the carrier frequency. GPS SVs transmit on two frequencies so that the delay may be quantified in the receiver, an error correction figure calculated and applied to the final fix solution. After all corrective data has been applied to the solution in a single frequency GPS receiver system, fix error due to the ionosphere is unlikely to exceed 10 m.

### *Tropospheric delay error*

Extending from the earth's surface to an altitude of 70 km, the troposphere also introduces a delay into the pseudo-range calculation. Unfortunately the error is independent of frequency, but it is predictable. GPS receivers hold a software solution in the form of a mathematical model to eliminate the effect of this delay. Figures for relative humidity, pressure and temperature are interfaced with the processor computer to produce corrective data which is then applied to fix calculation. Error from this source is unlikely to exceed 1 m.

Both ionospheric and tropospheric errors are reduced if ranges are measured from SVs showing a high elevation from the receiver. Modern receivers are capable of automatically selecting SVs with the highest elevation or those exceeding pre-set limits.

### *Multipath error*

This results from the reception of the same SV signal from more than one source. A major contributor to this error is the reflected wave from an object close to the receiving antenna. Each receiver position is unique and therefore the error is not consistent. Final fix errors in the region of 1 metre can be produced by this effect. Careful positioning of the antenna will eliminate this error.

### *Relativity error*

A commonly referred error is that produced by the effects of relativity. It is entirely predictable and is effectively cancelled in the GPS but it is briefly described here.

Albert Einstein stated that time is compressed by the mass of the earth. Time on the surface of the globe is compressed by $1.4 \times 10^{-9}\,\mathrm{ms}^{-2}$ compared to time in free space. It is evident that as one travels further away from the earth's surface towards free space, the compression of time is of less significance. At the altitude of a GPS SV, time compression is calculated to be $0.4 \times 10^{-9}\,\mathrm{ms}^{-2}$. An effective rate range time error of 1 ns therefore exists between the time on board the SV and that in the receiver. At the accepted propagation velocity of radio waves, i.e. $300 \times 10^{6}\,\mathrm{ms}^{-1}$, an error of 1 ns corresponds to a range error of 0.3 m. In addition, a second time error is produced by time compression caused as the SV moves at $26.61\,\mathrm{kms}^{-1}$ through space. To compensate for all relativity errors, the SV clock oscillator frequency is slightly offset. By the time that the radio wave arrives at the receiving antenna the effects of relativity will have been cancelled and the pseudo-range can be more accurately calculated.

These are by no means the only factors that affect the accuracy of the GPS system but they are often referred to in papers on this subject. A combined position error produced by all the above factors is unlikely to exceed 12 m.

### *User Range Error (URE)*

This is a parameter for the estimated error in range calculation due to unknown factors. These include multipath, unmodelled atmospheric effects, operator error and unpredictable orbital errors. The URE figure is sent from SVs to GPS receivers and may be displayed in metres.

## 5.8 Differential GPS (DGPS)

As has already been stated, the accuracy of GPS fixes can be vastly improved using differential techniques. Experimental differential systems have been in use for some years as part of earlier hyperbolic earth-based navigation systems. DGPS is merely an improvement of those now outdated systems. The principle, as shown in Figure 5.18, is that GPS data from SVs are downloaded to both a mobile station and a fixed station at a precise location. A computer at the fixed site calculates the pseudo-range from GPS SVs and then compares it with the known ranges for that precise geographic location. It then computes a range error figure which is transmitted to mobile stations where it is used to correct the pseudo-range system errors.

The use of DGPS does not eliminate errors introduced by multipath reception or receiver noise.

For maritime use, DGPS differential monitor stations have been established around the coast of some 28 countries. As examples, the US Coast Guard maintains DGPS transmission stations around the continental coastline of the USA (see Figure 5.19 and Table 5.6), and in the UK beacons are operated by Trinity House and the General Lighthouse Authority (see Figure 5.20 and Table 5.7).

Corrective data are transmitted from the beacons on frequencies in the lower medium frequency band and as a result the range over which they can be reliably received is limited to between 100 and 250 km. But DGPS can and does assist in waters where freedom to manoeuvre is restricted.

The US Coast Guard and the International Association of Lighthouse Authorities (IALA) support the International Telecommunications Union (ITU) Recommendation M.823 which allows for DGPS data to be transmitted as supplementary information on the radiobeacon band 283.5–315 kHz (285–325 kHz in some parts of the world). The transmission protocol RTCM SC-104 (developed by the Radio Technical Commission for Maritime Services Special Committee 104) is used to determine the speed and data format of the transmission. DGPS data is phase shift keyed onto the carrier at a rate of 100 or 200 bits per second.

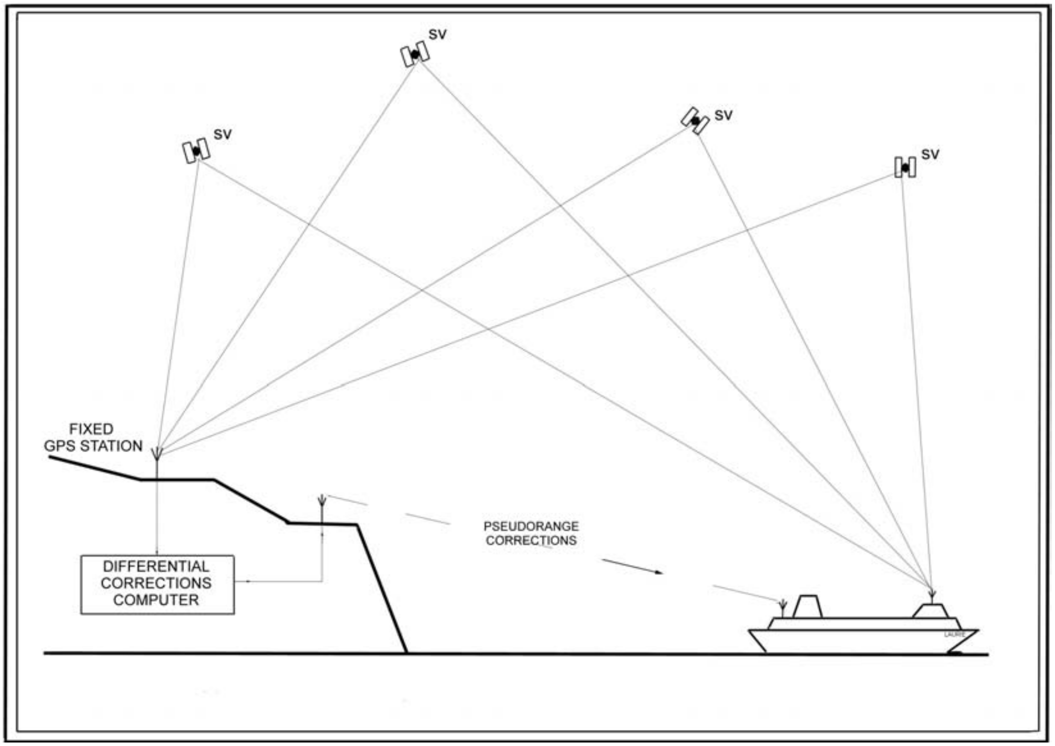


**Figure 5.18** Principle of operation of DGPS.

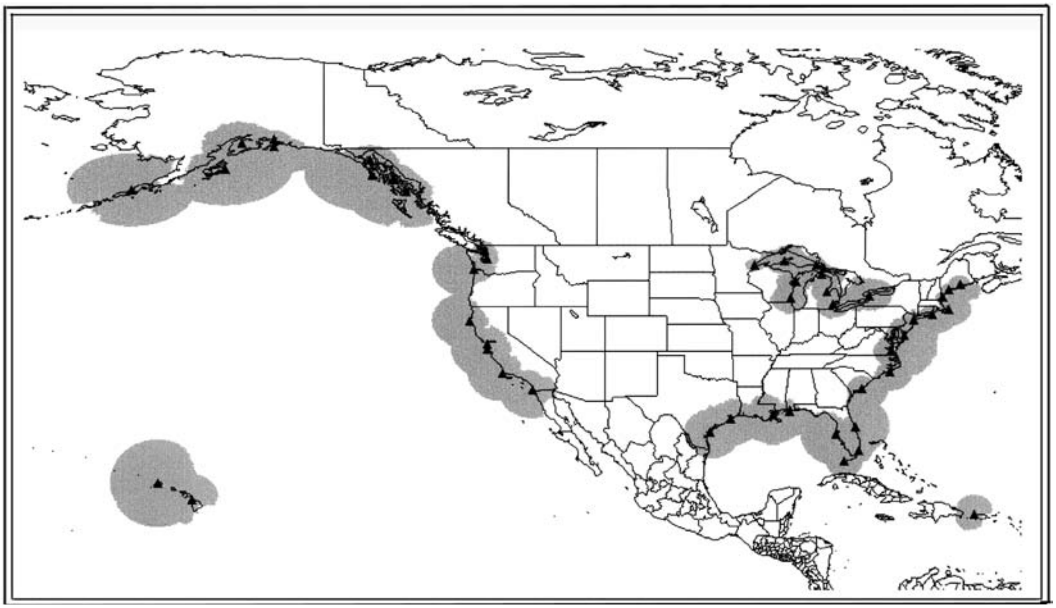


**Figure 5.19** Maritime DGPS coverage of the United States. (Reproduced courtesy of the United States Coast Guard.)

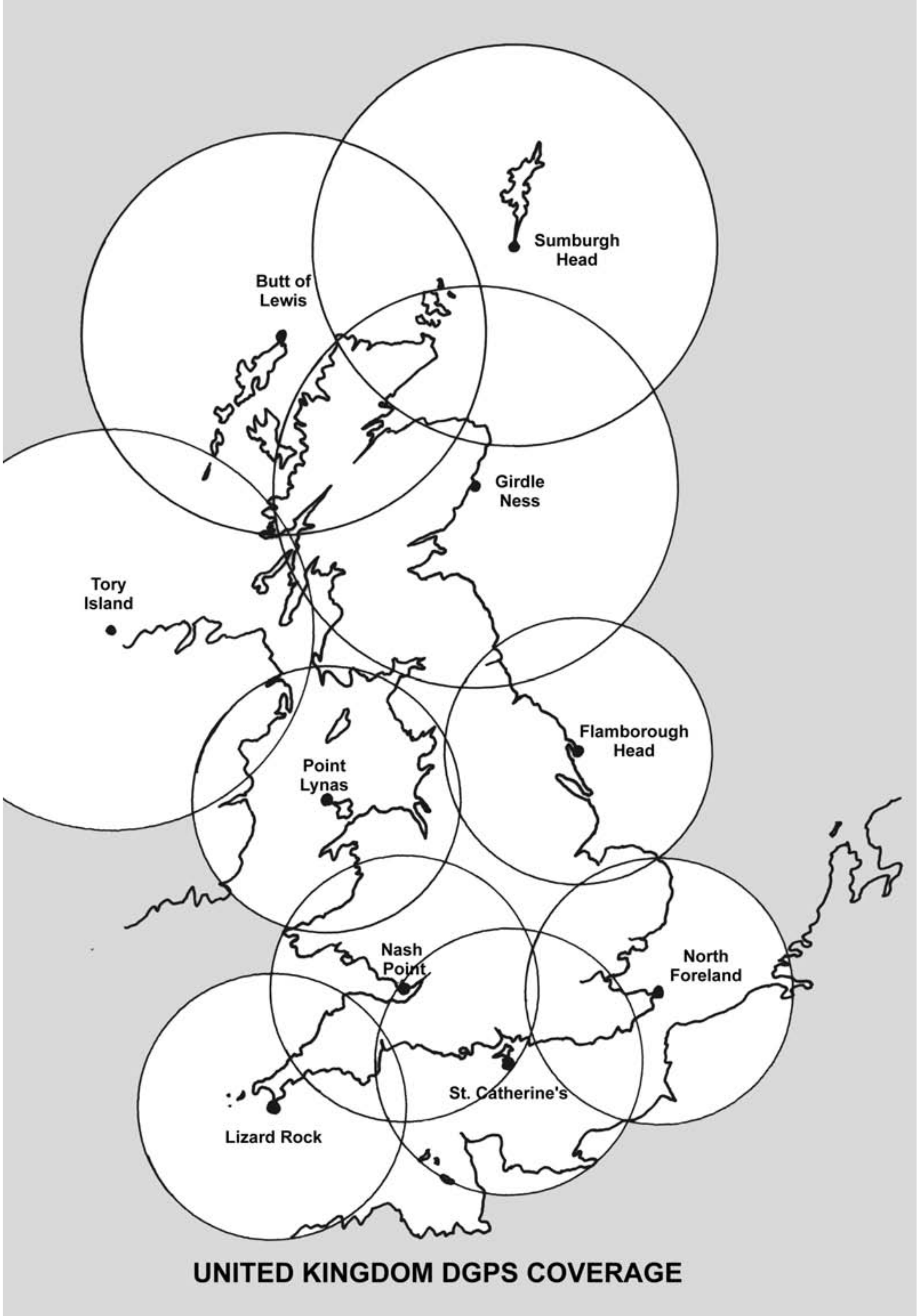


**Figure 5.20** DGPS coverage of the UK coastline.

**Table 5.6** Florida differential GPS stations data

| *Station* | *Location* | *Frequency (kHz)* | *Nominal range (km)* |
|---|---|---|---|
| Cape Canaveral | 28.27 N 80.32 W | 289 | 200 |
| Miami | 25.43 N 80.09 W | 322 | 75 |
| Key West | 24.34 N 81.39 W | 286 | 75 |
| Egmont Key | 27.36 N 82.45 W | 312 | 200 |

Source: United States Coast Guard.

**Table 5.7** UK differential GPS station data

| *Station* | *Location* | *Frequency (kHz)* | *Nominal range (km)* |
|---|---|---|---|
| Sumburgh Head | 59.51 N 01.16 W | 304.0 | 275 |
| Butt of Lewis | 58.31 N 06.16 W | 294.0 | 275 |
| Girdle Ness | 57.08 N 02.03 W | 311.0 | 275 |
| Tory Island | 55.16 N 08.15 W | 313.5 | 275 |
| Flamborough Head | 54.07 N 00.05 W | 302.5 | 185 |
| Point Lynas | 53.25 N 04.17 W | 305.0 | 185 |
| Nash Point | 51.24 N 03.33 W | 299.0 | 185 |
| North Foreland | 51.23 N 01.27 E | 310.5 | 185 |
| St. Catherine's | 50.35 N 01.18 W | 293.5 | 185 |
| Lizard Rock | 49.58 N 05.12 W | 284.0 | 185 |

Source: Trinity House

### 5.8.2 Wide Area Differential GPS (WDGPS)

WDGPS is a real-time global differential system currently under consideration for future implementation. Using the INMARSAT communications network, differential data will be transmitted to ships throughout the world enabling better fixes to be made. It is still in the discussion stage.

## 5.9 GPS antenna systems

Arguably the antenna is the most critical part of any radiocommunications system but unfortunately it is the piece of hardware that is most often ignored. Carefully designed and constructed, an antenna sits, open to the elements, on board a vessel's superstructure in a position where routine maintenance can be difficult. GPS antennas are small and rigidly constructed and to ensure that they survive the elements they are protected by a raydome.

In common with the INMARSAT communications antenna, a GPS antenna ideally requires an unobstructed view through 360° from the horizon up to 90° in elevation. Radiated energy from other microwave transmission systems can damage sensitive pre-amplifier circuitry inside the GPS protective dome. It is wise, therefore, to mount the GPS antenna below the INMARSAT raydome and outside the radar transmission beamwidth as shown in Figure 5.21.

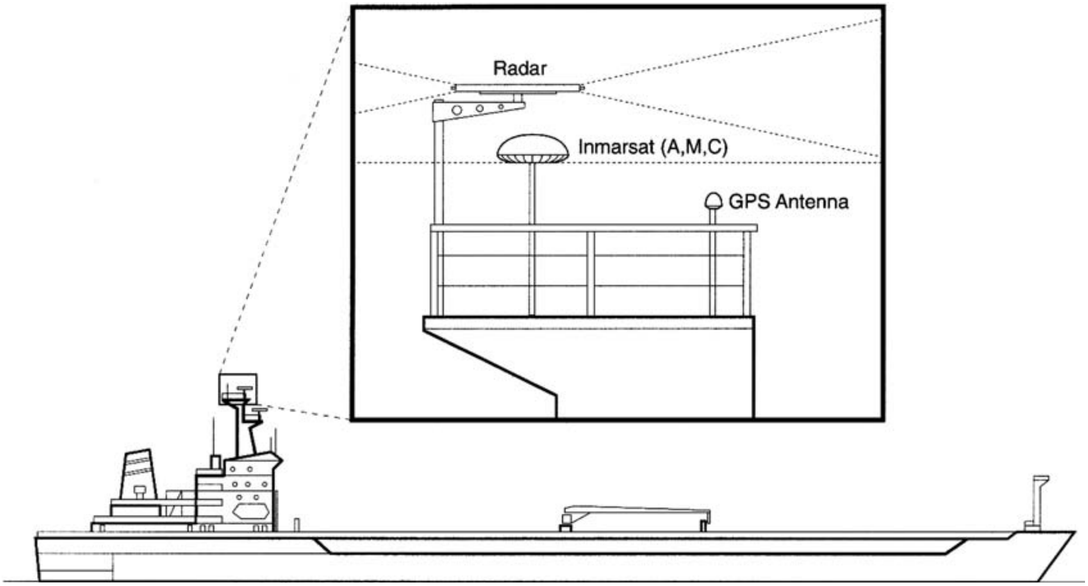


**Figure 5.21** A GPS antenna mounted below other microwave system antennas on the superstructure of a merchant vessel. (Reproduced courtesy of Trimble Navigation Ltd.)

Other factors to be considered when siting an antenna are as follows.

- Mounting the antenna on the top of a tall mast will accentuate range errors caused by the vessel's motion especially if DGPS is used. The range error is dependent upon the extent of the vessel's motion and is therefore unpredictable.
- No special ground plane is required, but a large open deck space below the antenna will reduce the error caused by reflected multipath signals.
- Stays, masts and dry sails in the path between the SV signal and the antenna will have little effect of the received signal.
- GPS systems use an active (containing some electronic circuitry) antenna head which can be affected by severe vibration. Mount the antenna away from other antennas, engine housings or exhaust stacks.

## 5.10 GPS receiver designation

Because GPS is freely available to all users throughout the world, the range of available user equipment is vast. There are thousands of manufacturers producing a bewildering range of fixed and mobile equipment, all of which must comply with GPS standards. GPS receiver architecture varies depending upon how it is to be used. The following list itemizes the most popular GPS receiver systems currently produced. The more commonly found commercial receivers are listed first.

### *Multiplex (MUX) Receivers*

Amongst the cheapest GPS receiver architecture, MUX receivers are commonly found in the commercial sector. A MUX receiver continuously tracks multiple SVs by continuously switching its single channel between them. Time measurements and data streams are held in memory algorithms and ‘topped-up’ when data is made available by the MUX switch rate. Receiver architecture is less complex and consequently cheaper. MUX receivers are only used on slow moving platforms such as merchant vessels.

### *Sequential Receivers*

Receiver architecture is designed to track one SV at a time and calculate the pseudo-range. The data is held in memory until four SVs have been interrogated, when the position–velocity–time (PVT) fix is calculated. These receivers are the least expensive and possess the slowest time-to-first-fix (TTFF) performance.

### *Single Channel Sequential Receivers*

As the title suggests, these receivers use a single channel to sequentially measure the pseudo-ranges from four SVs. Each SV is fully interrogated in sequence and the final fix made from stored data. Any uncorrected movement of the receiver during this process reduces the fix accuracy.

### *Dual Channel Sequential Receivers*

The only advantage of this type of receiver is that, in using two channels, it reduces the time it takes to calculate a fix. They tend to be used on medium velocity platforms, such as aircraft.

### *All-in-View Receivers*

An All-in-View receiver has the necessary hardware to search the sky and track all the SVs that it finds. Whilst four SVs are needed to give a good PVT fix, it is likely that satellites will be lost before they can be fully interrogated. This type of receiver architecture can track seven or eight SVs continuously so if some SVs drop out of its view the PVT fix should still be good. If satellite data is not lost during tracking, a fix is produced from the data of more that four SVs. In general, the more satellites that provide data for a fix, the better the fix.

### *Continuous Tracking Receivers*

This type of GPS receiver possesses multiple channels to track four SVs simultaneously whilst acquiring new satellites. TTFF figures are the lowest for any receiver architecture and PVT fix accuracy can be maintained on high velocity platforms such as fighter aircraft and missiles. Continuous tracking receivers offer the best performance and versatility but, as you would expect, they are the most expensive.

### *Differential GPS Receiver*

DGPS receivers are now in common use on maritime vessels that require better PVT fix accuracy than can be obtained with a basic receiver. Vessel’s trading in confined waters use DGPS receivers. They are more expensive, but the cost is justified. (See the section on DGPS.)

*Time Transfer Receiver*

This type of GPS receiver provides an accurate time source. It may be integrated into one of the receiver systems previously described or the time figure may be used in other navigation fix solutions.

## 5.11 Generic GPS receiver architecture

This section includes the description of a simple receiver and then goes on to consider specific modern systems. Figure 5.22 shows a generic GPS receiver system.

### 5.11.1 SV selection and acquisition

If the receiver can immediately 'recognize' a SV it will target that satellite and begin a tracking sequence. This is possible if the receiver has already downloaded almanac data from any SV, if not it will enter a 'search' mode and systematically hunt the sky looking for a recognizable PRN code. Once this is received, tracking will be initiated, lock will be achieved and the navigation message can be interrogated. The current almanac will then be cross-examined and the health status of all the other satellites will be determined. The computer then selects the best subset of visible SVs, or, all-in-view. In practice, data from a minimum of four SVs is required to provide a reliable navigation fix, but the greater the number that can be tracked and accessed, the better.

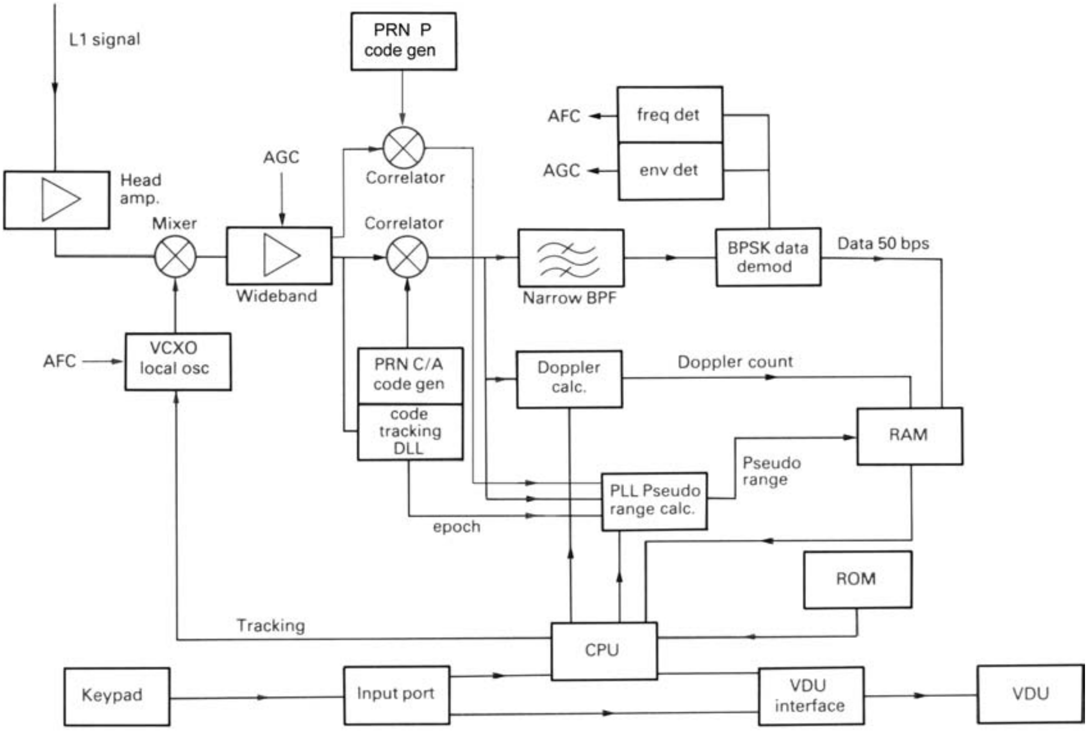


**Figure 5.22** A generic GPS receiver system.

Because of limited satellite transmitter power, spread spectrum modulation techniques and ionospheric attenuation, the satellite signal power received at the earth's surface is far less than the receiver's natural or thermal noise level. This minute signal is received by a compact, fixed, above-deck unit using an isotropic antenna with ground plane radial reflectors, a low noise pre-amplifier and filters. Circularly polarized radio waves from the SV, are received by the isotropic antenna whilst the radial reflectors reduce the problem of multipath errors caused by the earth's surface reflected signals. The head unit should be mounted in such a way that the antenna has a clear view of the whole area in azimuth from the zenith to the horizon. Input to the receiver is therefore the amplified SV signal at 1575.42 MHz, plus a slight Doppler frequency shift and possessing a very poor signal-to-noise ratio.

The single signal mixer down-converts the $L_1$ carrier to an intermediate frequency. Frequency conversion is achieved using a Variable Frequency Local Oscillator (VCXO) under the control of both the Central Processing Unit (CPU) and a signal derived Automatic Frequency Control (AFC). CPU input to the VCXO enables initial SV tracking to be achieved and the tiny direct current AFC, derived from the received signal, maintains this lock. A wideband IF amplifier is used to permit reception of the 20.46 MHz bandwidth P code enabling future modification of the receiver to be made if required. Output from this amplifier is coupled to a correlator along with the locally generated PRN C/A code.

It is essential that the receiver tracks the received signal precisely despite the fact that it is at an amplitude which is hardly above the locally generated noise level. To achieve tracking the received signal is applied to a Delay Lock Loop (DLL) code tracking circuit that is able to synchronize the locally generated PRN code, by means of the EPOCH datum point, with the received code to produce the reconstituted code to the narrow bandpass filter. The DLL is able to shift the local PRN code so that it is early or late (ahead or behind) when compared to the received code. A punctual (Pu) line output to the correlator is active only when the two codes are in synchronism. PRN codes are described in more detail at the end of this chapter.

Output of the correlator is the autocorrelation function of the input and local PRN C/A codes. The bandwidth of the narrow band bandpass filter is 100 Hz so that data is passed only to the BPSK data demodulator where code stripping occurs. The autocorrelated C/A code is also used for both Doppler and pseudo-range measurement. The PLL used for pseudo-range measurement has a clock input from the CPU to enable clock correction and an EPOCH input each millisecond for alignment.

All receiver functions are controlled by a microprocessor interfaced with a keypad and a VDU display. The use of a microprocessor ensures economy of design. In this outline description most of the control lines have been simplified for clarity. The receiver operating sequence is given in Table 5.8.

### 5.11.2 Autocorrelation of random waveforms

The main function of the correlator in this receiver is to determine the presence of the received PRN code that is severely affected by noise. Correlation is a complex subject and the brief description that follows attempts to simplify the concept. Both the C/A and P codes are 'chain codes' or 'pseudo-random binary sequence' (PRBS) codes that are actually periodic signals. Within each period the code possesses a number of random noise-like qualities and hence is often called a 'pseudo-random noise code' (PRN code). The PRN binary sequence shown assumes that the code has a period of 15 samples, i.e. it repeats every 15 bits. The GPS P code possesses a period of 267 days and the C/A code a period of 1 ms. It is obvious therefore that a PRN code can possess any period.

To establish the autocorrelation function, both the received C/A code and the locally generated C/A code are applied to the correlator. Consider the local code to be shifted three stages ahead or behind (early or late) on the received code by a time period ($t$) known as parametric time. To obtain the product of the two codes, add each received bit to a locally generated bit shifted in time, as shown in Figure 5.23.

**Table 5.8** Receiver operating sequence

| | |
|---|---|
| 01 | Initialize |
| 02 | Search for an SV |
| 03 | Identify $L_1$ carrier |
| 04 | Acquire $L_1$ C/A code |
| 05 | Track $L_1$ C/A code |
| 06 | Strip data |
| 07 | Measure pseudo-range |
| 08 | Measure Doppler frequency shift |
| 09 | Store data |
| 10 | Commence next SV search and repeat steps 03–09 |
| 11 | Commence next SV search and repeat steps 03–09 |
| 12 | Commence next SV search and repeat steps 03–09 |
| 13 | Compute navigation position |
| 14 | Output position data to display |

The product is achieved by adding bits of data using the terms:

$(+1) + (+1) = +1$
$(-1) + (-1) = +1$
$(+1) + (-1) = -1$
$(-1) + (+1) = -1$

The average value of the products thus produced is –1/15. If the local code is now shifted one bit to the right and the products are noted again, the average value of the products is –3/15. When the two

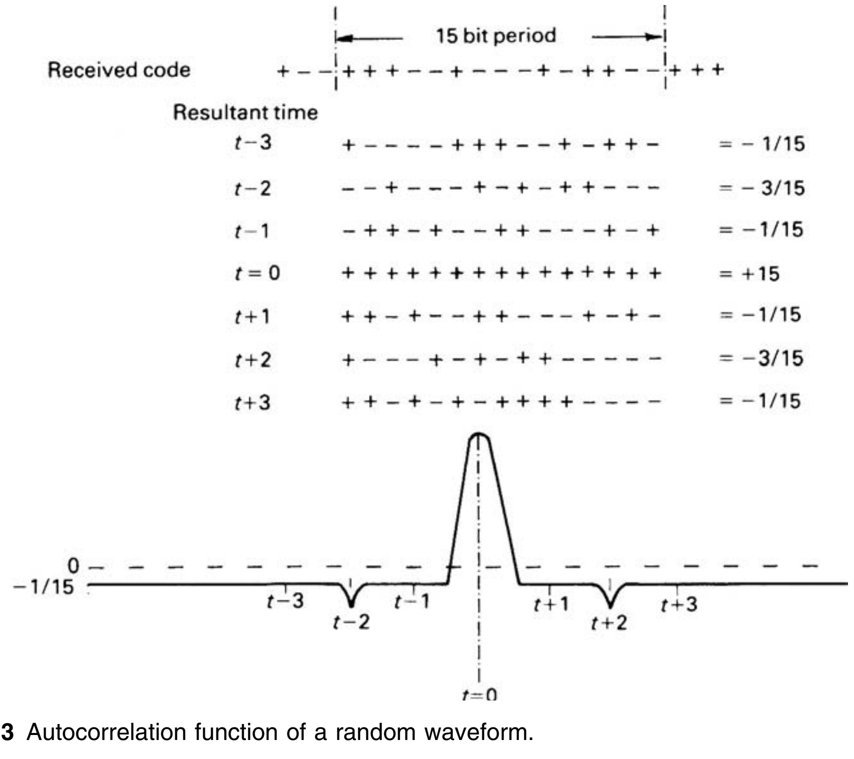


**Figure 5.23** Autocorrelation function of a random waveform.

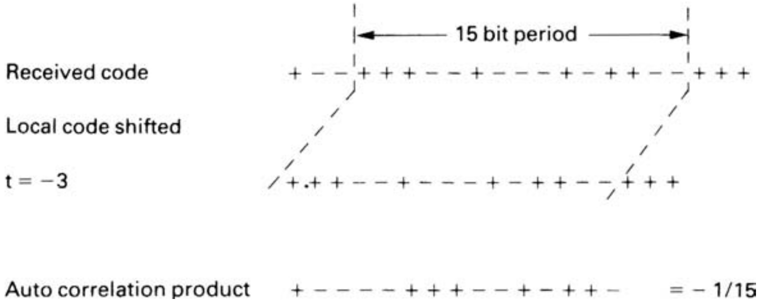


**Figure 5.24** The autocorrelation product of a random waveform.

codes are synchronized the product of all bits is +1. Therefore the average value of the products is also +1. This is the only time per code period when all the code products are +1. The peak thus produced is called the autocorrelation function (see Figure 5.24) and enables the received code to be identified, even in the presence of noise which is essentially an amplitude variation.

The PRBS is periodic, therefore the autocorrelation function is periodic and repeats at the rate of the original signal. It is possible to determine the period of the received code by noting the periodicity of the peaks produced in parametric time. Thus the C/A code can be acquired even when it is severely affected by noise. The autocorrelation function peak also indicates the power density spectrum of the received code signal. A signal with a wide bandwidth (the P code) produces a sharper narrower correlation spike, whereas a wide correlation spike indicates a narrow bandwidth signal (C/A code). Obviously the width of the correlation spike is inversely proportional to the bandwidth of the received signal code.

The user equipment just described demonstrates many of the principles of GPS reception. However, equipment manufacturers will have their own ideas about how a GPS receiver should be configured.

## 5.12 GPS user equipment

The GPS is the undisputed leader in modern position fixing systems and, when interfaced with various shipboard sensors, GPS equipment forms the heart of a precise navigation system offering a host of facilities. Modern equipment is computer controlled, and this fact along with a versatile human interface and display means that the equipment is capable of much more than that produced for earlier position fixing systems.

There is a huge selection of GPS equipment available from a large number of manufacturers. Much of this equipment is designed for the small craft market, more is specifically designed for geodesy and earth mapping, still more is designed for the aeronautical market, and more for trucking operators. In fact it appears that the GPS has found a range of diverse uses in every corner of the globe. This book is written for the maritime navigation sector of this huge market and equipment is described to demonstrate the versatility and flexibility of modern GPS receivers.

Two huge companies that offer a full range of GPS equipment and services are Trimble Navigation Ltd. based in the heart of silicon valley at Sunnyvale, California, and Garmin based at Olathe, Kansas in the USA.

### 5.12.1 Trimble GPS receiver specifications

At the top of the Trimble's GPS range is the NT300D, a 12-channel parallel GPS receiver, capable of tracking up to 12 satellites simultaneously and also containing a dual-channel differential beacon receiver. The equipment is capable of submetre accuracy derived from carrier-phase filtered $L_1$ pseudo-range calculations. In addition, vessel velocity is obtainable from differentially corrected Doppler measurements of the $L_1$ carrier. Position information is displayed on a backlit LCD screen in one of two main navigation modes.

Interfacing with other navigation equipment is via one of the two serial RS-422 data ports using a variety of protocols including NMEA-0183 output and RTCM SC-104 in/out. Speed data output is available at the standard rate of 200 ppnautical mile.

### *Receiver operation*

At switch on, the equipment automatically begins to acquire satellites and calculate range error to produce a position fix. TTFF varies between 30 s and 2–3 min depending upon the status of the GPS almanac, ephemeris data stored in the NT GPS's memory, and the distance travelled while the unit was switched off. During the acquisition process, the equipment operates on dead reckoning and shows this by displaying a DR in the top right corner of the display.

Figure 5.25 shows the user interface of the Trimble Navigation GPS NT200D. The buttons/keypads data input controls have been ergonomically designed to be easily operated and user friendly. A 15 cm (6 inch diagonal), high resolution, 320 × 240 pixel, backlit, LDC displays navigation data that can be easily read in most lighting conditions. Referring to Figure 5.25, the numbered functions are as follows.

1 Power key
2 Display
3 Brightness and contrast keys. Standard up/down scrolling key for screen viewing parameters.
4 Numeric keypad. Used to enter numeric data as well as controlling chart information layers when in the chart mode of operation.
5 Cursor controls. Arrow keys permitting movement of the cursor on those screens where it is present. When inputting data they are used to move through the programming functions.
6 Function keys. Used to access various functions.
SETUP: used when customizing the operation of the equipment.
STATUS: used to display various GPS parameters such as signal strength.
NAV: toggles between NAV1 and NAV2 displays.
SAVE: pressing this displays current position and time and gives the user a choice of entering the position as a waypoint or selecting the position as an emergency destination – the 'man overboard' function.
WAYPT: used to access waypoint and route libraries.
7 Soft keys. So named because the functions they perform changes from screen to screen.
8 Menu key. Toggles the soft key labels on and off.
9 Plot key. Toggles between an electronic chart display and a Mercator grid display.

The NAV 1 screen shown is a graphic depiction of the vessel's relationship to the intended course. The intended course, represented by the central lane in the graphic, is based on the active route and current leg. The next waypoint is shown, by number and name, in the box located above the central lane.

At the top of the page, the screen header displays the current mode of operation. This may be DGPS, GPS, DR or EXT (external). External mode indicates that the equipment is receiving updates from an external device.

In the centre of the display is a circular symbol with crossed lines representing the ship's position. An arrow intersecting the screen centre indicates the ship's current heading (course over ground

(COG)) relative to the destination. When this arrow points at the next waypoint (course to waypoint (CTW)), the ship is heading in the correct direction; COG = CTW.

A right or left offset of the ship's symbol signifies the cross-track error (XTE). No error exists when the symbol is shown in the centre of the lane. XTE limits can be set using the main Setup screen. The relative velocity of the ship is indicated by the rate of advance of the horizontal lines located outside the central lane.

Other data fields may be selected for display. In Figure 5.25 the following have been selected: true course over ground (COG), speed over ground (SOG) in knots, XTE in NmR, and the ship's true heading (HDG) in degrees. Other options are CTW, speed (SPD), distance to waypoint (DTW), distance to destination (DTD), velocity made good (VMG), and distance made good (DMG).

An alternative display, NAV 2 in Figure 5.26, shows a graphic representation of a compass displaying the vessel's course COG and the bearing to the next waypoint CTW. The compass card graphic consists of an inner ring with a COG arrow and an outer ring with a CTW indicator arrow. When the two arrows are in alignment, COG = CTW, the vessel is on course. The compass graphic defaults to a north-up presentation but may be changed to a head-up display.

At the bottom of the display a steering indicator, labelled XTE, shows any cross track error in nautical miles. When the two arrowheads are in alignment at the centre of the bar, XTE is zero.

As a further indication of the capabilities of a modern electronic system, the Trimble NT GPS range may be fitted with a Smart Card Reader to read Navionics chart cards.

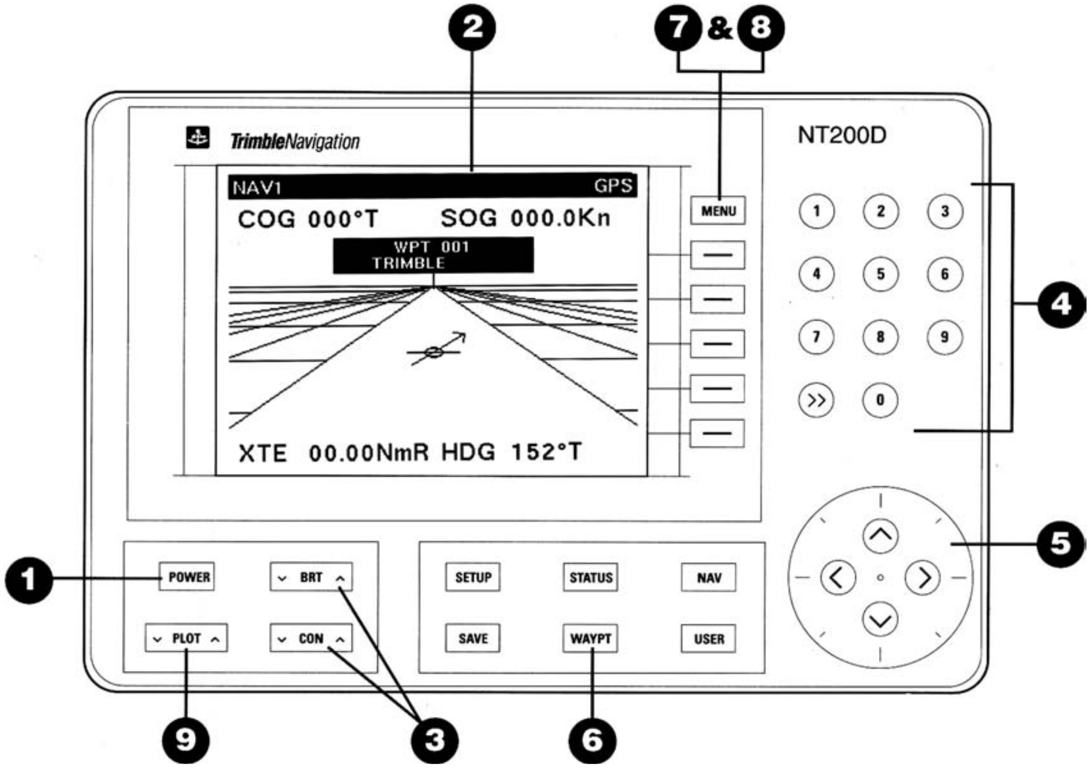


**Figure 5.25** The NT200D GPS receiver displaying the NAV1 navigation display. (Reproduced courtesy of Trimble Navigation Ltd.)

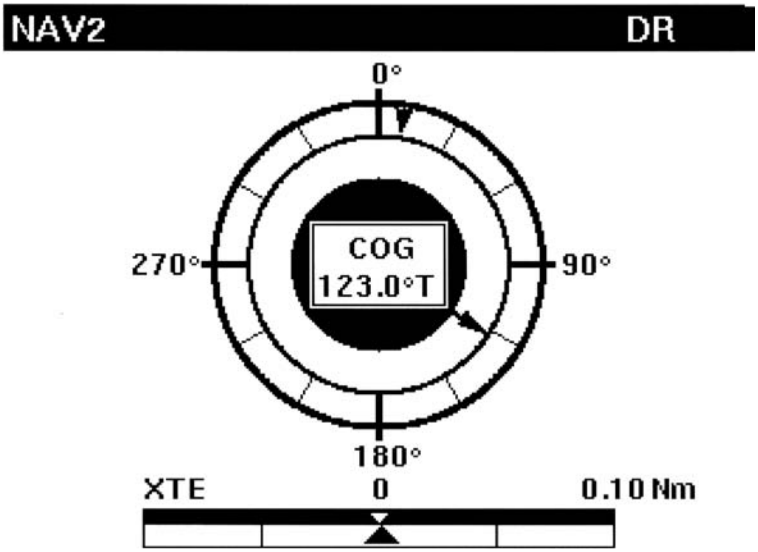


**Figure 5.26** NAV 2 display. (Reproduced courtesy of Trimble Navigation Ltd.)

Each Navionics card holds the data necessary to give a screen display in the form of a maritime chart for a specified geographical area. The display then integrates the GPS data with the chart data, producing a recognizable nautical chart and the vessel's course and speed. Figure 5.27 shows a vessel (a flashing icon) with a track (a solid line) taking it under the western part of the Bay Bridge and a residual course (a line of dots) extending back to Alameda.

To avoid cluttering the chart, not all available data is shown on the Bay Area chart in Figure 5.27. Additional key commands are able to bring up the following information: depth contours, XTE lines, COG indicator, names (of cities, ports, bodies of water etc.), track, lighthouses and buoys, waypoints, landfill (for a clearer display of coastlines), maps, and much more. It is also possible to zoom in/out to show greater detail.

Another navigation screen display is the Mercator grid plot (Figure 5.28) showing the vessel's current position, the track history and the waypoints and legs in the active route. There are several scale or zoom levels ranging from 010 to 1000 km plus nautical miles or Mi increments.

Modern equipment is capable of much more than simply calculating and displaying position and track information and the NT200D is no exception. The versatility of its display coupled with adequate computing power and reliable data processing circuitry means that a wealth of other information can be accessed and presented to users. Set-up screens, system health checks, interface information, status displays, waypoint information, routes and more can be selected for display. Two displays in the status directory (Figure 5.29), of interest to students, present information about the satellites in view.

In Figure 5.29(a), the vessel is at the centre of concentric circles with a radial arrow indicating the current COG. The outer ring of the plot represents the horizon (0° elevation) and the inner rings, 30° and 60° elevation, respectively. Satellites in the centre of the plot are directly overhead (90° elevation). A satellite's true position in azimuth is shown relative to the north-up plot or may be determined relative to the vessel's COG.

Blackened icons indicate satellites being tracked by the receiver. Received data from the others falls below the parameters selected for their use. The table on the right shows the number of the SV and

**Figure 5.27** Chart display of San Francisco Bay and approaches using data input from a smart card. (Reproduced courtesy of Trimble Navigation Ltd.)

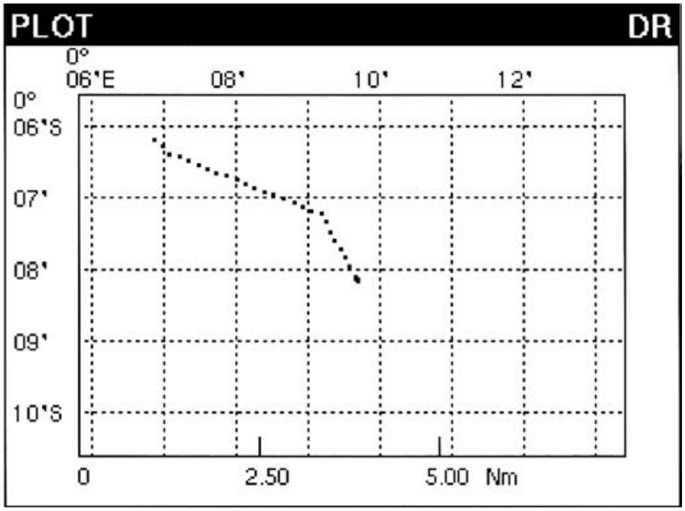
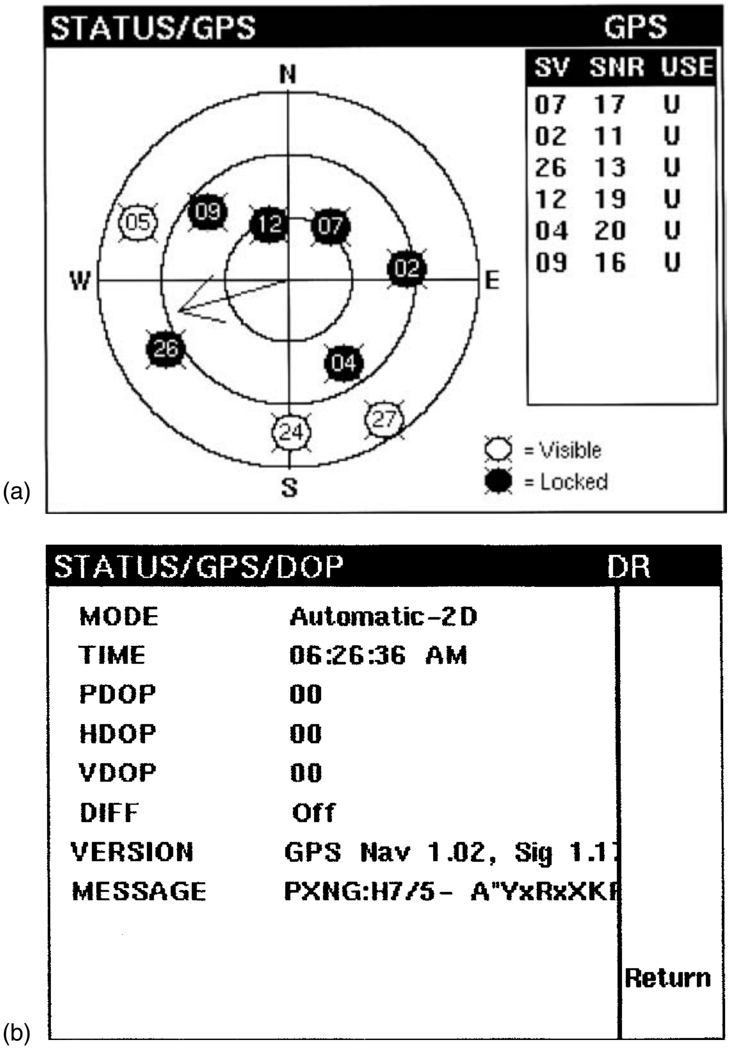


**Figure 5.28** The Mercator grid plot screen display of the GPS receiver DR track. The vessel's current position is indicated by a flashing icon in the centre of the screen. (Reproduced courtesy of Trimble Navigation Ltd.)



| SV | SNR | USE |
|---|---|---|
| 07 | 17 | U |
| 02 | 11 | U |
| 26 | 13 | U |
| 12 | 19 | U |
| 04 | 20 | U |
| 09 | 16 | U |

| | |
|---|---|
| MODE | Automatic-2D |
| TIME | 06:26:36 AM |
| PDOP | 00 |
| HDOP | 00 |
| VDOP | 00 |
| DIFF | Off |
| VERSION | GPS Nav 1.02, Sig 1.1 |
| MESSAGE | PXNG:H7/5- A"YxRxXK |

**Figure 5.29** Satellite status/GPS display. Darkened icons are the numbered satellites currently being tracked by the receiver. Light icons represent received satellites that fall below the parameters selected for their use. The vessel is in the centre of the display and its course-over-ground is indicated by an arrow. (Reproduced courtesy of Trimble Navigation Ltd.)

the signal-to-noise ratio (SNR) for each satellite tracked. A SNR of 15 is considered good, 10 is acceptable and a SNR below 6 indicates that the satellite should not be relied upon for a position solution. A 'U' shows that an SV is being used and a 'D' that the equipment is receiving differential correction data for the satellite.

The second Status/GPS display is the dilution of precision screen (Figure 5.29(b)). PDOP, HDOP, and VDOP are numerical values based on the geometry of the satellite constellation used in a position solution. A figure of unity, 1.0, is the best DOP achievable. The most important of these parameters is the PDOP, the position dilution of precision. The lower the PDOP figure, the more precise the solution will be and the better the position fix. In practice a PDOP figure greater than 12 should be used with caution. A PDOP in the range 1–3 is excellent, 4–6 is good, 7–9 acceptable, 10–12 marginal and 12+ should be used with caution.

HDOP represents the accuracy of the latitude and longitude co-ordinates in two- or three-dimensional solutions, and VDOP is the accuracy of the altitude in a three-dimensional solution.

The display also shows the current GPS operating mode, the time of the last GPS fix, the current DGPS operating mode DIFF, the receiver firmware version, and the GPS system message.

For further information about Trimble GPS products see www.trimble.com

### 5.12.2 Garmin GPS receiver specifications

Amongst a range of GPS equipment designed for the maritime market, Garmin offers a 12-channel GPS receiver (with an optional DGPS receiver) combined with a navigation plotter. This versatile equipment, known as the GPSMAP 225, is representative of the way that system integration is making life easier for the maritime navigator. The GPSMAP 225 effectively presents an electronic charting/navigation system based on a 16-colour active-matrix TFT display that modern navigators will feel comfortable with.

Figure 5.30 shows the front panel of the receiver including the main operator controls and a sample chart showing own ship as a wedge icon. Note that the equipment is operating in a simulation mode.

**Figure 5.30** Front panel of the Garmin GPSMAP 225 system showing operator controls and a sample navigation map generated in the simulation mode. (Reproduced courtesy of Garmin.)

#### *Operator controls*

| | |
|---|---|
| ZOOM key | Changes the map display scale to one of 16 settings, or the highway display scale to one of five settings. |
| CTR key | Eliminates the cursor and centres own vessel on the screen. |
| ARROW keys | Controls the movements of the cursor and selects screen options and positions. |
| ENT key | Used to confirm data entry and execute various on-screen function prompts. |
| MAPS key | Returns the display to the Map page and/or displays the outlines of chart coverage in use. |

| | |
|---|---|
| PAGE key | Scrolls through the main screen pages in sequence. |
| DATA key | Turns the data window on or off in map mode and toggles the displayed data on other pages. |
| MENU key | Turns the softkey menu on or off in the map mode. |
| MARK key | Captures present position for storage as a waypoint. |
| MOB key | Marks present GPS position and provides a return course with steering guidance. |
| GOTO key | Enables waypoints or target cursor position as a destination and sets a course from current position. |
| SOFT keys | Perform route, waypoint and set-up functions. Also enable custom set-ups and many navigation functions from the map display. |

### *Navigation and plotting functions*

By using the built-in simulator mode for full route and trip planning, the GPSMAP system is capable of relieving a navigator of some of the more mundane navigation exercises. The system also includes the following specification to assist with the day-to-day navigation of a vessel.

- Over 1900 alphanumeric waypoints with selectable icons and comments.
- Built-in worldwide database usable from 4096 to 64 nautical miles scales.
- 20 reversible routes with up to 50 waypoints each.
- Graphic softkeys for easy operation of the chart display.
- G-chart$^{\text{TM}}$ electronic charting for seamless, worldwide coverage (see Figure 5.33).
- On-screen point-to-point distance and bearing calculations.
- 2000 track log points with time, distance or resolution settings.
- Built-in simulator mode for full route and trip planning.
- Conversion of GPS position to Loran-C TD co-ordinates.

### *Loran-C TD conversion*

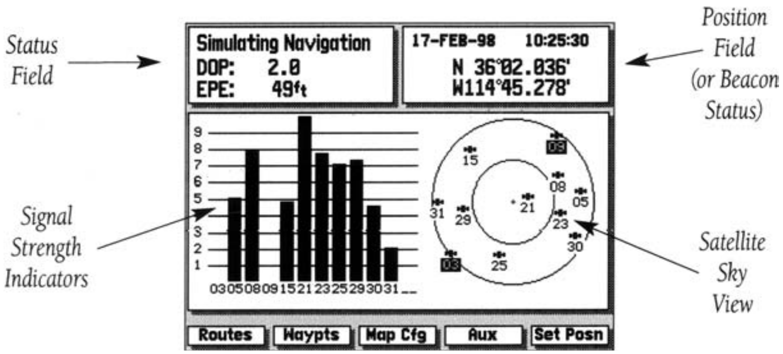
The GPSMAP unit automatically converts GPS co-ordinates to Loran-C TDs (time delay) for users who have a collection of Loran fixes stored as TDs. When the unit is used in this mode, it simulates the operation of a Loran-C receiver. Position co-ordinates may be displayed as TDs, and all navigation functions may be used as if the unit was actually receiving Loran signals. The expected accuracy is approximately 30 m.

### *GPSMAP system operation*

At power-up, the satellite status page will appear. This gives a visual reference of satellite acquisition and status, with a signal bar graph and satellite sky view in the centre of the screen. In Figure 5.31, satellites 5, 8, 15, 21, 23, 25, 29, 30, and 31 are all currently being tracked, with the corresponding signal strength bars indicating the relative strength of the signals. Satellites 3 and 9 (shown with highlighted numbers) are visible but are not being tracked. The Dilution of Precision (DOP) figure is shown as 2 giving an estimated position error (EPE) of 49 feet.

The outer circle of the satellite sky view represents the horizon (north-up), the inner circle 45° above the horizon, and the centre point at a position directly overhead.

The GPSMAP Map page (see Figure 5.32), the primary navigation page, provides a comprehensive display of electronic cartography, plotting and navigational data. The Map page is divided into three main sectors: chart display, data window and softkey menu.


Simulating Navigation
DOP: 2.0
EPE: 49ft

17-FEB-98 10:25:30
N 36°02.036'
W114°45.278'

**Figure 5.31** The satellite status display of the Garmin GPSMAP 225 system.

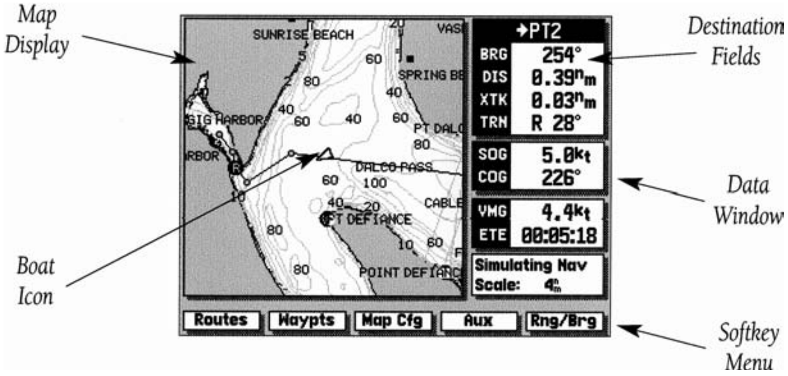

→PT2
BRG 254°
DIS 0.39nm
XTK 0.03nm
TRN R 28°
SOG 5.0kt
COG 226°
VMG 4.4kt
ETE 00:05:18
Simulating Nav
Scale: 4nm

**Figure 5.32** The MAP page, the main navigation display of the Garmin GPSMAP 225 system showing own vessel and track.

The chart display shows the user's vessel on an electronically generated chart, complete with geographical names, navaids, depth contours and a host of other chart features. A wedge icon represents the vessel's position, with its track plot shown as a solid yellow line. Routes and waypoints that have been created are also displayed. An on-screen cursor permits panning and scrolling to other map areas showing distance and bearing to a selected positions and waypoints as required. The GPSMAP system, using Garmin G-chart$^{TM}$ data cartridges, has a worldwide database to 64 nautical miles and a global coverage as shown in Figure 5.33.

The Map page also displays a wealth of navigation data in digital form. The destination fields show the bearing (BRG), in this case 254°, and the distance (DIS) 0.39 nautical miles to a destination waypoint or to the cursor. Cross-track error (XTE, 0.03 nautical miles) and turn (TRN, R 28°) information for an active destination is also displayed. The XTE value is the distance the vessel is off a desired course (left or right), whilst TRN represents the direction (left or right) in degrees between the vessel's course-over-ground (COG) and the bearing to the destination. The present speed-over-ground (SOG) is 5.0 knots and course-over-ground (COG) is 226°. This information and the terms used are illustrated in Figure 5.32.

Below this is the arrival and status field. The velocity-made-good (VMG), in this case 4.4 knots, is the speed of the vessel on a destination along a desired track, and the estimated time en route (ETE),

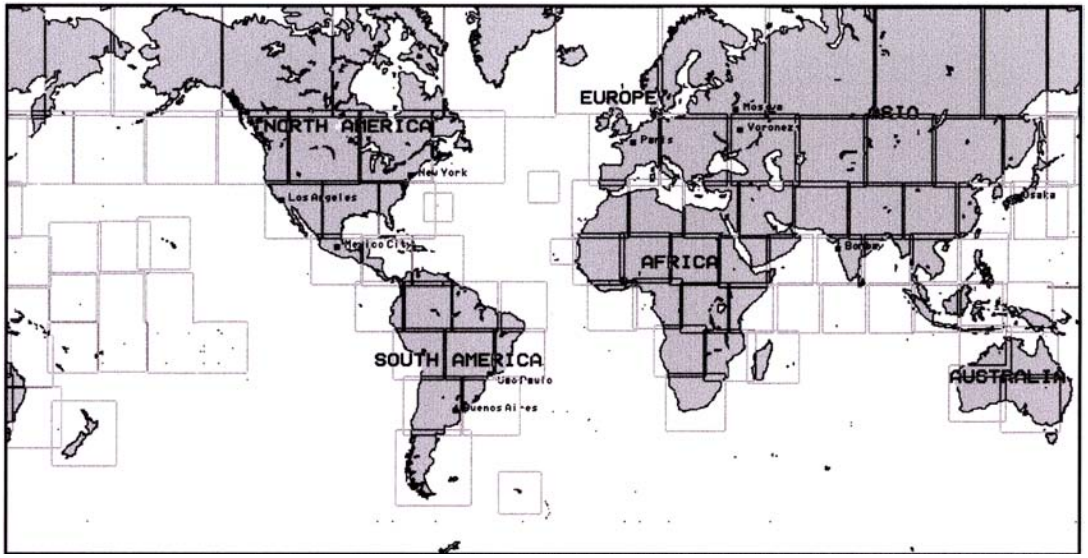
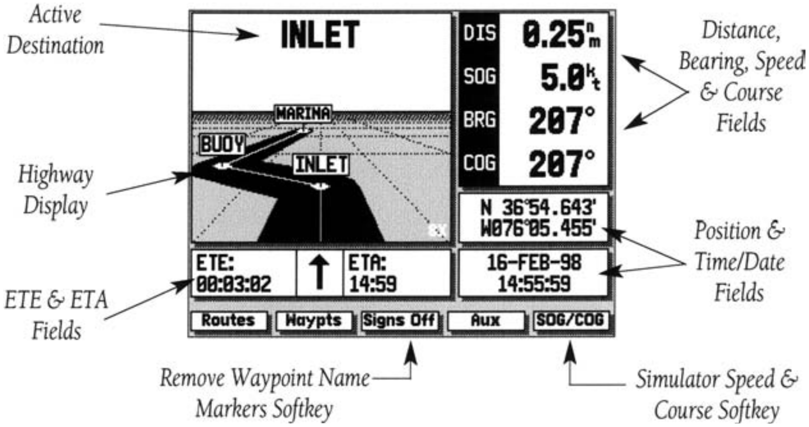


**Figure 5.33** Global coverage chart showing the Garmin GPSMAP's built-in database for chart coverage down to 64 nautical miles.

00:05:18, is the estimated time remaining on the voyage leg. The status field indicates the operating mode, in this case simulating navigation, and the scale shows the map display depth, 4 nautical miles.

The GPSMAP's built-in worldwide database includes chart coverage down to 64 nautical miles (120 km) for the areas shown in Figure 5.33.

Switching to the GPSMAP Highway page (see Figure 5.34) provides a large character display of navigation data and graphic steering guidance to an active waypoint via a planned highway. The active destination point is displayed at the top of the screen with the ETE and ETA based on the present speed and course shown at the bottom.


INLET

DIS 0.25 nm
SOG 5.0 kt
BRG 207°
COG 207°

N 36°54.643'
W076°05.455'

ETE: 00:03:02
ETA: 14:59

16-FEB-98
14:55:59

**Figure 5.34** A sample Highway page of the Garmin GPSMAP 225 system used when navigating a route to an active waypoint.

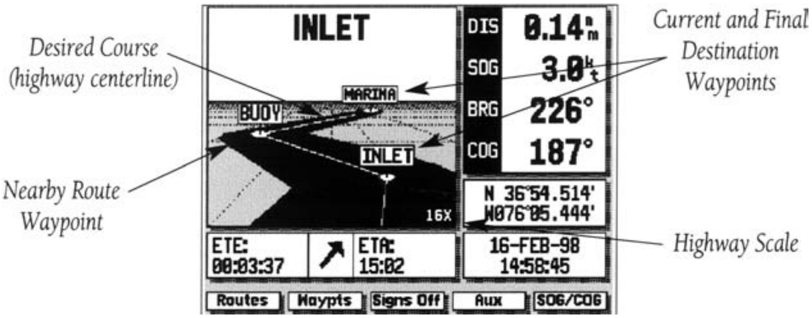


**Figure 5.35** A sample Highway page of the Garmin GPSMAP 225 system used when navigating a route to an active waypoint.

The distance and bearing to the destination waypoint, along the present SOG and COG, are shown along the right-hand side. The SOG and COG fields may be changed to display the velocity-made-good and the turn value (VMG and TRN). The position field shows the present GPS position and the date/time field displays the current date and time as calculated from GPS satellites.

The Highway page's graphic display occupies the majority of the screen (see Figure 5.35). It provides visual guidance to the destination waypoint and keeps the vessel on the intended course line. The vessel's course is represented by a centre line down the middle of the graphic highway. As the vessel progresses towards its destination, the highway perspective changes to indicate progress and which direction should be steered to remain on course. When navigating a route, the highway display shows each route waypoint in sequence. Nearby waypoints not in the steered route also will be displayed.

This brief description demonstrates that GPS receivers have moved away from the simple positional display in latitude and longitude. In future the use of more powerful computers and further integration will no doubt see GPS as merely a small but valuable input to a huge electronic charting system (for further details see Chapter 7).

### *Interface details*

The following interface formats are supported by the GPSMAP system for connection to up to three NMEA devices.

NMEA 0180
NMEA 0182
NMEA 0183 version 1.5

Approved sentences-
GPBWC, GPGLL, GPRMB, GPRMC, GPXTE, GPVTG, and GPWPL

Proprietary sentences-
PGRMM (map datum), PGRMZ (altitude) and PSLIB (beacon receiver control input)

NMEA 0183 version 2.0

Approved sentences-
GPGGA, GPGSA, GPGSV, GPRMB, GPRMC, GPRTE and GPWPL

Proprietary sentences-
PGRME (estimated error), PGRMM (map datum), PGRMZ (altitude) and PSLIB (beacon receiver control input)

For further information and explanation about the NMEA format see Appendix 3. For further information about Garmin GPS products see www.garmin.com

## 5.13 GPS on the web

GPS enjoys massive coverage on the world wide web and there are simply far too many sites to list here. However, some of the better sites are worth a visit and are listed below.

http://www.navcen.uscg.mil
An essential site for all navigators. United States Coast Guard site with numerous pages of data on GPS, Loran-C and US coastal navigation notices.

ftp://tycho.usno.navy.mil/pub/gps
Massive amounts of detail about GPS time transfer, current constellation status and health.

http//:www.spatial.maine.edu/~leick/alpha.htm
GPS and GLONASS alphabetical index link site to dozens of other relevant sites.

http://www.apparent-wind.com/gps.html
Another index site with useful links to other GPS and maritime sites.

http://www.trimble.com
GPS tutorials, fact sheets, satellite plots etc. from one of the biggest GPS equipment manufacturers. One of the best sites on the net.

http://www.trinityhouse.co.uk/dgps.htm
Details of the differential GPS beacons, parameters and availability around the UK coast.

http://www.utexas.edu/depts/grg/gcraft/notes/gps/gps.html
Extensive high-tech education notes on the GPS system from the University of Texas. Intended for use by university students.

http://www.igeb.gov
Interagency GPS Executive Board site. Includes the latest news about the GPS.

http://www.ngs.noaa.gov
National Oceanic and Atmospheric Administration (NOAA) and the National Geodetic Survey Site. Lots of detailed statistics about GPS health and status.

http://www.notams.faa.gov
The FAA's site holding Notices to Airmen (NOTAMs) listed interruptions in the GPS service. Coastal area NOTAMs are of use to mariners.

http://www.garmin.com
A huge informative site belonging to a major manufacturer of GPS equipment, holding a wealth of information about a huge range of equipment.

GPS continues to be updated and improved. It has been announced that two new civilian signals designed to carry data to enhance the civilian and commercial service will be added to the GPS. Furthermore, 18 additional satellites are to be used to support the system.

## 5.14 Global Orbiting Navigation Satellite System (GLONASS)

The Russian Federation's GLONASS was developed in parallel with GPS to serve the same primary function, that is, as a weapons navigation and guidance system. And like GPS, GLONASS has been released for international position fixing use, albeit in a downgraded form.

GLONASS is owned and operated by a Military Special Forces team at the Russian Ministry of Defence. SV time synchronization, frequency standards and receiver technology development are controlled from The Russian Institute of Navigation and Time in St. Petersburg. The system possesses similar architecture to the GPS and is equally capable of highly accurate position fixing.

### 5.14.1 Space segment

Work on the system began in the early 1970s and the first satellites were launched into orbit in 1982. Since then a full constellation has been established and GLONASS became fully operational in early 1996.

The space segment is based on 24 SVs, eight in each of three, almost circular orbital planes spaced at 120° intervals and inclined at 64.8° and at an altitude of 25 440 km. Each SV completes one earth orbit in 11 h 25 min and of course two orbits in 22 h 50 min in real time. Taking into account the length of a sidereal day, the westerly shift of each orbit brings all SVs back to an earth epoch point every 8 days, and the entire cycle repeats naturally.

All GLONASS SVs transmit on two frequencies to allow for correction of ionospheric signal delay, but unlike the GPS system, each SV uses different frequencies. Phase modulated onto the two carrier frequencies are a Coarse/Acquisition (C/A), a Precise code (P) and navigation data frames.

### 5.14.2 Ground segment

All ground control stations are located in former Soviet Union territory. The Ground Control and Operations Centre and Time Standard Centre are in Moscow. SV telemetry and tracking stations are located in Eniseisk, Komsomolsk-na-Amure, St. Petersburg and Ternopol.

### 5.14.3 Signal parameters

Initially all SVs were designed to transmit on different carrier frequencies, but in 1992, following the World Administrative Radio Conference (WARC-92) frequencies were grouped. Then in 1998 they were again changed. Currently, the $L_1$ transmission frequency band is 1598.0625–1609.3125 MHz and the $L_2$ band 7/9ths below this between 1242.9375 and 1251.6875 MHz (see Table 5.9).

Both $L_1$ and $L_2$ carriers are BPSK-modulated at 50 bauds with the navigation message. $L_1$ also carriers a PRN Coarse/Acquisition (C/A) code and $L_2$ both a Precision (P) code and the C/A code. The P code has a clock rate of 5.11 MHz and the C/A code is 0.511 MHz.

As in the GPS, the GLONASS navigation message contains timing, SV position and tracking data. All SVs transmit the same message (see Table 5.10).

### 5.14.4 Position fixing

GLONASS navigation fixes are obtained in precisely the same way as those for GPS. Pseudo-range calculations are made and then corrected in the receiver to obtain the user location in three dimensions. Precise timing is also available.

**Table 5.9** SV carrier frequency designation

| *Channel no.* | *L1 carrier (MHz)* | *L2 carrier (MHz)* |
|---|---|---|
| –7 | 1598.0625 | 1242.9375 |
| –6 | 1598.6250 | 1243.3750 |
| –5 | 1599.1875 | 1243.8125 |
| –4 | 1599.7500 | 1244.2500 |
| ↓ | | |
| +13 | 1609.3125 | 1251.6875 |

Expression for channel increment:

L1 = 1598.0625 + 0.5625 MHz
L2 = 1242.9375 + 0.4375 MHz

Note: The ratio of L2/L1 channels is 7/9.

**Table 5.10** GPS – GLONASS system comparison

| *Parameter* | *GPS* | *GLONASS* |
|---|---|---|
| Orbital | | |
| Altitude: | 20 180 km | 19 130 km |
| Period: | 11 h 58 min | 11 h 15 min 40 s |
| Inclination: | 55° | 64.8° |
| Planes: | 6 | 3 |
| Number of SVs | 24 | 24 |
| Carrier frequency | | |
| L1: | 1575.420 MHz | 1598.6250–1609.3125 MHz |
| L2: | 1227.600 MHz | 1242.9375–1251.6875 MHz |
| Code clock rate | | |
| C/A: | 1.023 Mbit $s^{-1}$ | 0.511 Mbit $s^{-1}$ |
| P: | 10.23 Mbit $s^{-1}$ | 5.11 Mbit $s^{-1}$ |
| Time reference | UTC | UTC |
| Navigation message | | |
| Rate: | 50 bit $s^{-1}$ (baud) | 50 bit $s^{-1}$ (baud) |
| Modulation: | BPSK NRZ | BPSK Manchester |
| Frame duration: | 12 min 30 s | 2 min 30 s |
| Subframe: | 6 s | 30 s |
| Almanac content | Timing and orbital parameters | Timing and orbital parameters |

### 5.14.5 User equipment

Because of the initial secrecy surrounding the system and the scarcity of detailed parameters, it is to be expected that there is little user equipment available. In the past, western manufacturers have had little incentive to invest heavily in the development of receivers when the GPS has been freely available. However, this situation could well change in the future.

## 5.15 Project Galileo

At the time of writing, the European Commission has produced a working paper for a European-based Global Navigation Satellite Service (GNSS) called the Galileo. It is to be designed to be totally independent of both GPS and GLONASS and thus will end the reliance of countries within the European Commission on systems beyond their control. It remains to be seen if the finance and indeed the impetus to create the system will be forthcoming.

## 5.16 Glossary

| | |
|---|---|
| **Almanac data** | Satellite constellation information including location and health status. |
| **Apogee** | The furthest point away from the earth reached by a satellite in orbit. |
| **Azimuth** | The direction vector drawn to a satellite from a fixed point on earth. |
| **BNM** | USCG Broadcast Notice to Navigators. |
| **BPSK** | Bi-phase shift keying. |
| **BRG** | Bearing. |
| **C/A code** | Coarse/Acquire code. A PRN code operating at 1.023 Mbit $s^{-1}$ |
| **CEP** | Circular area probable. An accuracy figure achievable for 50% of the time in two dimensions; latitude and longitude. |
| **COG** | Course over ground. |
| **CSOC** | Consolidated Space Operations Centre. |
| **dB** | A unit for measuring power in a communications system. |
| **DTK** | Desired track. The compass course between the start and finish waypoints. |
| **DGPS** | Differential GPS. A method to improve the accuracy of a GPS fix by the use of corrective data transmitted on medium frequency to coastal shipping. |
| **DMA** | US Defence Mapping Agency. |
| **DoD** | US Department of Defence. |
| **DOP** | Dilution of Precision. A term used for expressing the mathematical quality of a solution. |
| $d_{\mathrm{RMS}}$ | A circle around the true position containing 95% of the fix calculations. |
| **ECEF** | Earth-centred-earth-fix. A GPS fix solution is quoted in ECEF co-ordinates. |
| **EPE** | Estimated position error. |
| **ETA** | Estimated time of arrival. |
| **ETE** | Estimated time en route. The time remaining to a destination. |
| **FAA** | US Federal Aviation Authority. |
| **GDOP** | Geometric dilution of precision. A measure of the quality of a solution. |
| **GLONASS** | Global Orbiting Navigation Satellite System. The Russian Federation system. |
| **GMT** | Greenwich mean time. Often referred to as Zulu. |
| **GNSS** | Global Navigation Satellite System. |
| **GPS** | Global Positioning System. |
| **Ground speed** | The vessel's velocity referenced to the ocean floor. |
| **HDOP** | Horizontal dilution of position. A measure of the quality of a solution in terms of latitude/longitude. |
| **Inclination** | The angle formed between the eastern end of the equatorial plane and a satellite orbit. For GPS orbits it is 55°. |
| **Kepler's laws** | Satellites in orbit follow an ellipse as defined by Johannes Kepler. |
| $L_1$ | The GPS primary transmission frequency; 1575.42 MHz. |

| | |
|---|---|
| **$L_2$** | The GPS secondary transmission frequency; 1227.6 MHz. |
| **MCS** | The GPS Master Control Station situated at Colorado Springs. |
| **NMEA** | National Maritime Electronics Association. An organization of manufacturers and distributors responsible for agreeing the standards of interfacing between various electronic shipboard systems. |
| **NOTAM** | FAA's Notice to Airmen regarding GPS service interruption. |
| **P code** | Precision code. A PRN code operating at 10.23 MHz. |
| **PDOP** | Precision dilution of position. |
| **Perigee** | The closest point of approach to the earth reached by an orbiting satellite. |
| **PPS** | GPS Precise Positioning Service. |
| **PRN** | Pseudo-random noise. |
| **RTCM** | Radio Technical Commission for Maritime Services. |
| **SEP** | An accuracy that is achievable 50% of the time in all dimensions. |
| **SOG** | Speed over ground. |
| **SPS** | GPS Standard Positioning Service. |
| **SV** | Space vehicle – a satellite. |
| **TDOP** | Time dilution of precision. |
| **TTFF** | Time to first fix. Used to identify how long a GPS receiver takes before a fix is available. |
| **TTSF** | Time to subsequent fix. |
| **TRN** | Turn. |
| **URE** | User range error. |
| **UERE** | User equivalent range error. Determined by summing the squares of the individual range errors and then taking the square root of the total. |
| **USCG** | United States Coast Guard. |
| **USNS** | United States NOTAM Service. |
| **ULS** | Satellite uplink station. |
| **UTC** | Universal time co-ordinated. |
| **UTM** | Universal transverse mercator. A grid co-ordinate system that projects global sections onto a flat surface. |
| **VDOP** | Vertical dilution of precision. |
| **VMG** | Velocity made good. |
| **WADGPS** | Wide area differential GPS. An experimental system for improving the accuracy of GPS fixes globally. |
| **WGS-84** | World Geodetic Survey 1984. |
| **XDOP** | Cross-track dilution of precision. |
| **XTE** | Cross-track error. |

## 5.17 Summary

- The GPS has replaced the Navy Navigation Satellite System (NNSS).
- Satellites, called space vehicles (SVs), follow elliptical orbits conforming to Kepler's laws of astrophysics.
- The GPS, occasionally called NAVSTAR, has three segments: Space, Control and User.
- There are 24 operational SVs, four in each of six orbital planes inclined at 55°.
- SVs orbit the earth at an altitude of 20 200 km and possess an approximate 12-h orbital period.

- SVs transmit two codes to enable receivers to acquire the signal. The Coarse and Acquire (C/A) code is a pseudo-random noise (PRN) code stream operating at 1.023 Mbit $s^{-1}$. The precise (P) code is also a PRN stream operating at the faster rate of 10.23 MHz.
- The C/A code epochs every 1 ms and has been designed to be easily acquired while the P code has an epoch every 267 days and is difficult to acquire.
- Navigation data is transmitted at 50 bit $s^{-1}$ and is modulated onto both codes.
- The $L_1$ signal carrier frequency (1575.42 MHz) is modulated with the C/A code, the P code and the navigation message, whilst the $L_2$ carrier carries only the P code and the navigation message.
- There are two levels of fix available. The Precise Positioning Service (PPS) and the Standard Positioning Service (SPS). Until May 2000 the SPS was downgraded by a factor of 10, but on that date downgrading, called Selective Availability, was removed and now SPS and the PPS fixes have virtually the same accuracy.
- GPS position fixes are achieved by the precise measurement of the distance between a number of SVs and a receiver at an instant in time and/or by phase measurement. It is assumed that the receiver clock is in error and therefore the range measured is called a pseudo-range (false range). The receiver processor corrects the range measurement to produce a precise fix.
- The fix, in XYZ co-ordinates (latitude/longitude and altitude) is converted to earth-centred co-ordinates called ECEF (earth-centred-earth-fix).
- Dilution of precision (DOP) is the term used for expressing the mathematical quality of a fix solution. TDOP, HDOP, VDOP, and PDOP are also used in the GPS.
- System errors may cause an imprecise fix. Fix error and thus GPS accuracy is quoted using one of the figures CEP, SEP, $d_{\mathrm{RMS}}$ and UERE.
- Differential GPS (DGPS) is a system whereby SV signals are received at a fixed location, errors are corrected and the new data is transmitted on MF to vessels in the local area.
- GPS uses an active antenna with a ground plane to reduced the effect of reflected signals.
- There is a huge range of GPS equipment available ranging from simple hand-held units to sophisticated dual-channel systems used for survey purposes.
- The Russian Federation's satellite navigation system, GLONASS, is operational but is not compatible with GPS.

## 5.18 Revision questions

1 What are the basic principles of Kepler's laws of astrophysics?
2 How are the orbital period and the velocity of a space vehicle (SV) related?
3 How many SVs are used in a full GPS constellation and how many are there in each orbital plane?
4 What are the GPS transmission frequencies?
5 Why do Navstar SVs transmit on two frequencies?
6 How long does it take an SV to transmit an entire navigation data message of 25 frames?
7 The GPS uses two codes, the P code and the C/A code, for encryption purposes. Why is this?
8 Why is the P code more difficult for a receiver to lock onto than the C/A code?
9 Why is it essential to maintain SV transmit frequency stability?
10 PPS fixes require the use of more complex receiving equipment. Why is this?
11 What is a pseudo-range measurement?
12 How does the choice of SVs used for a fix affect the PDOP?
13 What is ECEF XYZ?
14 Which of the error-inducing factors is likely to introduce the largest error?

15 How is the figure for UERE derived?

16 The use of DGPS offers improved fix accuracy. Over what range would you expect to receive DGPS data?

17 Why does a GPS antenna need a ground plane?

18 Why is the C/A code generated (or held in memory) in a receiver and applied to the correlator?

19 Autocorrelation is used in the signal processing stages of a GPS receiver. Why is this?

20 The Russian Federation satellite navigation system, GLONASS, offers similar features and accuracy of position fixing to the GPS. Are the two systems compatible?

# Chapter 6

# Integrated bridge systems

## 6.1 Introduction

The 20th century saw many milestones in terms of nautical events and much was learnt from such events for the benefit of those seafarers that came afterwards. Starting with events such as the sinking of the Titanic in 1912 with its impact on the Safety of Life at Sea, the use of wireless telegraphy and, continuing throughout the century, the increasing use of electronics and satellites for navigation and communication purposes.

During that time there was a realization for the need to set up international bodies with a view to the harmonization, and the international recognition, of standards for ships involved in international trading. Bodies set up during the 20th century to monitor and influence these trends included the following.

### 6.1.1 International Maritime Organization (IMO)

Originally set up as the Inter-Governmental Maritime Consultative Committee (IMCO) in 1958, the name was changed in 1982. Its first task was to adopt a new version of the International Convention for the Safety of Life at Sea (SOLAS) and this was completed in 1960. The best known of the responsibilities of the IMO is the adoption of maritime legislation. About 40 conventions and protocols have been adopted by the organization and amended as necessary to keep pace with the changes in world shipping. The IMO has 158 member states and is based in London, England.

### 6.1.2 The International Standards Organization (ISO)

This is a non-governmental organization established in 1947 with a view to promoting the development of standardization in the world, facilitating the international exchange of goods and services, and developing co-operation in the areas of intellectual, scientific, technological and economic activity. The work of the organization results in international agreements, which are published as International Standards. There are more than 130 countries represented within the organization which is based in Geneva, Switzerland.

### 6.1.3 The International Electrotechnical Commission (IEC)

Established in 1906, the organization has more than 50 member countries covering 85% of the world's population. Standards established are used in more than 100 countries and there are approximately 200 Technical Committees (TCs) of which TC80 is an important part (see Section 6.3). The IEC collaborates with the ISO in matters of mutual interest and both organizations co-operate on a joint basis with the International Telecommunications Union (ITU). Like the ISO, the IEC is a non-governmental body while the ITU is part of the United Nations organization with governments as its members. The IEC is based in Geneva, Switzerland.

## 6.2 Design criteria

In the 1960s Planned Ships Bridges were available from at least one manufacturer and fitted on some vessels. This was probably the first attempt to construct a bridge within design concepts that took into consideration the operational requirements of the vessel. Integrated navigation systems and integrated bridge systems have evolved from those days and the concept is now accepted, with a variety of systems available from many different manufacturers.

Certain classification societies have initiated terms of carriage requirements if particular notations are specified for a vessel. A leading influence has been Det Norske Veritas (DNV) of Norway, a member of the International Association of Classification Societies (IACS). The Association was formed in 1968 and claims that ‘At the heart of ship safety, classification embodies the technical rules, regulations, standards, guidelines and associated surveys and inspections covering the design, construction and through-life compliance of a ship’s structure and essential engineering and electrical systems.’ More than 90% of the world’s merchant tonnage is classed by the 10 members and two associates of IACS. IACS members include the American Bureau of Shipping (ABS), Germanischer Lloyd and Lloyds Register of Shipping, together with societies from China, France, Italy, Japan, Korea, Norway and Russia. IACS has held consultative status with the IMO since 1969 and is the only non-government organization with observer status able to develop rules. The DNV rules for ships are as follows.

- To reduce the risk of failure in bridge operation, causing collisions, groundings and heavy weather damages and to minimize the consequences to ship and complement, should an accident occur.
- To include relevant requirements and recommendations from the IMO.
- To include relevant international standards within the rules or indicating the points in which they differ.

The various classification societies have adopted different standards although discussions on establishing international performance standards for integrated bridge systems have progressed under the direction of the IEC’s Technical Committee 80 (TC80). Progress has been made on type approval and system notation.

The integrated bridge system should be designed and installed as a physical combination of equipment or systems using interconnected controls and displays. Workstations should provide centralized access to all nautical information. The type of operational function carried out from the bridge would include navigation, communications, automation and general ship operation. Manufacturers can provide shipbuilders and potential shipowners with computer-generated drawings of how a particular bridge layout would look when installed. One such diagram produced by Litton Marine Systems is shown in Figure 6.1.

In the absence of any internationally-agreed operating standards, from either the IMO or national authorities, reliance must be placed on industry guidelines and standards which do exist for bridge layout and equipment. These include the ISO standard for ‘bridge layout and associated equipment.’

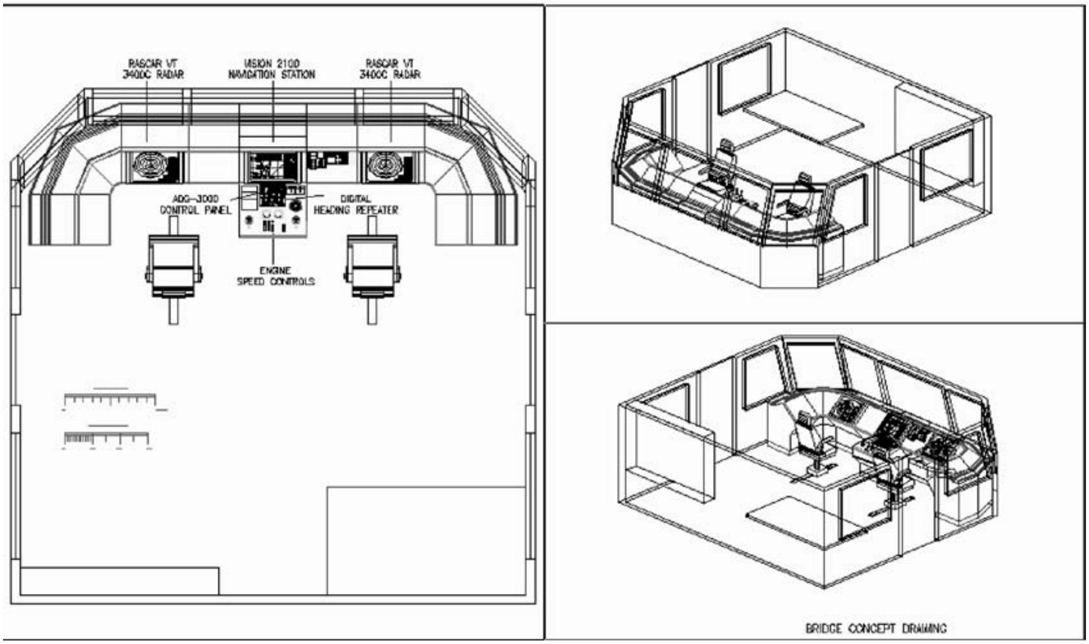


**Figure 6.1** Line drawing of an integrated bridge system. (Reproduced courtesy of Litton Marine Systems.)

An IEC definition of an integrated bridge system states that such a system must be capable of carrying out at least two of the following functions:

- navigation planning
- passage execution and manoeuvring
- collision and stranding avoidance
- communications
- machinery control and monitoring
- loading and discharge of cargo
- safety and security
- management.

The integrated bridge system that meets these requirements must provide: redundancy in the event of system failure; the use of standardized equipment interfacing; the centralization of all nautical data and alarms; and the use of suitable displays to allow the monitoring of sensor data. The fact that current trends involve a reduction in manning levels suggests that the few members of a crew on the bridge must be capable of interpreting and responding to the multitude of information and alarms being presented to them. This would involve improvements in training and system documentation for the crews.

The DNV rules specify design criteria for particular workstations namely:

- traffic surveillance/manoeuvring
- navigation

- route planning
- manual steering
- safety operations
- docking operations
- conning operations.

In each case the tasks that have to be performed are specified and the siting of relevant instruments/equipment required for those tasks is defined. As an example, the workstation for navigation is specified to enable the following tasks to be performed:

- determine and plot the ship's position, course, track and speed;
- effect internal and external communications related to navigation;
- monitor time, course, speed and track, propeller revolutions, pitch indicator and rudder angle.

The following instruments and equipment should be installed within reach:

- navigation radar display and controls
- chart table
- relevant position fixing systems (GPS and Loran-C)
- VHF unit
- whistle control.

Instruments, indicators and displays providing information considered essential for the safe and efficient performance of tasks at the navigation workstation should be easily readable from the workstation. These instruments, indicators and displays should include:

- gyro repeater
- rudder angle indicator
- depth indicator
- clock
- propeller RPM indicator
- pitch indicator (where fitted)
- speed and distance indicator.

Means to be used at intervals for securing safe course and speed in relation to other ships and safety of bridge operation should also be easily accessible from the navigation workstation. Such means include:

- instruments and equipment installed at the workstation for traffic surveillance/manoeuvring
- internal communications equipment
- central navigation alarm panel (if provided)
- wipers and wash controls for the windows within the required field of vision.

DNV specification for one-man bridge systems in an unbounded voyage area, known as DNV-W1, requires an Automatic Navigation and Track-keeping System known as ANTS. The specification requires integration of the following:

- Electronic Chart Display and Information System (ECDIS)
- automatic steering system (including software for calculation/execution of adjustments for the maintenance of pre-planned routes)

- differential GPS (2)
- gyrocompass (2)
- speed over ground (SOG) and speed through water (STW)
- course alteration warnings and acknowledgement
- automatic safety contour checking and alarming during voyage planning and execution
- capacity to create own electronic charts from paper charts for areas not covered by ENCs issued or certified by official sea chart authorities.

In addition to the above functional requirements, ANTS also places great emphasis on suitable technical documentation.

The requirements for ANTS place additional demands on certain aspects of the system. For example, the accuracy of the ship's heading should be a value that has been corrected for any errors typical of the source of the heading input, and at least one of the gyrocompasses should be provided with an automatic system for the correction of errors caused by speed and latitude. The steering system should also keep automatic track-keeping of the ship within the limits set on both sides of the pre-planned track and should provide the capability to steer the ship along a route consisting of straight and curved lines by both automatic and manual input of turn orders. The speed input should have sufficient accuracy to safeguard the quality of position fixing by dead reckoning. The system should be provided with a filtered position from the GPS receiver and when performing turns, the system should be provided with the most accurate real-time position. The quality of the integrated position fixing system should be monitored and a warning should appear if the quality is below an acceptable limit.

The need for integration has meant that there has been a tendency to move away from sourcing equipment from a variety of manufacturers and attempts to integrate disparate pieces of equipment, to single-sourcing a package of equipment from just one manufacturer. Many manufacturers, aware of the requirement, now offer complete systems with all the necessary interfacing requirements guaranteed. The use of standard modules and interfaces, not only for navigation but also for other bridge functions, such as communications, engine monitoring and control, power supply etc., is likely to produce cost savings and reduce the amount of equipment required. Factors such as the reduced number of consoles, reduced installation and interfacing costs, more cost-effective design, installation and testing requirements have to be taken into account.

## 6.3 Standards

Those organizations involved in the production of world standards are the International Standards Organization (ISO), the International Electrotechnical Commission (IEC), and the International Telecommunications Union (ITU). The first two organizations work closely together and, as they both have their headquarters in Geneva, some facilities have been amalgamated.

The International Maritime Organization (IMO) is responsible for defining the requirements for marine equipment but it does not provide sufficient specification detail for manufacturers to design specified equipment or for national maritime authorities to provide test and approval facilities for the equipment. Thus, the IEC and ISO standards are designed to allow the necessary specification requirements for design, testing and approval.

The IEC has several Technical Committees working in specialized technical areas. The IEC Technical Committee 80 (IEC TC80) covers the area of 'Marine Navigation and Radio communication Equipment and Systems' and was formed in 1980. IEC TC80 responsibility is to concern itself with the development of international technical standards for the navigation and radio communication equipment designated by the IMO for mandatory carriage on vessels covered by the SOLAS (Safety of Life at Sea) Conventions.

IEC TC80 currently has 10 working groups:

| | |
|---|---|
| WG1 | radar and ARPA |
| WG1A | Track control |
| WG4 | Terrestrial position-fixing aids |
| WG4A | Global Navigation Satellite Systems (GNSS) |
| WG5 | General requirements |
| WG6 | Digital interfaces |
| WG8 | Global Maritime Distress and Safety System (GMDSS) |
| WG8A | Automatic shipborne Identification Systems (AIS) |
| WG10 | Integrated navigation systems |
| WG11 | Voyage data recorders. |

Until fairly recently there were two other TC80 working groups: WG7 Electronic chart display and information system (ECDIS) and WG9 Integrated bridge systems for ships. The latter group was responsible for the publication in April 1999 of IEC 61209 ‘Maritime navigation and radio communication equipment and systems – Integrated bridge systems (IBS) – operational and performance requirements, methods of testing and required test results’. This document covers features such as: data exchange, displayed information, system configuration, human factors, alarms, training facilities, power supplies and failure analysis. This latter point is doubly important as it has implications in other areas such as training facilities.

## 6.4 Nautical safety

All aspects of bridge operation have evolved because of the requirement for the safety of the ship, crew and, where applicable, the passengers. The safety philosophy is encapsulated within the rules of Det Norske Veritas (DNV) and the following is reproduced from the DNV rules, part 6, chapter 8 with their kind permission.

### 6.4.1 Safety philosophy

To achieve optimum safety and efficiency in bridge operation the rules address the total bridge system, which is considered to consist of four essential parts.

- The technical system which deduces and presents information as well as enabling the proper setting of course and speed.
- The human operator who is to evaluate available information, decide on the actions to be taken and execute the decisions.
- The man/machine interface which safeguards that the technical system is designed with due regard to human abilities.
- The procedures which shall ensure that the total bridge system performs satisfactorily under different operating conditions.

### 6.4.2 Scope of rule requirements

These are set out in each section of the Rules for Nautical Safety and reflect the different factors that affect the performance of the total bridge system and are intended to regulate the following areas.

- Design of workplace, based on the analysis of functions to be performed under various operating conditions and the technical aids to be installed.
- Bridge working environment, based on factors affecting the performance of human operators.
- Range of instrumentation, based on information needs and efficient performance of navigational tasks.
- Equipment reliability applicable to all types of bridge equipment, based on common requirements to ensure their suitability under various environmental conditions.
- Specific requirements to different types of bridge equipment, based on the facilities required for the performance of their specific functions.
- Man/machine interface, based on the analysis of human limitations and compliance with ergonomic principles.
- Qualifications, based on the competence required for mastering rational navigational methods and relevant technical systems installed on board the ship.
- Operating procedures, based on the work organization needed to make the bridge system function under different operational situations.
- Information on the ship's manoeuvring characteristics, based on the manoeuvres commonly used in various operational situations.
- Tests and trials for new ships, based on the need to ensure that technical systems perform in accordance with their specifications before being relied upon and used in practical operation.
- Reporting system, from ships in service, on bridge instrument failures, based on the information needed to detect their factual reliability level.
- Survey schemes for ships in service, based on the follow-up and testing required to safeguard that bridge systems maintain their reliability.

## 6.5 Class notations

The Rules for Nautical Safety are divided into three class notations. Two class notations represent the minimum requirements within bridge design, instrumentation and procedures whereby NAUT-C covers bridge design and W1-OC, in addition, includes instrumentation and bridge procedures. The third class notation, W1, extends the basic requirements for bridge design and instrumentation and additionally requires information on the manoeuvring characteristics of the ship and an operational safety manual for safe watchkeeping and command of the ship.

NAUT-C covers bridge design, comprising the following main areas:

- mandatory and additional workstations
- field of vision from workstations
- location of instruments and equipment.

W1-OC covers bridge design, instrumentation and bridge procedures comprising the following main areas:

- NAUT-C
- range of instrumentation
- instrument and system performance, functionality and reliability
- equipment installation
- monitoring and alarm transfer system
- procedures for single-man watchkeeping.

W1 covers W1-OC and extensions within the following areas of W1-OC:

- design of one-man workstation
- field of vision astern
- range of instrumentation
- instrument performance
- automation level
- qualifications.

Also covered is information on the manoeuvring characteristics of the ship comprising the following main items:

- speed at different settings
- steering ability
- turning ability
- stopping ability.

There is also a requirement for an operational safety manual comprising the following main items:

- bridge organization and responsibilities
- watchkeeping procedures
- system fall-back procedures
- accident and emergency procedures.

## 6.6 Bridge working environment

Ships requesting class notation NAUT-C, W1-OC or W1 should comply with rules for bridge working environment which specifies vibration levels, noise, lighting, temperature, ventilation, surfaces, colours and the safety of personnel.

### 6.6.1 Equipment carriage requirements

Ships requesting class notation W1-OC are equipped with the following systems:

- course information systems (two gyrocompasses or one gyro + one TMC)
- steering systems (manual and automatic steering)
- speed measuring system (water speed, > 40 000 tons gross, dual axis)
- depth measuring system (over 250 m length, two transducers)
- radar systems (two radars, at least one X-band)
- traffic surveillance systems (ARPA)
- position fixing systems (Loran-C, GPS)
- watch monitoring and alarm transfer system
- internal communication systems
- nautical safety radio communication systems
- sound reception system (technical device to receive signals).

Additional equipment required for class notation W1 includes:

- steering system with rate of turn indicator
- course information system, which should have two independent gyrocompasses
- speed measuring system, through the water, which should provide information for traffic surveillance system
- Electronic Chart Display and Information System (ECDIS)
- Automatic Navigation and Track-keeping System (ANTS)
- conning information display
- central alarm panel
- wind measuring system.

### 6.6.2 General bridge equipment requirements

The rules specify the following:

- environmental conditions
- location and installation of equipment
- electrical power supply, alarms, performance confirmation and failure protection
- computer-based systems and software quality.

### 6.6.3 Specific requirements for different types of bridge equipment

Ships requesting class notation W1-OC shall comply with specific requirements for the following systems:

- course information system (speed and latitude correction)
- steering systems (manual override control and rate of turn display)
- speed measuring system (if bottom track then up to 200 m depth)
- depth measuring system
- radar systems (two floating EBLs, interswitch, ship track monitoring)
- traffic surveillance systems (ARPA with two guard zones)
- position fixing systems (performance standards)
- watch monitoring and alarm transfer system
- internal communication systems
- nautical communication systems
- sound reception system.

Class notation W1 requires in addition the following systems:

- Electronic Chart Display and Information System (ECDIS)
- Automatic Navigation and Track-keeping System (ANTS)
- conning information display
- central alarm panel.

### 6.6.4 Man/machine interface

Ships requesting class notation W1-OC or W1 must comply with the rules in this section. All instruments must be logically grouped according to their functions within each workstation. Their location and design should give consideration to the physical capabilities of the human operator and comply with accepted ergonomic principles. The amount of information to be presented for conducting the various tasks, as well as the methods of displaying the information needed, should give consideration to the capabilities of the human operator to understand and process the information made available. The rules specify the following:

- instrument location and design
- illumination and individual lighting of instruments
- requirements for the man/machine dialogue of computer-based systems.

## 6.7 Ship manoeuvring information

Ships requesting class notation W1 must comply with rules for manoeuvring information.
Information about the ship's manoeuvring characteristics, enabling the navigator to safely carry out manoeuvring functions, shall be available on the bridge. This section deals with: the manoeuvring information to be provided, and the presentation of the manoeuvring information.

The provision of manoeuvring information should include:

- speed ability
- stopping ability
- turning ability
- course change ability
- low-speed steering abilities
- course stability
- auxiliary manoeuvring device trial
- man-overboard rescue manoeuvre.

The presentation of manoeuvring information should include:

- pilot card
- wheelhouse poster
- manoeuvring booklet.

## 6.8 Qualifications and operational procedures

Class notation W1-OC specifies responsibilities of shipowner and ship operators, qualifications and bridge procedures. Class notation W1 has extensions to responsibilities, qualifications, bridge procedures, and a special requirement for operational safety standards.

### 6.8.1 Operational safety manual

This is a requirement for class notation W1 to obey the following guidelines.

1 Organization:
  - general
  - bridge organization

- responsibilities of shipowners and ship operators
- responsibilities of the master
- responsibilities of the officer in charge of single-man watchkeeping
- qualifications of bridge personnel
- manning
- safety systems – maintenance and training.

2 Daily routines:
- general
- look-out
- changing of the watch
- periodic checks of navigational equipment
- log-books
- communications and reporting.

3 Operation and maintenance of navigational equipment:
- general
- radars/ARPA
- automatic pilot
- gyro and magnetic compasses
- echo sounder
- speed/distance recorder
- electronic position fixing aid
- electronic navigational chart
- automatic navigation and track-keeping system
- hydrographic publications
- emergency navigation light and signal equipment.

4 Departure/arrival procedures:
- general
- preparation for sea
- preparation for arrival in port
- embarkation/disembarkation of pilot
- master/pilot information exchange.

5 Navigational procedures:
- general
- helmsman/automatic pilot
- navigation with pilot embarked
- navigation in narrow waters
- navigation in coastal waters
- navigation in ocean areas
- navigation in restricted visibility
- navigation in adverse weather
- navigation in ice
- anchoring.

6 System fall-back procedures:
  - general
  - bridge control/telegraph failure
  - gyrocompass failure
  - steering failure
  - auxiliary engine failure
  - main engine failure.

### 6.8.2 Contingency and emergency manual

1 Contingency and emergency organization:
  - general
  - duties and responsibilities.

2 Accident procedures:
  - general
  - collision
  - grounding
  - fire/explosion
  - shift of cargo
  - loss of buoyancy/stability.

3 Security procedures:
  - general
  - sabotage threat/sabotage
  - hijacking threat/hijacking
  - piracy
  - local war situation
  - criminal act committed on board
  - detention/arrest.

4 Emergency procedures:
  - general
  - emergency notification
  - abandon ship preparations
  - lifeboat evacuation
  - helicopter evacuation
  - use of other evacuation equipment.

5 Miscellaneous:
  - general
  - dead or injured person aboard
  - man overboard
  - search and rescue actions
  - stowaways
  - political refugees
  - missing or lost person
  - documentation and reporting
  - press releases.

## 6.9 Bridge equipment tests

Ships requesting class notation W1-OC or W1 must comply with rules for equipment tests. After installation of equipment, on-board testing shall be performed in order to ascertain that the equipment, as installed, operates satisfactorily.

It should be noted that reliable figures for all aspects of equipment performance/accuracy cannot be established by the on-board testing required for classification. Hence, to ensure that equipment performance is in accordance with specifications, shipowners are advised to choose equipment that is type approved.

A detailed test programme for the on-board testing of equipment should be submitted for approval at the earliest possible stage before sea trials. The following systems are tested according to general requirements for testing of equipment:

- gyrocompass
- automatic steering system
- rudder indicator(s)
- rate-of-turn indicator
- speed log
- echo sounder
- radar system
- ARPA system
- electronic position fixing systems
- watch monitoring and alarm transfer system
- internal communication systems
- nautical communication system
- sound reception system
- computer system(s)
- Electronic Chart Display and Information System (ECDIS)
- Automatic Navigation and Track-keeping System (ANTS)
- conning display.

## 6.10 Examples of integrated bridge systems

A variety of manufacturers offer a range of integrated bridge systems that can be tailored to fit the requirements of the user. Some of these systems will be described in this section. The systems selected come from leading manufacturers in this field.

### 6.10.1 Voyager by Furuno Electric Co. Ltd

An automatic navigation system designed by Furuno to meet the requirements for one-man bridge operation and the new ECDIS standards is the Voyager Integrated Bridge System. The system was designed to meet the class notation W1-OC of DNV, Norway. The system is modular which allows it to be set up to meet the requirements of the user and to provide capability for future expansion of the system as necessary. The complete system requirement comes from a single supplier with the claimed benefits of:

- increased safety
- increased cost-effectiveness
- increased navigation efficiency.

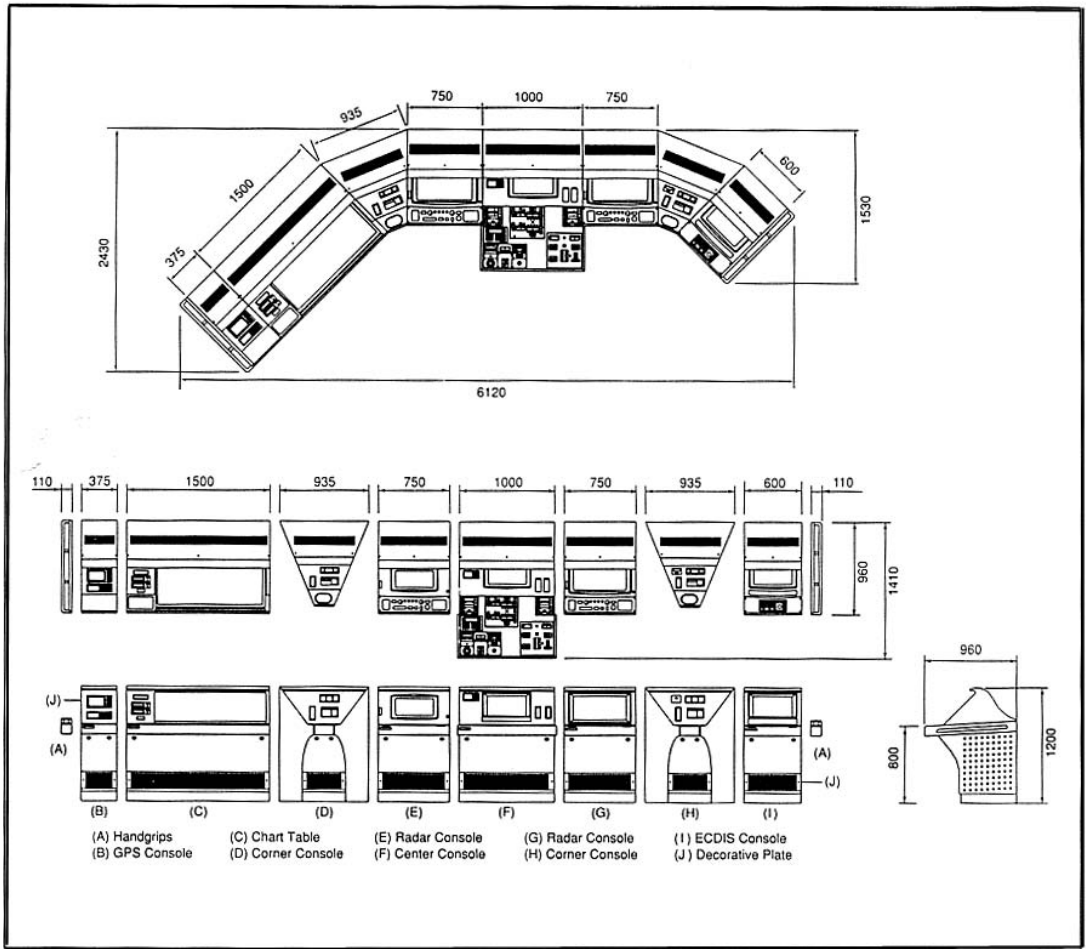


**Figure 6.2** Components of the Voyager integrated bridge system. (Reproduced courtesy of Furuno Electric Co. Ltd.)

The modular nature of the system components can be seen from Figure 6.2 which shows a possible bridge layout using the Voyager system. Figure 6.3 shows one module, that of the ARPA/Radar which is module E/G in Figure 6.2.

### *Main functions of Voyager*

There are three main functions of the system:

- electronic chart display and user interface
- position calculation and track steering
- automatic steering of the vessel.

Each of the main functions is performed using an individual processor as indicated in Figure 6.4. This guarantees real time data processing for critical applications such as positioning and steering.

**Figure 6.3** Voyager ARPA console. (Reproduced courtesy of Furuno Electric Co. Ltd.)

The system has built-in dual displays to satisfy the requirement for separate ECDIS and conning monitors. The ECDIS monitor provides the main display and user interface for the navigation system, while the conning monitors display the most important navigational sensor data in a graphical form, i.e. gyrocompass, speed log etc.

The navigation system is operated through a control panel that has dedicated function and execute keys for fast, easy operation. The steering functions are performed on their own operation control panel that integrates all functions for automatic steering. A block diagram that shows these control panels and also indicates all inputs to the navigation and track-keeping processor is shown in Figure 6.5. Figure 6.5 also indicates the type of interface connection that exists between a particular sensor and the processor.

### *Electronic chart display and user interface*

For this system the electronic chart functions are designed to meet the performance standards for the ECDIS as laid down by the IMO and the IHO. More details on these requirements can be found in

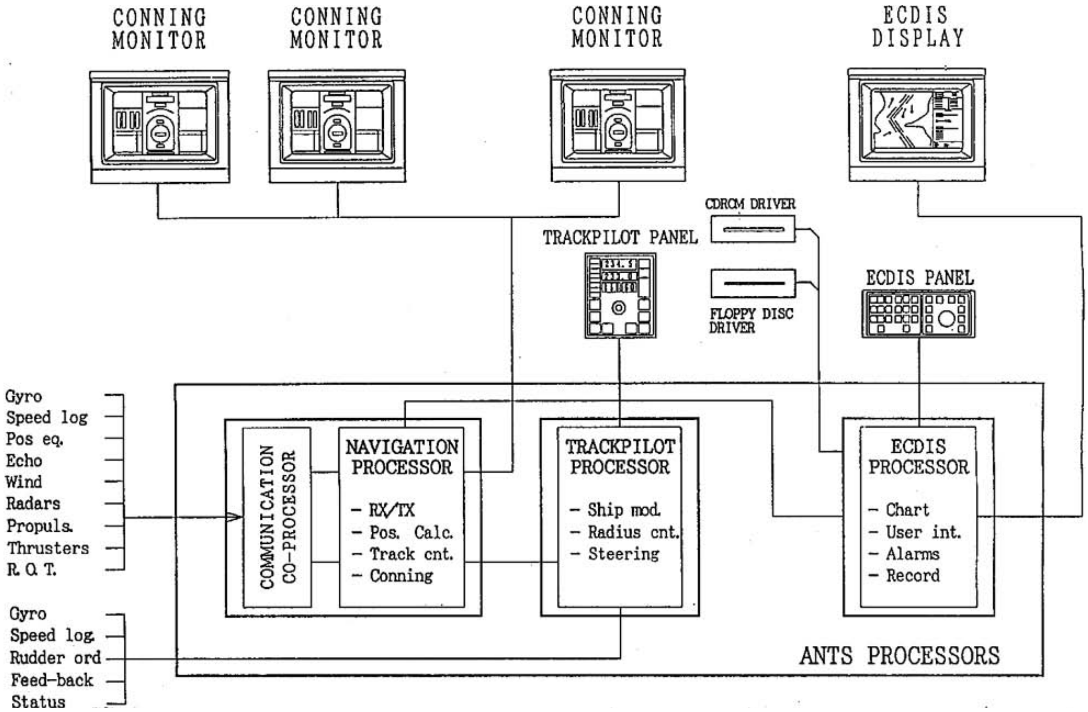


**Figure 6.4** Block diagram of the Voyager integrated bridge system. (Reproduced courtesy of Furuno Electric Co. Ltd.)

Chapter 7. ECDIS functions are performed on their own computer unit, housed in the same electronic cabinet, so as to optimize graphical performance and cost, especially when a second chart display is necessary.

The main features of the ECDIS are:

- presentation of an electronic version of a sea chart, based on the latest ENC format using a 21- (or 29-) inch high resolution colour display
- multiple navaid interface for GPS/DGPS, gyrocompass, speed log, echo-sounder etc.
- capable of use with both ENC and ARCS
- route planning and route monitoring
- primary and secondary route planning facilities
- grounding warnings
- user generated navigational safety lines which are overlaid on the radar screen
- user selectable chart layer presentation
- navigational tools such as VRM, EBL, track-ball
- display of ARPA targets
- voyage recording to meet standards
- user generated information note-books
- display of alarms
- MOB and event functions
- dedicated function keys for scale up/down, standard display, TM-reset and other functions which are the most often used functions.

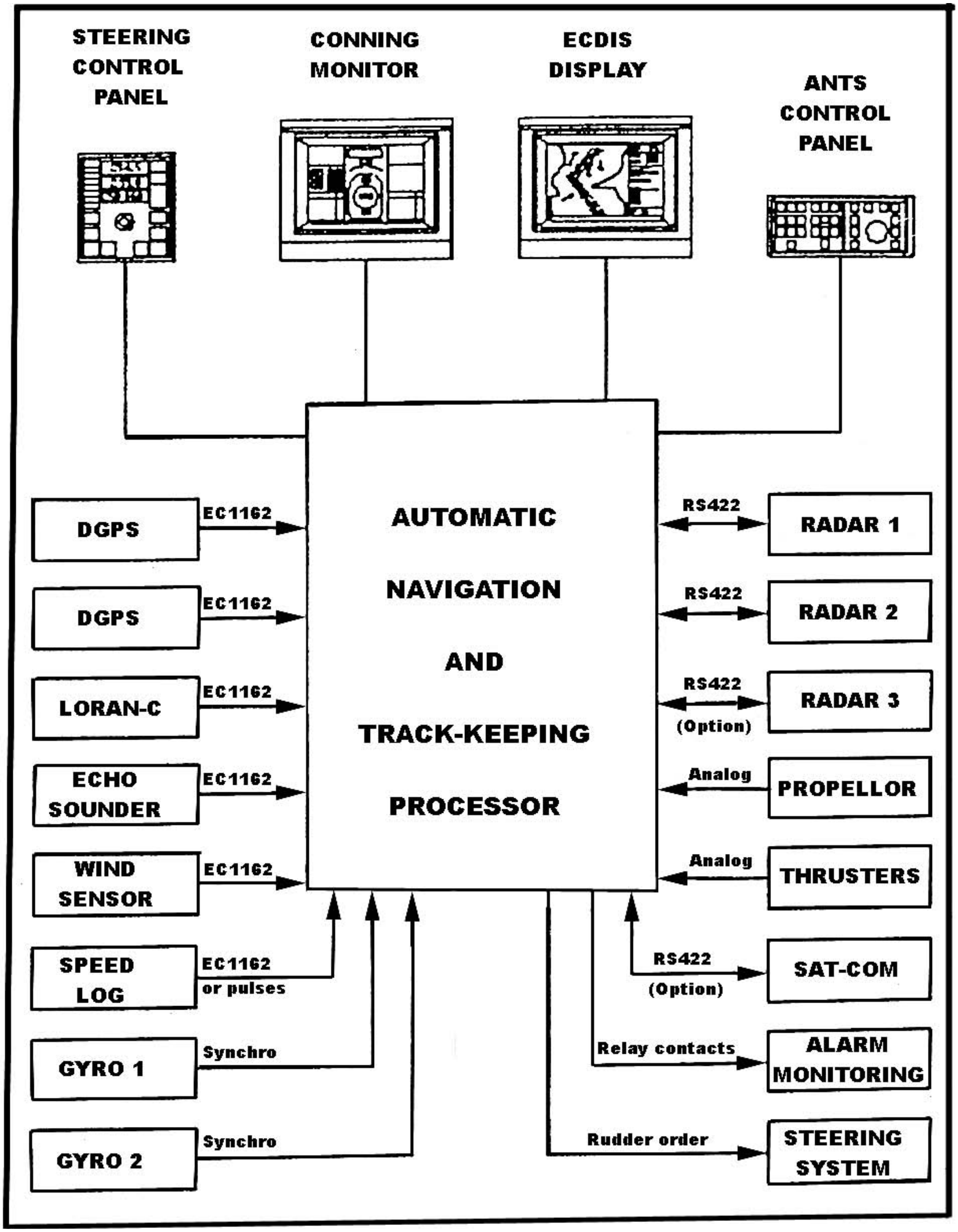


**Figure 6.5** Block diagram of Voyager automatic navigation and track-keeping system (ANTS). (Reproduced courtesy of Furuno Electric Co. Ltd.)

The option of fitting a second ECDIS computer and display, to meet the required back-up arrangements in case of an ECDIS failure, is available. If fitted, the second ECDIS computer is linked to the first through a local area network (LAN).

### *Position calculation and track steering*

The ship's position is calculated from the position sensors using the information from the gyrocompass and speed log. The position calculation is based on Kalman filter technology, which is capable of using different types of sensors and in operator-defined configurations.

Because of the need to allow for time-critical operations in position calculation and track steering, a separate processor is used for these functions. The main features of this processor are:

- interface to all external devices
- position calculation based on Kalman filter technology
- position quality calculation and alarm
- off-track calculation and alarm
- waypoint pre-warning and waypoint alarm
- graphical process and display for conning information.

### *Automatic steering function*

The system includes a complete radius/track controlled autopilot for safe and automatic steering of the vessel with the functions and operations meeting the DNV-W1 requirements. The autopilot is fully integrated into the system allowing it to be easily controlled and operated.
The main features of the automatic steering system are:

- speed adaptive operation
- radius controlled turns
- direct gyro and log inputs for accurate and reliable performance
- user selectable steering modes
- gyro mode (rudder limit controlled)
- radius mode (immediate course change)
- programmed radius mode (programmed course change)
- programmed track mode (position referenced course change)
- precision track steering with pre-memorized waypoints
- relaxed track steering with pre-memorized waypoints.

The autopilot system has its own operation control panel for logical, simple to use operation while two separate operation control panels can be installed for special applications.

### *Interface specifications*

The Voyager has a wide and flexible interface structure that allows for the system to be easily set up and configured for use. Both analogue and serial digital interfaces are available. The available interfaces to other systems are:

- gyrocompass: one analogue and one serial (NMEA) or two serial (NMEA)
- rate-of-turn gyro: analogue or serial (NMEA)
- speed log: pulse type or serial (NMEA)

- position receivers: up to five serial inputs (NMEA)
- echo sounder: serial input (NMEA)
- wind sensor: serial input (NMEA)
- rudder angle: analogue or serial (NMEA)
- propeller RPM/pitch: analogue or serial (NMEA)
- thrusters: up to four analogue inputs.

The autopilot interface requirements are:

- gyrocompass: two 1:1 synchros or high update rate serial inputs (NMEA)
- speed log: 200 p/nautical miles pulses or serial input (NMEA)
- rudder order: analogue output (0.25 V/degree) or solid-state solenoid outputs
- steering status: galvanically isolated contacts.

If a direct solenoid type of steering order is required then an optional feedback unit and solenoid drive distribution box is required.

### *Electrical specifications*

The following supplies are required with battery back-up in case of supply failure:

| | |
|---|---|
| navigation system | 24 V d.c. supply (250 W approx.) |
| alarm supply | 24 V d.c. supply (10 W approx.) |
| display monitors | 230 V a.c. or 110 V a.c. |

## 6.10.2 NINAS 9000 by Kelvin Hughes

Kelvin Hughes, the Naval and Marine division of Smiths Industries Aerospace, offer a fully integrated navigation system. Units from the Kelvin Hughes Nucleus Integrated Navigation System (NINAS) are used together with ancillary navigational equipment from specialist manufacturers.

The advantages claimed for the NINAS 9000 system include the following.

- Any number of auxiliary consoles can be added to the basic radar and navigation displays
- The use of modules gives flexibility in the final arrangement adopted by the ship owner and ship operator
- The centre consoles can be adapted to accept equipment from a number of Kelvin Hughes preferred third party suppliers
- The system is based around the proven nucleus2 6000 radar systems which are available with a variety of antennas and transmitters.

A possible bridge layout for a large passenger-carrying vessel is shown in Figure 6.6.

The wheelhouse layout consists of a centre-line steering console, two mid-position (manoeuvring and pilot) and two enclosed bridge wing consoles. The manoeuvring and pilot stations consist of a dedicated radar and a dedicated ECDIS/conning display, both being type approved CRT equipment. The centre-line station has two multifunctional LCD displays, which connect to any of three radar processors, for use as a remote operating station for either of the two ECDIS displays or as a remote operating station for any other function as required. The two stations at each wing bridge perform a similar function to that of the centre-line station.

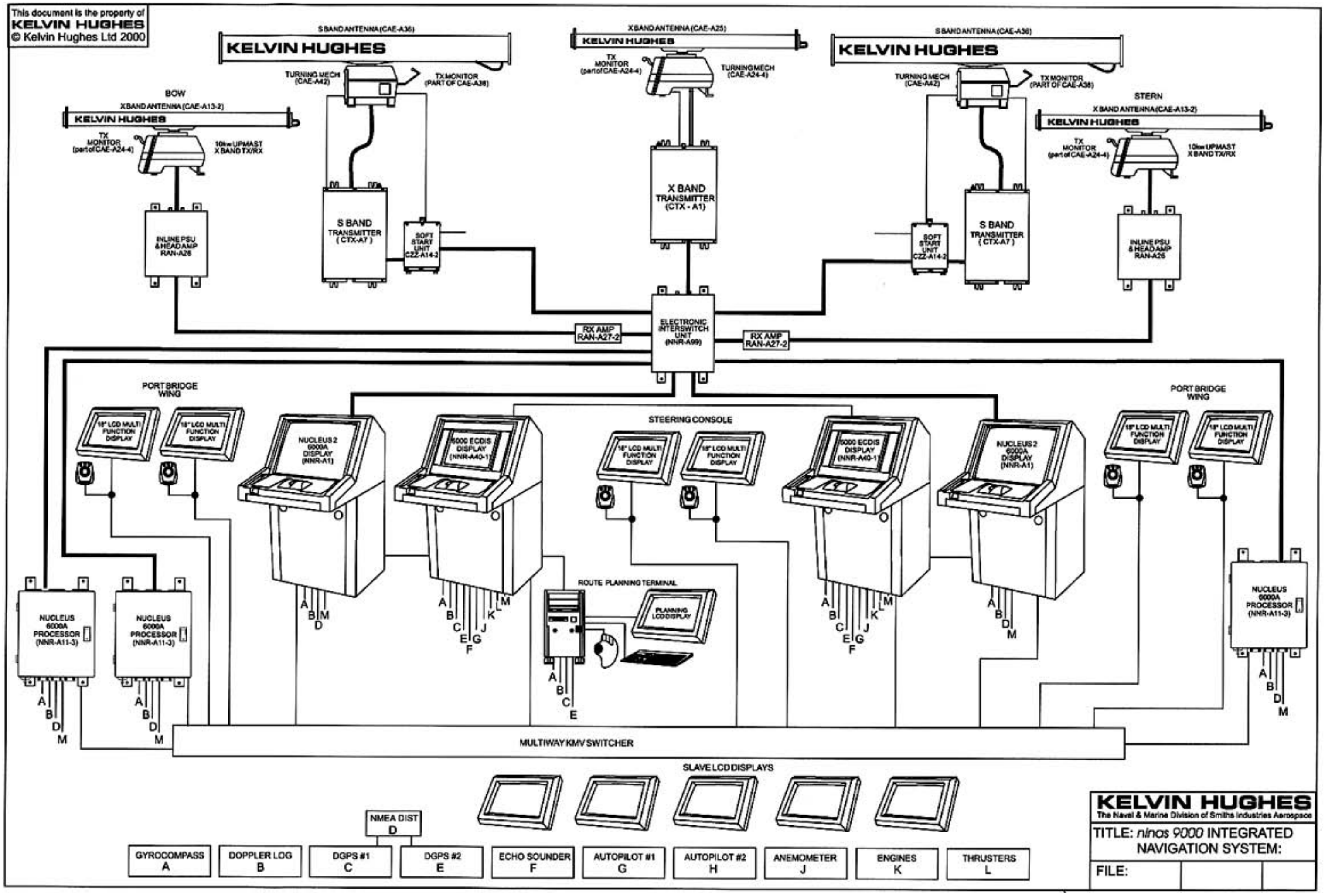


**Figure 6.6** NINAS 9000 integrated bridge system. (Reproduced courtesy of Kelvin Hughes.)

### *Display systems*

1 Radar displays

The two radar displays are 26-inch PPI, rasterscan ARPA radar displays with 10 range scales 0.25–96 nautical miles presented in relative motion, true motion and centred display true motion. There is auto tracking capability for up to 50 targets with a choice of manual or auto acquisition of targets using guard zones or footprint acquisition. The display has as standard parallel index lines, a flexible mapping system with a map storage capacity of 64K byte showing, for example, 100 maps of 80 elements.

The display has an interfacing capability of two RS232 bi-directional serial links and four NMEA opto-isolated inputs. The input capabilities are:

- GPS/Loran; waypoints; route; chart 'puck' position
- steering sequence; man overboard position; turning radius data
- serial link data from navigation display.

Output capabilities are tracker ball position and target data to ECDIS. A tracker ball and three buttons control all the radar display functions with external tracker-ball capability from each bridge wing.

2 ECDIS displays

The two ECDIS displays are IEC 1174 type approved 20-inch displays with the following functions.

- Operates with Windows-NT operating software with multi-window display showing S57 ed.3 ENC vector charts and/or ARCS/NOAA (BSB) raster charts. These may be viewed simultaneously or independently in variably sized windows.
- Graphic overlay of ownship symbol, route, waypoints, target vectors and trails on chart.
- Radar interlay of radar target echoes on chart. The interlay technique places the radar information video plane below that of the overlay to avoid obstruction of essential information.
- The ECDIS display can also act as a slave radar display by having its own radar video processing functions that allow independent control of the radar image on the ECDIS.
- North-up, course-up and head-up ENC chart presentation.
- Route safety zone function which provides a three-dimensional guard zone around own ship to monitor ship draft against chart depths and ships air draft against chart clearances to improve safety when on passage or route planning.
- Automatic plotting of time on chart with plot-on-demand function for special events.
- Passage calculator that allows route planning from the ECDIS screen. This allows calculation of distances, ETA, required speed for specific ETA and other navigational computations. This may be carried out locally or at a networked optional route planning workstation.
- Planning may be carried out visually with waypoints being dragged to modify legs and to allow the route to pass around obstacles.
- Uses ENC chart embedded database for interrogation feature, which allows the operator to request pop-up window information for any buoy, light etc. Also menu selection allows ECDIS or traditional chart symbols to be viewed for buoys and lights. There are six ENC colour palettes for optimal viewing in all light conditions.
- Continuous display of own ship heading, speed, position and depth on right side of the screen.
- Automatic Navigation and Tracking System (ANTS) interface to autopilot, allowing automated route sailing and constant radius turns.
- ECDIS display may be controlled either from the local tracker ball and three-button screen control unit (SCU) or from the remote display.

Additional functions within the ECDIS systems include a conning display, featuring the display of real-time vessel's position upon the chart in use, while displaying navigational and dynamic data in side panels. Data displayed includes:

- position
- heading
- speed (dual axis)
- depth
- wind (true and relative)
- route data
- engine RPM
- engines and thrusters.

3 Centre line console multi-function displays

Two 20-inch LCD displays that are capable of operating in the following modes.

- Fully independent radar displays capable of controlling any one of the five main radar transmitters.
- Remote radar displays capable of controlling any one of four main radar transmitters via another display (in the event of failure of the unit's own processor).
- Remote ECDIS/Conning display.

Additional functions that could also be allowed include:

- CCTV
- control and command monitoring
- alarm monitoring.

4 Bridge wing multi-function displays

Two 18-inch LCD displays that are capable of operating in the following modes.

- Fully independent radar displays capable of controlling any one of the five main radar transmitters.
- Remote radar displays capable of controlling any one of four main radar transmitters via another display (in the event of failure of the unit's own processor).
- Remote ECDIS/conning display.

Additional functions that could also be allowed include:

- CCTV
- control and command monitoring
- alarm monitoring.

5 Route planning terminal

A 17-inch LCD display with a dedicated processor designed in the same manner as an IEC 1174 type approved ECDIS display. The route planning terminal is installed as a slave unit to allow off-line route planning at the chart table position. The unit includes dedicated interfaces to log, gyro and GPS to allow it to act as a back-up ECDIS in the event of failure of the main units. Features are as for the type approved ECDIS, with the exception of radar interlay and target data.

Other components of the total system include the following.

- *Radar transmission system*. This comprises a five-way interswitched X and S band system allowing independent control of individual systems and complete interswitching of all radars.
- *Autopilot and steering system*. A system with full ANTS functionality when connected to the ECDIS. The system has inputs for both gyrocompass and magnetic compass heading data. During the normal operating mode the headings from both gyrocompass and magnetic compass are produced in the independent course monitor. In the event of a gyrocompass failure all major receivers of the gyrocompass heading, such as radar, Satcomm, GPS and digital repeaters, can be switched over immediately to the heading from the magnetic compass from the course monitor.
- *Gyrocompass system*. This is a microprocessor-controlled digital system designed as a single unit with control and display unit in the front cover. The control and display unit can be removed from the housing and installed at a position ( e.g. a bridge console) remote from the gyrocompass. The gyrocompass has an integrated TMC function, gives a rate-of-turn (ROT) output, has seven independent RS 422 and NMEA 0183 serial outputs and complies with DNV-W1.
- *Magnetic compass*. The system includes aluminium alloy binnacle, magnetic flat glass compass, a fluxgate pick-off with an integrated sine/cosine interface, bypass arrangements, azimuth devices, electronic compasses, and magnetic compass autopilots (TMC). Variation correction, gyro/TMC changeover etc. is incorporated in the gyrocompass monitor/changeover system. System uses gyro repeaters for indication when TMC is selected at the compass monitor.
- *Dual axis Doppler log*. The log is a two-axis system, the data obtained from the speed log is longitudinal and transversal bottom-track speed and depth, and longitudinal water-track speed. The log provides simultaneous W/T and B/T speeds of ±30 knots with 0.1 knot scale and depth. Bottom-track speed and depth are displayed from 3 to 300 m. Data from the log is transmitted to the log processing unit (LPU) which serves as a data concentrator/distributor in the system. The LPU is programmed according to the geometry of the ship and the position of the transducer. With this information the LPU computes transversal speeds of bow and stern. The system comprises two independent log systems each with a dedicated display at the chart table. Log selection for output to other repeaters, integrated bridge system etc. is via a selector switch at this position.
- *Echo sounder*. This unit can be operated as a single or dual frequency unit with up to four transducers. The display offers five basic ranges between 0 and 2000 m. The high resolution LCD display allows continuous observation of bottom recordings and shows all relevant navigation data. The display includes continuous indication of digital depth and range. Bottom alarm can be set at any required depth. The unit can store the last 24 h data together with the position so that a printout can be made if required.
- *DGPS*. The receiver automatically locates the strongest transmitting beacon station and lock on in seconds. In the case of signal loss it automatically switches over to an alternative station ensuring a strong signal at all times. A navtalk NMEA distribution unit is included which is fed with the output from both DGPS receivers and supplies 10 buffered outputs. In the event of failure of the primary DGPS the system automatically switches to the secondary.
- *Loran-C*. The system uses the Furuno LC-90 Mk-II receiver. Full details of this receiver can be found in Chapter 4.
- *Bridge alarm system*. This is a central alarm/dead man system which meets the highest current classification society bridge alarm specification. The system is capable of handling 40 opto-isolated switched inputs. Alarms are managed and displayed in order of priority. It is connected interactively to the integrated navigation system to allow the alarms to be repeated on the ECDIS.

### 6.10.3 Sperry Marine Voyage Management System – Vision Technology (VMS-VT)

The Sperry VMS-VT system, provided by Litton Marine Services, is a computer-based navigation, planning and monitoring system which typically consists of two or more computer workstations connected by a local area network (LAN). A typical arrangement for a VMS-VT system is shown in Figure 6.7.

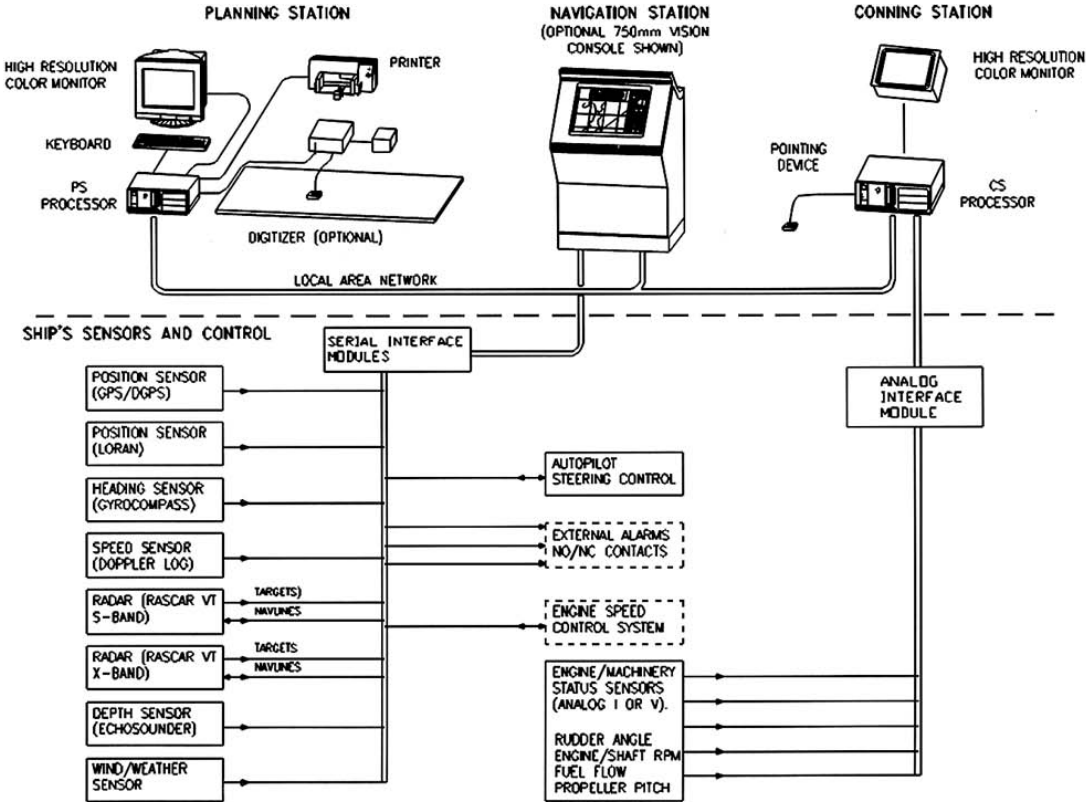


**Figure 6.7** Typical arrangement for the Voyage Management System – Vision Technology (VMS-VT). (Reproduced courtesy of Litton Marine Systems.)

Figure 6.7 shows three workstations, providing a navigation station, a planning station and a workstation designated as a conning station. The navigation station is usually located in the conning position. All VMS-VT functions are available at this station except chart digitizing and chart additions.

The planning station is usually located in the chart room and has a high-resolution monitor and printer which can provide hard copies of voyage data. Separating the planning station from the navigation station allows an operator to effect voyage planning or chart editing at the planning station without interfering with conning operations at the navigation station. The display at the navigation station is also available at the planning station so that the ship's position can be monitored at either location. A typical VMS-VT main display is shown in Figure 6.8.

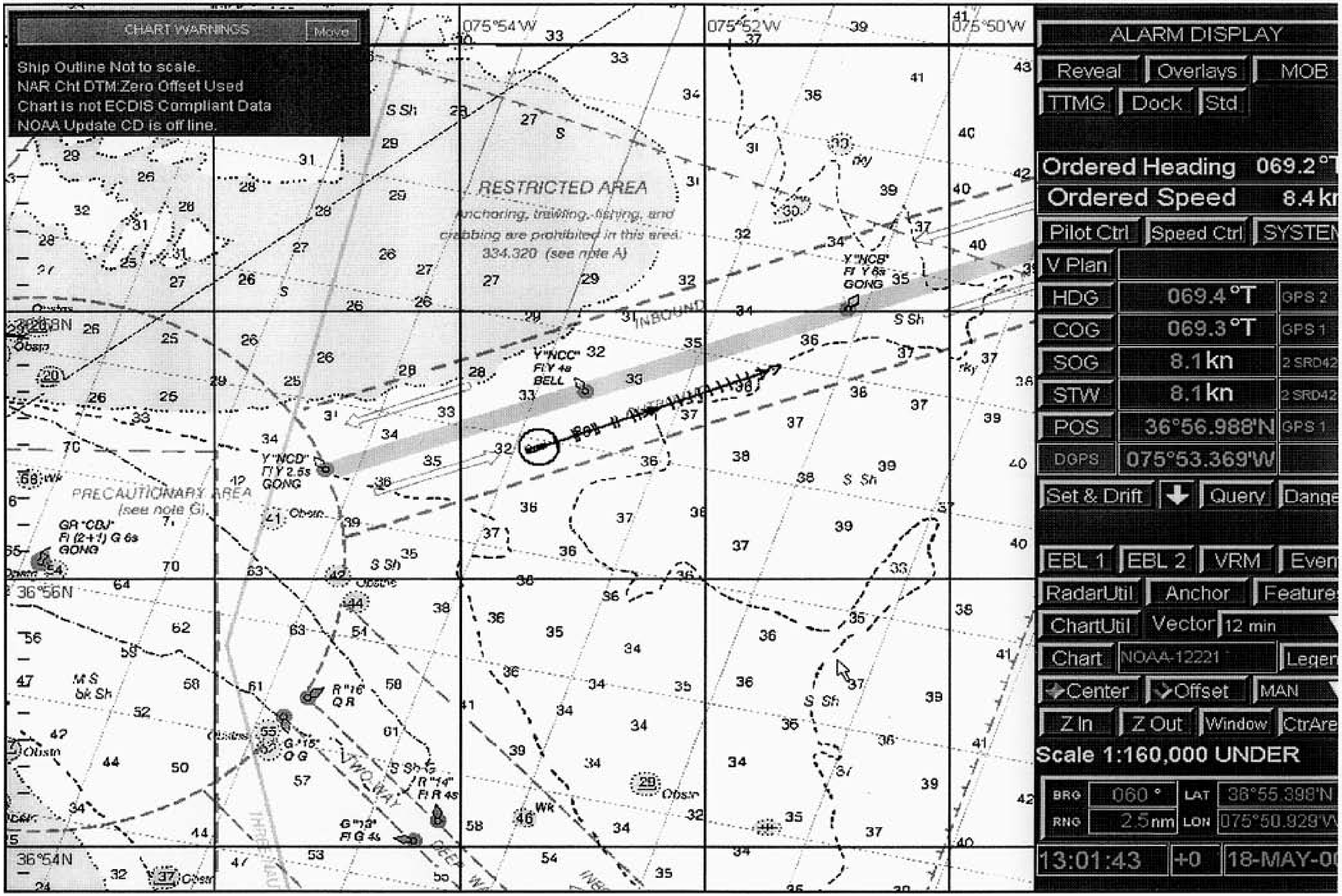


CHART WARNINGS

Ship Outline Not to scale.
NAR Cht DTM:Zero Offset Used
Chart is not ECDIS Compliant Data
NOAA Update CD is off line.

Ordered Heading 069.2°
Ordered Speed 8.4 k

HDG 069.4 °T GPS 2
COG 069.3 °T GPS 1
SOG 8.1 kn 2 SRD42
STW 8.1 kn 2 SRD42
POS 36°56.988'N GPS 1
DGPS 075°53.369'W

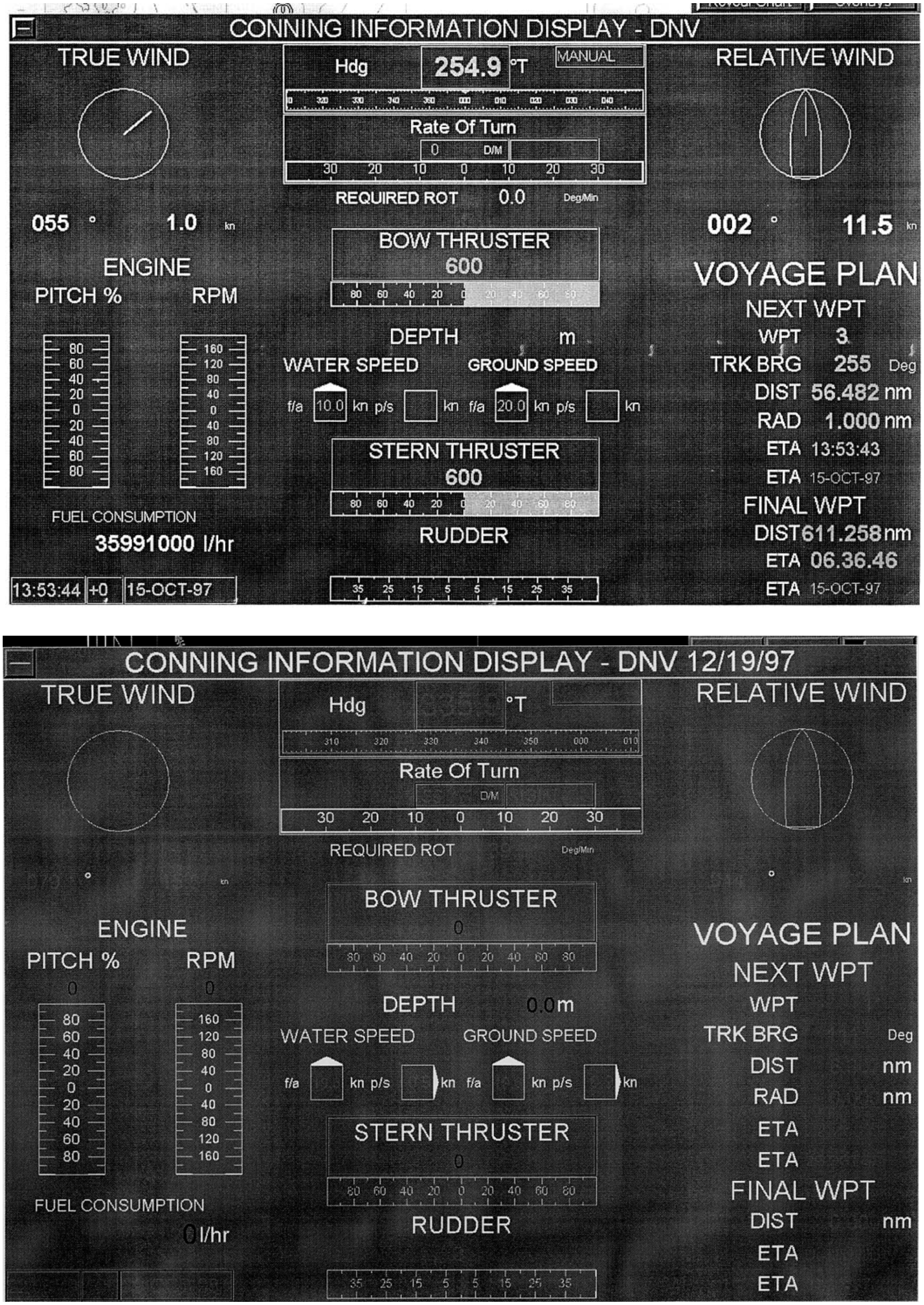
Vector 12 min
Chart NOAA-12221

Scale 1:160,000 UNDER

BRG 060 ° LAT 36°55.398'N
RNG 2.5 nm LON 075°50.929'W

13:01:43 +0 18-MAY-0

**Figure 6.8** VMS-VT main display screen showing own ship's position, heading and speed using an electronic chart. (Reproduced courtesy of Litton Marine Systems.)



CONNING INFORMATION DISPLAY - DNV

TRUE WIND 055 ° 1.0 kn

Hdg 254.9 °T MANUAL

Rate Of Turn 0 D/M

REQUIRED ROT 0.0 Deg/Min

RELATIVE WIND 002 ° 11.5 kn

ENGINE PITCH % RPM

BOW THRUSTER 600

DEPTH m

WATER SPEED f/a 10.0 kn p/s kn

GROUND SPEED f/a 20.0 kn p/s kn

STERN THRUSTER 600

RUDDER

FUEL CONSUMPTION 35991000 l/hr

13:53:44 +0 15-OCT-97

VOYAGE PLAN

NEXT WPT

WPT 3

TRK BRG 255 Deg

DIST 56.482 nm

RAD 1.000 nm

ETA 13:53:43

ETA 15-OCT-97

FINAL WPT

DIST 611.258 nm

ETA 06.36.46

ETA 15-OCT-97

CONNING INFORMATION DISPLAY - DNV 12/19/97

TRUE WIND ° kn

Hdg °T

Rate Of Turn D/M

REQUIRED ROT Deg/Min

RELATIVE WIND ° kn

ENGINE PITCH % 0 RPM 0

BOW THRUSTER 0

DEPTH 0.0m

WATER SPEED f/a kn p/s kn

GROUND SPEED f/a kn p/s kn

STERN THRUSTER 0

FUEL CONSUMPTION 0 l/hr

RUDDER

VOYAGE PLAN

NEXT WPT

WPT

TRK BRG Deg

DIST nm

RAD nm

ETA

ETA

FINAL WPT

DIST nm

ETA

ETA

**Figure 6.9** Examples of VMS-VT conning information display screens. (Reproduced courtesy of Litton Marine Systems.)

The conning station is usually configured to display a single page of specific navigation data as specified by regulatory group requirements. For this arrangement a pointing device is not provided since the display is non-interactive. At the conning station the screen is known as the conning information display (CID). Where possible the navigational and meteorological digital data is presented on the CID screen graphically to mimic analogue instruments in order to make it easier for an operator to assimilate and manage data quickly. The data presented is updated continuously and has a fixed layout pattern so that particular data is always available at the same location. A similar CID page is often available as a large display overlay screen at the VMS-VT navigation station and planning station (see Figure 6.9).

DNV on the screen displays of Figure 6.9 refers to the classification society Det Norske Veritas, Norway.

An engineering information display, as shown in Figure 6.10, can be provided as a display overlay screen at the VMS-VT navigation station and planning station or as a full-screen display at a dedicated monitor. The system can also be configured to display other pages such as a performance monitor window as shown in Figure 6.11.

As Figures 6.9–6.11 indicate, the main advantage of the VMS-VT system is its flexibility in presenting information that can be displayed in a manner that meets the customer's requirements.

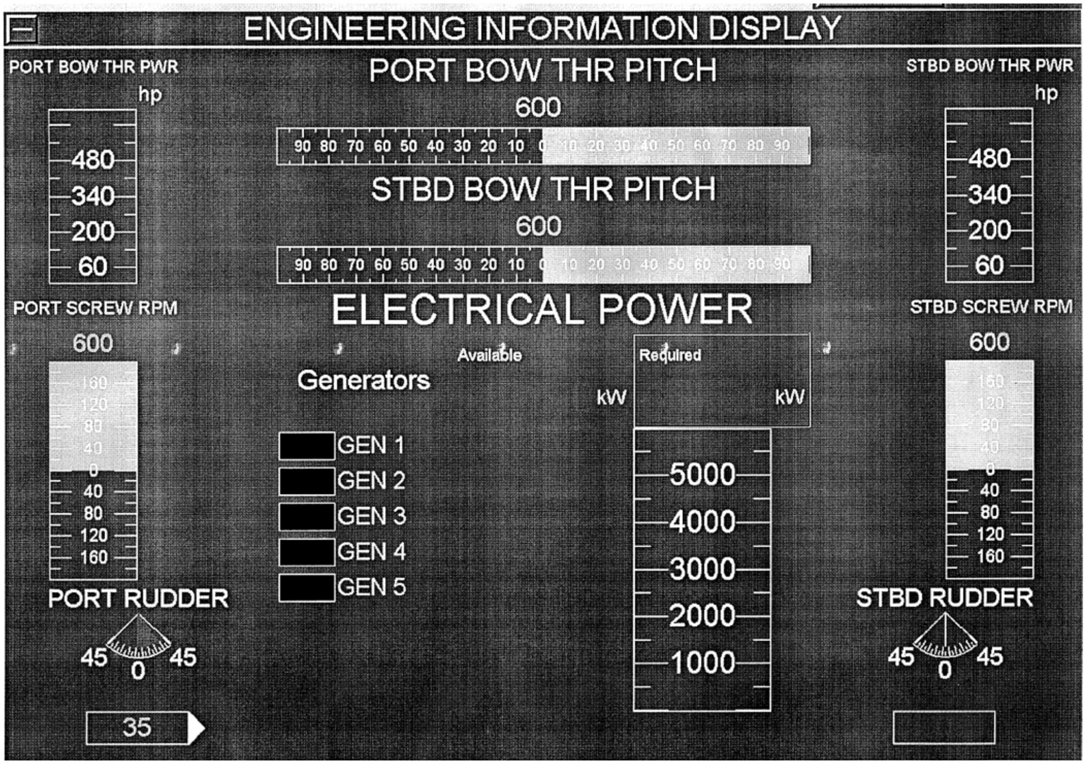
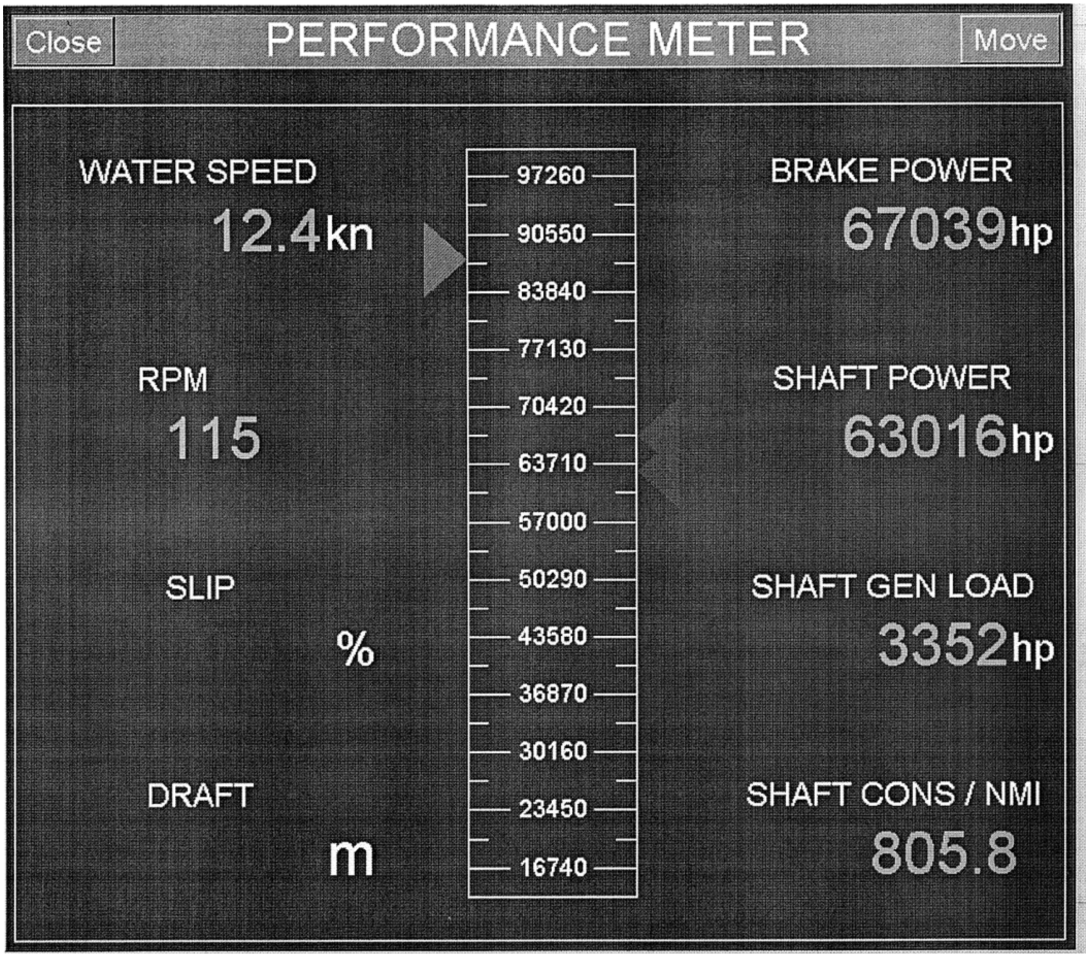


**Figure 6.10** Example of VMS-VT engineering information display screen. (Reproduced Courtesy of Litton Marine Systems.)



WATER SPEED 12.4kn

RPM 115

SLIP %

DRAFT m

BRAKE POWER 67039hp

SHAFT POWER 63016hp

SHAFT GEN LOAD 3352hp

SHAFT CONS / NMI 805.8

**Figure 6.11** Example of VMS-VT performance monitor window. (Reproduced courtesy of Litton Marine Systems.)

Basic VMS-VT functions include:

- integration of data from various sensors
- data sharing on a local area network (LAN)
- display of real-time sensor information
- display of electronic charts with ownship position
- creation of a voyage plan
- execution of a voyage plan
- display of electronic bearing lines (EBLs)
- display of variable range markers (VRMs)
- comprehensive alarm and operator message system
- printing of ship's navigation data.

Optional VMS-VT functions include:

- autopilot control
- speed order control
- display of radar target information
- DNV certified track keeping
- ECDIS S-57 or digital navigational chart (DNC) display
- interface to voyage recorder
- creation and editing of charts using the digitizer or chart additions editor
- providing data to docking displays
- providing precision manoeuvring displays
- man overboard display
- providing data to a conning station
- display of engine room data
- display of meteorological data.

Computers required for essential and important functions are only to be used for purposes relevant to vessel operation and the VMS-VT is normally configured to prevent the operator from installing or running any other application.

A VMS-VT application that includes some of the optional functions mentioned above is shown in Figure 6.12.

Among the displays shown in Figure 6.12 is an ECDIS that uses digital chart data to produce a chart display (see Chapter 7 for more information on ECDIS). The VMS-VT system has the capability to catalogue and display many types of chart formats including commercially available scanned charts produced by official hydrographic offices and/or commercially produced vector charts. Chart formats differ but VMS-VT can be configured at the factory or on the ship to use the chart format specified

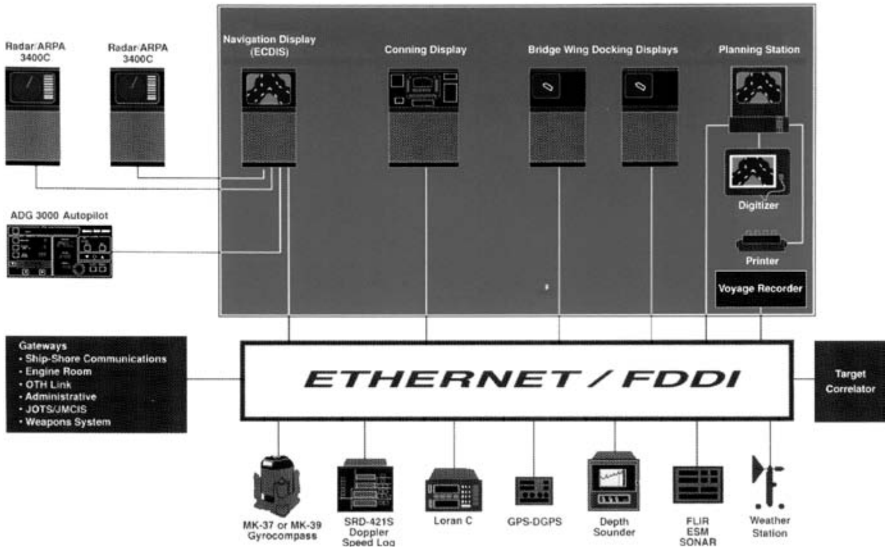


**Figure 6.12** Block diagram of the VMS-VT system. (Reproduced courtesy of Litton Marine Systems.)

by the customer. Reference to Chapter 7 will show that an ECDIS must use an electronic navigational chart (ENC) which possesses a single universal data format and they must be ‘official’ charts in that they are issued on the authority of a government authorized hydrographic office.

Available chart formats include: S57 charts; NIMA (National Imagery and Mapping Agency) DNC charts; British Admiralty ARCS raster charts (BA charts); BSB format charts such as those issued by the National Oceanic and Atmospheric Administration (NOAA); and digitized charts. Electronic charts can be retrieved from CD-ROM disks or from the computer hard disk if the required chart has been stored there.

The VMS-VT Planning Station may include a digitizer pad so that staff can create electronic charts. The digitizer can also be used to edit these electronic charts when a published Notice to Mariners updates the corresponding paper chart. The charts are stored as individual files in the VMS-VT workstations. Those charts digitized at the planning station can be copied to floppy disks for back-up storage and for transfer between ships. A standard 1.44M byte floppy disk can hold about 20 detailed charts. The digitizer can also be used to create navlines with a latitude/longitude reference, which can be transferred and displayed on the RASCAR radars.

### *Sensor data integration and display*

A major feature of the VMS-VT system is the ability to receive sensor data from the local area network and from direct hardware interfaces. The primary type of sensor data processed by the system is navigational information, which includes:

- heading
- speed over the ground
- speed through the water
- geographic position
- set and drift
- course over the ground.

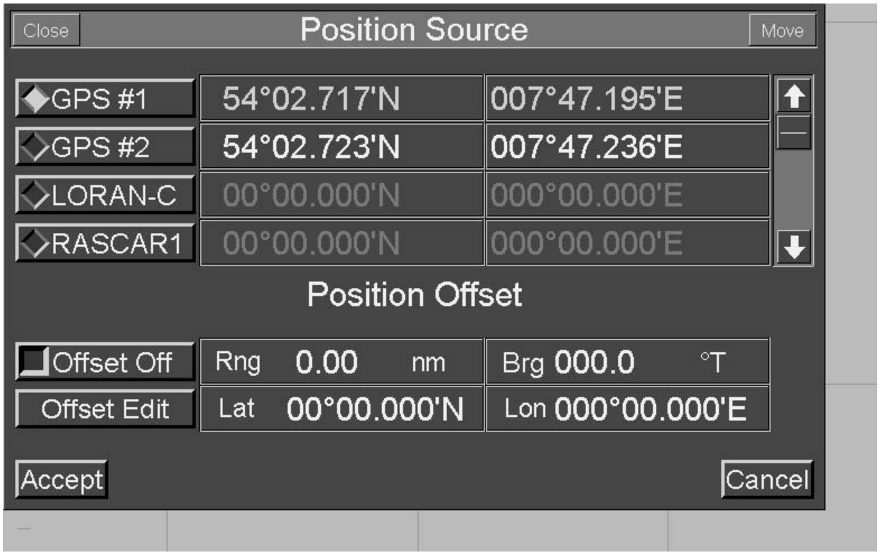
The VMS-VT sorts the data by type and provides a separate source window for each type of data. To display the source window for a particular data type requires the operator to select the appropriate button on the main menu. Each source window lists a group of sensors appropriate to the data type. The present data from each sensor is included in the window so that the best source can be selected from the list. As an example the position source window, as shown in Figure 6.13, is displayed by selecting the POS button on the main menu.

The position source window provides a list of all the configured position sensors along with the present data from each sensor. The operator may select the desired source of position data from this list or may open source windows for other types of navigational data in a similar manner.

### *Radar target data*

The VMS-VT system allows access and display of target information from multiple ARPA radars. The Litton Marine Systems RASCAR radar contains a target data logging switch for the target data logging option. If required, all the connecting RASCAR radars can send their target data allowing the operator to choose the source of ARPA target information. Radar data is automatically processed into a single target list so that if two radars have acquired the same target it will be displayed as one target at the VMS-VT. Symbols representing radar targets are displayed on the electronic chart. Each target symbol includes a speed vector, history dots and an identification number (ID).

A typical bridge layout with VMS-VT installed is shown in Figure 6.14.



Position Source

| | | |
|---|---|---|
| ◆GPS #1 | 54°02.717'N | 007°47.195'E |
| ◇GPS #2 | 54°02.723'N | 007°47.236'E |
| ◇LORAN-C | 00°00.000'N | 000°00.000'E |
| ◇RASCAR1 | 00°00.000'N | 000°00.000'E |

Position Offset

☐ Offset Off

Rng 0.00 nm

Brg 000.0 °T

Lat 00°00.000'N

Lon 000°00.000'E

**Figure 6.13** Example of VMS-VT position source window. (Reproduced courtesy of Litton Marine Systems.)

**Figure 6.14** A typical integrated bridge VMS-VT installation. (Reproduced courtesy of Litton Marine Systems.)

## 6.11 Glossary

| | |
|---|---|
| **ABS** | American Bureau of Shipping. |
| **AIS** | Automatic Identification System. |
| **ANTS** | Automatic Navigation and Track-keeping System. A system which automatically keeps a ship along a safe pre-planned track. |
| **ARCS** | Admiralty Raster Chart Service. The UKHO proprietary Raster Navigational Chart. |
| **ARPA** | Automatic Radar Plotting Aid. |
| **Bridge** | The area from which the navigation and control of the ship is managed. |
| **Bridge system** | The total system required for the performance of bridge functions, including bridge personnel, technical systems, man/machine interface and procedures. |
| **Bridge wing** | That part of the bridge on each side of the wheelhouse which extends to the ship's side. |
| **CCTV** | Closed Circuit Television. A system that allows monitoring of positions remotely by using cameras and monitor screens. |
| **Coastal waters** | Waters that encompass navigation along a coast at a distance less than the equivalence of 30 min of sailing with the relevant ship speed. The other side of the course line allows freedom of course setting in any direction for a distance equivalent to at least 30 min of sailing with the relevant speed. |
| **Conning position** | A place on the bridge with a commanding view, which provides the necessary information and equipment for a conning officer (pilot) to carry out his functions. |
| **Conning information display (CID)** | A display which clearly presents the state and/or value of all sensor inputs relevant to navigation and manoeuvring as well as all corresponding orders to steering and propulsion systems. |
| **Display** | The means by which a device presents visual information to the navigator. |
| **DGPS** | Differential Global Positioning System. |
| **DNV** | Det Norske Veritas. A member of IACS. |
| **Docking** | Manoeuvring the ship alongside a berth and controlling the mooring operations. |
| **EBL** | Electronic bearing lines. |
| **ECDIS** | Electronic Chart Display and Information System. The performance standard approved by the IMO and defined in publications from the IHO (Special Publications S-52 and S-57) and IEC document 1174. |
| **ENC** | Electronic Navigational Chart. Those charts, manufactured for use with ECDIS, which meet the ECDIS performance standards and are issued by or on the authority of government-authorized hydrographic offices. |
| **Ergonomics** | Application of the human factors implication in the analysis and design of the workplace and equipment. |
| **ETA** | Estimated time of arrival. |
| **GMDSS** | Global Maritime Distress and Safety System. |
| **GNSS** | Global Navigation Satellite System. The use of GPS for civilian purposes. |
| **GPS** | Global Positioning System. A satellite navigation system designed to provide continuous position and velocity data in three dimensions and accurate timing information globally. |

| | |
|---|---|
| **Helmsman** | The person who steers the ship when it is under way. |
| **IACS** | International Association of Classification Societies. Classification embodies the technical rules, regulations, standards, guidelines and associated surveys and inspections covering the design, construction and through-life compliance of a ship's structure and essential engineering and electrical systems. |
| **IEC** | International Electrotechnical Commission. The organization which produces world standards in the area of electrical and electronic engineering. |
| **IEC TC80** | A technical committee of the IEC that covers the area of Marine Navigation and Radio Communication Equipment and Systems. |
| **IHO** | International Hydrographic Organization. A grouping of national hydrographic offices responsible for promoting international standards in the fields of hydrographic surveying and chart production. |
| **IMO** | International Maritime Organization. A specialized agency of the United Nations and responsible for promoting maritime safety and navigational efficiency. |
| **ISO** | International Standards Organization. A non-governmental organization working to produce international agreements that are published as International Standards. |
| **ITU** | International Telecommunications Union. |
| **LAN** | Local area network. |
| **LCD** | Liquid crystal display. A form of display where the display elements are typically dark coloured alphanumerics on a grey screen. The display is easily read even in bright light conditions. |
| **MOB** | Man overboard. |
| **Narrow waters** | Waters that do not allow the freedom of course setting to any side of the course line for a distance equivalent to 30 min of sailing with the relevant ship speed. |
| **Navigation** | The determination of position and course of a ship and the execution of course alterations. |
| **NMEA** | National Marine Electronics Association. An organization comprising manufacturers and distributors. Responsible for agreeing standards for interfacing between various electronic systems on ships. NMEA 0183 version 2.3 is the current standard. |
| **Manoeuvring** | The operation of steering systems and propulsion machinery as required to move the ship into predetermined directions, positions or tracks. |
| **Monitoring** | The act of constantly checking information from instrument displays and environment in order to detect any irregularities. |
| **Ocean areas** | Waters that encompass navigation beyond the outer limits of coastal waters. Ocean areas do not restrict the freedom of course setting in any direction for a distance equivalent to 30 min of sailing with the relevant ship speed. |
| **PPI** | Plan position indicator. A type of radar display. |
| **Route planning** | Pre-determination of course and speed in relation to the waters to be navigated. |
| **Route monitoring** | Continuous surveillance of the ship's position and course in relation to a pre-planned route and the waters. |
| **RPM** | Revolutions per minute. |

| | |
|---|---|
| **Screen** | A device used for presenting visual information based on one or more displays. |
| **SOLAS** | Safety of Life at Sea. The International Convention for the Safety of Life at Sea, Chapter V Safety of Navigation, Regulation 20, Nautical Publications requires that 'All ships shall carry adequate and up-to-date charts, sailing directions, lists of lights, notices to mariners, tide tables and all other nautical publications necessary for the intended voyage.' SOLAS does not apply universally and some vessels, such as ships of war, cargo ships of less than 500 GRT, fishing vessels etc are exempt from the SOLAS requirements. |
| **VRM** | Variable range markers. |
| **Watchkeeping** | Duty undertaken by an officer of the watch. The officer of the watch is responsible for the safety of navigation and bridge operations until relieved by another qualified officer. |
| **Waypoint** | A point entered into a computer and used as a reference point for navigational calculations. Planned voyages would have a series of waypoints indicating legs of the voyage. A modern computer is capable of storing multiple waypoints. |
| **Wheelhouse** | Enclosed area of the bridge. |
| **Workstation** | A position at which one or more tasks constituting a particular activity are carried out. |

## 6.12 Summary

- Organizations such as the IMO, ISO and IEC have established international recognition of standards for ships involved in international trading.
- The integrated bridge system should be designed and installed as a physical combination of equipment or systems using interconnected controls and displays.
- Rules from classification societies, such as DNV, specify design criteria for bridge workstations, defining tasks to be performed and the siting of equipment to enable those tasks to be performed.
- The IEC Technical Committee (TC80) has produced a publication IEC 61209 covering operational and performance requirements, methods of testing and required test results for integrated bridge systems.
- To achieve optimum safety and efficiency in bridge operation, the classification society rules address the total bridge system that is considered to consist of four essential parts, namely the technical system, the human operator, the man/machine interface, and the procedures.
- The Rules for Nautical Safety are divided into three class notations: NAUT-C covers bridge design; W1-OC covers bridge design, instrumentation and bridge procedures; and W1 covers W1-OC and extensions within specified areas of W1-OC.
- Equipment carriage requirements are specified for ships according to the requested class notation.
- An operational safety manual is a requirement for class notation W1 and should obey the following guidelines: organization, daily routines, operation and maintenance of navigational equipment, departure/arrival procedures, navigational procedures, and system fall-back procedures.
- Ships requesting class notation W1-OC or W1 must comply with rules for bridge equipment tests. After installation of equipment on-board testing shall be performed in order to ascertain that the equipment, as installed, operates satisfactorily.

- A variety of manufacturers offer a range of integrated bridge systems that can be tailored to fit the requirements of the user.

## 6.13 Revision questions

1 Describe briefly the design criteria that define an integrated bridge system.
2 Describe briefly the equipment requirements for an automatic navigation and track-keeping system (ANTS).
3 Discuss the DNV rules for design criteria for bridge workstations and comment on the implications of such rules in terms of the tasks that have to be performed and the siting of relevant instruments/equipment required for those tasks.
4 What are the four essential parts that have to be considered to achieve optimum safety and efficiency in bridge operation.
5 Discuss the rule requirements set out in each section of the DNV Rules for Nautical Safety and comment on the different factors that affects the performance of the total bridge system.
6 What do you understand by class notations? Discuss the differences between the class notations NAUT-C, W1-OC and W1.
7 What do you understand by the term ‘general bridge equipment requirements’. What are the specific requirements for different types of bridge equipment?
8 Comment briefly on the rules for manoeuvring information. What type of information should be included in the provision of manoeuvring information? What form should the presentation of manoeuvring information take?
9 Describe the requirement for bridge equipment testing. Mention the type of equipment to be tested and discuss the reasons for the requirement for testing.
10 Refer to one of the examples of an integrated bridge system discussed in Section 6.10 and discuss how the system is organized to meet the requirements for such a system as specified in Sections 6.2, 6.3 and 6.4.

# Chapter 7

# Electronic charts

## 7.1 Introduction

Ever since man first went to sea there has been a requirement for some form of recognition of the sea-going environment to assist in the safe passage to the required destination. Knowledge of the coastline, safe channels for navigation which avoid wrecks, sandbanks etc., and tidal information all play their part in assisting the navigator. Paper charts giving information about particular areas have been around for centuries and hydrographers from various countries have explored the world's oceans to produce up-to-date charts which are an invaluable aid to the seafarer whether they are aboard commercial vessels plying their trade around the world or leisure craft sailing for pleasure and recreation.

In 1683 an official survey of British waters was initiated by Royal Command, although the surveys that were published some 10 years later were produced at the surveyor's expense. In the 18th century much hydrographic work around the world was done by British hydrographers, although they still had to have their work published at their own expense, gaining recompense only by selling the results of their efforts privately. It was not until 1795 that the office of Hydrographer to the Board of Admiralty was established, the French having established their Hydrographic Office some 75 years earlier. The United Kingdom Hydrographic Office (UKHO), as it is now called, has an enviable reputation as a supplier of high quality charts and provides worldwide coverage with a folio of some 3300 charts. The UKHO is a member of the International Hydrographic Organization (IHO), a body set up to co-ordinate the activities of national hydrographic offices, promote reliable and efficient hydrographic surveys and ensure uniformity of chart documentation.

It was in 1807 that the Office of Coast Survey was set up in the United States for the purpose of surveying the US coast. Various name changes followed over the years, becoming the National Ocean Survey under the newly established National Oceanic and Atmospheric Administration (NOAA) in 1970. In 1982 a further name change produced the National Ocean Service (NOS) which contained an Office of Charting and Geodetic Services which was renamed as the Coast and Geodetic Survey (C&GS) in 1991. C&GS disappeared in a 1994 restructuring but the former subordinate division, the Nautical Charting Division, re-emerged as the present Office of Coast Survey (OCS), responsible for NOAA's mapping and charting programmes. Divisions within the OCS include the Marine Chart Division, which collects the data to enable the production of nautical charts, and the Hydrographic Surveys Division, which is responsible for all areas of hydrographic survey operations.

The OCS produces about 1000 nautical charts and is also a member of the IHO and, together with the National Imagery and Mapping Agency (NIMA) share responsibilities associated with IHO membership. The IHO presently consists of 67 member states. Most of these chart only their own waters but there are three nations that can supply chart folios of the world and two more that have coverage that extends outside their own waters. The IHO is a force for chart standardization throughout the world and this is an important feature of the move towards digital production of chart data.

At the present time most hydrographic offices still operate with the paper chart as the basis of their operations. However, over the past few years electronics has moved into the sphere of charting and now digital chart data is becoming more popular and is likely to be the mainstay product of the hydrographic offices in the years to come. With this new technology the seafarer is provided with a means of viewing a chart using a monitor that can display, in colour, all the information present on a paper chart. The chart information is contained on a memory device such as a CD-ROM and can be stored on a computer hard disk. Suitable navigational software can enable the chart data to be viewed for the purpose of 'safe and efficient navigation'. The electronic chart is one where chart data is provided as a digital charting system and it is capable of displaying both geographical data and text to assist the navigator. An electronic chart may fall into one of two categories.

- Official, which describes those electronic charts which are issued by, or on the authority of, a national hydrographic office. The hydrographic offices are government agencies and are legally liable for the quality of their products regardless of whether those products are paper or digital. Such charts are updated at regular intervals in order to conform to the SOLAS (Safety of Life at Sea) requirement that charts should be 'adequate and up-to-date for the intended voyage'.
- Non-official, which describes those electronic charts which are issued by commercial organizations which may use data owned by a hydrographic authority but are not endorsed by that authority.

An electronic chart may be constructed using either of two types of data, raster or vector.

### 7.1.1 Raster data

Raster data is produced by scanning a paper chart. This process produces an image that is an exact replica of the paper chart and which comprises a number of lines that are composed of a large number of coloured dots, or pixels. This technique does not recognize individual objects, such as a sounding, which limits its ability to conform to certain international guidelines. However, the use of what is termed a vector overlay, which can display specified user data such as waypoints and system data such as radar overlays etc., can overcome this deficiency. The advantages of raster charts can be summarized as follows.

- User familiarity since they use the same symbols and colours as paper charts.
- They are exact copies of the paper charts with the same reliability and integrity.
- The user cannot inadvertently omit any navigational information from the display.
- Cost of production is less than their vector counterpart.
- Wide availability of official raster charts. ARCS charts, for example, have near worldwide coverage.
- By using vector overlays together with appropriate software, raster charts can be used for all standard navigational tasks normally undertaken using paper charts. They can also emulate some of the functions of an electronic display and information system (ECDIS).

Disadvantages of raster charts can be summarized as follows.

- The user cannot customize the display.
- When using vector overlays the display may appear cluttered.
- They cannot be interrogated without an additional database with a common reference system.

- They cannot, directly, provide indications or alarms to indicate a warning to the user.
- Unless data content is the same, more memory is required to store data compared to a vector chart.

### 7.1.2 Vector data

Apart from the electronic navigational chart (ENC), which is compiled using raw data, vector data may also be produced by scanning a paper chart. However, the raster image is then vectorized by digitally encoding individual charted objects and their attributes (structured encoding) and storing such data, together with the object's geographical location, in a database. The ENC is the designated chart for the ECDIS system and is discussed in the next section. Chart features may be grouped together and stored in thematic layers that individually categorize each group. For example, the coastline could form one layer while depth contours are found on another layer etc. The system operator can thus optimize the display to show only that data of interest and avoid the display becoming cluttered with unwanted data. The vector chart is intelligent in that it can provide information that allows a warning of impending dangers to be generated.

The process of producing vector charts is time consuming and expensive while verification of chart data is more complicated than its raster counterpart. The advantages of vector charts can be summarized as follows.

- Chart information is in layers which allows selective display of data.
- The display may be customized to suit the user.
- Chart data is seamless.
- It is possible to zoom-in without distorting the displayed data.
- Charted objects may be interrogated to give information to the user.
- Indications and alarms can be given when a hazardous situation, such as crossing a safety contour, occurs.
- Objects may be shown using different symbols to those used on paper or raster charts.
- Chart data may be shared with other equipment such as radar and ARPA.
- Unless data content is the same, less memory is required to store data compared to a raster chart.

Disadvantages of the vector chart can be summarized as follows.

- They are technically far more complex than raster charts.
- They are more costly and take longer to produce.
- Worldwide coverage is unlikely to be achieved for many years, if ever.
- It is more difficult to ensure the quality and integrity of the displayed vector data.
- Training in the use of vector charts is likely to be more time consuming and costly compared to that needed for raster charts.

The vehicle for the delivery of electronic chart data is the Electronic Chart Display and Information System (ECDIS) which is a navigation hardware/software information system using official vector charts. Such a system must conform to the internationally agreed standard adopted by the International Maritime Organization (IMO) as satisfying a vessel's chart-carrying requirements under SOLAS. The ECDIS hardware could be simply a computer with graphics capability or a graphics workstation provided as part of an integrated bridge system. The system has inputs from other sources, namely position sensors such as GPS or loran, course indication from the gyrocompass, speed from the ship's log etc.

The information is transmitted to the ECDIS using National Marine Electronics Association (NMEA) interfacing protocols. Radar information can also be superimposed using either raw data from a raster scan radar or as synthetic ARPA (automatic radar plotting aid) data. The ECDIS software must comprise a user interface and a component that allows charts to be displayed and data read. The chart data component of ECDIS is the electronic navigational chart (ENC) which must comply with ENC production specifications under the IHO's S-57 edition 3 data transfer standard. More details of this system can be found in Section 7.3.

## 7.2 Electronic chart types

There are many different types of electronic charts available that use different formats, different levels of content and attribution, and may, or may not, be official charts. As described above, all presently available electronic charts are either vector or raster. For the former, the chart may be based on the IHO S-57 format or some other format. Only if the level of content and attribution of the chart conforms to the IHO ENC product specification and is produced by, or on the authority of, a government authorized hydrographic office, can the chart be considered an ENC as defined by the IMO ECDIS performance standards.

Official vector charts issued by the relevant hydrographic offices should conform to the ENC product specification based on the IHO S-57 format. Privately produced vector charts (non-official) may, or may not, conform to the ENC product specification. However, the use of unofficial ENCs will render an ECDIS non-compliant. Finally, it is possible to obtain charts that do not use the IHO S-57 format and do not conform to the ENC product specification.

### 7.2.1 Privately produced vector charts

These are generally made from scanned hydrographic office paper charts. The image produced is then digitized by tracing lines and features on the chart. This vectorization process stores chart features in 'layers' which can be redrawn automatically at an appropriate size if the chart is zoomed into. Categories of data, such as spot depths, navigation marks etc., can be added/deleted as required. In some systems specific chart items can be interrogated to obtain more information.

The nature of the vector display is such that the chart data is not displayed electronically as it was compiled in its paper chart form. Most systems automatically decide on the information to be displayed, depending on the level of zoom, to avoid the image being cluttered. Thus a new operational regime has to be developed to take account of the implications of:

- adding/deleting layers of data
- zooming and seeing more/less data appear according to the level of zoom
- displaying the chart at a larger scale than the source paper chart.

One of the principal producers of digital format electronic charts is C-MAP of Norway with worldwide coverage of 7500 charts on a CD-ROM. Data is coded in a System Electronic Navigational Chart (SENC) format called CM-93/3 which is compliant with the IHO S-57 format. C-MAP 93/3 displays a -U- (for unofficial) on their privately produced S-57 compliant charts. Details of the use of a SENC in an ECDIS is discussed in Section 7.3.

### 7.2.2 Official raster chart

There are two official raster chart formats.

- BSB raster charts, which contain all the data found in NOAA paper charts, with updates published weekly. These updates are available via the Internet and are in-sync with the US Coast Guard (USCG), NIMA and Canadian notices. NOAA has 1000 official charts and all have been available in raster form since 1995. These raster charts are produced jointly by NOAA and Maptech Inc. under a co-operative research and development agreement. The growth of computer-based navigation systems, together with GPS and other positioning systems, has meant an increase in the sale of raster charts and today approximately twice as many raster charts are sold compared to paper charts. The raster charts are available in CD-ROM form with each CD-ROM containing about 55 charts together with other relevant navigational facilities.
- UKHO ARCS and Australian Hydrographic Office (AHO) Seafarer both produced in the UKHO's proprietary hydrographic chart raster format (HCRF). ARCS is updated weekly using a CD-ROM with the same information as the weekly Notice to Mariners used to correct paper charts. Seafarer is updated monthly on a similar basis. ARCS has near worldwide coverage with 2700 charts available on CD-ROM.

ARCS/Seafarer charts are produced from the same process used to print paper charts, i.e. a rasterized process is used either to print a paper chart or produce a raster chart. They are accurate representations of the original paper chart with every pixel referenced to a latitude and longitude. Where applicable, horizontal datum shifts are included with each chart to enable the chart, and any information overlaid on it, to be referenced back to WGS-84. Not all available charts have WGS-84 shift information and such charts must be used with caution when a GPS position fix is applied. Chart accuracy is discussed further in Section 7.4.

The UKHO ARCS production system involves the use of a raster base maintenance and on-line compilation system (ABRAHAM) which is used to update, manage and plot navigational chart bases. The ARCS production system is integrated with ABRAHAM as is shown in Figure 7.1.

In its simplest terms ABRAHAM is all processes that are necessary to create and maintain the high-resolution (25 µ/1016 dpi) monochrome raster bases from which paper charts are produced, and ARCS is all processes that turn the ABRAHAM bases into ARCS CD-ROMs including:

- processing bases into lower resolution (200 µ/127 dpi) colour images
- adding header and catalogue information
- quality assurance checks
- encrypting the data
- ending a CD-ROM master to a pressing plant
- checking the stock returned by the pressing plant.

The ARCS CD-ROM production can be subdivided into periodic processing cycles.

- Weekly. Updates for all charts affected by Notice to Mariners are generated, checked and placed on the weekly ARCS update CD-ROM. New charts and new editions published that week are also included on the update CD, as is the text of temporary and preliminary Notices to Mariners.
- Periodic. To prevent the update CD from filling up, accumulated updates are periodically moved onto reissues of the ARCS chart CD-ROMs. This results in the production of a reissued chart CD at the same time as the weekly update CD. Nominally, one chart CD requires to be reissued each

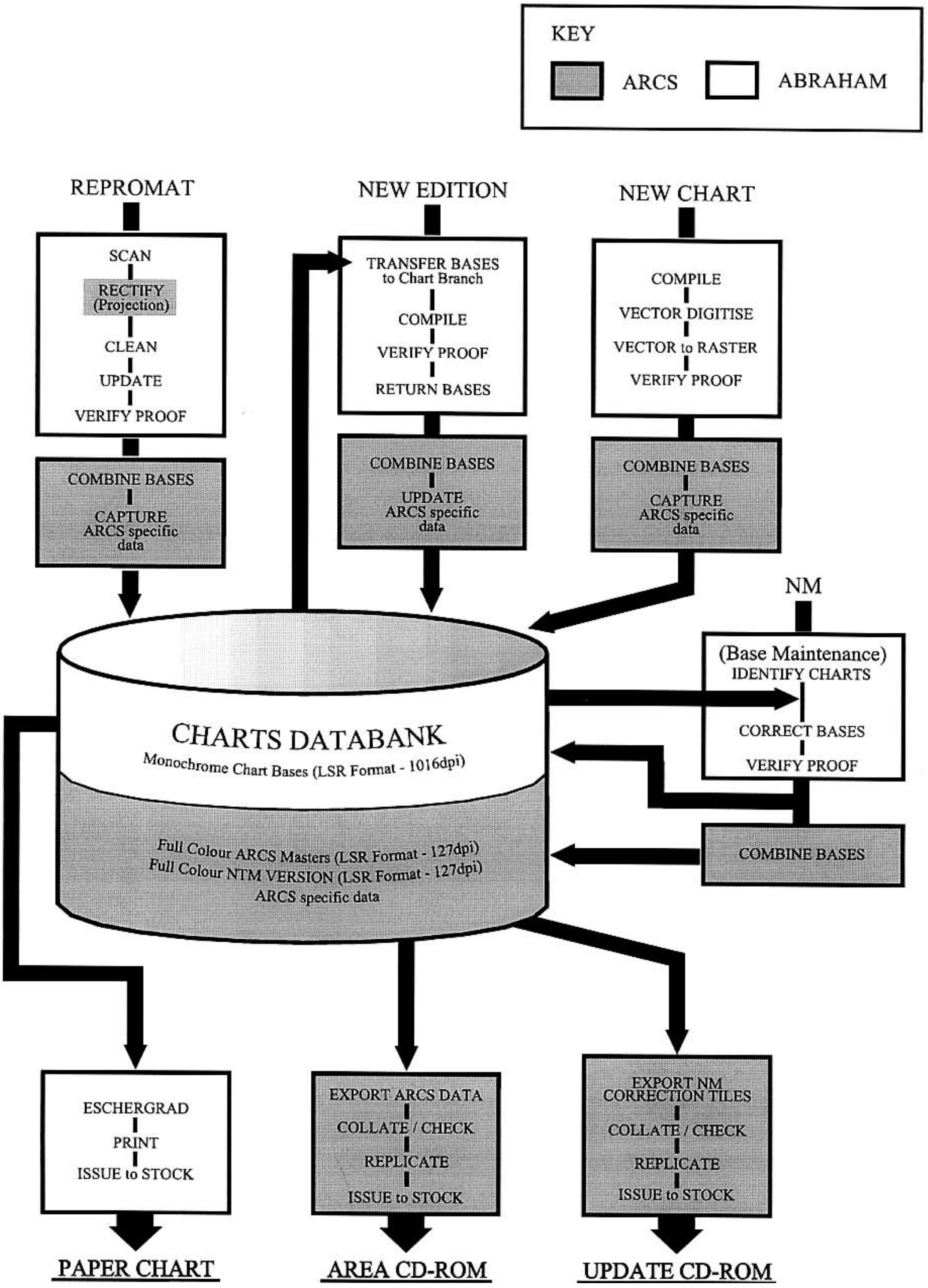
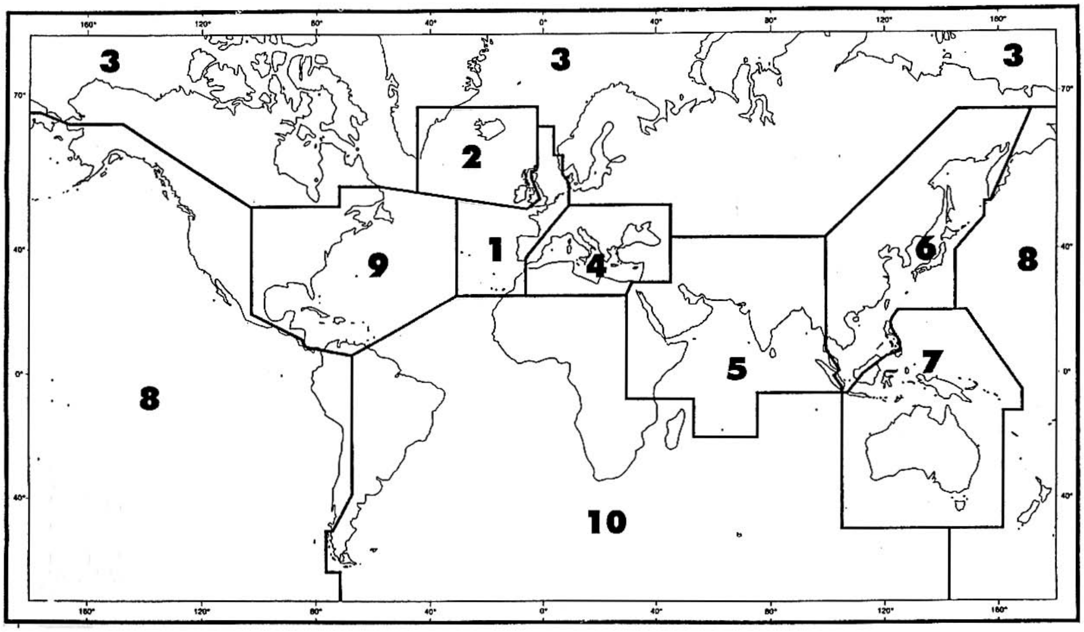


**Figure 7.1** ARCS/ABRAHAM production system. (Reproduced with the permission of the Controller of HMSO and the United Kingdom Hydrographic Office.)



| | | | |
|---|---|---|---|
| RC 1. | North Sea and English Channel to Gibraltar | RC 6. | Singapore to Japan |
| RC 2. | British Isles (west coast) and Iceland | RC 7. | Australasia, Borneo and Philippines |
| RC 3. | Northern waters and Baltic Sea | RC 8. | Pacific Ocean |
| RC 4. | Mediterranean and Black Seas | RC 9. | North America (east coast) and Caribbean |
| RC 5. | Indian Ocean (northern part) and Red Sea | RC 10. | South Atlantic and Indian Ocean (southern part) |
| | | RC 11. | Ocean Charts (1:3.500.000 and smaller) |

**Figure 7.2** Regional coverage of ARCS CD-ROMs. (Reproduced with the permission of the Controller of HMSO and the United Kingdom Hydrographic Office.)

month, but the schedule varies according to the number of corrections outstanding and the number of chart CDs in stock.

- Monthly. Cross-checks are carried out against the data held on the Sales Order Processing System and the Chart Information System (CIS).

The UKHO provides a near worldwide coverage with 2700 charts available as ARCS CD-ROMs. The regional coverage of these charts is shown in Figure 7.2.

Table 7.1 gives a comparison between the BSB and ARCS raster types.

### 7.2.3 Electronic navigational charts (ENC)

These are the designated charts for the ECDIS system and they possess a single universal data format. Such charts use vector data based on the IHO Special Publication S-57, edition 3, IHO Transfer Standard for Digital Hydrographic Data. Some of the major points which identify the unique property of these charts are as follows.

- They are issued by or on the authority of a government-authorized hydrographic office.
- Items on the chart must be attribute-coded and must be able to be interrogated to provide information.

**Table 7.1** Comparison between different raster chart types. (Reproduced courtesy of D. Edmonds of PC Maritime, UK)

| *Feature* | *BSB* | *ARCS* |
|---|---|---|
| Government authorized | Yes | Yes |
| Entire catalogue always up to date to latest notice to mariner | Yes | Yes |
| Update service | Weekly | Weekly |
| Original scan from: | Stable mylar film originals used for printing paper charts | Stable colour separates used for printing paper charts |
| Scan resolution | 762 dpi | 1016 dpi |
| Chart resolution | 256 dpi | 127 dpi |
| Anti-aliasing | Yes | Yes |
| No of points used to relate the chart images to Lat/Long conversion | 10–20, pixel to location conversions are also provided, accuracy depends on the printed chart | Pixel to position conversion is by calculation and is accurate to 1 pixel |
| Geodetic datum shifts | Yes | Yes |
| Integrity checks | Byte checksums are included in chart file | 32-bit CRC check on original and updated image |
| Liability | US government accepts liability for errors on NOAA charts | UK government accepts liability on UKHO products |

- The data is delivered in cells to provide seamless data for the task in hand. The cell structure changes according to the data set used.
- All chart data is referenced to a global geodetic datum, WGS-84, which is the datum used by GPS.

The data is fully scaleable and it only needs a view area to be defined for an appropriate level of data to be automatically presented to the operator. If it is required to add/delete data then information can be grouped into layers and turned on/off as required. Zooming can allow the chart image to be enlarged to provide greater ease of use. Zooming with a raster chart clearly shows when an image is presented at a scale greater than the compilation scale since the text and navigational symbols would be larger than their normal size rendering the chart unsafe for navigation. Over-scaling with an ENC has the problem that the navigational symbols remain the same size regardless of the scale used and this could cause a potential navigation hazard. The ECDIS is required to display an over-scale warning automatically if it has used zooming to produce an image beyond the compilation scale of the chart.

Individual contour lines can be defined as safety contours with anti-grounding warnings given based on the ship's closeness to them. Alarms will be generated automatically if the ECDIS detects a conflict between the vessel's predicted track and a hydrographic feature within the ENC that represents a potential hazard to the vessel.

The ECDIS can offer different chart information by displaying all ENC content, a subset of the ENC content (known as standard display) or a minimum permitted subset of ENC content (known as display base). The first two categories permit information to be added/deleted while the display base cannot have information deleted since it is stipulated as the minimum required for safe navigation. A System Electronic Navigational Chart (SENC) is that database obtained by the transformation of the ENC data, including any updates and data added by the user, by the ECDIS prior to display. It is the SENC that forms the basis for the display and the user decides what part of the SENC database is required for the display. It is a requirement that the ENC database must remain unaltered so that the SENC database could be reconstructed should it be debased in any way during operations.

The availability of ENCs will depend on key factors that affect the NHOs producing them. These factors include the following.

- Production experience. The rate of production should increase as staff gain more experience in the production of these charts.
- Data quality. Software tools necessary to underpin the quality assurance of the digital database have to be developed to ensure compliance with S57, edition 3 requirements. This will take time.
- Uniformity of data. There is a need for all hydrographic offices to ensure their ENCs are produced with consistency in the interpretation of the standard and to product specification. The use of regional co-ordinating centres is of use in facilitating this.
- Geographical cover. By concentrating on the geographical areas most used by shipping companies it should be possible to deliver the required charts ahead of others.

As an example of the development of ENCs, the UKHO awarded a contract to the Indian company, IIC Technologies, for data capture work in February 2000. This is the first step in the production of ENCs with the data sets produced by IIC to be quality assessed by the UKHO to ensure compliance with the required standards. The UKHO will also concentrate on stitching together the data set cells and matching the edges to produce a seamless ENC database. The contract is an enabling contract of up to four and a half years allowing the UKHO to request data sets in tranches with continuity of production.

The regional co-ordinating centres are an important means of distributing the ENCs to potential customers. The International Hydrographic Organization (IHO) proposed a system for supplying ENCs to be known as Worldwide ENC Database (WEND). Using this concept the world is divided into Regional ENC Co-ordinating Centres (RENCs). At present only one RENC has been set up,

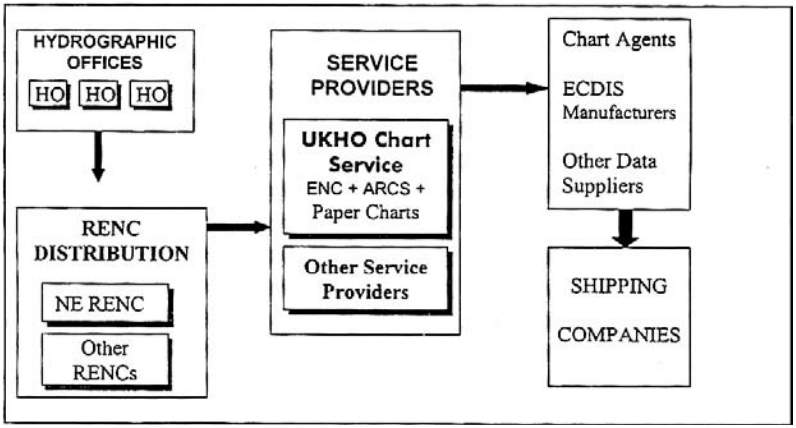


**Figure 7.3** RENC distribution system. (Reproduced with the permission of the Controller of HMSO and the UK Hydrographic Office.)

**Table 7.2** Equivalence to the paper chart. (Reproduced courtesy of D. Edmonds of PC Maritime, UK)

| *Privately produced vector charts* | *Official raster (RNCs)* | *ENCs* |
|---|---|---|
| Generally a copy of the paper chart | An exact replica of paper chart | All data merged into cells |
| A different image to the original paper chart is presented at all levels of zoom and scale | The same image as the paper chart is always presented. The chart is more equivalent to the paper chart than any vector chart including ENCs | No resemblence to the paper chart |
| Symbols and colour vary with manufacturer | Symbols and colour are the same as the paper chart equivalent | The IHO publication S–52 defines new colours and symbols for ENCs |
| Accuracy, reliability and completeness vary with manufacturer | RNCs are as accurate, reliable and complete as the paper version | ENCs should eventually be more accurate and reliable than the paper version |
| A new operational regime is required | The same operational regime as paper charts is followed. There are some changes, if only because of screen size | A new operational regime is required |

**Table 7.3** Chart integrity. (Reproduced courtesy of D. Edmonds of PC Maritime, UK)

| *Privately produced vector charts* | *Official raster (RNCs)* | *ENCs* |
|---|---|---|
| Produced by private companies | Produced by, or under the authority of government authorised hydrographic offices | Produced by, or under the authority of government authorised hydrographic offices |
| Unofficial | Official | Official |
| Generally no responsibility is accepted | Responsibility is accepted for chart data in terms of its completeness and accuracy in comparison with the equivalent paper chart | Responsibility is accepted for chart data in terms of its completeness and accuracy |
| Is unlikely to become legally equivalent to the paper chart | Is unlikely to become legally equivalent to the paper chart | Is legally equivalent to the paper chart |
| It may be possible to change original chart data | The chart data is tamper proof | The chart data is tamper proof |
| Charts can be zoomed (i.e., the display of a single chart is magnified or reduced without restriction. Chart detail varies depending on the level of zoom) | Chart zoom should be limited to a level that does not break up the image. Information displayed on the chart remains unaltered | Charts can be zoomed in or out without restriction. Chart detail varies depending on the level of zoom |
| Quality control varies with manufacturer | Quality control is government standard | Quality control is government standard |

**Table 7.4** Chart corrections. (Reproduced courtesy of D. Edmonds of PC Maritime, UK)

| *Privately produced vector charts* | *Official raster (RNCs)* | *ENCs* |
|---|---|---|
| Up-to-dateness of charts varies with manufacturer | Charts are up-to-date at the point of sale | Charts will be up-to-date at the point of sale |
| It is difficult to determine the up-dating policy of manufacturers | Chart data is maintained up-to-date to clearly stated standards | Chart data is maintained to a clearly defined standard |
| Varies with manufacturer | On demand updates for leisure users | Not applicable |
| Varies with manufacturer | Subscription updates for commercial users | Subscription updates available |
| Varies with manufacturer | Automatic integration of chart updates | Automatic integration of chart updates |

**Table 7.5** Safety. (Reproduced courtesy of D. Edmonds of PC Maritime, UK)

| *Privately produced vector charts* | *Official raster (RNCs)* | *ENCs* |
|---|---|---|
| Geodetic datum shift to WGS-84 may not be provided | Chart data includes geodetic datum shift to WGS-84, if known | All data is referenced to WGS-84 |
| Chart data can be removed from the display. Significant navigation information may be inadvertently removed | Chart data cannot be removed from the display. The user cannot inadvertently remove significant navigation information | Chart data can be removed from the display. Significant navigation information may be inadvertently removed |

namely the Northern Europe RENC known as PRIMAR. This is a co-operative arrangement between most of the national hydrographic offices in northern and western Europe. To date the hydrographic offices of Denmark, Finland, France, Germany, Netherlands, Norway, Portugal, Poland, Sweden and UK have signed the formal co-operation arrangement and other hydrographic offices have expressed an interest in joining. PRIMAR is operated by the UK Hydrographic Office and the Norwegian Mapping Authority's Electronic Chart Centre.

The ENCs will be sold through a network of distributors and should be able to provide worldwide cover by exchange of data with other RENCs once these are established in other parts of the world. A block diagram showing the RENC concept is shown in Figure 7.3.

Tables 7.2 to 7.5 summarize the features of each chart type in relation to each other.

## 7.3 Electronic chart systems

### 7.3.1 Electronic Chart Display and Information System (ECDIS)

There are several types of electronic chart systems available but only one performance standard has been approved by the International Maritime Organization (IMO) in November 1995. The IMO resolution A817(19) states that the ECDIS should 'assist the mariner in route planning and route monitoring and, if required, display additional navigation-related information'. The system approved is known as the Electronic Chart Display and Information System (ECDIS) and applies to vessels governed by Regulation V, Chapter 20 of the 1974 Safety of Life at Sea (SOLAS) convention. It complies with the carriage requirement for charts with an ECDIS system using Electronic Navigational Charts (ENCs). ECDIS is a navigational information system comprising hardware, display software and official vector charts and must conform to the ECDIS performance standards; amongst other aspects these performance standards govern chart data structure, minimum display requirements and minimum equipment specifications. Chart data used in an ECDIS must conform to the Electronic Navigational Chart (ENC) S-57, edition 3.0 specification and the performance standard for this was agreed by the International Hydrographic Organization (IHO) in February 1996. Any ENC must be issued on the authority of a government-authorized hydrographic office.

Back-up arrangements for ECDIS were agreed by the IMO in November 1996, becoming Appendix 6 to the Performance Standards and allowing ECDIS to be legally equivalent to the charts required under regulation V/20 of the 1974 SOLAS convention. It is an IMO requirement that the National Hydrographic Offices (NHOs) of Member Governments issue, or authorize the issue of, the ENCs, together with an updating service, and that ECDIS manufacturers should produce their systems in accordance with the Performance Standards. Other notable milestones leading to the ECDIS specification include the following.

- IHO Special Publication S-52 which specifies chart content and display of ECDIS. This includes appendices specifying the issue, updating and display of ENC, colour and symbol specification. The IHO Special Publication S-52 was produced in December 1996.
- IEC International Standard 61174. In this publication the International Electrotechnical Commission describes methods of testing, and the required test results, for an ECDIS to comply with IMO requirements. The standard was officially published in August 1998 and is to be used as the basic requirement for type approval and certification of an ECDIS which complies with the IMO requirements.

Some ECDIS definitions are summarized below.

- **Electronic Chart Display and Information System (ECDIS)** means a navigation system which, with adequate back-up arrangements, can be accepted as complying with the up-to-date chart required by regulation V/20 of the 1974 SOLAS Convention, by displaying selected information from a System Electronic Navigational Chart (SENC) with positional information from navigational sensors to assist the mariner in route planning, route monitoring and displaying additional navigational-related information if required.
- **Electronic Navigational Chart (ENC)** is the database, standardized as to content, structure and format, issued for use with ECDIS on the authority of government-authorized hydrographic offices.
- **System Electronic Navigational Chart (SENC)** is a database resulting from the transformation of the ENC by ECDIS for appropriate use, updates to the ENC by appropriate means, and other data added by the mariner.
- **Standard Display** means the SENC information that should be shown when a chart is first displayed on an ECDIS. The level of information provided for route planning and route monitoring may be modified by the mariner.
- **Display Base** means the level of SENC information which cannot be removed from the display, consisting of information which is required at all times in all geographical areas and all circumstances.

The basic ECDIS requirements can be summarized as follows.

- **ENC data.** This is to be supplied by government-authorized hydrographic offices and updated regularly in accordance with IHO standards.
- **Colours/Symbols.** These must conform to the specification outlined in IHO Special Publication S-52. Symbol size and appearance are specified and the mariner should be able to select colour schemes for displaying daylight, twilight and night-time conditions.
- **Own Ship's Position.** The ECDIS should show own ship's position on the display. Such a position is the result of positional input data received from suitable sensors and should be continuously updated on the display.
- **Change Scale.** The use of zoom-in and zoom-out should allow information to be displayed using different scales. ECDIS must display a warning if the information shown is at a scale larger than that contained in the ENC or if own ship's position is produced by an ENC at a larger scale than that shown by the display.
- **Display Mode.** The mariner should be able to select a 'north-up' or 'course-up' mode. Also the display should be able to provide true motion, where own ship symbol moves across the display, or relative motion where own ship remains stationary and the chart moves relative to the ship.
- **Safety Depth/Contour.** The mariner can select safety depth, whereby all soundings less than or equal to the safety depth are highlighted, or safety contour whereby the contour is highlighted over other depth contours.
- **Other Navigational Information.** Radar or ARPA data may be added to the display.

As emphasized earlier, one of the key requirements for ECDIS is to assist the user to plan a route and monitor the route while under way. This and other functions are listed below.

- **Route Planning.** The mariner should be able to undertake the planning of a suitable route, including the provision of waypoints which should be capable of being amended as required. It should be possible for the mariner to specify a limit of deviation from the planned route at which activation of an automatic off-track alarm occurs.
- **Route Monitoring.** ECDIS should show own ship's position when the display covers the area involved. The user should be able to 'look-ahead' while in this mode but be able to restore own ship's position using a 'single operator action'. The data displayed should include continuous indication of ship's position, course and speed and any other information, such as time-to-go, past track history etc., considered necessary by the user. Indication/alarms should feature using parameters set by the mariner.
- **Indication/Alarm.** ECDIS is required to give information about the condition of the system or a component of the system; an alarm should be provided when a condition requires urgent attention. An indication could be visual whereas an alarm could be visual but must also be audible. Indications should include, among others, information overscale, different reference system, route planned over a safety contour etc. Alarms should include, among others, system malfunction, deviation from route, crossing safety contour etc.
- **Record of Voyage.** ECDIS must be capable of recording the track of an entire voyage with timings not exceeding 4-hourly intervals. Also ECDIS should keep a record of the previous 12 h of a voyage; such a record should be recorded in such a way that the data cannot be altered in any way. Also during the previous 12 h of a voyage ECDIS must be capable of reproducing navigational data and verifying the database used. Information such as own ship's past track, time, position, speed and heading and a record of official ENC data used, to include source, edition, date, cell and update history, should be recorded at 1-min intervals.

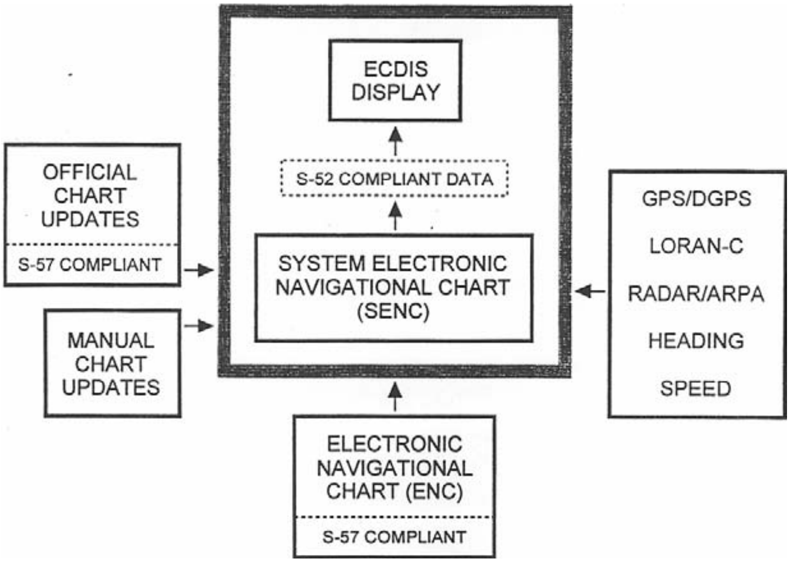


**Figure 7.4** Block diagram of an ECDIS. (Reproduced courtesy of Warsash Maritime Centre.)

- **Back-up Arrangements.** This is required in case of an ECDIS failure. The back-up system should display in graphical (chart) form the relevant information of the hydrographic and geographic environment necessary for safe navigation. Such a system should provide for route planning and monitoring. If the back-up system is electronic in form it should be capable of displaying at least the information equivalent to the standard display as defined by the performance standard.

A block diagram of an ECDIS is shown in Figure 7.4.

The production of ENCs is proceeding but it is a lengthy and costly business and it is likely that widespread coverage will not be available for some time and certain regions may never be covered at all. Because of the delay likely in implementing ECDIS, hydrographic offices around the world have proposed an alternative official chart solution that uses the raster chart and is known as the Raster Chart Display System (RCDS).

### 7.3.2 Raster Chart Display System (RCDS)

This is a system capable of displaying official raster charts that meets the minimum standards required by an appendix to the ECDIS Performance Standard. The raster nautical chart (RNC) is a digital facsimile of the official paper chart and provides a geographically precise, distortion-free image of the paper chart.

The IHO proposed a raster chart standard that ‘should form a part of the ECDIS performance standards where it would logically fit’. This was approved by the IMO’s Maritime Safety Committee in December 1998 as a new appendix to the existing ECDIS Performance Standard, entitled ‘RCDS Mode of Operation’. It is now permissible for ECDIS to operate in RCDS mode using official RNCs when ENCs are not available. The use of ECDIS in RCDS mode can only be considered providing there is a back-up folio of appropriate up-to-date paper charts as determined by national administrations.

Raster charts for these systems have been developed in recent years by major hydrographic offices and include the British Admiralty Raster Chart Service (ARCS) and the NOAA's BSB raster chart. The United States started raster scanning in 1991 and evaluated a prototype of the scheme in 1992. NOAA began converting its charts to raster format in 1993 and completed the task in 1994. The United Kingdom Hydrographic Office (UKHO) started the raster scanning of its Admiralty charts in 1994 and shipboard trials of ARCS began in 1995; the service becoming commercial in 1996. Other nations have also developed their own RCDS charts.

Raster charts are offered as an interim measure while awaiting the arrival of the ENCs and are designed to offer a performance specification that closely follows that of the ENCs and includes important requirements such as:

- continuous chart plotting and chart updating
- at minimum, the same display quality as the hydrographic office paper chart
- extensive checking, alarms and indicators relating to the integrity and status of the system
- route planning and voyage monitoring.

The IMO has drawn mariners' attention to the fact that the RCDS mode of operation lacks some of the functionality of ECDIS. Some of the limitations of RCDS mode compared to ECDIS mode include the following.

- The raster navigational chart (RNC) data will not itself trigger automatic alarms although some alarms can be generated by the RCDS from information inserted by the user.
- Chart features cannot be altered or removed to suit operational requirements. This could affect the superimposition of radar/ARPA.
- It may not be possible to interrogate RNC features to gain additional information about charted objects.
- An RNC should be displayed at the scale of the paper chart and RCDS capability could be degraded by excessive use of the zoom facility.
- In confined waters the accuracy of the chart data may be less than that of the position fixing system in use. ECDIS provides an indication in the ENC that permits determination of the quality of the data.

### 7.3.3 Dual fuel systems

Because of the adoption by the IMO of the amendments to the performance standards for ECDIS to include the use of RCDS, an ECDIS is now able to operate in two modes:

- ECDIS mode when ENC data is used
- RCDS mode when ENC data is unavailable.

Thus the dual fuel system is one that is either an ECDIS or RCDS depending on the type of chart data in use. At the present there are only few ENCs so the ability to use ECDIS is restricted. RNCs are plentiful and can provide two vital functions:

- provide official electronic chart coverage for areas not covered by ENCs
- provide link coverage between the ENCs that are available.

### 7.3.4 Electronic chart systems (ECS)

Where a system does not conform to either ECDIS or RCDS performance standards it is classified as an ECS system. There are no official performance standards for this system. The IMO had been considering the production of advisory guidelines but at the 1998 meeting of the IMO Navigation Safety Subcommittee it was decided that guidelines for ECS were not necessary and the matter will not be pursued further. As a general rule, a system is an ECS if:

- it uses data which is not issued under the authority of a government-authorized hydrographic office
- vector chart data is not in S-57 format
- the system does not meet the standards of either ECDIS or RCDS performance standards.

An ECS may not be used as a substitute for official paper charts, and ships fitted with an ECS are legally required to carry suitable up-to-date official paper charts. Examples of ECS include radar systems incorporating video maps, stand-alone video plotters and all systems while using commercial raster charts and vector charts systems.

## 7.4 Chart accuracy

Any chart is only as good as the original survey data allows and the accuracy with which that data is recorded on the chart by the cartographer. A navigational chart is referenced to two data: horizontal, for latitude and longitude; and vertical, for depth and height.

Since the beginning of mapmaking, local maps were based on the earth's shape in that area and, since the earth is not a perfect sphere, the shape does vary from location to location. Figure 7.5 shows a representation of a vertical slice through the earth. The diagram shows an uneven surface to the

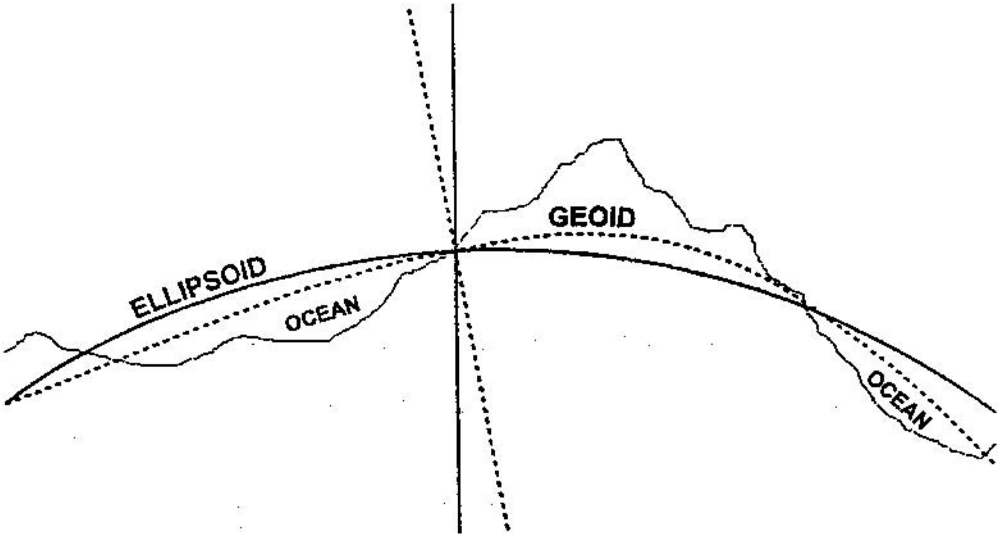


**Figure 7.5** View of the earth's surface showing the geoid and ellipsoid. (Reproduced courtesy of Warsash Maritime Centre.)

earth, a dotted line representing a geoid and a solid line representing an ellipsoid. The geoid represents a surface with equal gravity values and where the direction of gravity is always perpendicular to the ground surface. For mapping purposes it is necessary to use a geodetic datum which is a specifically orientated reference ellipsoid. The surface of a geoid is irregular while that of an ellipsoid is regular.

Many different ellipsoids have been used to represent the best fit to the geoid in a particular area. The use of an ellipsoid for positional calculations must first be referenced to the geoid and that relationship defines what is known as a datum. The accuracy of a particular datum may be fine for the local area for which it was intended but the accuracy may suffer as the deviation from that area increases. There are scores of different data such as Ordnance Survey Great Britain 1936 (OSGB36), the European Datum 1950 (ED50), the Australian Geodetic System 1984, North American Datum 1983 (NAD83), etc. Charts drawn for a particular area therefore may contain datum information that is localized.

The use of satellite systems has involved the use of a global datum and GPS uses the World Geodetic System 1984 (WGS-84) which uses a model of the complete earth. The ellipsoid for this system is centred on the Earth's centre of mass and, over the earth as a whole, is a better fit to the geoid than other ellipsoids, although the local datum may give a better fit within their own small area. Ideally all charts should be referenced to WGS-84 but this is not expected to occur for many years to come. Reasons for the delay include:

- the time necessary to replace current charts with new versions using WGS-84
- lack of data necessary to calculate datum shifts and, in some cases, the datum used for the chart is either unknown or poorly defined.

As far as the UKHO is concerned, about 20% of its charts are referenced to the WGS-84 datum, a further 40% use datum when the shift is known, while some 40% use unknown datum. When the shift to WGS-84 is known the UKHO charts have a 'Satellite Derived Positions' note that provides shift values in minutes of latitude and longitude which allows GPS-determined positions, referenced to WGS-84, to be correctly adjusted before they are plotted on the chart. Currently about 40% of the UKHO charts contain shift values.

Electronic chart systems using raster chart displays can use the datum shift values indicated in the 'Satellite Derived Positions' note on the chart to convert the WGS-84 co-ordinates to the local datum. The shift values are mean values for the area covered by the chart but the shift variation across the chart is within manual plotting tolerance at the scale of the chart and can be ignored. However, the quoted shift values on an adjacent chart could well be different.

For electronic chart systems using vector charts it is a requirement that the charts are referenced directly to WGS-84. Since so few official paper charts are referenced directly to WGS-84 it follows that vector chart producers must use a mathematical model to shift the data on certain charts to WGS-84. Users of the system should always check to see whether the official paper chart is referenced directly to WGS-84. If the official chart has a 'Satellite Derived Positions' note giving datum shift values then it could safely be assumed that errors introduced by the conversion to WGS-84 will be small at the scale of the official chart. If WGS-84 shift values do not appear on the paper chart it would suggest that the existing data is insufficient to establish accurate datum shifts and GPS-derived positions cannot be used with confidence.

With ECDIS and the use of ENCs, all references are to WGS-84 so there should be no problem with datum shifts. However, as discussed earlier, there could be a problem of geodetic datum shifts using paper charts, RNCs and privately produced vector charts if positional information is received based on one datum and such data is plotted on a chart which is based on another datum. Figure 7.6

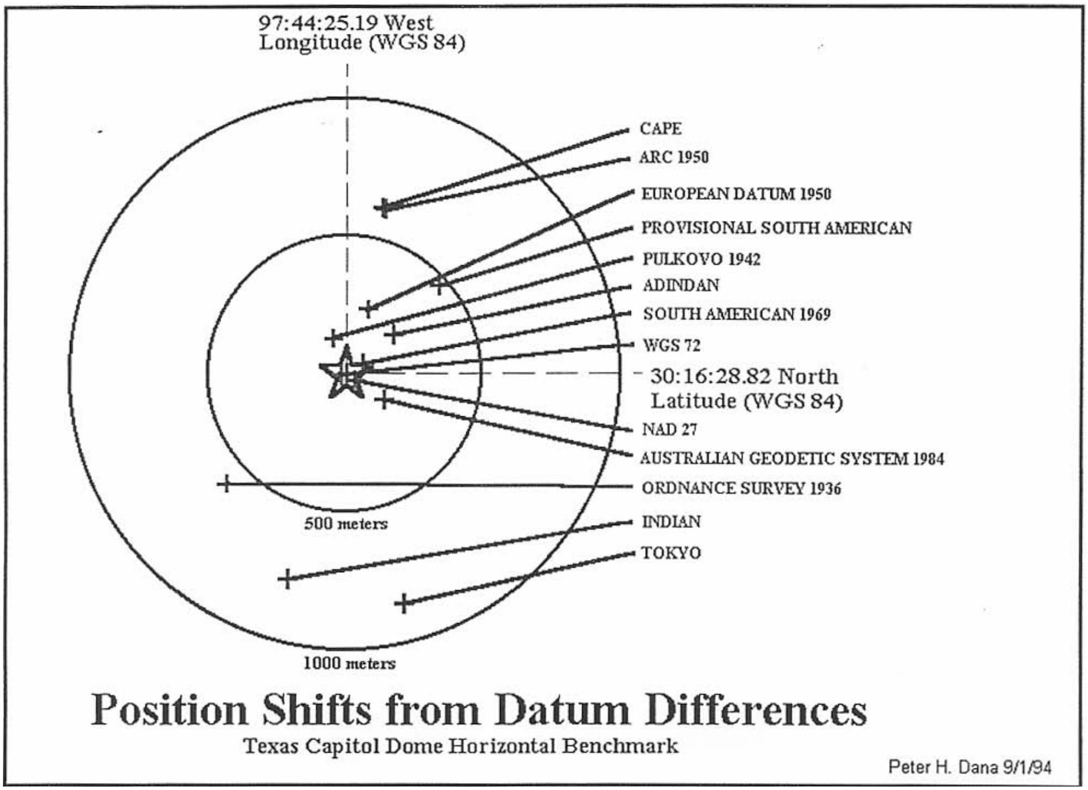


**Figure 7.6** World geodetic datums. (Reproduced courtesy of PC Maritime.)

illustrates the variation in latitude and longitude positions that could be derived for the same real location depending on the datum used.

Consider another example of datum differences in the English Channel. The Admiralty charts covering the English coastline are in OSGB36 whereas the Admiralty charts covering the French coastline are in ED50. The OSGB36 datum is used for charts covering the coastline of England, Wales and Scotland while the ED50 was developed for military mapping in Central Europe. UKHO charts covering both sides of the channel tend to be in OSGB36. Thus if an operator working in the channel plotted a position on an OSGB36 chart and then moved to a European 1950 chart without allowing for a datum shift, there will be a positional error as indicated in Figure 7.7.

In some regions of the world the difference between WGS-84 and the local datum can be quite large and this is illustrated in Figure 7.8.

The solution to the problem is obviously to obtain positional information in WGS-84 and to apply the published shift every time a change of paper chart is made. It must be remembered that GPS accuracy has tolerance values and any inaccuracy derived from GPS may be exacerbated by plotting charts of different datum. Most GPSs have built-in datum transformations so that the system can output positions in a local datum but this has certain disadvantages.

- Because there are no standards applicable to the transformation formulae, two different GPSs may use different formulae and give different results. The solutions produced are averaged over a wide

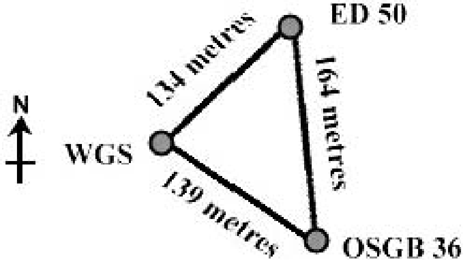


**Figure 7.7** An example of datum shift in the English Channel. (Reproduced courtesy of PC Maritime.)

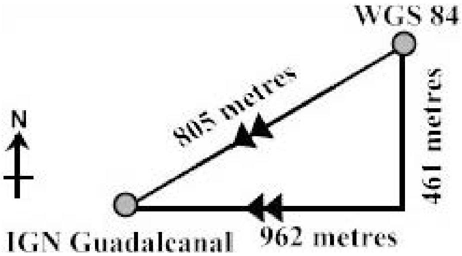


**Figure 7.8** An example of datum shift in the Pacific. (Reproduced courtesy of PC Maritime.)

area and any transformation error may range from, say 25 m to much more at the fringes of the area covered by the datum.

- It is difficult to ensure the GPS is switched to the correct datum every time a chart is changed.
- GPS positions may be fed simultaneously to other equipment, such as ARPA, autopilot etc., which expect to receive data in WGS-84 co-ordinates.
- Some GPSs apply the data transformation to all waypoint positions held in memory when a datum other than WGS-84 is selected for the display of positions.

It may be better to maintain the output of GPS in WGS-84. As stated earlier, for the UKHO paper charts, a shift from WGS-84 to the local datum is printed on the chart. Any figure printed on the chart indicates that the original survey has been referenced to WGS-84 and the published shifts can be used with confidence. If the chart contains no shift data then no referencing to WGS-84 has been made and any plotted positions made must be treated with caution because of possible shift errors.

An advantage of modern charts and the use of software is that the management of datum shifts can be automated. A system such as ARCS has the shift data included and thus an RCDS can keep track of the data of positions of all types, including vessel position and track, waypoints and any other overlaid points on the chart, and adjust them all to the local geodetic datum as required.

## 7.5 Updating electronic charts

As mentioned on page 228 with reference to the UKHO's ARCS system, updates for all charts affected by Notice to Mariners (up to about 200 a week) are generated, checked and placed on a weekly ARCS Update CD-ROM which includes temporary and preliminary notices. This provides error-free automatic corrections and provides cumulative updates with only the latest update CD-ROM required. The CD-ROMs are sent to chart agents who then send them to shipping companies as required.

NOAA provides continuous updating to all 1000 charts using information from the USCG, NIMA and the Canadian Hydrographic Service, and Maptech makes the necessary raster chart updates. Maptech uses modern technology to update only those parts of a chart identified as needing correction. This so-called ‘patch’ technique compares the existing chart file and its corrected counterpart on a pixel-by-pixel basis. A difference file is produced which can be manipulated so that it registers exactly with the existing raster file to which it applies. A raster chart can therefore be updated by displaying it, using the relevant CD-ROM, and using the patch file to alter the pixels on the old chart as necessary to incorporate the corrections.

The updating service became available on subscription in January 2000. Customers receive a weekly e-mail that contains a hot link to the update computer server. Clicking on the hot link begins the transmission of the update patches to the computer; the updates in the transmission are cumulative updates for all charts on a CD-ROM. Downloading takes from a few seconds to up to 5 min depending on the modem speed. Once the file reception is completed the charts may be corrected and stored on the computer’s hard drive. It is anticipated that dynamic updating should soon be available. With this technique the charts and patches are kept separate and the patch is applied to the chart in real time allowing the user to see the changes produced by the patch.

Dynamic patching is the preferred method under the international standards for ECDIS where it is a requirement that mariners should not change the original data files. It is expected that in future single chart updates may be made available rather than a complete CD’s worth and that the procedure will be extended to ENCs when they become available.

## 7.6 Automatic Identification System (AIS)

Automatic Identification System (AIS) is a shipborne transponder system capable of broadcasting continuously, using the VHF marine band, information about the ship. Such information could include:

- ship identification data, i.e. ship name, call sign, length, breadth, draught etc.
- type of cargo carried and whether it was hazardous in nature
- course and manoeuvring data
- position to GPS accuracy limits.

Such broadcast information would be capable of reception by other AIS-equipped ships and by shore sites such as Vessel Traffic System (VTS) stations within broadcast range. Data received by a ship or shore station could be relayed to an ECS and AIS targets could be displayed, with GPS or DGPS accuracy, with a velocity vector indicating speed and heading. By ‘clicking’ on a target, other information such as ship identification data etc. could be displayed. A typical AIS scenario is illustrated in Figure 7.9.

An AIS transponder system requires a GPS or DGPS receiver, a VHF transmitter, two VHF TDMA receivers, a VHF DSC receiver and a standard marine electronic communications connection to the ship’s display system. Position and timing information is derived from the GNSS (GPS) receiver. Information, such as ship’s heading, course and speed over ground, is normally broadcast using AIS but other information such as destination, ETA etc. could also be promulgated if available.

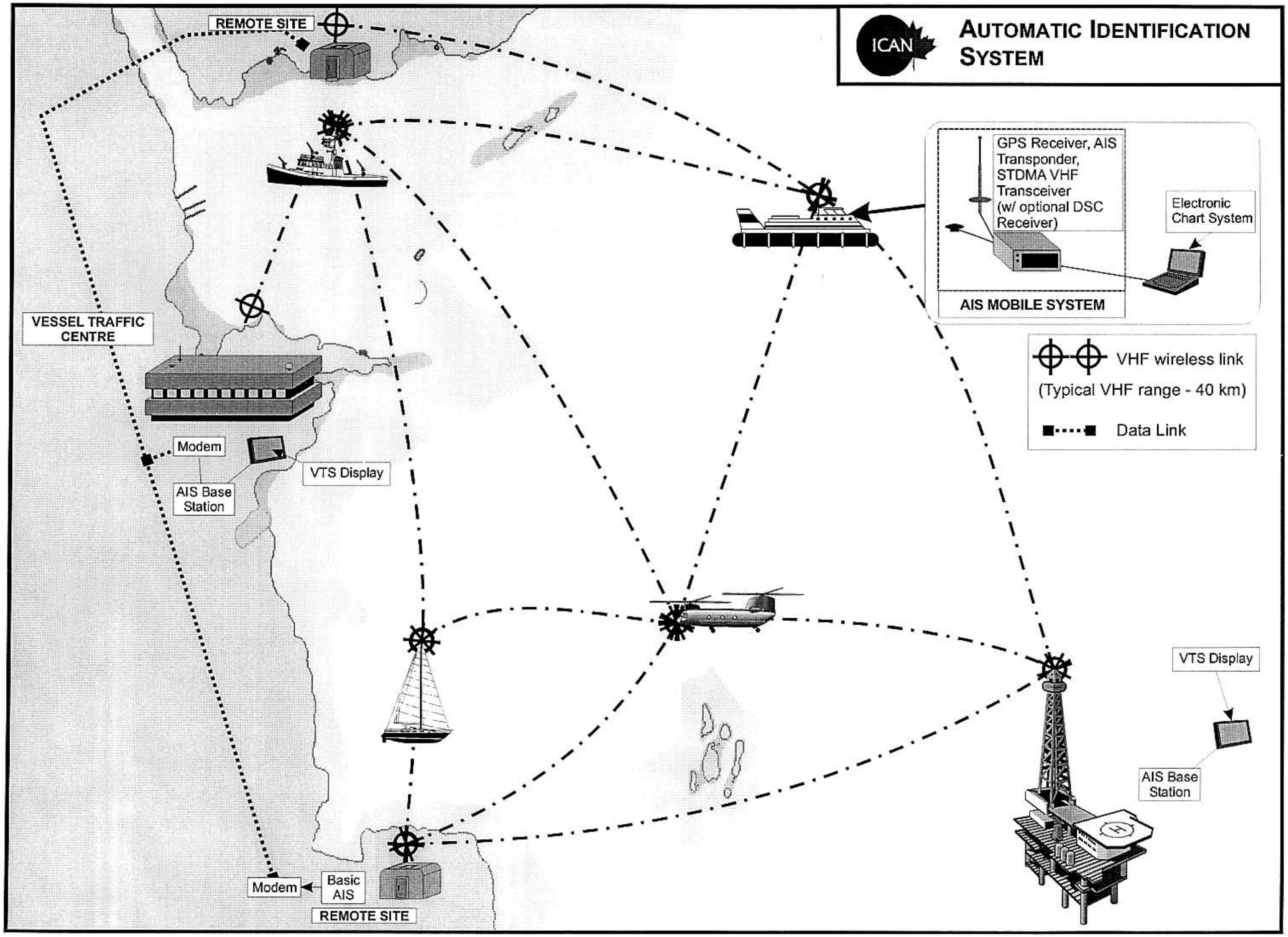


**Figure 7.9** Automatic Identification System (AIS). (Reproduced courtesy of ICAN.)

The AIS transponder transmits using 9.6 kbyte GMSK (Gaussian minimum shift keying) FM modulation over 25 kHz or 12.5 kHz channels using HDLC packet protocols. The channel bandwidth of 25 kHz is for use on the high seas and the 12.5/25 kHz channel bandwidth used as defined by the appropriate authority in coastal waters. There are two radio channels available for transmission/reception that minimize RF interference, provide increased capacity and allow channels to be shifted without loss of communication from other ships. The ITU has allocated frequencies with AIS channel 1 using 161.975 MHz (ch87B) and AIS channel 2 using 161.025 MHz (ch88B).

Each transponder self-allocates time slots for its position reports and such reports occur at time intervals that correspond to the traffic situation. This method of communication is known as self-organizing time division multiple access (SOTDMA). The SOTDMA broadcast mode allows the system to be overloaded by up to 500% while still providing nearly 100% communication capacity for ships within 10 nautical miles of each other in ship-to-ship mode. If system overload tends to occur, then targets at the longer ranges will tend to drop out of the system leaving only closer range targets, which are the ones of greater interest to the navigator. There are 2250 time slots established every 60 s for each AIS channel; this gives a time slot duration of 26.67 ms and as each slot has 256 bits the data transmission rate is 9600 bit $s^{-1}$.

AIS stations continuously synchronize with other stations to obviate any slot transmission overlap. Slot selection by an AIS station is randomized within a defined interval and triggered with a random timeout of between 0 and 8 frames. When a station changes its slot assignment the new location and associated timeout is pre-announced, thus allowing new stations to be received.

Although the AIS concept has been around for many years and trials have taken place at many geographical locations, there is still much work to be done to produce an internationally-agreed standard. Some of the detail of what has been achieved to date is listed below.

### *IMO Resolution MSC.74(69). Annex 3, Recommendation on Performance Standards for a Universal Shipborne Automatic Identification System (AIS)*

The 43rd session of the IMO Navigation Subcommittee, which met in July 1997, completed a draft performance standard on shipborne automatic identification systems (transponders). This performance standard describes the operational requirements for the device but does not define the telecommunications protocol the device must use. The 69th session of the IMO Maritime Safety Committee formally adopted the standard without change in May 1998.

A report from the Subcommittee on Safety of Navigation on its 45th session included the following items.

1 All ships of 300 gross tonnage and upwards (engaged on international voyages), cargo ships of 500 gross tonnage and upwards (not engaged on international voyages), and passenger ships, irrespective of size, shall be fitted with AIS, as follows:
  1.1 ships constructed on or after 1 July 2002;
  1.2 ships engaged on international voyages constructed before 1 July 2002;
  1.2.1 in the case of passenger ships irrespective of size and tankers of all sizes, not later than 1 July 2003;
  1.2.2 in the case of ships, other than passenger ships and tankers, of 50000 gross tonnage and upwards, not later than 1 July 2004;
  1.2.3 in the case of ships, other than passenger ships and tankers, of 10000 gross tonnage and upwards but less than 50000 gross tonnage, not later than 1 July 2005;
  1.2.4 in the case of ships, other than passenger ships and tankers, of 3000 gross tonnage and upwards but less than 10000 gross tonnage, not later than 1 July 2006;

- 1.2.5 in the case of ships, other than passenger ships and tankers, of 300 gross tonnage and upwards but less than 3000 gross tonnage, not later than 1 July 2007; and
- 1.3 ships not engaged on international voyages constructed before 1 July 2002, not later than 1 July 2008.

2 The Administration may exempt ships from the application of the requirements of this paragraph when such ships will be taken permanently out of service within two years after the implementation date specified in paragraph 1.

3 AIS shall:

- 3.1 provide automatically to appropriately equipped shore stations, other ships and aircraft information, including the ship's identity, type, position, course, speed, navigational status and other safety-related information;
- 3.2 receive automatically such information from similarly fitted ships;
- 3.3 monitor and track ships; and
- 3.4 exchange data with shore-based facilities, the requirements of this paragraph shall not be applied to cases where international agreements, rules or standards provide for the protection of navigational information. AIS shall be operated taking into account the guidelines adopted by the Organization.

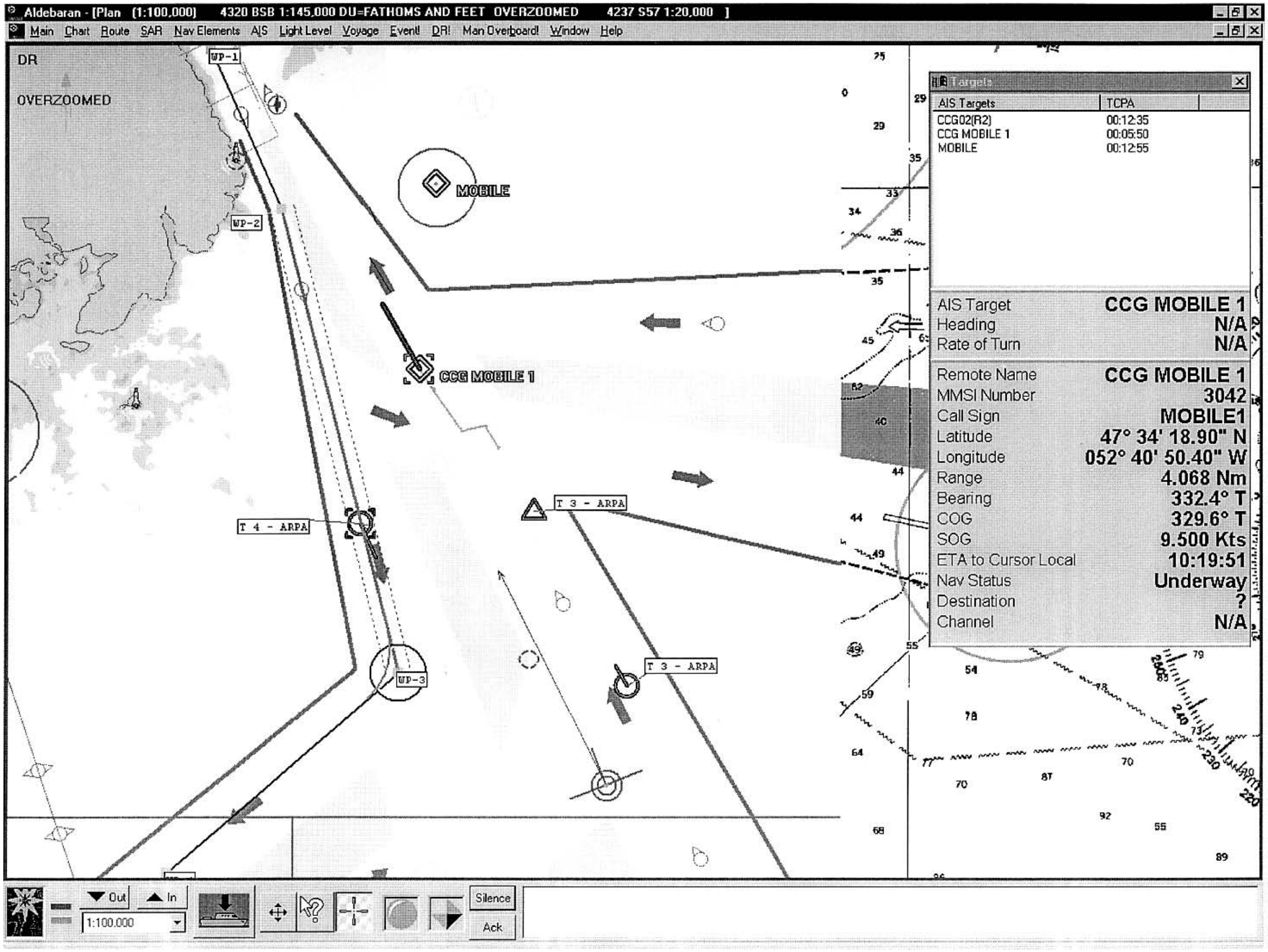
### *ITU-R Recommendation M.1371, Technical Characteristics for a Universal Shipborne Automatic Identification System Using Time Division Multiple Access in the Maritime Mobile Band*

The International Telecommunications Union Sector for Radiocommunication (ITU-R) met in March 1998 to define the technology and telecommunications protocol for this device. The draft recommendation completed by Working Party 8B was approved by Study Group 8, which met in July 1998. The recommendation was formally adopted in November 1998 and the publication is now available for a fee (see website www.itu.org). The International Association of Lighthouse Authorities (IALA) has been the main organization co-ordinating the development of the Universal AIS Transponder and a revision of this standard is being prepared by IALA for submission to the ITU-R Working Party 8B in October 2000. If adopted it will become ITU-R Recommendation M.1371–1.

### *IEC Standard 61993–2 on AIS*

In July 1998, the International Electrotechnical Commission TC80/WG8-U.AIS started work on the performance, technical, operational and testing standard for the Universal AIS Transponder. The working group is expected to meet regularly and complete its work during the year 2000 with an expected publication date for the standard of December 2001. This standard will supersede IEC Standard 61993–1 on digital selective calling AIS transponders. This new standard will define testing and interfacing requirements for AIS systems. Commercially-produced systems should meet all the three standards described above.

ICAN have developed an AIS module which is an add-on to their 'Aldebaran' Electronic Charting System. The module has been developed for use with Saab TransponderTechs AIS hardware for which ICAN is the exclusive Canadian agent. The AIS module enables ICAN's ECS to display broadcast AIS information on screen in InfoPanels and as overlays. A typical screen display with this feature is shown in Figure 7.10.



| AIS Targets | TCPA |
|---|---|
| CCG02(R2) | 00:12:35 |
| CCG MOBILE 1 | 00:05:50 |
| MOBILE | 00:12:55 |

| | |
|---|---|
| AIS Target | CCG MOBILE 1 |
| Heading | N/A |
| Rate of Turn | N/A |
| Remote Name | CCG MOBILE 1 |
| MMSI Number | 3042 |
| Call Sign | MOBILE1 |
| Latitude | 47° 34' 18.90" N |
| Longitude | 052° 40' 50.40" W |
| Range | 4.068 Nm |
| Bearing | 332.4° T |
| COG | 329.6° T |
| SOG | 9.500 Kts |
| ETA to Cursor Local | 10:19:51 |
| Nav Status | Underway |
| Destination | ? |
| Channel | N/A |

**Figure 7.10** Display showing AIS target information. (Reproduced courtesy of ICAN.)

Features of the ICAN AIS module are as follows.

1 AIS Target Monitoring
   - Unlimited on-screen AIS targets.
   - AIS Tracking InfoBox sorted based on TCPA and RCPA.
   - Targets can be individually centred on screen.
   - Single target activation.
   - Messages can be addressed or sent via broadcast as binary or ASCII data on specific channels.
   - Automatic (scheduled) and manual data transmissions.
   - Binary transmissions include: man overboard, ARPA, markers and points of interest (SAR, waypoints, routes and zones).
   - Displayed AIS transponder channels.
   - On screen CPA display.
   - Alarms and indications based on configurable CPA properties.

2 Long Range AIS Monitoring
   - Microsoft's MAPI (Mail Application Programming Interface) based mail set-up.
   - Office and remote monitoring through Inmarsat terminal or service provider.
   - Filtered sender information.
   - Multiple e-mail address transmissions (single Inmarsat message).
   - Configurable gateway formats.

3 AIS Module Configuration
   - Remote target properties (shape, labels).
   - Name, call sign, ship type, MMSI, IMO no., draught, trip, destination and ETA to destination.
   - Own ship transponder transmission information (Nav sensor, antenna location, UTC date time and channel designation).
   - Distinguishable transponder characteristics (R2 vs R3 labelling).
   - ECS back-up positioning device (transponder GPS).
   - Transponder GNSS status.
   - Closest point of approach (time and range based).
   - Channel polygons.

4 Data logging and Distribution
   - Unfiltered logging of serial inputs, including AIS transponder information.
   - File-based distribution of logged data.
   - TCP/IP distribution of serial inputs.
   - Playback of recorded data.

5 ICAN ECS Environment
   - Seamless display of charts of S-57, NTX, BSB and MRE formats (other formats in development include ARCS and CM-93).
   - Point-of-interest feature allows constant update of range/bearing to any point, marker or waypoint (station keeping).
   - Ability to add other software modules including high resolution radar overlay, useful for coastline mapping (scanner up to 120 rpm, 8-bit radar image, raw radar data recording capable).

Information on this AIS module and other useful products offered by ICAN are available on their website www.ican.nf.net.

## 7.7 Navmaster Electronic Navigation System

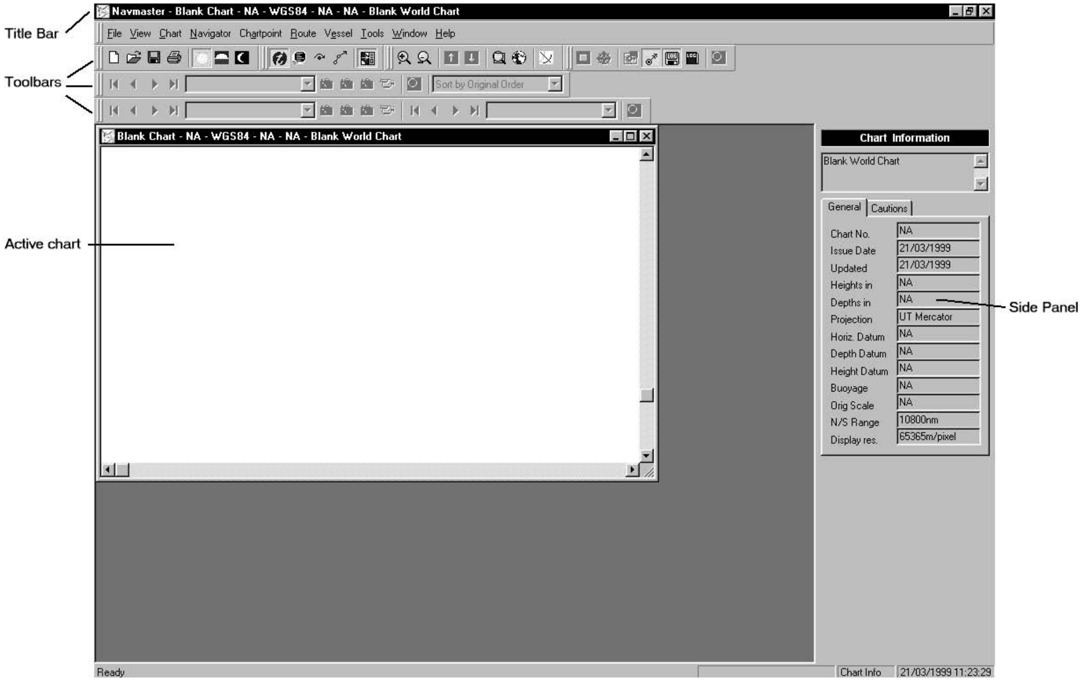
There are a multitude of suppliers of software suitable for implementing an electronic navigation system, requiring only the hardware and suitable electronic charts to produce an ECDIS or an ECDIS in RCDS mode. The 'Navmaster Professional' from PC Maritime of Plymouth, UK is used as a basis for showing how the software can assist the navigator in passage planning, position logging and navigation management, providing as it does a continuous display of vessel positions received from GPS and plotted on official electronic charts. The minimum system requirements for Navmaster are: a computer operating with a Pentium 133, or better, processor; Windows 95/98/NT/2000; 10-Mbyte hard drive for minimum installation; CD-ROM and floppy-disk drives; 32-Mbyte RAM; and a monitor with 800 × 600 resolution with 256 colours or more. Input/output requirements are one serial and one parallel port. The software is supplied on a CD-ROM. The system uses electronic chart data, which is the copyright of various national hydrographic offices; chart data is protected by a security key that allows access to the charts only via a user PIN number.

Once the software has been loaded into the computer then starting with Navmaster the display will be similar to the one shown in Figure 7.11.

The toolbars and side panels can be moved around the screen, hidden and displayed as required. The main window contains the following.

### *Title bar*

The title bar displays the program control icon, the activation or active image title and the standard minimize/maximize/restore/close buttons.



Blank World Chart

Chart No. NA
Issue Date 21/03/1999
Updated 21/03/1999
Heights in NA
Depths in NA
Projection UT Mercator
Horiz. Datum NA
Depth Datum NA
Height Datum NA
Buoyage NA
Orig Scale NA
N/S Range 10800nm
Display res. 65365m/pixel

21/03/1999 11:23:29

**Figure 7.11** The Navmaster start-up window. (Reproduced courtesy of PC Maritime.)

**Figure 7.12** Navmaster menu bar headings. (Reproduced courtesy of PC Maritime.)

### *Menu bar*

The menu bar displays the menu headings as shown in Figure 7.12. These are as follows.

- File. Contains standard menu commands for file management, printing and workspace, opening charts, and opening and saving chartpoint and route databases.
- Edit. Provides standard menu commands.
- View. Provides menu commands to select modes of operation, turn on or off the toolbars, side panel and status bar.
- Chart. Provides menu commands to change the chart display, install chart permits and updates, set the location of charts and updates and set chart-related options.
- Navigator. Provides menu commands to: turn position plotting on or off and set DR parameters; turn position logging on or off and make log entries; upload routes and waypoints to GPS; access diagnostic windows for equipment interfacing; open the Autoscroll monitor window; set position and navigation-related options.
- Target. Provides menu commands to Activate/Deactivate ARPA and Tender tracking and set related options.
- Chartpoint. Provides menu commands related to chartpoints.
- Route. Provides menu commands related to routes.
- Vessel. Provides menu commands to enter vessel information for use when calculating plans.
- Tools. Provides menu commands to display tidal atlas and activate the Range and Bearing tool, customize toolbars and set workspace and tidal atlas options.
- Window. Provides menu commands to manipulate windows.
- Help. Provides Help and information on obtaining technical support.

### *Toolbars*

The toolbars provide buttons that access some of the frequently used commands in the menus. If a command is unavailable, its button appears greyed-out. Toolbars and their button functions are shown in Figure 7.13.

### *Side panels*

These panels represent each of the main functions of the Navmaster system, i.e. monitoring position, storing chartpoints and creating and calculating routes. Switching between functions is achieved by pressing a button on the display toolbar or by selecting an item in the View menu.

**1 Monitor mode.** In this mode it is possible to monitor and plot the vessel's passage. The screen consists of three main areas: the chart area, the side panel and the toolbars. The chart area provides a view of the current chart, which may be manipulated as required. A typical side panel in monitor mode is shown in Figure 7.14.

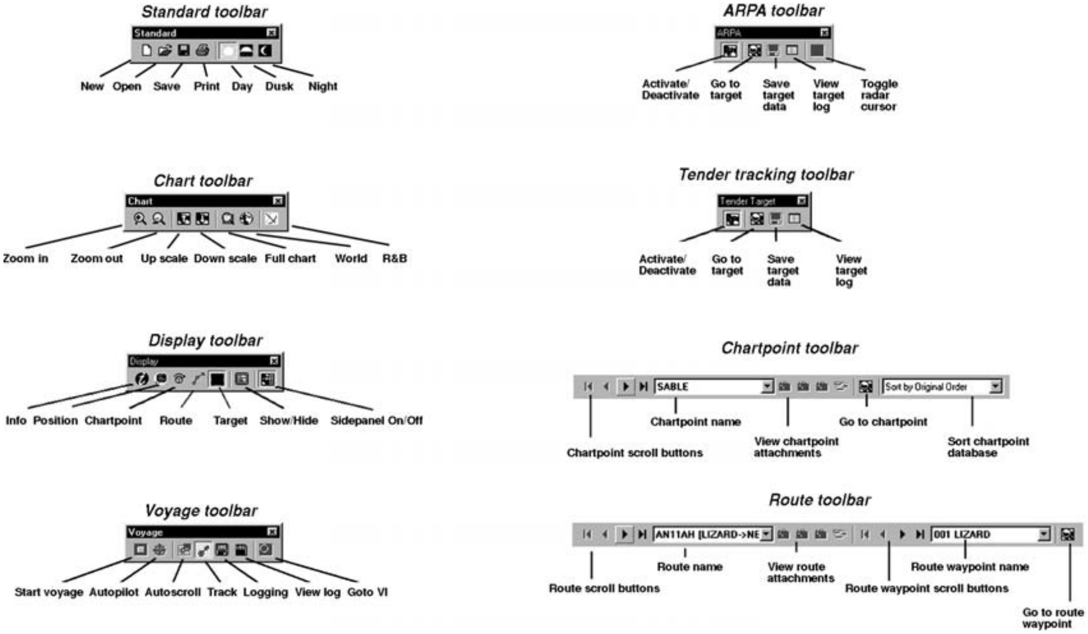
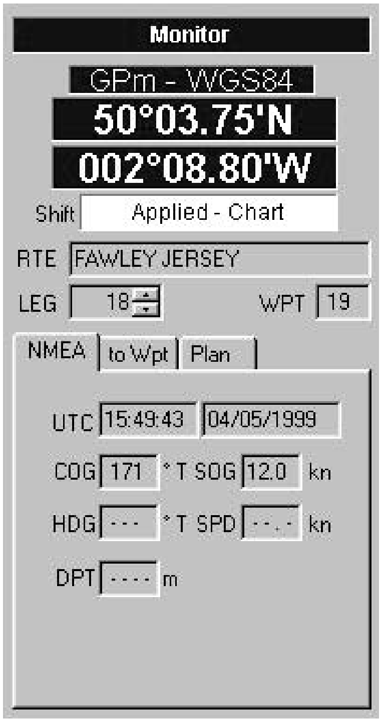


**Figure 7.13** Toolbars used in the Navmaster display. (Reproduced courtesy of PC Maritime.)



GPm - WGS84
50°03.75'N
002°08.80'W
Shift Applied - Chart
RTE FAWLEY JERSEY
LEG 18 WPT 19
UTC 15:49:43 04/05/1999
COG 171 °T SOG 12.0 kn
HDG --- °T SPD --.- kn
DPT ---- m

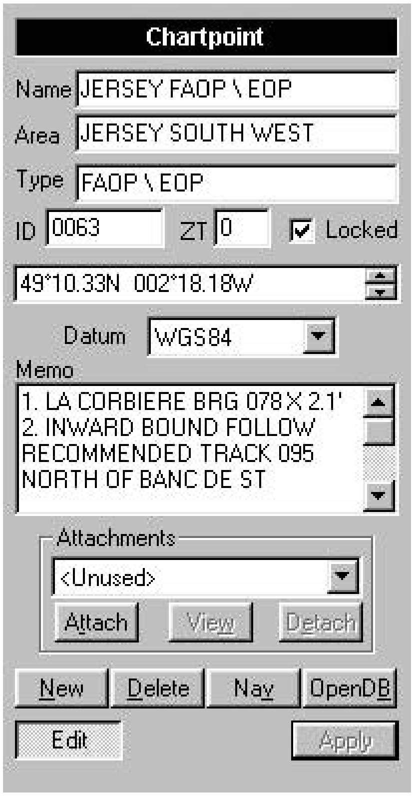
**Figure 7.14** Navmaster monitor mode side panel. (Reproduced courtesy of PC Maritime.)

The panel repeats the position obtained from the GPS, provides information on any datum shift that has been applied, and displays the current route name and active leg. Three tabs provide further information:

| | |
|---|---|
| NMEA | repeats information from electronic instruments; |
| to Wpt | provides calculated information from current position to the next waypoint in the route; |
| Plan | repeats information for the leg from the passage plan if one has been calculated. |

The Autocheck box activates/deactivates automatic leg advance.

**2 Chartpoint mode.** In this mode it is possible to add, delete, edit or save chartpoints. A chartpoint is the latitude and longitude of a geographical position stored in a database; a chartpoint on a chart is shown as a blue circle. Each chartpoint has database fields which allow the user to add other information which may be of assistance. Any number of chartpoint databases can be created and each database can contain any number of chartpoints. To enter chartpoint mode, the chartpoint button on the display toolbar (see Figure 7.13) is pressed and, provided side panel display is activated, the chartpoint side panel will be displayed (see Figure 7.15). The panel provides information about the current chartpoint. Each field within the panel can be edited.



Chartpoint

Name JERSEY FAOP \ EOP

Area JERSEY SOUTH WEST

Type FAOP \ EOP

ID 0063 ZT 0 ☑ Locked

49°10.33N 002°18.18W

Datum WGS84

Memo

1. LA CORBIERE BRG 078 X 2.1'
2. INWARD BOUND FOLLOW
RECOMMENDED TRACK 095
NORTH OF BANC DE ST

Attachments <Unused>

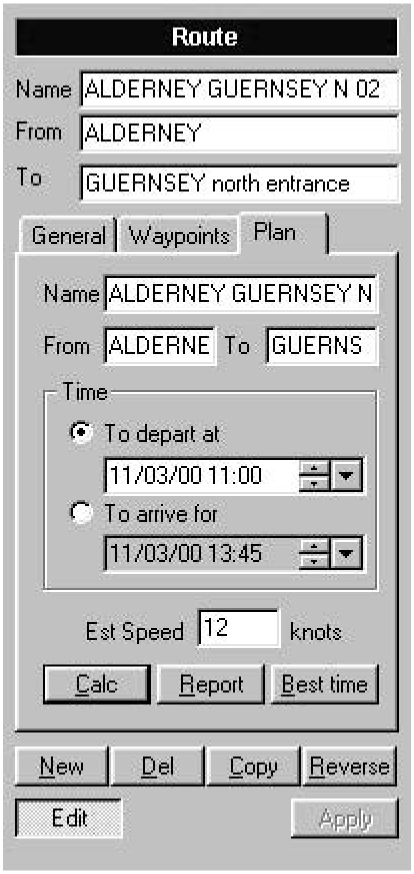
**Figure 7.15** Navmaster chartpoint mode side panel. (Reproduced courtesy of PC Maritime.)

Navmaster stores chartpoints in WGS-84 co-ordinates where possible and, provided a selected chartpoint is on the currently selected chart, it is possible to view and edit the chartpoint to match the local chart datum. A chartpoint can simply be used as a marker on a chart or, if used to indicate points on a route, they are known as waypoints.

**3 Route mode.** This mode enables the user to create new routes, edit existing routes and copy or reverse routes. A route is a sequence of waypoints built up from previously stored chartpoints or created by clicking on a chart. The route is drawn on the chart for evaluation and possible amendment. Routes are stored in databases and there can be many routes stored.

Route mode can be accessed by pressing the route button on the display toolbar (see Figure 7.13). The route side panel will be displayed, provided the side panel is switched on. Routes may be created, and edited, using waypoints from a chartpoint database or by drawing the route directly on the screen chart, or by a combination of both methods. Whatever route method is used, each waypoint in the route is inserted into the box on the route side panel and a line will connect the route waypoints on the chart. This line can be adjusted depending on whether the user adds, deletes or moves waypoints using the route side panel.

Routes are stored in a route database and any number of route databases can be created, containing any number of routes. A typical side panel in route mode is shown in Figure 7.16.



Route

Name ALDERNEY GUERNSEY N 02

From ALDERNEY

To GUERNSEY north entrance

Name ALDERNEY GUERNSEY N

From ALDERNE To GUERNS

Time

● To depart at

11/03/00 11:00

○ To arrive for

11/03/00 13:45

Est Speed 12 knots

**Figure 7.16** Navmaster route mode side panel. (Reproduced courtesy of PC Maritime.)

Three tabs provide further information:

| | |
|---|---|
| General | enables the user to enter and display a textual note relating to the route; |
| Waypoints | lists the waypoints in the route and provides a means to select waypoints for amendment or deletion or to locate a new waypoint; |
| Plan | gives the ability to calculate a passage plan based on the route which the user can print or view on the screen. |

Other side panels, which are available but not illustrated, are Target Tracking, which provides information on ARPA and Tender targets, and Information which gives information on the selected chart.

### 7.7.1 Installing charts

Navmaster supports the UKHO ARCS and the Australian Hydrographic Office Seafarer charts and will support ENC charts by the end of 2000. To install these charts the user needs:

- the floppy disk containing a licence file and chart permit file
- one (or more) chart CD-ROM
- one update CD-ROM with the latest chart corrections.

Each chart CD-ROM contains all the charts available for a particular region. A chart permit is a code that unlocks a specific chart. Charts can be installed from the chart permit disk or by entering the chart permit number manually. The user PIN number must be entered before a chart can be loaded or installed. When Navmaster loads a chart it also applies any chart updates at that time. A chart can be displayed without its update but a warning will be displayed indicating the fact that corrections are missing.

The ARCS or Seafarer chart is supplied as two, independent images namely a low-resolution (LR) image and a high-resolution (HR) image. The LR image provides an overview of the chart while the HR image is the one recommended for navigation and is updated with Notice to Mariner corrections. Navmaster provides further zooming in and out of the LR and HR images to give five levels of display for each chart.

The chart can be manipulated so that it is centred on a selected cursor position and the chart can be panned by using the scroll bars at the sides of the chart window.

**Table 7.6** The five levels of display for each chart

| *Resolution* | *Zoom level* | *Warning* |
|---|---|---|
| Low resolution | Zoom out (LR-out) | Underscale |
| | Normal (LR) | Underscale |
| High resolution | Zoom out (HR-out) | Underscale |
| | Normal (HR) | None |
| | Zoom in (HR-in) | Overscale |

### 7.7.2 Using Navmaster

When using Navmaster the recommended sequence to follow is:

- create chartpoints
- create a route
- calculate a plan
- monitor by plotting track, viewing data in the Navigation Monitor panel and comparing progress with the plan.

Navmaster is a multi-window application. Charts, the log, waypoint lists etc. all have their own window and windows can be tiled, cascaded or kept in the background as required. Turning on Autoscroll opens a dedicated window which displays the vessel's position in the centre of a chart. For safety reasons the Autoscroll window cannot be minimized so that the user is fully aware of the vessel's position. However, the window can be resized to allow more room for other charts, or it can be covered by a maximized window.

If the Autoscroll window is closed then Autoscroll is turned off. The remaining windows give complete flexibility to organize the charts to suit the task in hand. For example, new chart windows could be opened to:

- look ahead by displaying the vessel's position on a smaller scale chart than the Autoscroll chart
- view charts for other segments of the route
- view the approaches or harbour charts for intended destination
- plan new routes or chartpoints.

While the above is going on it is still possible to view a continuously updating vessel position on the largest scale chart available.

The maximum number of chart windows that can be opened is limited to three plus the Autoscroll window. The number can be increased but the default value of three is chosen to prevent users inadvertently opening too many windows.

The Chart Information panel in Navmaster displays information on the selected chart and indicates the following.

- Chart Description. The Hydrographic Office description of the chart.
- Chart No. The Hydrographic Office chart number.
- Orig Scale. The scale of the paper version of the chart.
- Edition Date. The date the chart was first issued.
- Updated. The date of the last update.
- Heights In. The units of height used.
- Depths In. The units of depth used.
- Projection. The type of projection used in the production of the chart.
- Horiz Datum. The geodetic datum of the chart. EG OSGB36 – The Ordnance Survey of Great Britain (1936) datum.
- Depth Datum. The datum to which depths are referred.
- Height Datum. The datum to which heights are referred.
- Buoyage. The buoyage system in use on the chart.
- N/S Range. The vertical distance in nautical miles of the portion of the chart currently displayed in the chart window.
- Display Resolution. The number of metres represented by each pixel on the computer display, which will alter depending on the zoom level of the chart.

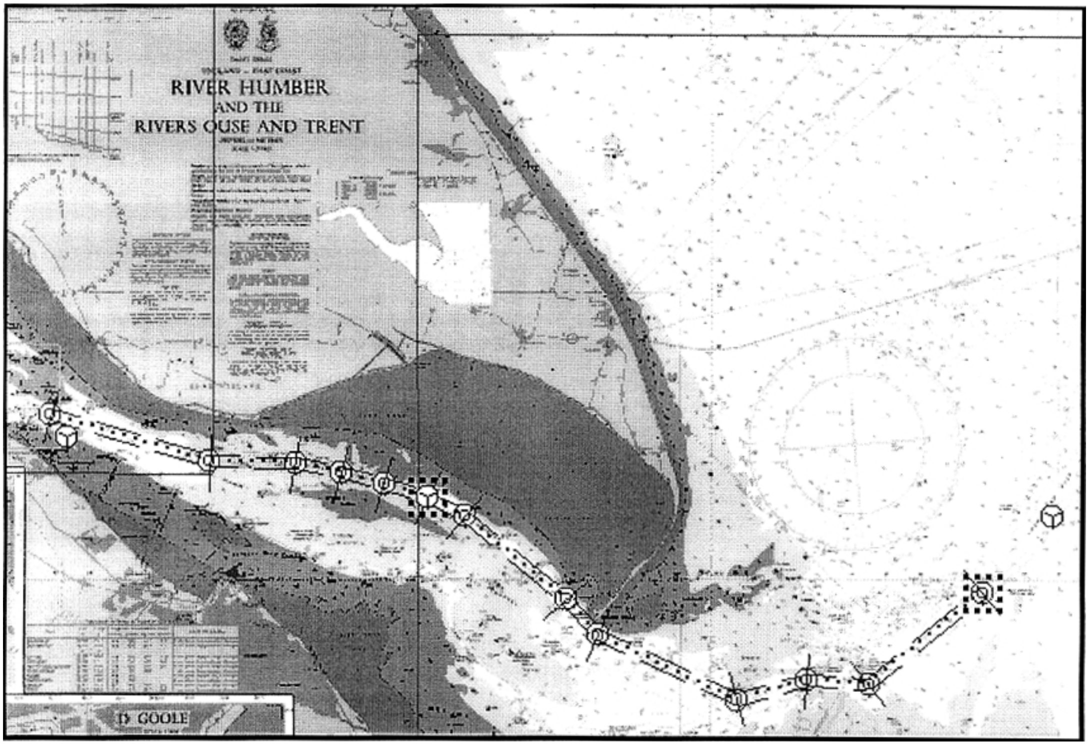


**Figure 7.17** ARCS chart 109, River Humber and the Rivers Ouse and Trent OSGB36. (Reproduced courtesy of PC Maritime.)

### *Passage plans*

Having created a route, the user can enter estimated speed, desired departure/arrival times and calculate for each leg of the route:

- course to steer, allowing for variation, deviation and tidal stream (if required)
- distance
- estimated time.

The user can view the plan on screen, change variables as required and then print a copy of the plan. As an example of a chart overlaid with a route Figure 7.17 shows ARCS chart 109, with a route approaching the Humber River, illustrating waypoints entered for the planned route.

Route monitoring options can be chosen so that it is possible to:

- automatically increment route legs as the vessel passes through waypoints so that Navmaster calculations on range and bearing to the next waypoint are relevant, and up-to-date information is sent to the Autopilot
- monitor the vessel's progress against the planned route.

Other options include the following.

- Automatic leg advance. Choosing this option allows the route legs to increment automatically as the vessel passes through the waypoint detection parameters set by the user.

- Waypoint detection. Choosing this option and setting a radius for the route leg to increment to the next leg when vessel position enters the circle. On entry a warning is given. The position and time of entry, and waypoint name are recorded in the log. See Figure 7.18.
- Passing perpendicular. Choosing this option allows the route leg to increment to the next leg when vessel position crosses a line drawn at right angles to the current leg. On passing, a warning is given. Position, time of entry and waypoint name are entered in a log. See Figure 7.18.
- Limits of deviation. Choosing this option allows a deviation limit to be set. If the vessel position exceeds this limit a warning message is displayed and remains until the vessel returns inside the

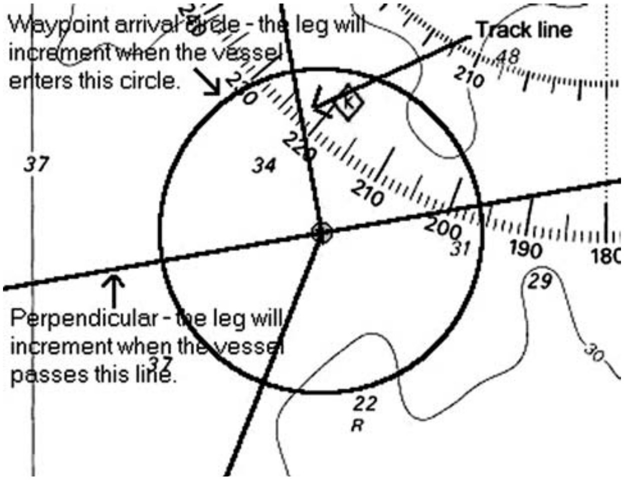


**Figure 7.18** Use of waypoint arrival circle and passing perpendicular. (Reproduced courtesy of PC Maritime.)

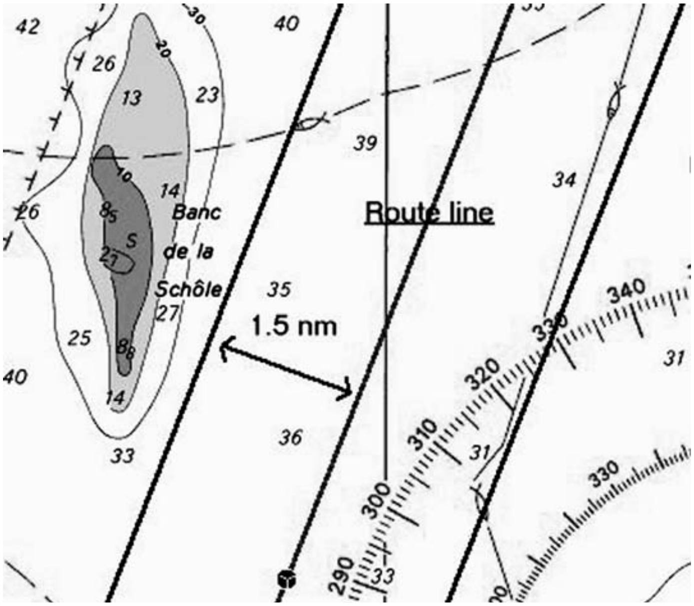
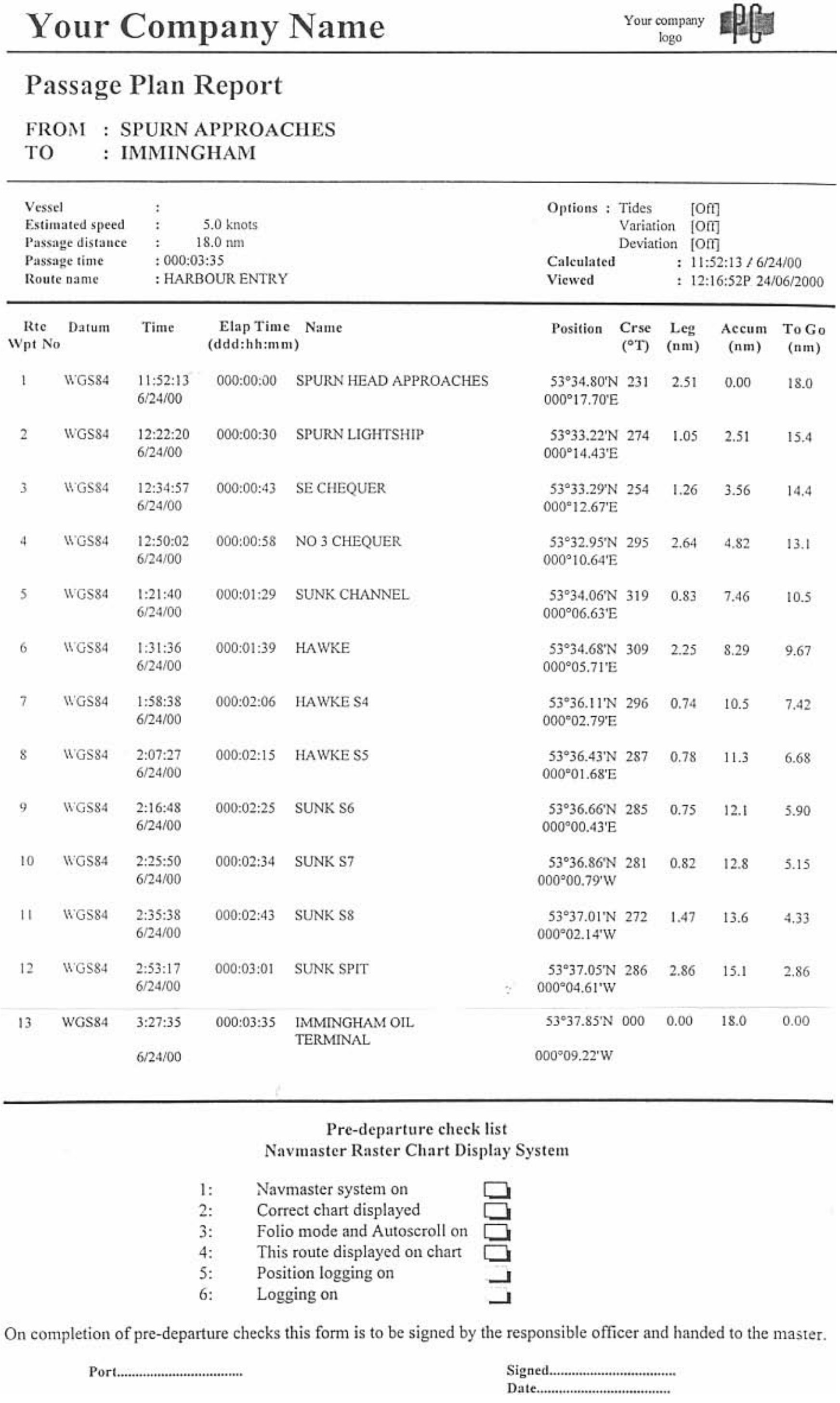


**Figure 7.19** Use of limits of deviation. (Reproduced courtesy of PC Maritime.)



**Your Company Name**

Your company logo

**Passage Plan Report**

FROM : SPURN APPROACHES
TO : IMMINGHAM

Vessel :
Estimated speed : 5.0 knots
Passage distance : 18.0 nm
Passage time : 000:03:35
Route name : HARBOUR ENTRY

Options : Tides [Off]
Variation [Off]
Deviation [Off]
Calculated : 11:52:13 / 6/24/00
Viewed : 12:16:52P 24/06/2000

| Rte Wpt No | Datum | Time | Elap Time (ddd:hh:mm) | Name | Position | Crse (°T) | Leg (nm) | Accum (nm) | To Go (nm) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | WGS84 | 11:52:13<br>6/24/00 | 000:00:00 | SPURN HEAD APPROACHES | 53°34.80'N<br>000°17.70'E | 231 | 2.51 | 0.00 | 18.0 |
| 2 | WGS84 | 12:22:20<br>6/24/00 | 000:00:30 | SPURN LIGHTSHIP | 53°33.22'N<br>000°14.43'E | 274 | 1.05 | 2.51 | 15.4 |
| 3 | WGS84 | 12:34:57<br>6/24/00 | 000:00:43 | SE CHEQUER | 53°33.29'N<br>000°12.67'E | 254 | 1.26 | 3.56 | 14.4 |
| 4 | WGS84 | 12:50:02<br>6/24/00 | 000:00:58 | NO 3 CHEQUER | 53°32.95'N<br>000°10.64'E | 295 | 2.64 | 4.82 | 13.1 |
| 5 | WGS84 | 1:21:40<br>6/24/00 | 000:01:29 | SUNK CHANNEL | 53°34.06'N<br>000°06.63'E | 319 | 0.83 | 7.46 | 10.5 |
| 6 | WGS84 | 1:31:36<br>6/24/00 | 000:01:39 | HAWKE | 53°34.68'N<br>000°05.71'E | 309 | 2.25 | 8.29 | 9.67 |
| 7 | WGS84 | 1:58:38<br>6/24/00 | 000:02:06 | HAWKE S4 | 53°36.11'N<br>000°02.79'E | 296 | 0.74 | 10.5 | 7.42 |
| 8 | WGS84 | 2:07:27<br>6/24/00 | 000:02:15 | HAWKE S5 | 53°36.43'N<br>000°01.68'E | 287 | 0.78 | 11.3 | 6.68 |
| 9 | WGS84 | 2:16:48<br>6/24/00 | 000:02:25 | SUNK S6 | 53°36.66'N<br>000°00.43'E | 285 | 0.75 | 12.1 | 5.90 |
| 10 | WGS84 | 2:25:50<br>6/24/00 | 000:02:34 | SUNK S7 | 53°36.86'N<br>000°00.79'W | 281 | 0.82 | 12.8 | 5.15 |
| 11 | WGS84 | 2:35:38<br>6/24/00 | 000:02:43 | SUNK S8 | 53°37.01'N<br>000°02.14'W | 272 | 1.47 | 13.6 | 4.33 |
| 12 | WGS84 | 2:53:17<br>6/24/00 | 000:03:01 | SUNK SPIT | 53°37.05'N<br>000°04.61'W | 286 | 2.86 | 15.1 | 2.86 |
| 13 | WGS84 | 3:27:35<br>6/24/00 | 000:03:35 | IMMINGHAM OIL TERMINAL | 53°37.85'N<br>000°09.22'W | 000 | 0.00 | 18.0 | 0.00 |

**Pre-departure check list**
**Navmaster Raster Chart Display System**

1: Navmaster system on ☐
2: Correct chart displayed ☐
3: Folio mode and Autoscroll on ☐
4: This route displayed on chart ☐
5: Position logging on ☐
6: Logging on ☐

On completion of pre-departure checks this form is to be signed by the responsible officer and handed to the master.

Port.................................

Signed.................................
Date.................................

**Figure 7.20** Passage plan for a route into Immingham on the River Humber. (Reproduced courtesy of PC Maritime.)

limit. When the vessel exceeds the limit a log entry is made, with time and position. A further log entry is made when the vessel returns inside the limit. See Figure 7.19.

### *Creating a passage plan*

A passage plan can be created as follows.

1 Prepare the route.
2 Select the plan tab on the Route panel (a typical Route side panel is shown in Figure 7.16).
3 Enter departure/arrival time and estimated speed.
4 Set any options required.
5 Click on Calc.
6 Click on Report to see the plan.

A typical passage plan report for the route of Figure 7.17 is shown in Figure 7.20.

Because Navmaster calculates routes almost instantly it is a simple matter to change parameters such as vessel speed, date and options.

The above has been extracted, with permission, from the Navmaster User Guide and only gives a very limited overview of the facilities available with the system. More detail can be obtained from the manufacturers PC Maritime, Brunswick House, Brunswick Road, Plymouth PL4 0NP, UK. E-mail: marketing@pcmaritime.co.uk and website: www.pcmaritime.co.uk.

## 7.8 Glossary

| | |
|---|---|
| **AHO** | Australian Hydrographic Office. |
| **AIS** | Automatic Identification System, see Transponder. |
| **ARCS** | Admiralty Raster Chart Service. The UKHO proprietary RNC. |
| **ARPA** | Automatic Radar Plotting Aid. |
| **Chart cell** | The smallest unit for geographical data. Each cell has a unique address in memory and may possess different data volume and size characteristics. |
| **Chart symbol** | A graphical representation of an object or characteristic. |
| **CIS** | Chart Information System. |
| **'Course-up' display** | A display where the heading of own ship is upwards on the screen and the chart moves relative to own ship. |
| **CPA** | Closest Point of Approach. |
| **Database** | A set of stored data used for a particular application which can be assessed as required. |
| **Datum** | See Geodetic datum. |
| **DGPS** | Differential Global Positioning System. |
| **ECDIS** | Electronic Chart Display and Information System. The performance standard approved by the IMO and defined in publications from the IHO (Special Publications S-52 and S-57) and IEC document 1174. |
| **ECS** | Electronic Chart System. A system that, unlike ECDIS, has no obligation to conform to the ECDIS performance standards. |
| **Ellipsoid** | A regular geometric shape which closely approximates to the shape of a geoid, having a specific mathematical expression, and can be used for geodetic, mapping and charting purposes. |

| | |
|---|---|
| **ENC** | Electronic Navigational Chart. Charts, manufactured for use with ECDIS, which meet the ECDIS performance standards and are issued by or on the authority of government-authorized hydrographic offices. |
| **ETA** | Estimated time of arrival. |
| **Geodetic datum** | A specifically orientated reference ellipsoid requiring typically eight parameters to define it. Two parameters relate to the dimensions of the ellipsoid, three parameters specify its centre with respect to the Earth's centre of mass while the remainder specify ellipsoid orientation with respect to the average spin axis of the Earth and Greenwich reference meridian. Provides a horizontal datum. |
| **Geoid** | An undulating but smooth representation of equal values of the Earth's gravitational field coinciding most closely with mean sea level. The geoid is the primary reference surface for heights. |
| **GMSK** | Gaussian Minimum Shift Keying. |
| **GNSS** | Global Navigation Satellite System. The use of GPS for civilian purposes. |
| **GPS** | Global Positioning System. A satellite navigation system designed to provide continuous position and velocity data in three dimensions and accurate timing information globally. |
| **Hardware** | The physical part of a computer system that provides the processing capability; includes peripheral devices and cabling. |
| **HCRF** | Hydrographic Chart Raster Format. Developed by the UKHO and used by them for the Admiralty Raster Chart Service (ARCS) and by the AHO for its Seafarer Chart Service. Other HOs are expected to adopt the format. |
| **HDLC** | High-Level Data Link Control, specified by ISO/IEC 3309, 5th edition 1993. |
| **IEC** | International Electrotechnical Commission. The organization which produces world standards in the area of electrical and electronic engineering. |
| **IHO** | International Hydrographic Organization. A grouping of national hydrographic offices responsible for promoting international standards in the fields of hydrographic surveying and chart production. |
| **IMO** | International Maritime Organization. A specialized agency of the United Nations and responsible for promoting maritime safety and navigational efficiency. |
| **ITU-R** | International Telecommunications Union Sector for Radiocommunication. |
| **MMSI** | Maritime Mobile Service Identities. An international system of automatic identification for all ships. |
| **NHO** | National Hydrographic Office. |
| **NIMA** | National Imagery and Mapping Agency. |
| **NMEA** | National Marine Electronics Association. An organization comprising manufacturers and distributors. Responsible for agreeing standards for interfacing between various electronic systems on ships. NMEA 0183 version 2.3 is the current standard. |
| **NOAA** | National Oceanic and Atmospheric Administration. |
| **'North-up' display** | A display configuration where north is always in the up direction. This corresponds to the orientation of nautical charts and is the normal display for an ECDIS. |
| **Notice to Mariners** | A notice issued by hydrographic offices, on a periodic or occasional basis, relating to matters that affect nautical charts, sailing directions, light lists and other nautical publications. |

| | |
|---|---|
| **NOS** | National Ocean Service. |
| **OCS** | Office of Coast Survey. |
| **Own ship** | Used to define the vessel on which the electronic chart system is operating. |
| **Performance standard** | Used to define the minimum performance requirements for a system to meet the requirements of the SOLAS Convention. |
| **Pixel** | An abbreviation for picture element. It is the smallest element that can be resolved by electronic raster devices such as a scanner, display and plotter. |
| **PRIMAR** | A series of regional distribution centres (RENCs) will be set up for the distribution of ENCs, and PRIMAR is the first of these centres. |
| **RCDS** | Raster Chart Display System. A navigation system which can be accepted as complying with the paper version of the up-to-date chart requirements of regulation V/20 of the SOLAS Convention, by displaying RNCs with position information from navigation sensors to assist the mariner in route planning and route monitoring, and if required display additional navigation related material. |
| **RCPA** | Range to closest point of approach. |
| **RENC** | Regional ENC Co-ordinating Centre. |
| **RNC** | Raster navigational chart. A facsimile of a paper chart. Both the paper chart and the RNC are originated by, or distributed on the authority of, a government authorized-hydrographic office. |
| **Route monitoring** | A function required of an ECDIS whereby own ship present position can be displayed on the chart and viewed relative to the chart data. |
| **Route planning** | A function required of an ECDIS whereby the mariner can study the intended route on a display and select an intended track, marking it with waypoints and other navigational data. |
| **S-52** | IHO Special Publication S-52. Specification for chart content and display aspects of ECDIS. |
| **S-57** | IHO Special Publication S-57. IHO transfer standard for digital hydrographic data, edition 3. It describes the data model and format to be used for ENCs. |
| **Safety contour** | The contour selected by the mariner, using the SENC data, to determine soundings which, relative to own ship's draught, provide safe water channels. The ECDIS can use the information to generate anti-grounding alarms. |
| **Safety depth** | The depth, selected by the mariner, which defines own ship's draught plus under-keel clearance which can be used by the ECDIS to indicate soundings on the display which may be equal or less than the defined value. |
| **SENC** | System Electronic Navigational Chart. This is the database produced by chart suppliers which meets the requirements of the IHO Special Publication S-57. |
| **Software** | This includes all the programs that can be used on a computer. Software can be subdivided into the operational software required for the computer to function and the application software developed for specific user applications. |
| **SOLAS** | Safety of Life at Sea. The International Convention for the Safety of Life at Sea Chapter V Safety of Navigation, Regulation 20, Nautical Publications requires that 'All ships shall carry adequate and up-to-date charts, sailing directions, lists of lights, notices to mariners, tide tables and all other nautical publications necessary for the intended voyage'. SOLAS does not apply universally and some vessels, such as ships of war, cargo ships of less than 500 GRT, fishing vessels etc are exempt from the SOLAS requirements. |

| | |
|---|---|
| **SOTDMA** | Self-organizing time division multiple access. Used by mobile stations operating in autonomous and continuous mode. The protocol offers an access algorithm that quickly resolves conflicts without intervention from controlling stations. |
| **Standard display** | The SENC information that should be displayed when a chart is first accessed by the ECDIS. The level of data contained can be customized to suit the mariner. |
| **TCPA** | Time to closest point of approach. |
| **TDMA** | Time division multiple access. |
| **Transponder (AIS)** | A shipborne transmit/receive system which broadcasts continuously, on VHF frequencies, details about ship's identity, ship characteristics, type of cargo, destination, course and speed. The ECDIS can be used to display AIS targets together with their speed and course vectors. |
| **UKHO** | United Kingdom Hydrographic Office. |
| **USCG** | US Coast Guard. |
| **UTC** | Co-ordinated universal time. Developed to meet the requirements of scientists to provide a precise scale of time interval and navigators surveyors and others requiring a time scale directly related to the earth's rotation. |
| **VTS** | Vessel Traffic System. A system for managing shipping traffic in congested areas such as ports and inland waterways. |
| **Waypoint** | A point entered into a computer and used as a reference point for navigational calculations. Planned voyages would have a series of waypoints indicating legs of the voyage. A modern computer is capable of storing multiple waypoints. |
| **WEND** | Worldwide ENC database. A model, developed by the IHO, to act as a distribution network to supply ENCs to ECDIS compliant ships. |
| **WGS-84** | World Geodetic System 1984. A global datum system for horizontal datum used as a standard in ECDIS. |
| **Zoom** | A method of changing the scale of the displayed chart information on the screen. Zoom-in or zoom-out facilities are usually provided at the touch of a button. |

## 7.9 Summary

- An electronic chart is one where chart data is provided as a digital charting system capable of displaying both geographical data and text.
- An electronic chart is 'official' if it is issued by or on the authority of a national hydrographic office. All other charts are 'non-official'.
- An electronic chart may use raster data or vector data.
- Delivery of electronic chart data is via an Electronic Chart Display and Information System (ECDIS) which is a navigational information system, comprising hardware, software and official vector charts and must conform to ECDIS Performance Standards.
- Chart types available include privately produced vector, official raster and Electronic Navigational Chart (ENC). The ENC is the designated chart system for ECDIS.
- A Raster Chart Display System (RCDS) is one that displays official raster navigational charts (RNCs).

- A dual fuel system is one that operates as an ECDIS or RCDS mode according to the type of chart data in use.
- Chart accuracy may depend on local datum that may differ from that used by satellite systems which use a global datum, e.g. WGS-84. Corrections may be necessary before a position is plotted on a chart.
- Electronic charts are updated regularly to ensure conformity with the SOLAS requirement that charts should be 'adequate and up-to-date for the intended voyage'.
- Automatic Identification System (AIS) is a shipborne transponder system that broadcasts information about a ship fitted with the system. The data generated may be used by other AIS-fitted ships and/or shore stations and such data may be passed to an electronic charting system where AIS-fitted ships could appear as 'targets' on the electronic chart. Such targets could be interrogated to generate information such as ship's speed, heading and other data.
- For any ECDIS system to operate, suitable software must be available to enable the function of an ECDIS system to meet performance standards as laid down by the regulatory bodies. A particular system examined is the Navmaster Electronic Navigation System of PC Maritime.

## 7.10 Revision questions

1. What do you understand by the term 'electronic chart'? What is the definition of an 'official' electronic chart?
2. Explain briefly the difference between a chart produced using raster data and one produced using vector data. Give advantages/disadvantages associated with each type of chart.
3. Explain briefly what defines an electronic navigational chart (ENC) used with ECDIS. What are the advantages of an ENC in terms of chart information that can be displayed?
4. Describe what you understand by the term Electronic Chart Display and Information System (ECDIS). What are the basic requirements of an ECDIS?
5. Describe what you understand by the term Raster Chart Display System (RCDS) and state briefly how a RCDS could be used in a dual fuel system.
6. Explain why there may be a difference between local datum used for a particular chart and the global datum used in ECDIS. How would a position, determined from a GPS or DGPS input, be affected if plotted on a chart based on a different datum?
7. Describe briefly the concept of an Automatic Identification System (AIS). Explain the advantages to be gained by fitting ships and specific shore stations with AIS.
8. The Navmaster Electronic Navigation System (Section 7.7) uses on-screen side panels that represent main functions of the system. Describe briefly the function of the following side panels:
   (a) monitor mode
   (b) chartpoint mode
   (c) route mode.
9. Using the Navmaster Electronic Navigation System (Section 7.7) describe how charts may be installed in the system. What information is displayed in the chart information panel for a selected chart?
10. Using the Navmaster Electronic Navigation System (Section 7.7) as a basis, describe the recommended sequence to be followed for route planning and monitoring. Define what is meant by a chartpoint and describe how chartpoints could be used in route planning.

# Chapter 8

# The ship's master compass

## 8.1 Introduction

Of all the navigation instruments in use today, the master compass is the oldest and probably the one that most navigators feel happiest with. However, even the humble compass has not escaped the advance of microelectronics. Although modern gyrocompasses are computerized the principles upon which they work remain unchanged.

## 8.2 Gyroscopic principles

At the heart of a marine gyrocompass assembly is a modern gyroscope consisting of a perfectly balanced wheel arranged to spin symmetrically at high speed about an axis or axle. The wheel, or rotor, spins about its own axis and, by suspending the mass in a precisely designed gimbals assembly, the unit is free to move in two planes each at right angles to the plane of spin. There are therefore three axes in which the gyroscope is free to move as illustrated in Figure 8.1:

- the spin axis
- the horizontal axis
- the vertical axis.

In a free gyroscope none of the three freedoms is restricted in any way. Such a gyroscope is almost universally used in the construction of marine gyrocompass mechanisms. Two other types of gyroscope, the constrained and the spring-restrained are now rarely seen.

In order to understand the basic operation of a free gyroscope, reference must be made to some of the first principles of physics. A free gyroscope possesses certain inherent properties, one of which is inertia, a phenomenon that can be directly related to one of the basic laws of motion documented by Sir Isaac Newton. Newton's first law of motion states that 'a body will remain in its state of rest or uniform motion in a straight line unless a force is applied to change that state'. Therefore a spinning mass will remain in its plane of rotation unless acted upon by an external force. Consequently the spinning mass offers opposition to an external force. This is called 'gyroscopic inertia'. A gyroscope rotor maintains the direction of its plane of rotation unless an external force of sufficient amplitude to overcome inertia is applied to alter that direction. In addition a rapidly spinning free gyroscope will maintain its position in free space irrespective of any movement of its supporting gimbals (see Figure 8.2).

Also from the laws of physics it is known that the linear momentum of a body in motion is the product of its mass and velocity (*mv*). In the case of a freely spinning wheel (Figure 8.3), it is more

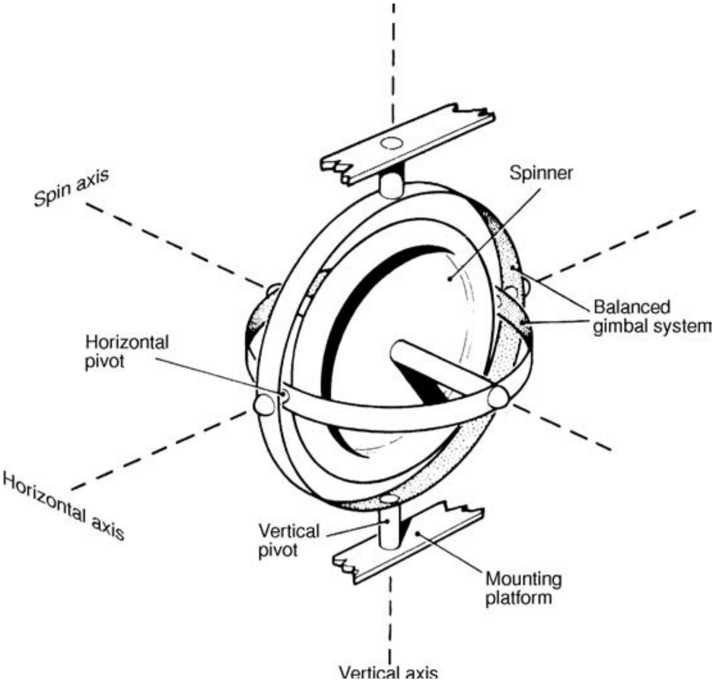


**Figure 8.1** A free gyroscope. (Reproduced courtesy of S. G. Brown Ltd.)

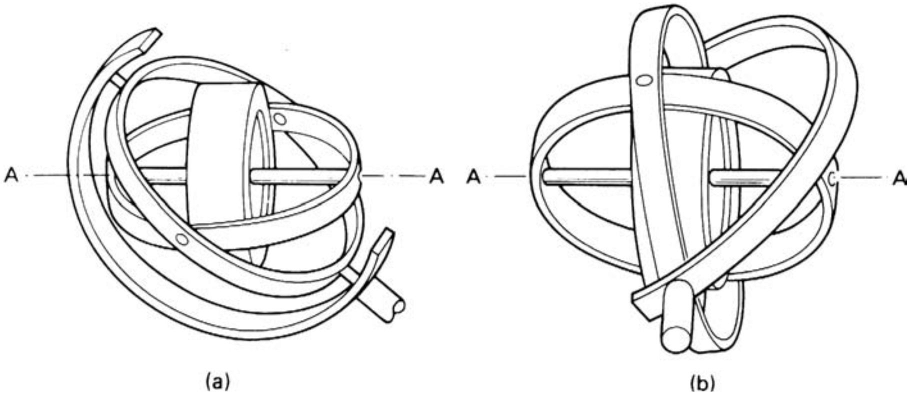


**Figure 8.2** The gyrospin axis is stabilized irrespective of any movement of the supporting gimbals. (Reproduced courtesy of Sperry Ltd.)

convenient to think in terms of angular momentum. The angular momentum of a particle spinning about an axis is the product of its linear momentum and the perpendicular distance of the particle from the axle:

$$\text{angular momentum} = mv \times r$$

where $r$ = rotor radius.

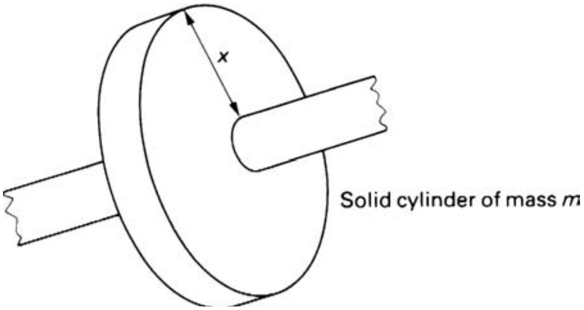


**Figure 8.3** A spinning rotor possessing a solid mass.

The velocity of the spinning rotor must be converted to angular velocity ($\omega$) by dividing the linear tangential velocity ($v$) by the radius ($r$). The angular momentum for any particle spinning about an axis is now:

$$m\omega r^2$$

For a spinning rotor of constant mass where all the rotating particles are the same and are concentrated at the outer edge of the rotor, the angular momentum is the product of the moment of inertia ($I$) and the angular velocity:

$$\text{angular momentum} = I\omega$$

where $I = 0.5\ mr^2$.

It can now be stated that gyroscopic inertia depends upon the momentum of the spinning rotor. The momentum of such a rotor depends upon three main factors:

- the total mass, $M$ of the rotor (for all particles)
- the radius $r$ summed as the constant $K$ (for all the particles) where $K$ is the radius of gyration
- the angular velocity $\omega$.

The angular momentum is now proportional to $\omega MK^2$. If one or more of these factors is changed, the rotor's gyroscopic inertia will be affected. In order to maintain momentum, a rotor is made to have a large mass, the majority of which is concentrated at its outer edge. Normally the rotor will also possess a large radius and will be spinning very fast. To spin freely the rotor must be perfectly balanced (its centre of gravity will be at the intersection of the three axes) and its mounting bearings must be as friction-free as possible. Once a rotor has been constructed, both its mass and radius will remain constant. To maintain gyroscopic inertia therefore it is necessary to control the speed of the rotor accurately. This is achieved by the use of a precisely controlled servo system.

### 8.2.1 Precession

Precession is the term used to describe the movement of the axle of a gyroscope under the influence of an external force. If a force is applied to the rotor by moving one end of its axle, the gyroscope will be displaced at an angle of 90° from the applied force. Assume that a force is applied to the rotor in Figure 8.4 by lifting one end of its axle so that point A on the rotor circumference is pushed

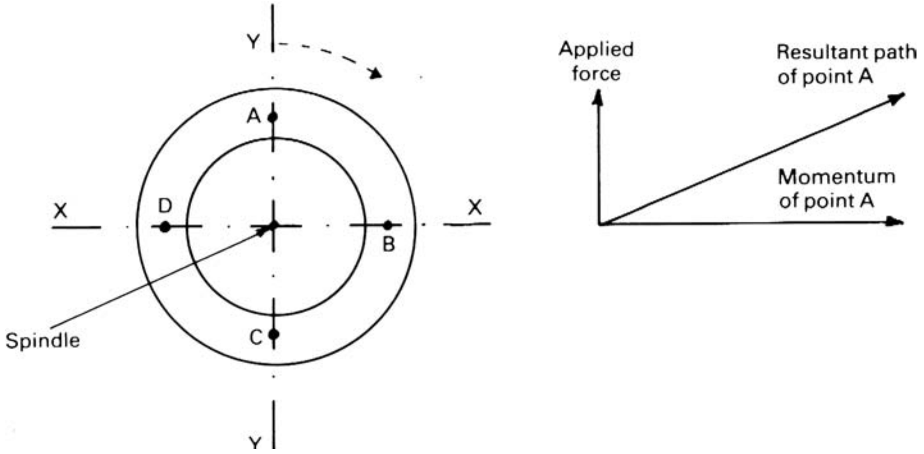


**Figure 8.4** Gyro precession shown as a vector sum of the applied forces and the momentum.

downwards into the paper. The rotor is rapidly spinning clockwise, producing gyroscopic inertia restricting the effective force attempting to move the rotor into the paper. As the disturbing force is applied to the axle, point A continues its clockwise rotation but will also move towards the paper. Point A will therefore move along a path that is the vector sum of its original gyroscopic momentum and the applied disturbing force. As point A continues on its circular path and moves deeper into the paper, point C undergoes a reciprocal action and moves away from the paper. The plane of rotation of the rotor has therefore moved about the H axis although the applied force was to the V axis.

The angular rate of precession is directly proportional to the applied force and is inversely proportional to the angular momentum of the rotor. Figure 8.5 illustrates the rule of gyroscopic precession.

### 8.2.2 The free gyroscope in a terrestrial plane

Now consider the case of a free gyroscope perfectly mounted in gimbals to permit freedom of movement on the XX and YY axes. In this description, the effect of gravity is initially ignored. It should be noted that the earth rotates from west to east at a rate of 15°/h and completes one revolution in a 'sidereal day' which is equivalent to 23 h 56 min 4 s. The effect of the earth's rotation beneath the gyroscope causes an apparent movement of the mechanism. This is because the spin axis of the free gyroscope is fixed by inertia to a celestial reference (star point) and not to a terrestrial reference point. If the free gyro is sitting at the North Pole, with its spin axis horizontal to the earth's surface, an apparent clockwise movement of the gyro occurs. The spin axis remains constant but as the earth rotates in an anticlockwise direction (viewed from the North Pole) beneath it, the gyro appears to rotate clockwise at a rate of one revolution for each sidereal day (see Figure 8.6).

The reciprocal effect will occur at the South Pole. This phenomenon is known as gyro drift. Drift of the north end of the spin axis is to the east in the northern hemisphere and to the west in the southern hemisphere. There will be no vertical or tilting movement of the spin axis. Maximum gyro tilt occurs if the mechanism is placed with its spin axis horizontal to the equator. The spin axis will be stabilized in line with a star point because of inertia. As the earth rotates the eastern end of the spin axis appears to tilt upwards. Tilt of the north end of the spin axis is upwards if the north end is to the east of the meridian and downwards if it is to the west of the meridian. The gyro will appear to execute one

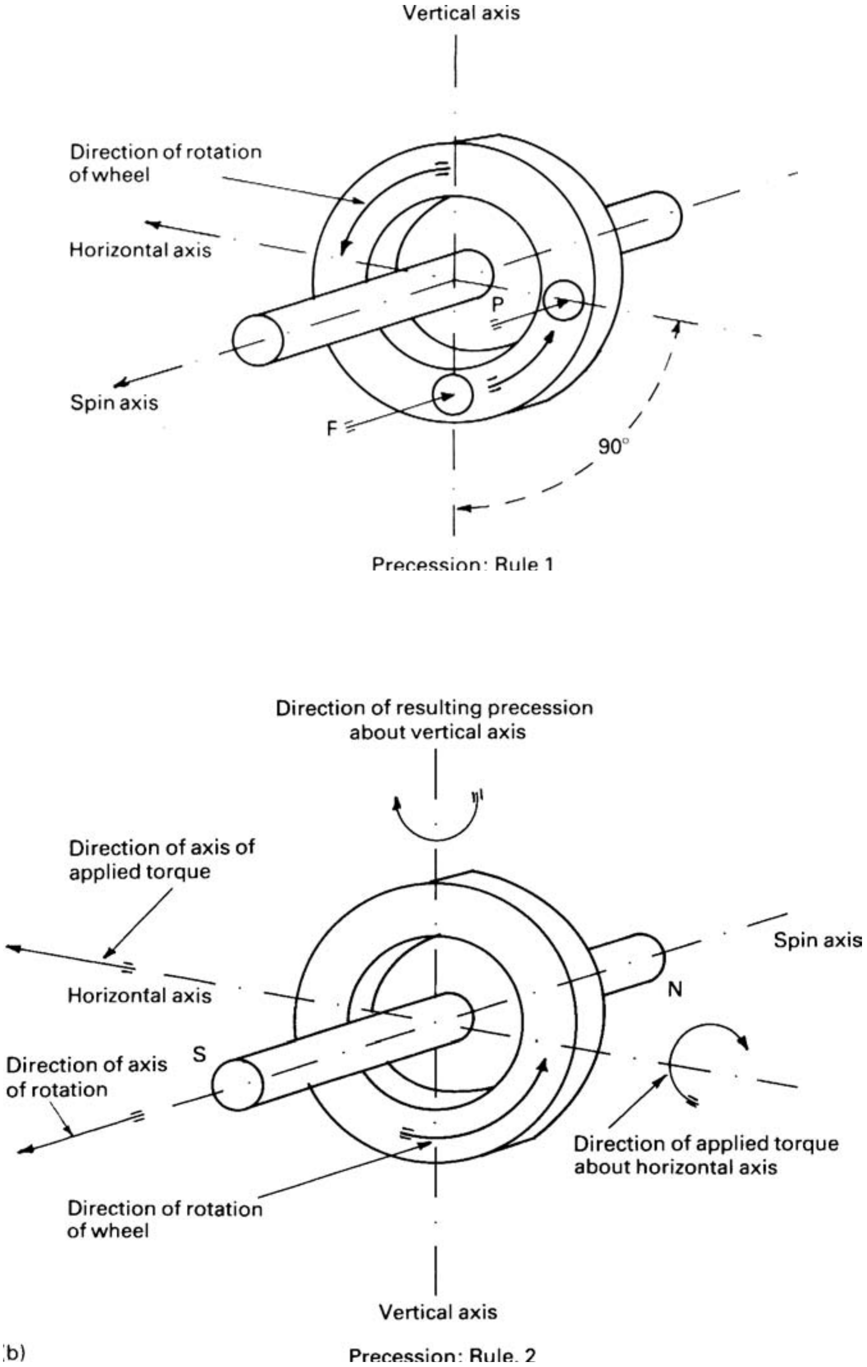
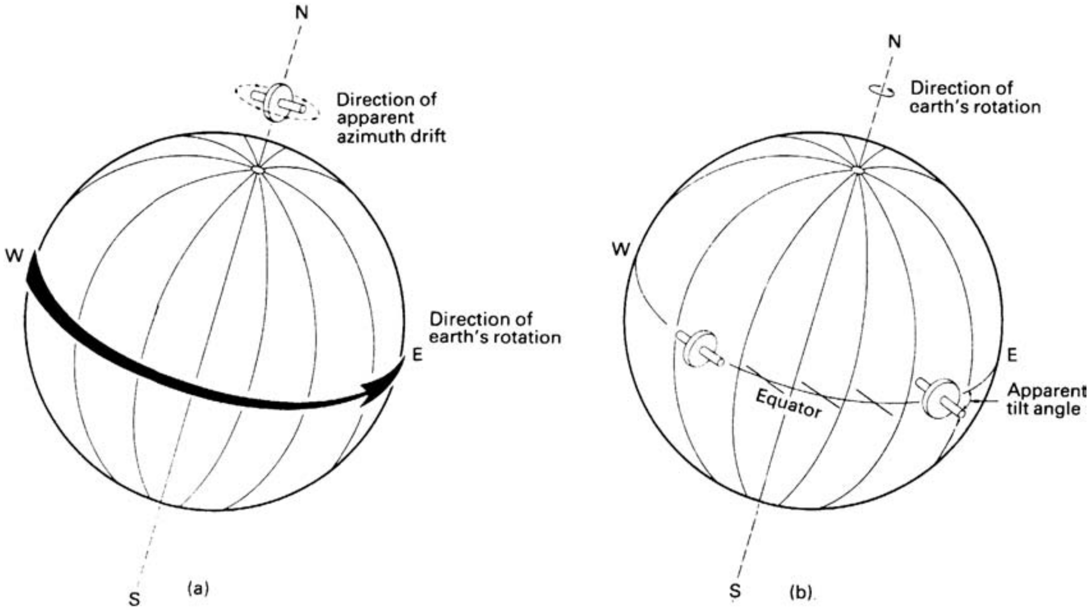


**Figure 8.5** (a) Resulting precession P occurs at 90° in the direction of spin from the applied force F. This direction of precession is the same as that of the applied force. (Reproduced courtesy of Sperry Ltd.) (b) The direction of axis rotation will attempt to align itself with the direction of the axis of the applied torque. (Reproduced courtesy of Sperry Ltd.)



Drift of the N end of the spin axis is to the E in the northern hemisphere and to the W in the southern hemisphere

Tilt of the N end of the spin axis is upwards if the N end is to the east of the meridian and downwards if it is to the W of the meridian

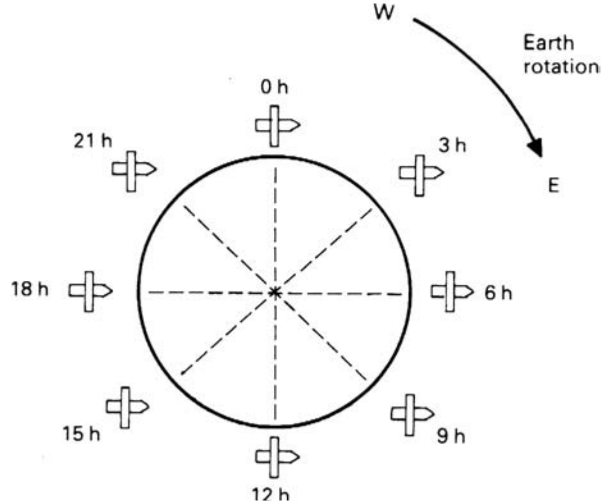


**Figure 8.6** (a) Effect of earth rotation on the gyro. (Reproduced courtesy of Sperry Ltd.) (b)View from the South Pole. The earth rotates once every 24 h carrying the gyro with it. Gyroscopic inertia causes the gyro to maintain its plane of rotation with respect to the celestial reference point. However, in relation to the surface of the earth the gyro will tilt.

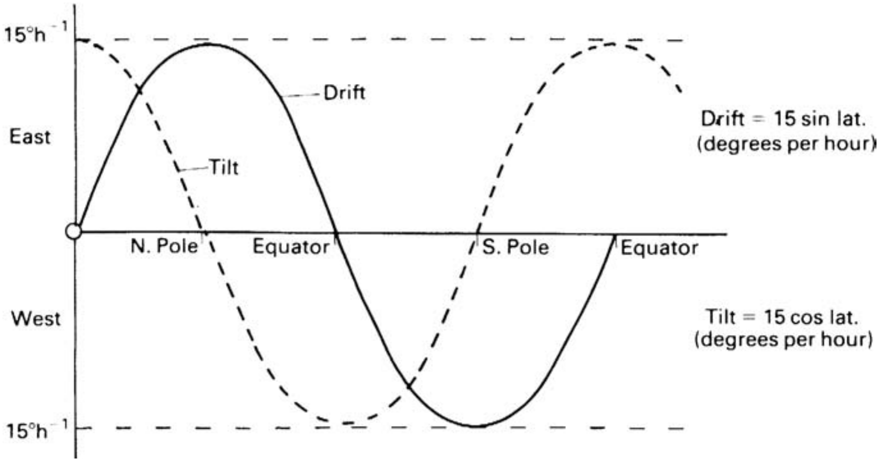


**Figure 8.7** The graphical relationship between drift and tilt.

complete revolution about the horizontal axis for each sidereal day. No drift in azimuth occurs when the gyro is directly over the equator. The relationship between drift and tilt can be shown graphically (see Figure 8.7).

Figure 8.7 shows that gyro drift will be maximum at the poles and zero at the equator, whilst gyro tilt is the reciprocal of this. At any intermediate latitude the gyro will suffer from both drift and tilt with the magnitude of each error being proportional to the sine and cosine of the latitude, respectively.

When a gyro is placed exactly with its spin axis parallel to the spin axis of the earth at any latitude, the mechanism will maintain its direction relative to the earth. There is no tilt or azimuth movement and the gyro may be considered to be Meridian stabilized. As the earth rotates the gyro will experience a movement under the influence of both tilt and azimuth motion. The rate of tilt motion is given as:

$$\text{tilt} = 15^\circ \cos \text{latitude (degrees per hour)}$$

where 15° is the hourly rate of the earth's rotation. The azimuth drift is:

$$\text{azimuth drift} = 15^\circ \sin \text{latitude (degrees per hour)}$$

### 8.2.3 Movement over the earth's surface

The free gyroscope, as detailed so far, is of no practical use for navigation since its rotor axis is influenced by the earth's rotation and its movement over the earth's surface. The stabilized gyroscopic change in position of longitude along a parallel of latitude requires a correction for the earth's rotary motion. Movement in latitude along a meridian of longitude involves rotation about an axis through the centre of the earth at right angles to its spin axis. Movement of the mechanism in any direction is simply a combination of the latitudinal and longitudinal motions. The faster the gyroscope moves the greater the rate of angular movement of the rotor axle attributable to these factors.

## 8.3 The controlled gyroscope

It has been stated that a free gyroscope suffers an apparent movement in both azimuth and tilt of the rotor axis depending upon its latitudinal location. When fitted to a vessel the latitude is known and consequently the extent of movement in azimuth and tilt is also known. It is possible therefore to calculate the necessary force required to produce a reciprocal action to correct the effect of apparent movement. A force can be applied to the gyro that will cause both azimuth and tilt precession to occur in opposition to the unwanted force caused by the gyro's position on the earth. The amplitude of the reciprocal force must be exactly that of the force producing the unwanted movement, otherwise over or under correction will occur. If the negative feedback is correctly applied, the gyro will no longer seek a celestial point but will be terrestrially stabilized and will assume a fixed attitude.

If the gyro is drifting in azimuth at '*N*' degrees per hour in an anticlockwise direction, an upward force sufficient to cause clockwise precession at a rate of '–*N*' degrees per hour must be applied vertically to the appropriate end of the rotor axle. The result will be that the gyro drift is cancelled and the instrument points to a fixed point on earth. Gyro tilt movement can also be cancelled in a similar way by applying an equal and opposite force horizontally to the appropriate end of the rotor axle. Although the gyro is now stabilized to a terrestrial point it is not suitable for use as a navigating compass for the following reasons.

- It is not north-seeking. Since the recognized compass datum is north, this factor is the prime reason why such a gyro is not of use for navigation.
- It is liable to be unstable and will drift if the applied reciprocal forces are not precise.
- A complex system of different reciprocal forces needs to be applied due to continual changes in latitude.
- Because of precessional forces acting upon it through the friction of the gimbal bearings, the mechanism is liable to drift. This effect is not constant and is therefore difficult to compensate for.

## 8.4 The north-seeking gyro

The gyrospin axis can be made meridian-seeking (maintaining the spin axis parallel to the earth's spin axis) by the use of a pendulum acting under the influence of earth gravity. The pendulum causes a force to act upon the gyro assembly causing it to precess. Precession, the second fundamental property of a gyroscope, enables the instrument to become north-seeking. As the pendulum swings towards the centre of gravity, a downward force is applied to the wheel axle, which causes horizontal precession to occur. This gravitational force acting downward on the spinner axle causes the compass to precess horizontally and maintain the axle pointing towards true north.

The two main ways of achieving precessional action due to gravity are to make the gyro spin axis either bottom or top heavy. Bottom-heavy control and a clockwise rotating gyro spinner are used by some manufacturers, whereas others favour a top-heavy system with an anticlockwise rotating spinner. Figure 8.8(a) illustrates this phenomenon.

With bottom-heavy control, tilting upwards of the south end produces a downward force on the other end, which, for this direction of spinner rotation, produces a precession of the north end to the west. In a top-heavy control system, tilting upwards of the north end of the gyro produces a downward force on the south end to causes a westerly precession of the north end. The result, for each arrangement, will be the same.

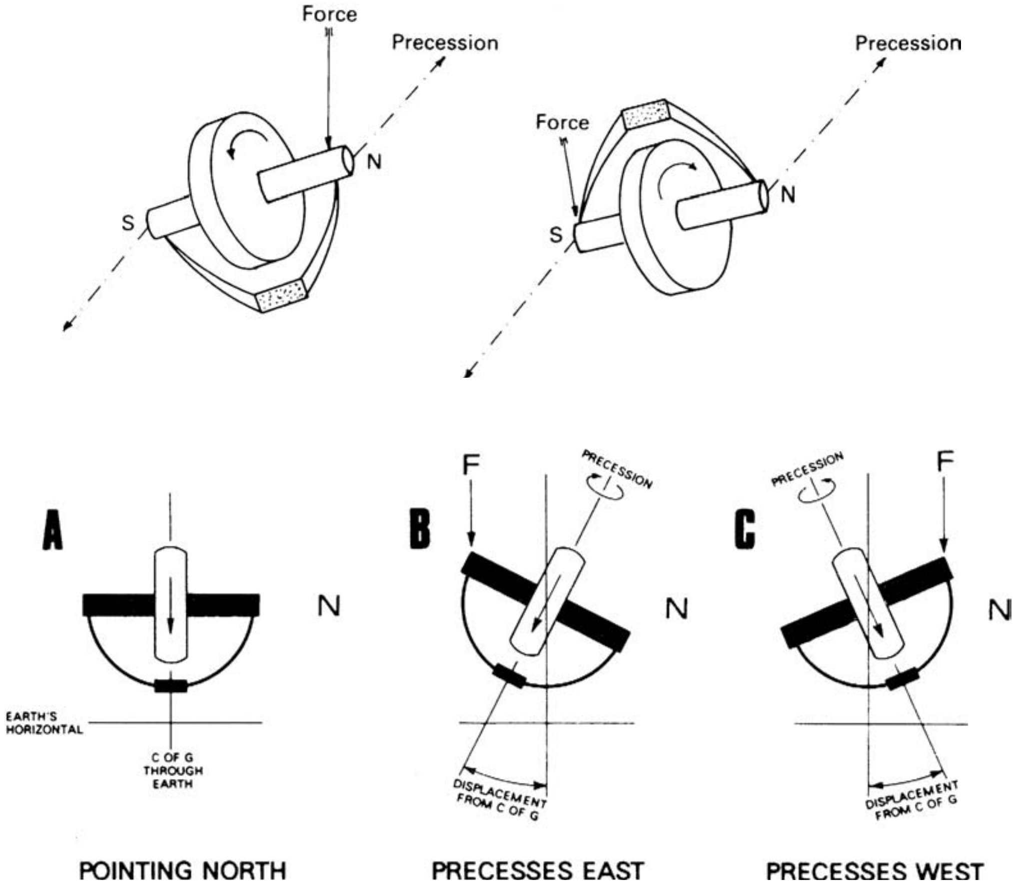


**Figure 8.8** (a) Methods of gravity control: bottom-heavy principal and top-heavy control. (b) Principle of gravity control. (Reproduced courtesy of S. G. Brown Ltd.)

### 8.4.1 Bottom-heavy control

Figure 8.8(b) illustrates the principle of precession caused by gravity acting on the bottom-weighted spin axis of a gyroscope. The pendulous weight will always seek the centre of gravity and in so doing will exert a torque about the gyro horizontal axis. Because of the earth's rotation and gyro rigidity, the pendulum will cause the gravity control to move away from the centre of gravity. The spinner is rotating clockwise, when viewed from the south end, and therefore, precession, caused by the gravitational force exerted on the spin axis, will cause the northeast end of the spin axis to move to the east when it is below the horizontal. A reciprocal action will occur causing the northeast end of the spin axis to precess towards the west when above the horizontal. The spin axis will always appear to tilt with its north end away from the earth (up) when to the east of the meridian, and its north end towards the earth (down) when to the west of the meridian (see Figure 8.9).

This action causes the north end of the spin axis, of a gravity-controlled undamped gyro, to describe an ellipse about the meridian. Because it is undamped, the gyro will not settle on the meridian. Figure 8.9 shows this action for a gyro with a clockwise rotating spinner. The ellipse produced will be anticlockwise due to the constant external influences acting upon the gyro. The extent of the ellipse will, however, vary depending upon the initial displacement of the gyro spin axis from the meridian and from the earth's horizontal. The term 'north-seeking' is given to the undamped gravity controlled gyro mechanism because the northeast end of the spin axis describes an ellipse around the North Pole but never settles. Obviously such a gyro is not suitable for use as a precise north reference compass aid.

### 8.4.2 The north-settling gyro

The ellipse described by the previous gyro mechanism possesses a constant ratio of the major and minor axes. Clearly, therefore, if the extent of one axis can be reduced, the length of the other axis will be reduced in proportion. Under these conditions the gyro spin axis will eventually settle both on the meridian and horizontally. If the gyro axis is influenced by a second force exerting a damping torque about the vertical axis, so as to cause the spin axis to move towards the horizontal, it is obvious from Figure 8.10 that the minor axis of the ellipse will be reduced.

As the north end of the spin axis moves to the west of the meridian, the earth's rotation will cause a downward tilt of the axis. This effect and the torque (Tv) will cause the gyro axis to meet the earth's horizontal at point H, which is a considerable reduction in the ellipse major axis. As Figure 8.10 clearly shows this action continues until the gyro settles in the meridian and to the surface of the earth, point N.

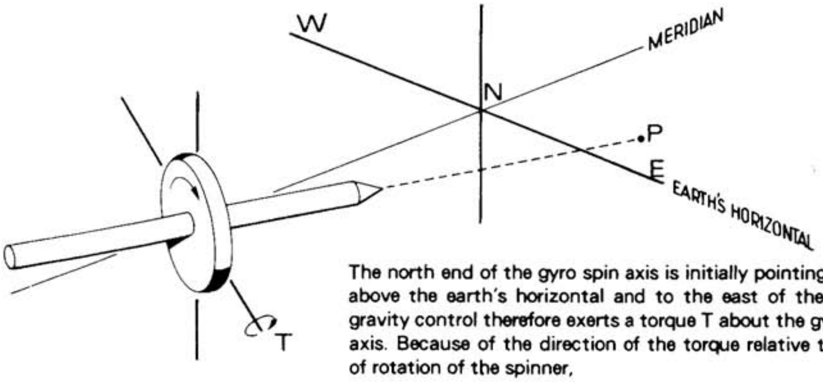
### 8.4.3 Top-heavy control

Whereas the previous compass relies on a bottom-weighted spin axis and a clockwise spinning rotor to produce a north-settling action, other manufacturers design their gyrocompasses to be effectively top-weighted and use an anticlockwise spinning rotor. But adding a weight to the top of the rotor casing produces a number of undesirable effects. These effects become pronounced when a ship is subjected to severe movement in heavy weather. To counteract unwanted effects, an 'apparent' top weighting of the compass is achieved by the use of a mercury fluid ballistic contained in two reservoirs or ballistic pots.

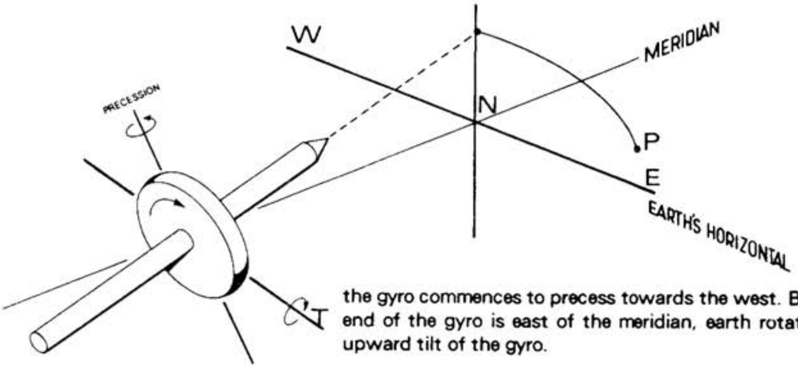
As shown in Figure 8.11, each ballistic pot, partly filled with mercury, is mounted at the north and south sides of the rotor on the spin axis. A small-bore tube connects the bases of each pot together providing a restricted path for the liquid to flow from one container to the other. The ballistic system is mounted in such a way that, when the gyro tilts, the fluid will also tilt and cause a displacement of mercury. This action produces a torque about the horizontal axis with a resulting precession in azimuth.

Consider a controlled gyroscope to be at the equator with its spin axis east west as shown in Figure 8.12. As the earth rotates from west to east the gyro will appear to tilt about its horizontal axis and the east end will rise forcing mercury to flow from pot A to pot B. The resulting imbalance of the ballistic will cause a torque about the horizontal axis. This in turn causes precession about the vertical axis and the spin axis will move in azimuth towards the meridian. The right-hand side of the gyro spin axis now moves towards the north and is referred to as the north end of the spin axis. Without the application of additional forces, this type of gyro is north-seeking only and will not settle in the meridian. The north end of the spin axis will therefore describe an ellipse as shown in Figure 8.9.

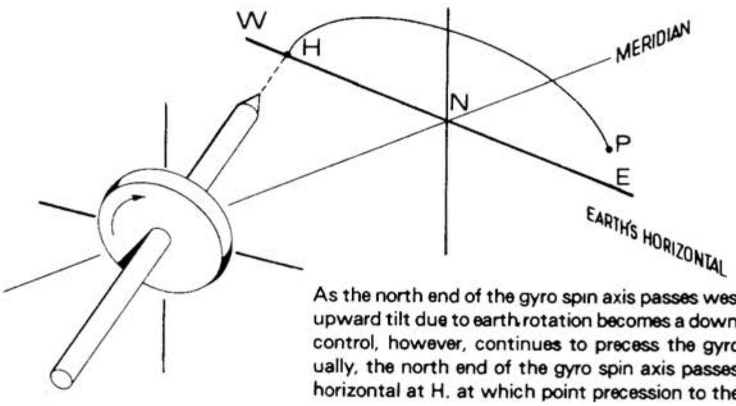
As the extent of the swings in azimuth and the degree of tilt are dependent upon each other, the gyro can be made to settle by the addition of an offset control force.



The north end of the gyro spin axis is initially pointing at P, which is above the earth's horizontal and to the east of the meridian. The gravity control therefore exerts a torque T about the gyro's horizontal axis. Because of the direction of the torque relative to the direction of rotation of the spinner,

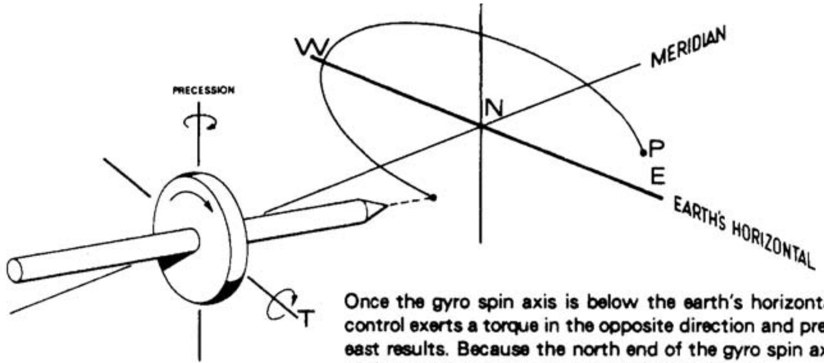


the gyro commences to precess towards the west. Because the north end of the gyro is east of the meridian, earth rotation results in an upward tilt of the gyro.

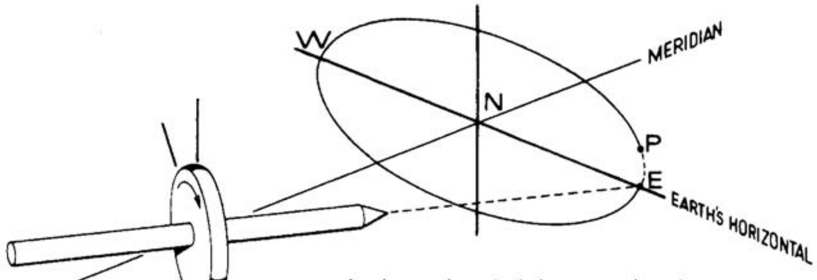


As the north end of the gyro spin axis passes west of the meridian, the upward tilt due to earth rotation becomes a downward tilt. The gravity control, however, continues to precess the gyro westwards. Eventually, the north end of the gyro spin axis passes through the earth's horizontal at H, at which point precession to the west ceases.

**Figure 8.9** Behaviour of the gravity-controlled gyro (undamped). (Reproduced courtesy of S.G. Brown Ltd.)



Once the gyro spin axis is below the earth's horizontal, the gravity control exerts a torque in the opposite direction and precession to the east results. Because the north end of the gyro spin axis is still west of the meridian, it continues to tilt downwards.



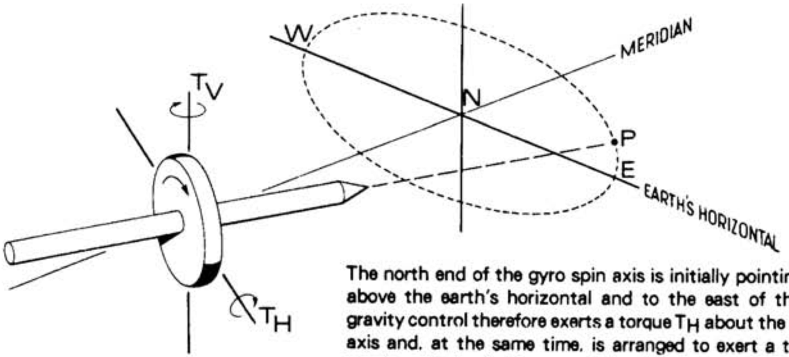
As the north end of the gyro spin axis passes east of the meridian again, the downward tilt due to earth rotation becomes an upward tilt. The gravity control, however, continues to precess the gyro towards the east until it is once more passing through the earth's horizontal. The gyro then commences to precess towards the west again (still tilting upwards) and will eventually pass through point P. The ellipse about North is thus completed and continues.

**Figure 8.9** Continued

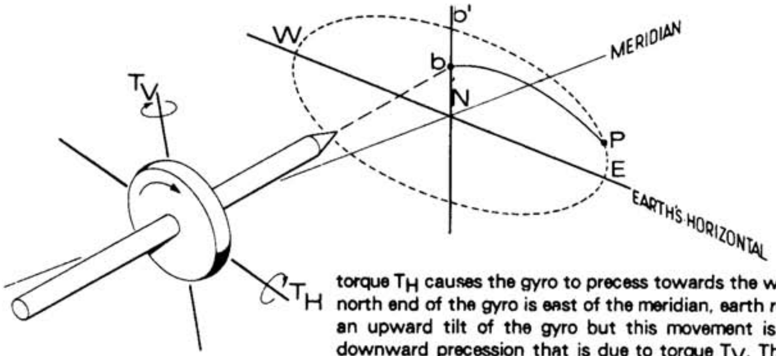
## 8.5 A practical gyrocompass

The apparent tilting of the gyroscope can be reduced by producing an offset controlling force, which in effect creates 'anti-tilt' precession allowing the unit to settle in the meridian. This is achieved by creating a force about the vertical axis to cause precession about the horizontal axis. This is achieved, in this gyro system, by offsetting the mercury ballistic controlling force slightly to the east of the vertical. The point of offset attachment must be precise so that damping action causes the gyro to settle exactly in the meridian. A comparatively small force is required to produce the necessary anti-tilt precession for the gyrocompass to be made suitable for use as a navigation instrument.

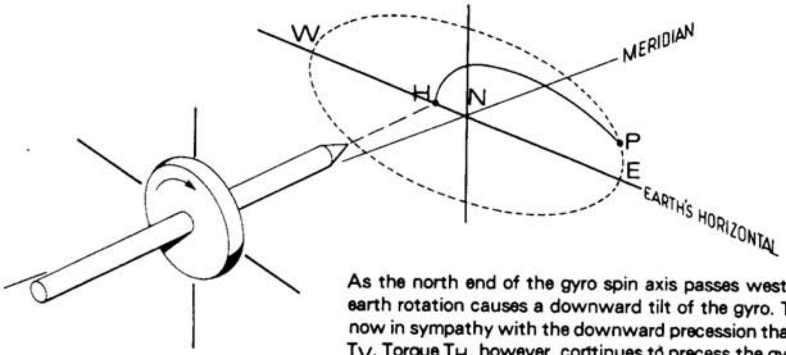
Figure 8.10 shows the curve now described by the north end of the damped gyrocompass which will settle in the meridian. An alternative and more commonly used method of applying anti-tilt damping is shown in Figure 8.13.



The north end of the gyro spin axis is initially pointing at P, which is above the earth's horizontal and to the east of the meridian. The gravity control therefore exerts a torque $T_H$ about the gyro's horizontal axis and, at the same time, is arranged to exert a torque about the gyro's vertical axis. Because of the direction of rotation of the spinner,

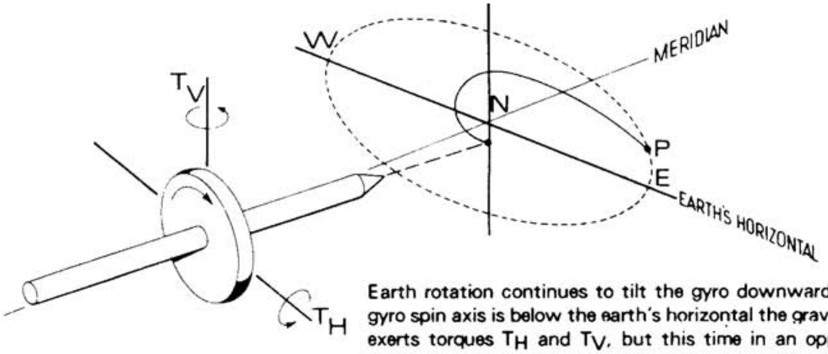


torque $T_H$ causes the gyro to precess towards the west. Because the north end of the gyro is east of the meridian, earth rotation results in an upward tilt of the gyro but this movement is damped by the downward precession that is due to torque $T_V$. The resultant path traced by the gyro spin axis is therefore Pb and not Pb′, as would be the case without the damping torque $T_V$. Note that the downward (damping) precession becomes more evident as the north end of the gyro spin axis approaches the meridian. Note also that the minor axis of the ellipse is decreased by the amount bb′

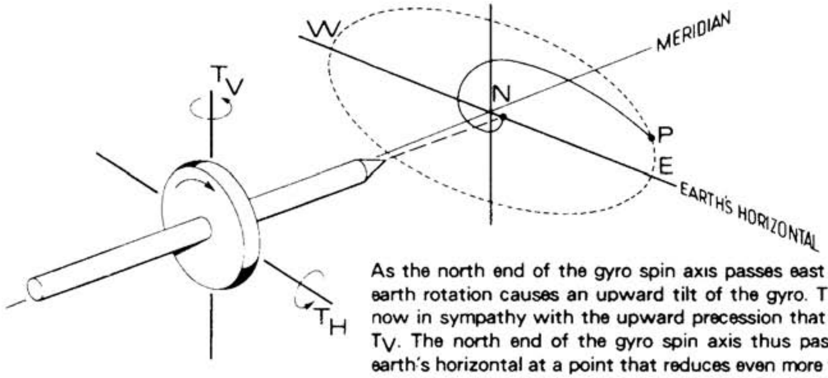


As the north end of the gyro spin axis passes west of the meridian, earth rotation causes a downward tilt of the gyro. This movement is now in sympathy with the downward precession that is due to torque $T_V$. Torque $T_H$, however, continues to precess the gyro westwards so that the gyro spin axis passes through the earth's horizontal at H, at which point torques $T_H$ and $T_V$ cease. Note that the major axis of the ellipse is considerably decreased.

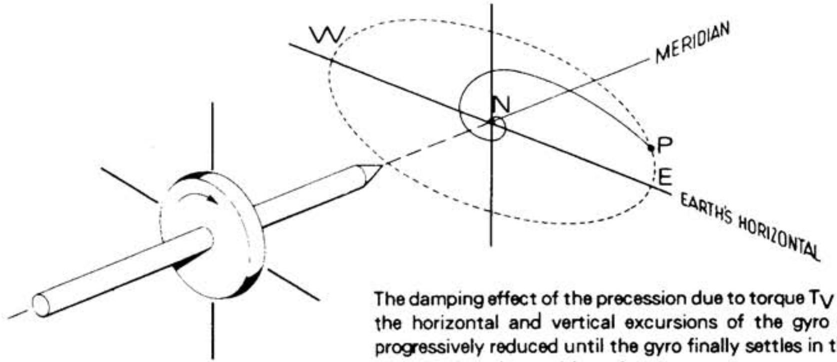
**Figure 8.10** Behaviour of the gravity-controlled gyro (damped). (Reproduced courtesy of S.G. Brown Ltd.)



Earth rotation continues to tilt the gyro downwards, and once the gyro spin axis is below the earth's horizontal the gravity control again exerts torques $T_H$ and $T_V$, but this time in an opposite direction. Torque $T_H$ now causes precession eastwards, whilst torque $T_V$ causes precession upwards and thus damps the downward tilt that is due to earth rotation. The resultant movement of the gyro causes the north end of the gyro spin axis to pass through the meridian at a point that reduces even more the minor axis of the ellipse.



As the north end of the gyro spin axis passes east of the meridian, earth rotation causes an upward tilt of the gyro. This movement is now in sympathy with the upward precession that is due to torque $T_V$. The north end of the gyro spin axis thus passes through the earth's horizontal at a point that reduces even more the major axis of the ellipse.



The damping effect of the precession due to torque $T_V$ continues, with the horizontal and vertical excursions of the gyro spin axis being progressively reduced until the gyro finally settles in the meridian and horizontal to the earth's surface.

**Figure 8.10** Continued

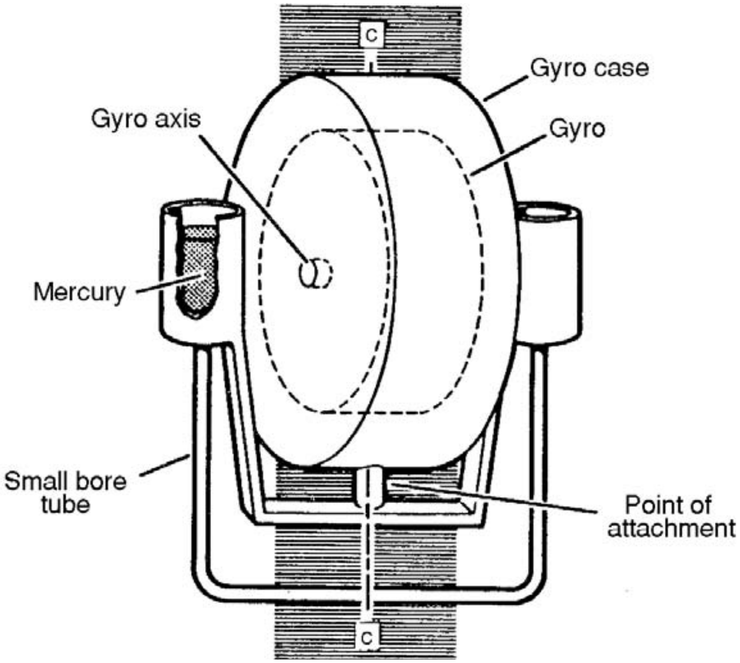
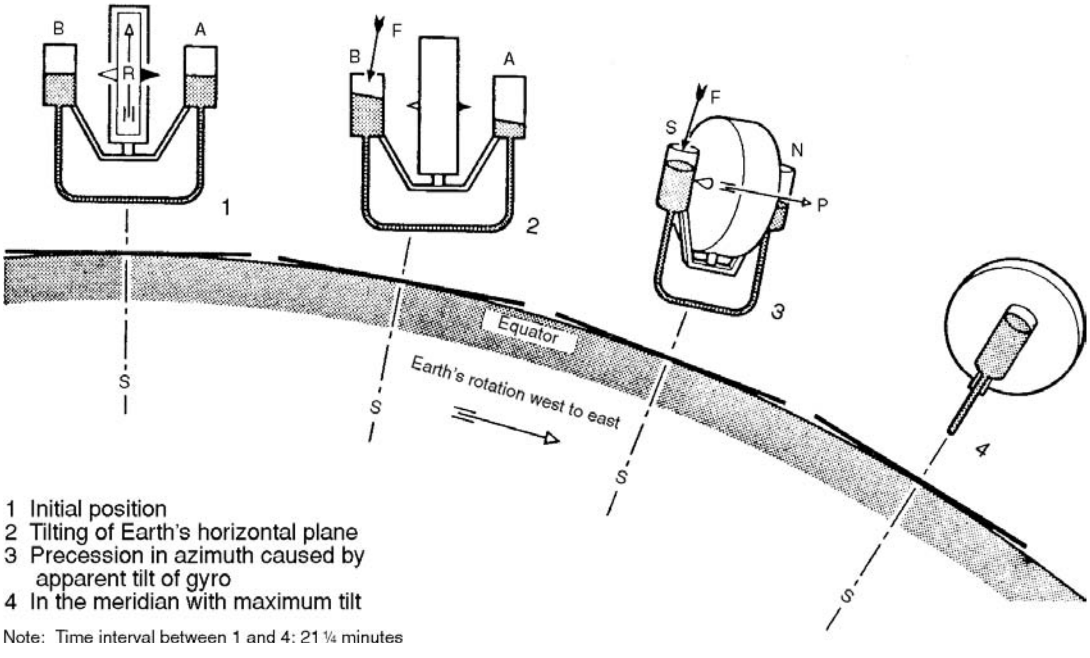


**Figure 8.11** A method of applying ‘offset damping’ to the gyro wheel. (Reproduced courtesy of Sperry Ltd.)



Note: Time interval between 1 and 4: 21 ¼ minutes

**Figure 8.12** Precession of a controlled gyroscope at the equator.

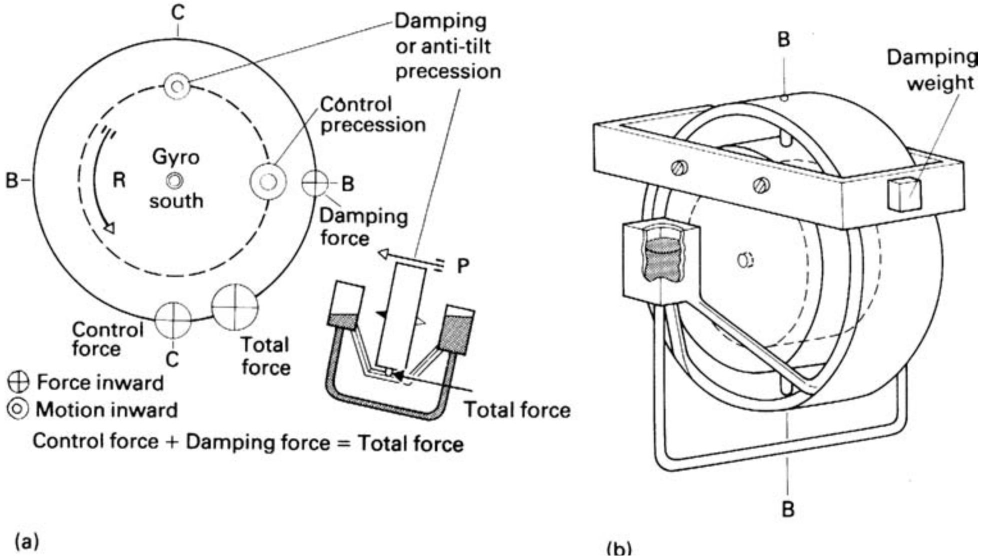


**Figure 8.13** (a) Effect of control force plus damping force.(b) An alternative method of applying offset damping. (Reproduced courtesy of Sperry Ltd.)

Damping gyroscopic precession by the use of weights provides a readily adjustable system for applying damping. The period of gyro damping is directly related to the size of the damping force, and thus the weight. If the weight is increased, the damping percentage will be increased. The effect of alternative damping application is illustrated in Figure 8.14.

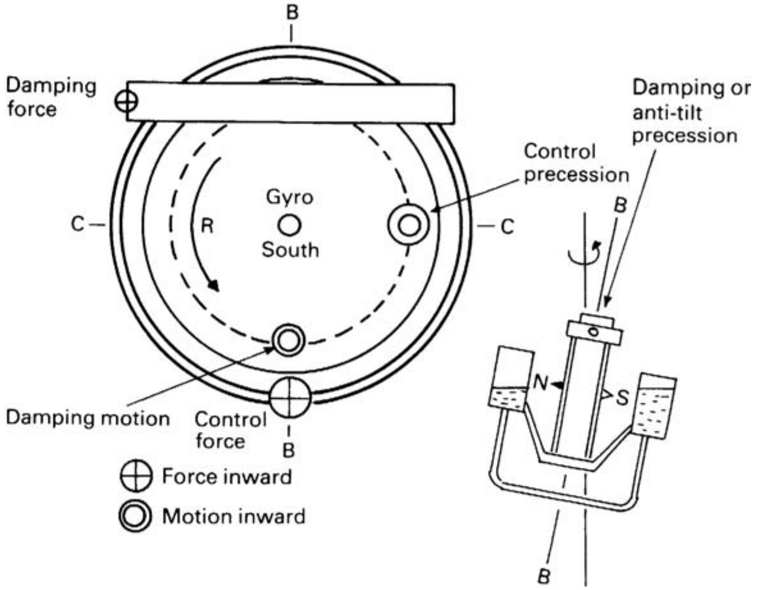


**Figure 8.14** The effects of alternative damping application.

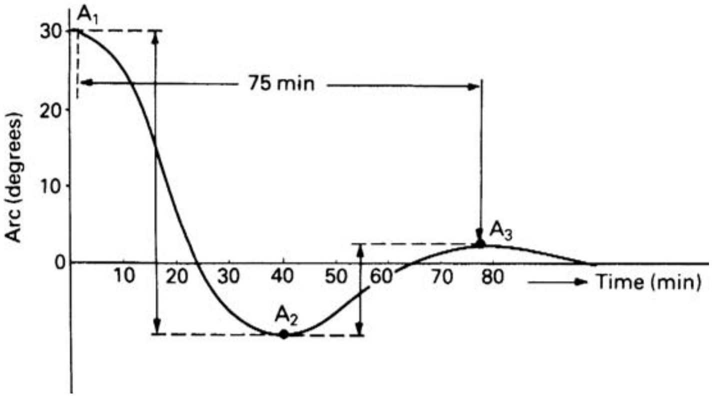


**Figure 8.15** The settling curve of a typical gyro compass with a 75-min period.

The amount of damping required depends upon the rate of tilt of the gyro axle and as such will be affected by latitude. As has been shown previously, tilt is a maximum at the equator. It follows, therefore, that damping should also be a maximum at the equator. However, the damping period will always remain constant, at approximately 86 min for some gyros, despite the change of amplitude of successive swings to east and west of the gyro axle. All gyrocompasses therefore require time to settle. Figure 8.15 shows a typical settling curve for a gyro possessing a damping period of greater than 80 min. The time taken for one oscillation, from Al to A3 is termed the natural period of the compass.

### 8.5.1 The amount of tilt remaining on a settled gyro

The settling curve traced by the north end of the gyrospin axis illustrated in Figure 8.10 assumes that the gyrocompass is situated at the equator and will, therefore, not be affected by gyro tilt. It is more likely that a vessel will be at some north/south latitude and consequently drift must be taken into account.

It has been stated that for a gyrocompass in northern latitudes, the gyrospin axis will drift to the east of the meridian and tilt upwards. For any fixed latitude the easterly drift is constant. Westerly precession, however, is directly proportional to the angle of tilt of the rotor axle from the horizontal, which itself is dependent upon the deviation between it and the meridian. At some point the easterly deviation of the north end of the spin axis produces an angle of tilt causing a rate of westerly precession that is equal and opposite to the easterly drift. The north end, although pointing to the east of the meridian, is now stabilized in azimuth.

As the north end moves easterly away from the meridian both the rate of change of the tilt angle and the angle itself are increasing. The increasing angle of tilt produces an increasing rate of downward damping tilt until a point is reached where the upward and downward rates of tilt cancel. The north end of the axle is above the horizontal although the rotor axle is stabilized. Figure 8.16 shows that the gyrocompass has settled, at point 0, to the east of the meridian and is tilted up.

The extent of the easterly and northerly (azimuth and tilt) error in the settled position is determined by latitude. An increase in latitude causes an increase in both the easterly deviation from the meridian and the angle of tilt above the horizontal. It is necessary therefore for latitude error, as the discrepancy is called, to be corrected in a gyrocompass.

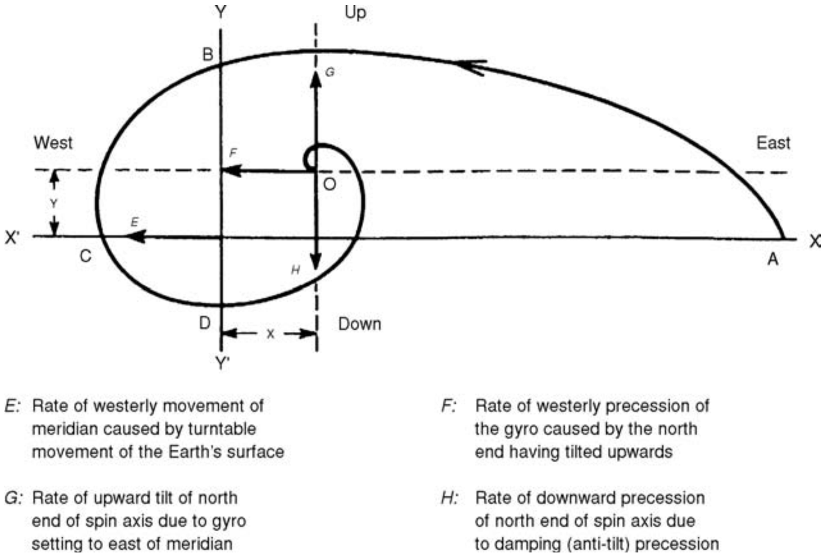


**Figure 8.16** A curve showing error to the east and tilt caused by latitude on a settled gyrocompass. X is the angle away from the meridian and Y is the angle with the horizon (tilt). (Reproduced courtesy of Sperry Ltd.)

As latitude increases, the effect of the earth's rotation becomes progressively less and consequently tilting of the rotor axle becomes less. It follows, therefore, that the rate of damping precession needed to cancel the rate of tilt, will also be less.

## 8.6 Follow-up systems

A stationary gravity-controlled gyrocompass will adequately settle close to the horizontal and near to the meridian, provided that it has freedom to move about the horizontal and vertical axes. However, if the gyrocompass is to be mounted on a ship, the base (phantom) ring needs to be capable of rotating through 360° without introducing torque about the vertical axis.

Freedom about the vertical axis is particularly difficult to achieve without introducing torque to the system. The most common way of permitting vertical-axis freedom is to mount the gyro in a vertical ring with ball bearings on the top and base plates. Obviously the weight of the unit must be borne on the lower bearing, which can create considerable friction and introduce torque. A number of methods have been developed to eliminate torque about the vertical axis. These include the use of high tensile torsion wires and buoyancy chambers, as described for each compass later in this chapter.

## 8.7 Compass errors

The accuracy of a gyrocompass is of paramount importance, particularly under manoeuvring situations where the compass is interfaced with collision-avoidance radar. An error, either existing or produced, between the actual compass reading and that presented to the radar could produce potentially catastrophic results. Assuming that the compass has been correctly installed and aligned, the static compass errors briefly listed below, should have been eliminated. They are, however, worthy of a brief mention.

### 8.7.1 Static errors

An alignment error can be:

- an error existing between the indicated heading and the vessel's lubber line
- an error existing between the indicated lubber line and the fore and aft line of the vessel.

Both of these errors can be accurately eliminated by critically aligning the compass with the ship's lubber line at installation.

#### *Transmission error*

An error existing between the indicated heading on the master compass and the heading produced by any remote repeater is a transmission error. Transmission errors are kept to a minimum by the use of multispeed pulse transmission.

#### *Variable errors*

Variable compass errors can effectively be classified into two groups.

- Dynamic errors that are caused by the angular motion of the vessel during heavy weather and manoeuvring.
- Speed/latitude errors that are caused by movement of the vessel across the earth's surface.

The magnitude of each error can be reduced to some extent as shown in the following text.

### 8.7.2 Dynamic errors

#### *Rolling error*

The gyrocompass is made to settle on the meridian under the influence of weights. Thus it will also be caused to shift due to other forces acting upon those weights. When a vessel rolls, the compass is swung like a pendulum causing a twisting motion that tends to move the plane of the sensitive element towards the plane of the swing. For a simple explanation of the error consider the surge of mercury caused in both the north and south reservoirs by a vessel rolling. If the ship is steaming due north or south, no redistribution of mercury occurs due to roll and there will be no error (see Figure 8.17).

But with a ship steaming due east or west, maximum lateral acceleration occurs in the north/south direction causing precession of the compass. However, rolls to port and starboard are equal, producing equivalent easterly and westerly precession. The resulting mean-error is therefore zero, as illustrated in Figure 8.18.

If the ship is on an intercardinal course the force exerted by the mercury (or pendulum) must be resolved into north/south and east/west components (see Figure 8.19).

The result of the combined forces is that precession of the compass occurs under the influence of an effective anticlockwise torque. Damping the pendulum system can dramatically reduce rolling error. In a top-heavy gyrocompass, this is achieved by restricting the flow of mercury between the

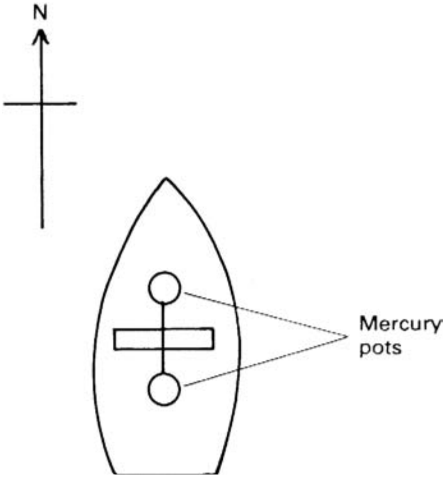


**Figure 8.17** A ship steaming due north or south produces no roll error.

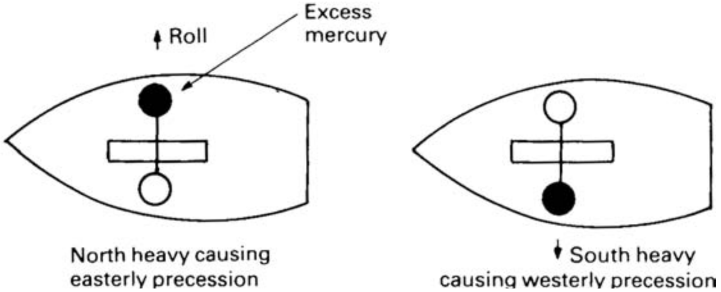


**Figure 8.18** Precession rates created by a rolling vessel on an east/west course are equal and will cancel.

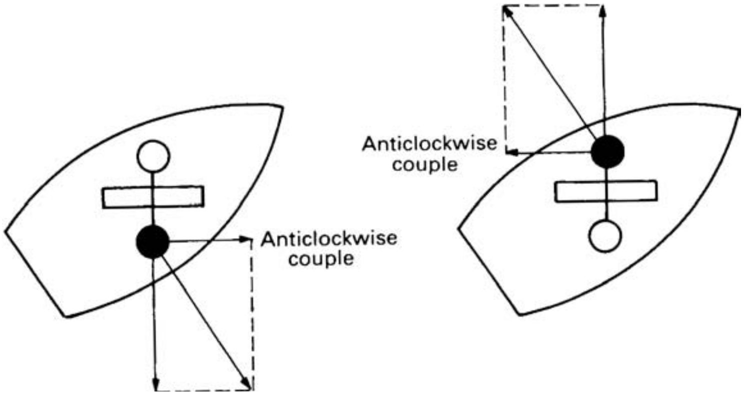


**Figure 8.19** For a vessel on an intercardinal course, rolling produces an anticlockwise torque.

two pots. The damping delay introduced needs to be shorter than the damping period of the compass and much greater than the period of roll of the vessel. Both of these conditions are easily achieved.

Electrically-controlled compasses are roll-damped by the use of a viscous fluid damping the gravity pendulum. Such a fluid is identified by a manufacturer's code and a viscosity number. For example, in the code number 200/20, 200 refers to the manufacturer and 20 the viscosity. A higher second number indicates a more viscous silicon fluid. One viscous fluid should never be substituted for another bearing a different code number. Additionally since roll error is caused by lateral acceleration, mounting the gyrocompass low in the vessel and as close as possible to the centre of roll will reduce this error still further.

### *Manoeuvring (ballistic) error*

This error occurs whenever the ship is subject to rapid changes of speed or heading. Because of its pendulous nature, the compass gravity control moves away from the centre of gravity whenever the vessel changes speed or alters course. Torque's produced about the horizontal and vertical axis by manoeuvring cause the gyro mechanism to precess in both azimuth and tilt. If the ship is steaming due north and rapidly reducing speed, mercury will continue to flow into the north pot, or the gravity pendulum continues to swing, making the gyro spin axis north heavy and thus causing a precession in azimuth.

In Figure 8.20 the decelerating vessel causes easterly precession of the compass. Alternatively if the ship increases speed the compass precesses to the west.

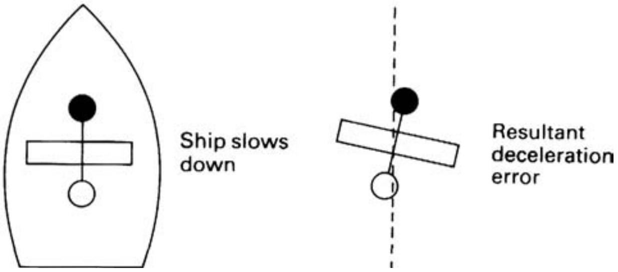


**Figure 8.20** Resultant easterly error caused by the vessel slowing down.

### *Latitude (damping) error*

Latitude error is a constant error, the magnitude of which is directly proportional to the earth's rotation at any given latitude. It is, therefore, present even when the ship is stationary. As has previously been stated, a gyrocompass will always settle close to the meridian with an error in tilt. To maintain the gyro pointing north it must be precessed at an angular rate varying with latitude. At the equator the earth's linear speed of rotation is about 900 knots and rotation from west to east causes a fixed point to effectively move at 900 × cos (latitude) knots in an easterly direction. For any latitude ($\lambda$) the rate of earth spin is $\omega = 15°\,h^{-1}$. This may be resolved into two components, one about the true vertical at a given latitude ($\omega \sin \lambda$) and the other about the north/south earth surface horizontal at a given latitude ($\omega \cos \lambda$) as illustrated in Figure 8.21.

The component of the earth's rotation about the north/south horizontal may be resolved further into two components mutually at right angles to each other. The first component is displaced $a°$ to the east

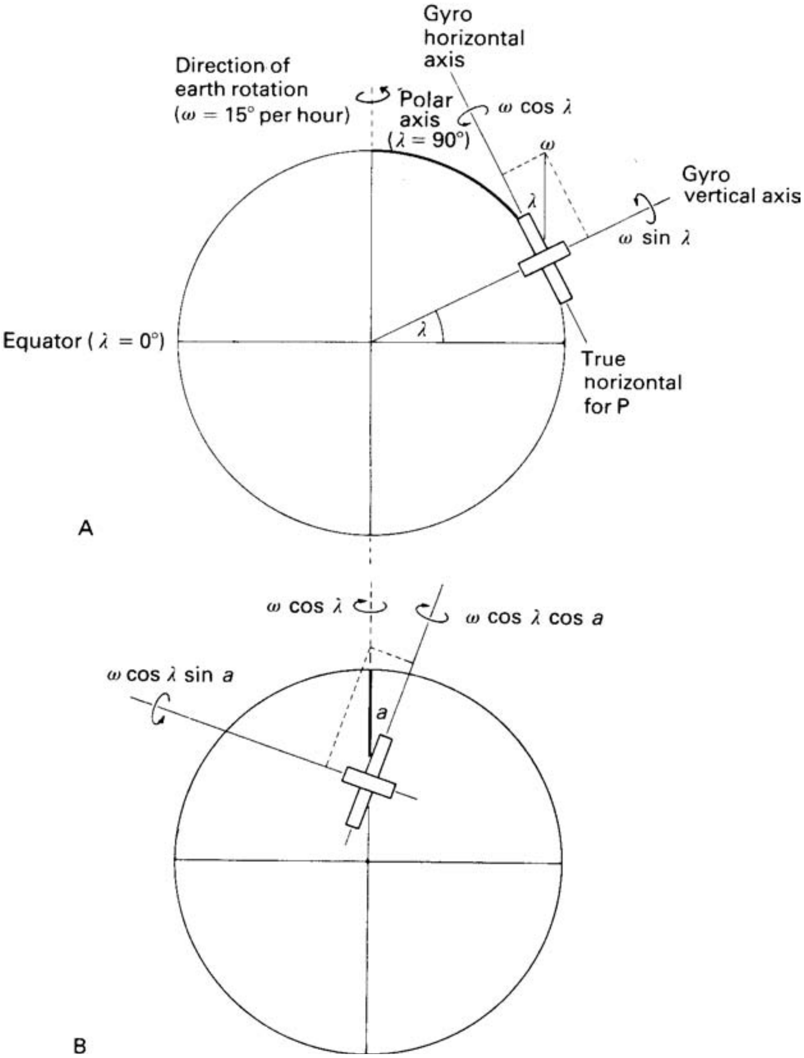


**Figure 8.21** Apparent movement of a gyro. (Reproduced courtesy S. G. Brown Ltd.)

of the meridian producing a rate of spin ω cos λ sin *a*°, whilst the other is 90 – *a*° to the west of north to produce a rate of spin ω cos λ cos *a*°.

Correction for latitude error requires that a torque be applied to precess the gyro at an angular rate, varying with latitude, to cancel the error. This will be an external correction that can be either mechanical or electronic. For mechanical correction, a weight on the gyro case provides the necessary torque. The weight, or 'mechanical latitude rider', is adjustable thus enabling corrections to be made for varying latitudes. Another method of mechanical correction is to move the lubber line by an amount equal to the error. Latitude correction in a bottom-weighted compass is achieved by the introduction of a signal proportional to the sine of the vessel's latitude, causing the gyro ball to precess in azimuth at a rate equal and opposite to the apparent drift caused by earth rotation.

### *Speed and course error*

If a vessel makes good a northerly or southerly course, the north end of the gyro spin axis will apparently tilt up or down since the curvature of the earth causes the ship to effectively tilt bows up or down with respect to space. Consider a ship steaming due north. The north end of the spin axis tilts upwards causing a westerly precession of the compass, which will finally settle on the meridian with some error in the angle, the magnitude of which is determined by the speed of the ship. On a cardinal course due east or west, the ship will display a tilt in the east/west plane of the gyro and no tilting of the gyro axle occurs – hence no speed error is produced. The error varies, therefore, with the cosine of the ship's course. Speed/course gyrocompass error magnitude must also be affected by latitude and will produce an angle of tilt in the settled gyro. Hence latitude/course /speed error is sometimes referred to as LCS error.

## 8.7.3 Use of vectors in calculating errors

With reference to Figure 8.22,

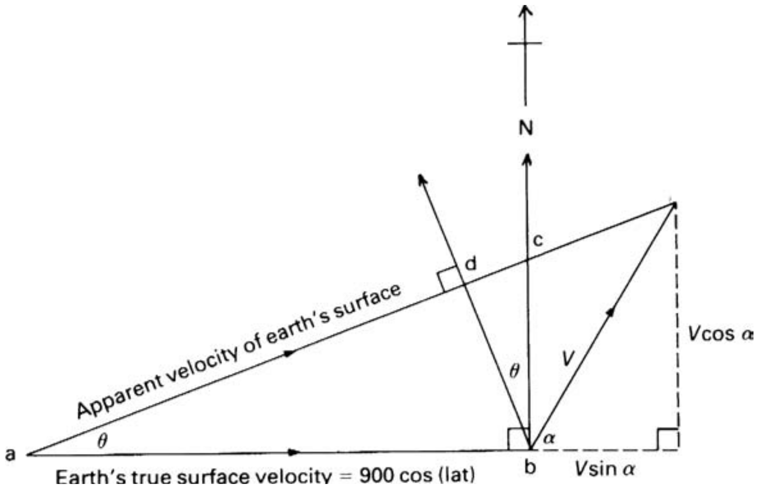


**Figure 8.22** Use of vectors in calculating errors

$V$ = ships speed in knots
$V \sin \alpha$ = easterly component of speed
$\alpha$ = ships course
$V \cos \alpha$ = northerly component of speed
angle acb = angle dcb
angle abc = angle bdc = 90°
angle bac = angle cbd = $\theta$ = error

In triangle abc:

$$\text{Error in degrees} = \text{angle bac} = \theta = \tan^{-1} \frac{V \cos(\text{course})}{900 \cos(\text{latitude}) + V \sin(\text{course})}$$

Obviously the ship's speed is very much less than the earth's surface velocity therefore:

$$\tan\theta \simeq \frac{V\cos(\text{course})}{900\cos(\text{latitude})}$$

The angle $\theta$ may be approximately expressed in degrees by multiplying both side of the equation by a factor of 60. Now:

$$\text{approximate error in degrees} = \frac{V\cos(\text{course})}{15\cos(\text{latitude})}$$

## 8.8 Top-heavy control master compass

Produced before the move towards fully sealed gyro elements, the Sperry SR120 gyrocompass (Figure 8.23) is a good example of an early top-heavy controlled system. The master compass consists of two main assemblies, the stationary element and the movable element.

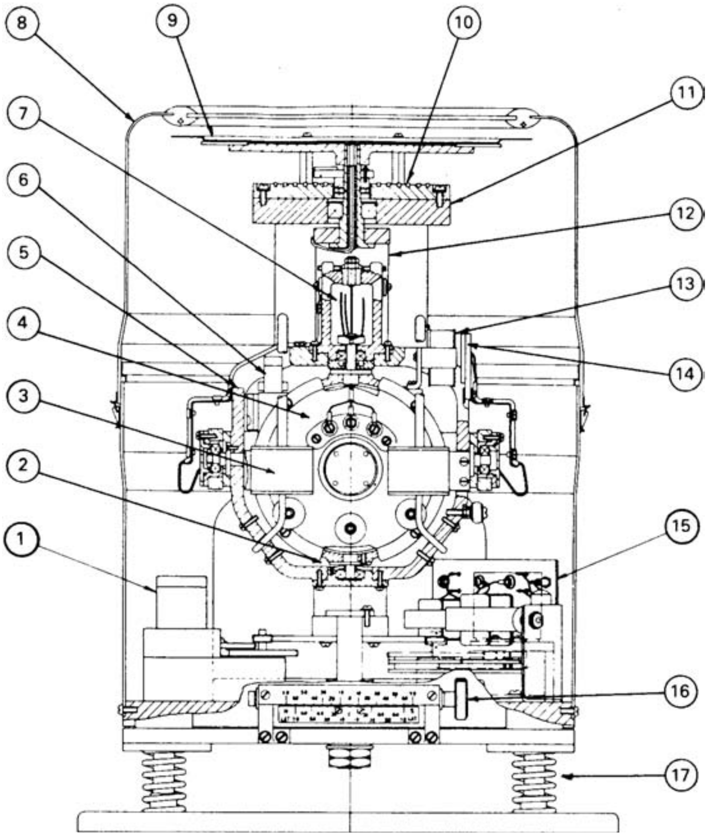


**Figure 8.23** A south elevation sectional view of a Sperry master compass . Key:1. Stepper transmitter; 2. Support ball bearings; 3. Ballistic pots; 4.Rotor (encased); 5.Rotor case; 6. Damping weight; 7. Suspension wire; 8. Cover; 9.Compass card; 10.Slip rings; 11. Main support frame; 12. Phantom ring support assembly (cutaway); 13. Follow-up primary transformer; 14. Follow-up secondary transformer; 15. Follow-up amplifier; 16. Latitude corrector; 17. Spring/shock absorber assembly.

### 8.8.1 The stationary element

This is the main supporting frame that holds and encases the movable element. It consists of the main frame and base, together with the binnacle and mounting shock absorbers. The top of the main support frame (11) (Figure 8.23) holds the slip rings, lubber line and the scale illumination circuitry, whilst the main shaft, connected to the phantom ring (12), protrudes through the supporting frame to hold a compass card that is visible from above.

A high quality ball bearing race supports the movable element on the base of the main support frame in order that movement in azimuth can be achieved. The base of the whole assembly consists of upper and lower base plates that are connected at their centre by a shaft. Rotation of the upper plate in relation to the lower plate enables mechanical latitude correction to be made. The latitude corrector (16) is provided with upper and lower latitude scales graduated in 10 units, up to 70° north or south latitude, either side of zero. Latitude correction is achieved by mechanically rotating the movable element relative to the stationary element thus producing a shift in azimuth. The fixed scale of the latitude adjuster (16) is secured to the stationary element with a second scale fixed to the movable element. To set the correction value, which should be within 5° of the ship's latitude, is simply a matter of aligning the ship's latitude on the lower scale with the same indication on the upper scale of the vernier scale.

Also supported by the base plate are the azimuth servomotor and gear train, and the bearing stepper transmitter.

### 8.8.2 The movable element

With the exception of the phantom ring, the movable element is called the sensitive element (Figure 8.24). At the heart of the unit is the gyro rotor freely spinning at approximately 12 000 rpm. The rotor is 110 mm in diameter and 60 mm thick and forms, along with the stator windings, a three-phase induction motor. Gyroscopic inertia is produced by the angular momentum of the rapidly spinning heavy rotor. Rotation is counter clockwise (counter earthwise) when viewed from the south end.

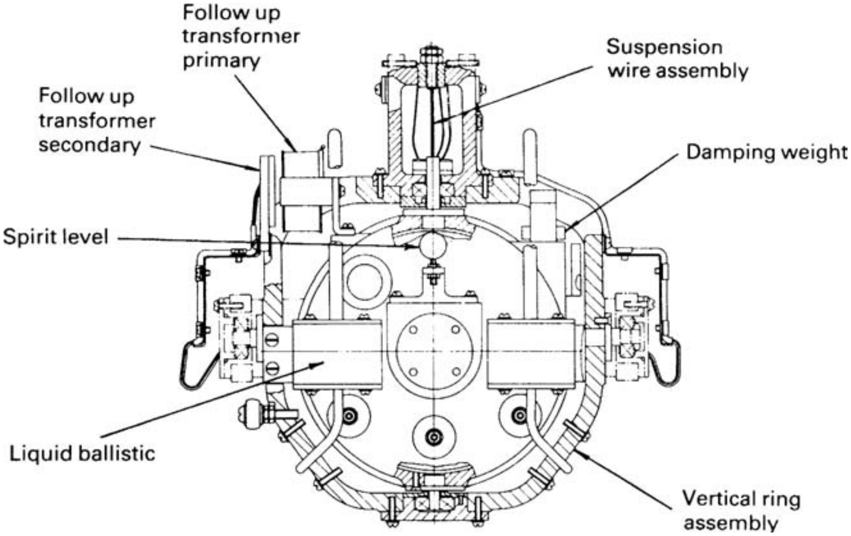


**Figure 8.24** The compass sensitive element.

A sensitive spirit level graduated to represent 2 min of arc, is mounted on the north side of the rotor case. This unit indicates the tilt of the sensitive element. A damping weight is attached to the west side of the rotor case in order that oscillation of the gyro axis can be damped and thus enable the compass to point north.

The rotor case is suspended, along the vertical axis, inside the vertical ring frame by means of the suspension wire (7). This is a bunch of six thin stainless steel wires that are made to be absolutely free from torsion. Their function is to support the weight of the gyro and thus remove the load from the support bearings (2).

### 8.8.3 Tilt stabilization (liquid ballistic)

To enable the compass to develop a north-seeking action, two ballistic pots (3) are mounted to the north and south sides of the vertical ring. Each pot possesses two reservoirs containing the high density liquid 'Daifloil'. Each north/south pair of pots is connected by top and bottom pipes providing a total liquid/air sealed system that operates to create the effect of top heaviness.

Because the vertical ring and the rotor case are coupled to each other, the ring follows the tilt of the gyro spin axis. Liquid in the ballistic system, when tilted, will generate a torque which is proportional to the angle of the tilt. The torque thus produced causes a precession in azimuth and starts the north-seeking action of the compass.

### 8.8.4 Azimuth stabilization (phantom ring assembly)

Gyro freedom of the north/south axis is enabled by the phantom ring and gearing. This ring is a vertical circle which supports the north/south sides of the horizontal ring (on the spin axis) by means of high precision ball bearings.

A small oil damper (6) is mounted on the south side of the sensitive element to provide gyro stabilization during the ship's pitching and rolling.

The compass card is mounted on the top of the upper phantom ring stem shaft and the lower stem shaft is connected to the support ball bearings enabling rotation of the north/south axis. The azimuth gearing, located at the lower end of the phantom ring, provides freedom about this axis under a torque from the azimuth servomotor and feedback system.

### 8.8.5 Azimuth follow-up system

The system shown in Figure 8.25 enables the phantom ring to follow any movement of the vertical ring. The unit senses the displacement signal produced by misalignment of the two rings, and amplifies the small signal to a power level of sufficient amplitude to drive the azimuth servo rotor. Movement of the azimuth servo rotor causes rotation, by direct coupling, of the phantom ring assembly in the required direction to keep the two rings aligned.

The sensing element of the follow-up system is a transformer with an 'E'-shaped laminated core and a single primary winding supplied with a.c., and two secondary windings connected as shown in Figure 8.25. With the 'E'-shaped primary core in its central position, the phase of the e.m.f.s induced in the two secondaries is such that they will cancel, and the total voltage produced across R1 is the supply voltage only. This is the stable condition during which no rotation of the azimuth servo rotor occurs. If there is misalignment in any direction between the phantom and the vertical rings, the two e.m.f.s induced in the two secondaries will be unbalanced, and the voltage across R1 will increase or decrease accordingly.

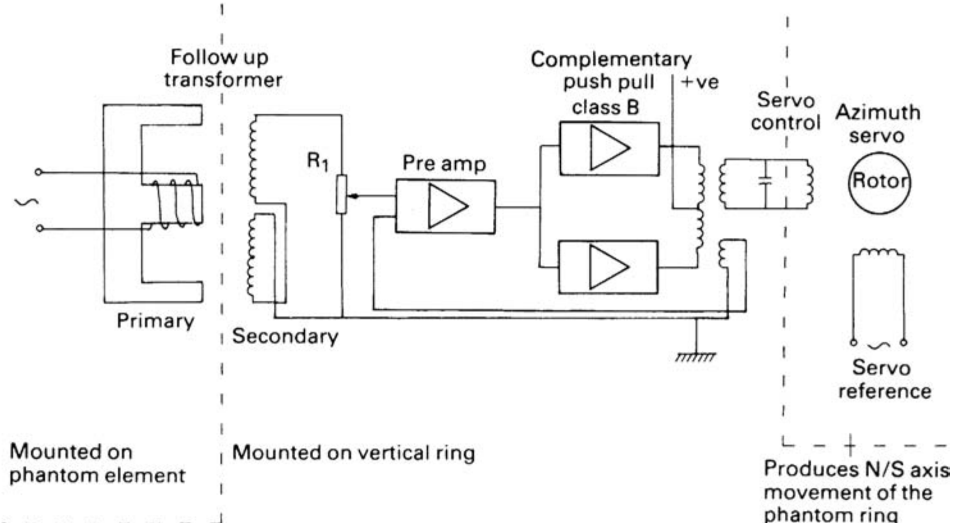


**Figure 8.25** The Sperry compass azimuth follow-up circuit.

This error signal is pre-amplified and used to drive a complementary push/pull power amplifier producing the necessary signal level to cause the azimuth servo to rotate in the required direction to re-align the rings and thus cancel the error signal. Negative feedback from T2 secondary to the pre-amplifier ensures stable operation of the system.

Another method of azimuth follow-up control was introduced in the Sperry SR220 gyrocompass (Figure 8.26).

In practice only a few millimetres separate the sphere from the sensitive element chamber. The point of connection of the suspension wire with the gyrosphere, is deliberately made to be slightly above the centre line of the sphere on the east–west axis. At the north and south ends of the horizontal axis are

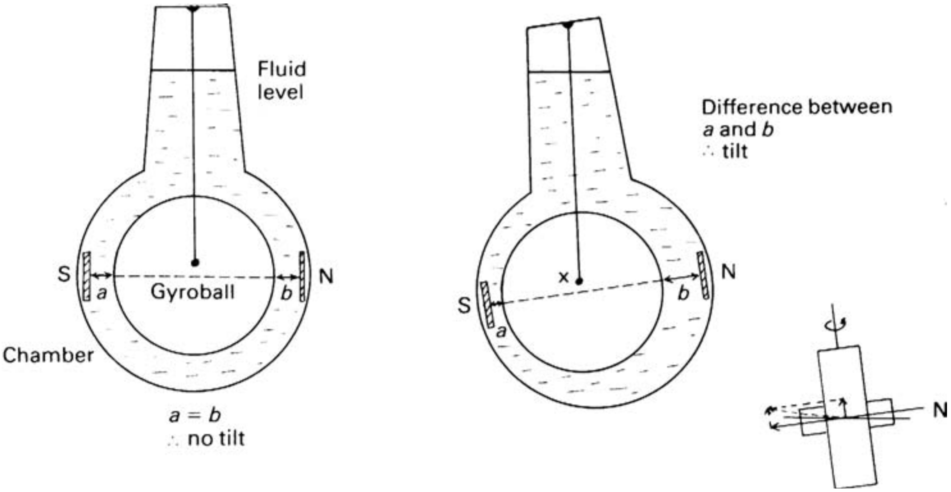


**Figure 8.26** Simplified diagrams of the gyroball action in the Sperry SR220 gyrocompass.

mounted the primary coils of the follow-up pick-off transformers. With no tilt present, the sphere centre line will be horizontal and central causing distance *a* to be equal to distance *b* producing equal amplitude outputs from the follow-up transformers which will cancel. Assuming the gyrocompass is tilted up and to the east of the meridian, the gyrosphere will take up the position shown in Figure 8.26. The sphere has moved closer to the south side of the chamber producing a difference in the distances *a* and *b*. The two pick-off secondary coils will now produce outputs that are no longer in balance. Difference signals thus produced are directly proportional to both azimuth and tilt error.

Each pick-off transformer is formed by a primary coil mounted on the gyrosphere and secondary pick-off coils mounted on the sensitive element assembly. The primary coils provide a magnetic field, from the 110 V a.c. supply used for the gyrowheel rotor, which couples with the secondary to produce e.m.f.s depending upon the relationship between the two coils.

Figure 8.27 shows that the secondary coils are wound in such a way that one or more of the three output signals is produced by relative movement of the gyrosphere. $X$ = a signal corresponding to the distance of the sphere from each secondary coil; $\phi$ = a signal corresponding to vertical movement; and $\theta$ = a signal corresponding to horizontal movement

In the complete follow-up system shown in Figure 8.28, the horizontal servomechanism, mounted on the west side of the horizontal ring, permits the sensitive element to follow-up the gyrosphere about the horizontal axis. This servo operates from the difference signal produced by the secondary pick-off coils, which is processed to provide the amplitude required to drive the sensitive element assembly in

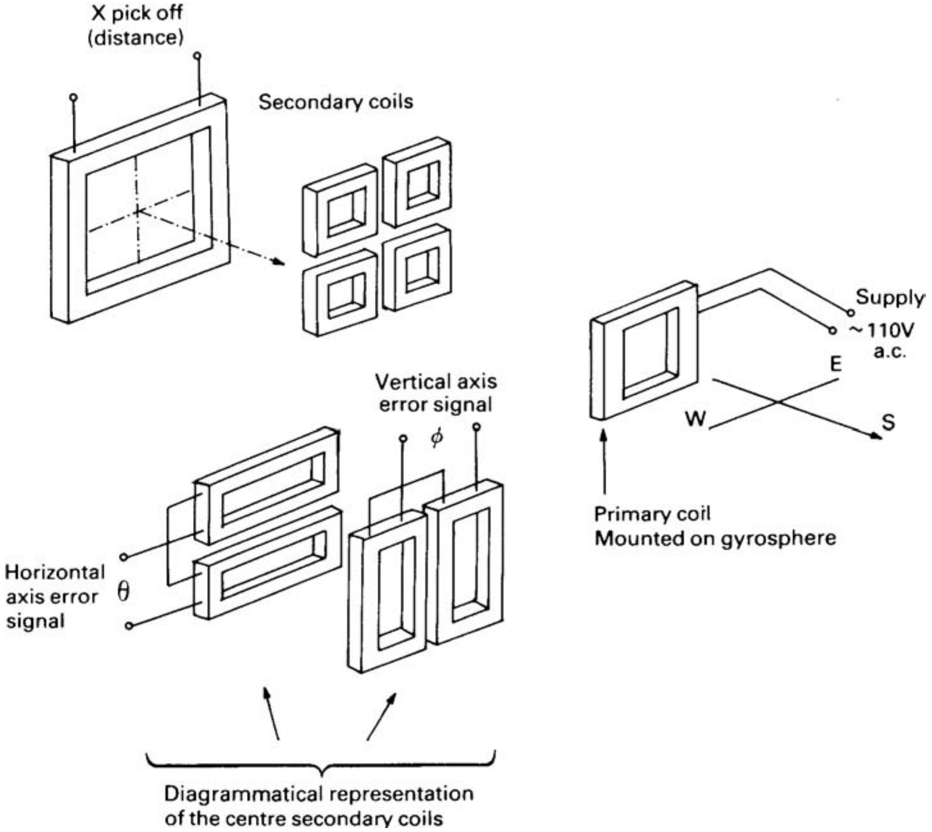


**Figure 8.27** Follow-up signal pick-off coils.

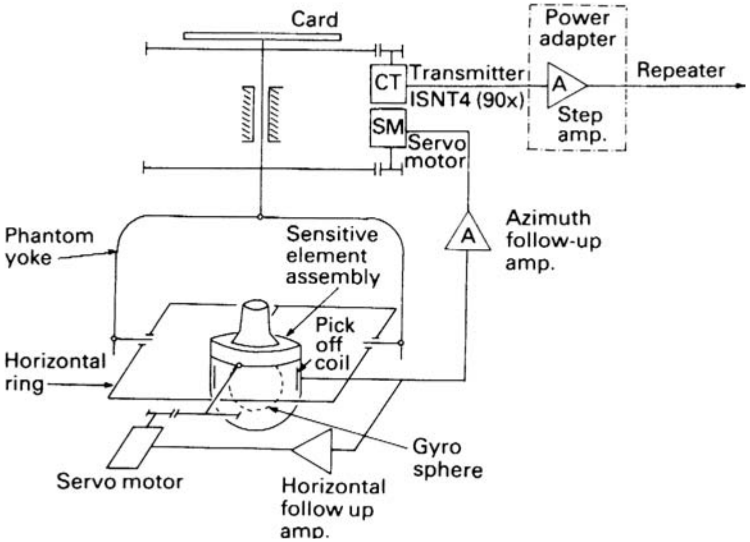


**Figure 8.28** The Sperry SR220 follow-up system.

azimuth by rotating the phantom yoke assembly in the direction needed to cancel the error signal. In this way the azimuth follow-up circuit keeps the gyrosphere and sensitive element chamber in alignment as the gyro precesses.

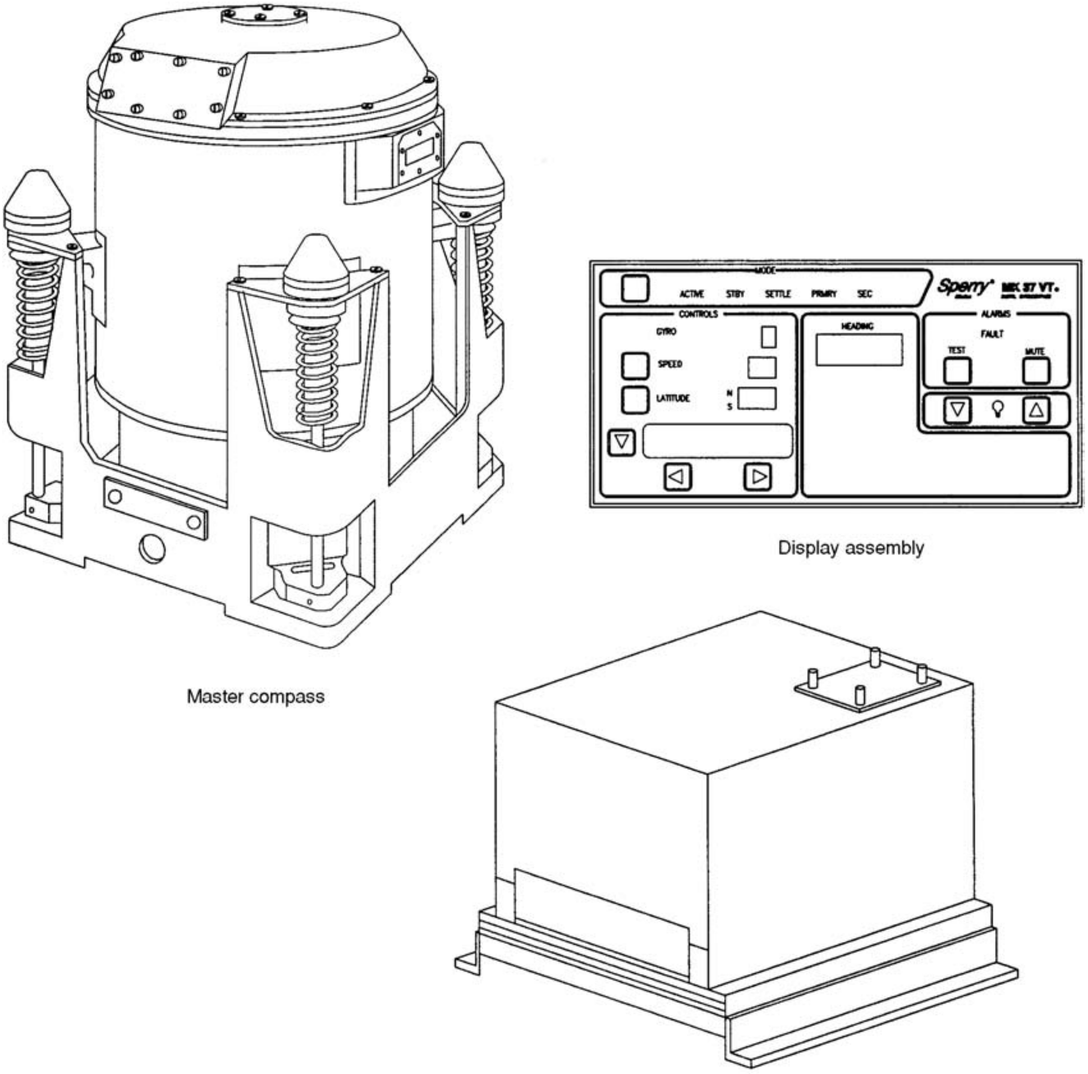
## 8.9 A digital controlled top-heavy gyrocompass system

In common with all other maritime equipment, the traditional gyrocompass is now controlled by a microcomputer. Whilst such a system still relies for its operation on the traditional principles already described, most of the control functions are computer controlled. The Sperry MK 37 VT Digital Gyrocompass (Figure 8.29) is representative of many gyrocompasses available. The system has three main units, the sealed master gyrocompass assembly, the electronics unit and the control panel.

The master compass, a shock-mounted, fluid-filled binnacle unit, provides uncorrected data to the electronics units which processes the information and outputs it as corrected heading and rate of turn data. Inside the three-gimbals mounting arrangement is a gyrosphere that is immersed in silicone fluid and designed and adjusted to have neutral buoyancy. This arrangement has distinct advantages over previous gyrocompasses.

- The weight of the gyrosphere is removed from the sensitive axis bearings.
- The gyrosphere and bearings are protected from excessive shock loads.
- Sensitivity to shifts of the gyrosphere's centre of mass, relative to the sensitive axis, is eliminated.
- The effects of accelerations are minimized because the gyrosphere's centre of mass and the centre of buoyancy are coincident.

The system's applications software compensates for the effects of the ship's varying speed and local latitude in addition to providing accurate follow-up data maintaining yoke alignment with the gyrosphere during turn manoeuvres.



Master compass

Display assembly

Electronics control unit

**Figure 8.29** Sperry Mk 37 VT digital gyrocompass equipment. (Reproduced courtesy of Litton Marine Systems.)

### 8.9.1 Control panel

All command information is input via the control panel, which also displays various data and system indications and alarms (see Figure 8.30).

The Mode switch, number 1, is fixed when using a single system, the Active indicator lights and a figure 1 appear in window 13. Other Mode indicators include: 'STBY', showing when the gyrocompass is in a dual configuration and not supplying outputs; 'Settle', lights during compass start-up; 'Primary', lights to show that this is the primary compass of a dual system; and 'Sec', when it is the secondary unit.

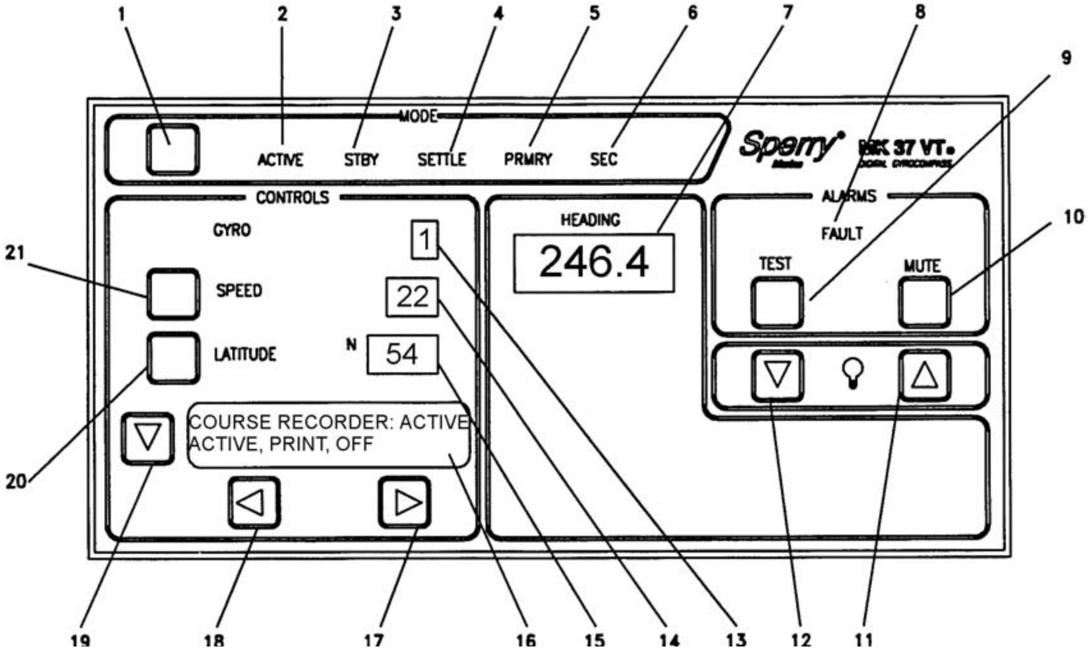


**Figure 8.30** Sperry MK 37 VT control panel. (Reproduced courtesy of Litton Marine Systems.)

Number 7 indicates the Heading display accurate to within 1/10th of a degree. Other displays are: number 14, speed display to the nearest knot; number 15,latitude to the nearest degree; and 16, the data display, used to display menu options and fault messages. Scroll buttons 17, 18 and 19 control this display. Other buttons functions are self-evident.

### 8.9.2 System description

Figure 8.31 shows, to the left of the CPU assembly, the gyrosphere with all its control function lines, and to the right of the CPU the Display and Control Panel and output data lines.

The gyrosphere is supported by a phantom yoke and suspended below the main support plate. A 1-speed synchro transmitter is mounted to the support plate, close to the azimuth motor, and is geared to rotate the compass dial. The phantom yoke supports the east–west gimbal assembly through horizontal axis bearings. To permit unrestricted movement, electrical connections between the support plate and the phantom yoke are made by slip rings. The east–west gimbal assembly supports the vertical ring and horizontal axis bearings. See Figure 8.32.

#### *The gyrosphere*

The gyrosphere is 6.5 inches in diameter and is pivoted about the vertical axis within the vertical ring, which in turn is pivoted about the horizontal axis in the east–west gimbal assembly. At operating temperature, the specific gravity of the sphere is the same as the liquid ballistic fluid in which it is immersed. Since the sphere is in neutral buoyancy, it exerts no load on the vertical bearings. Power to drive the gyro wheel is connected to the gyrosphere from the vertical ring through three spiral hairsprings with a fourth providing a ground connection.

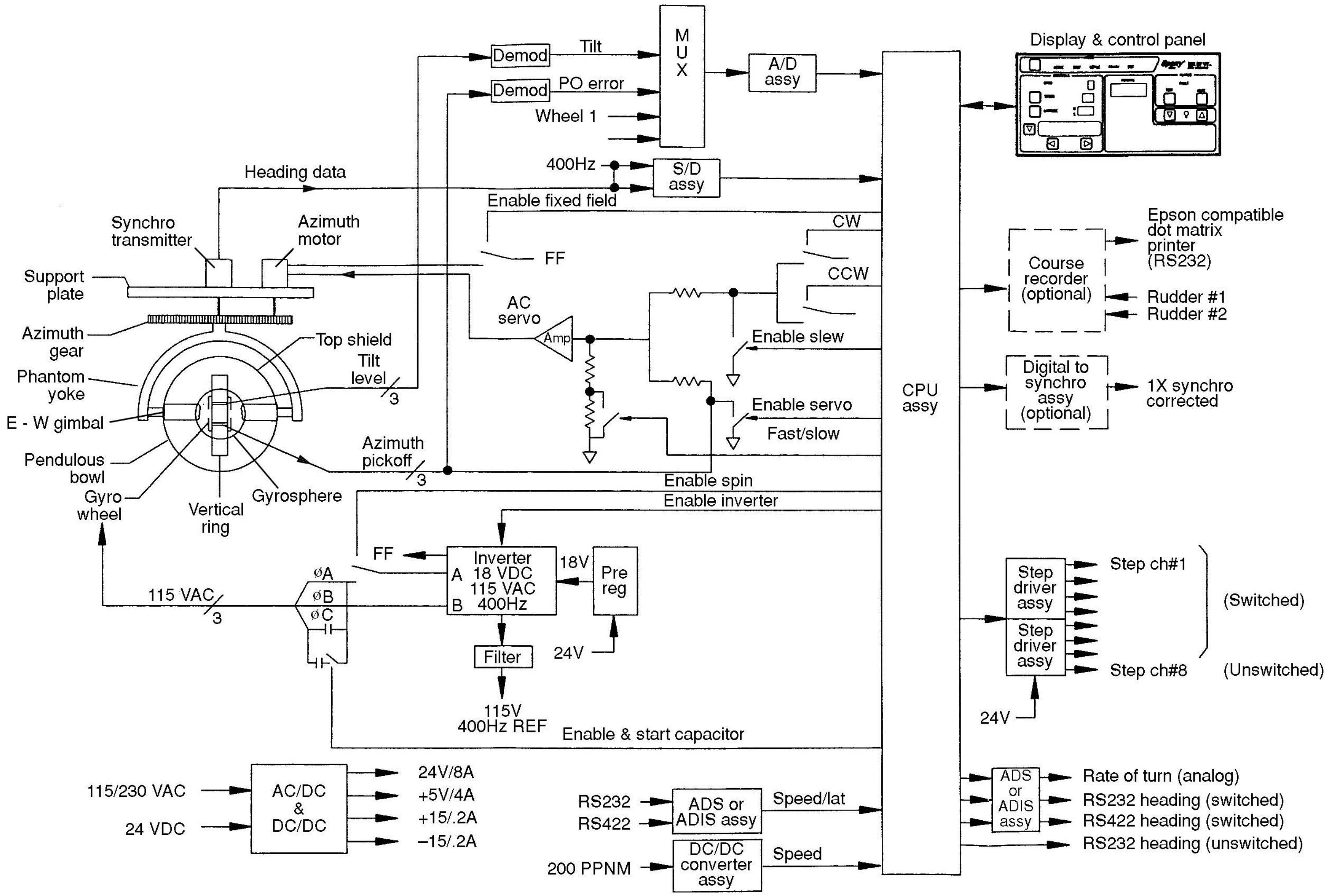


**Figure 8.31** Overall functional block diagram. (Reproduced courtesy of Litton Marine Systems.)

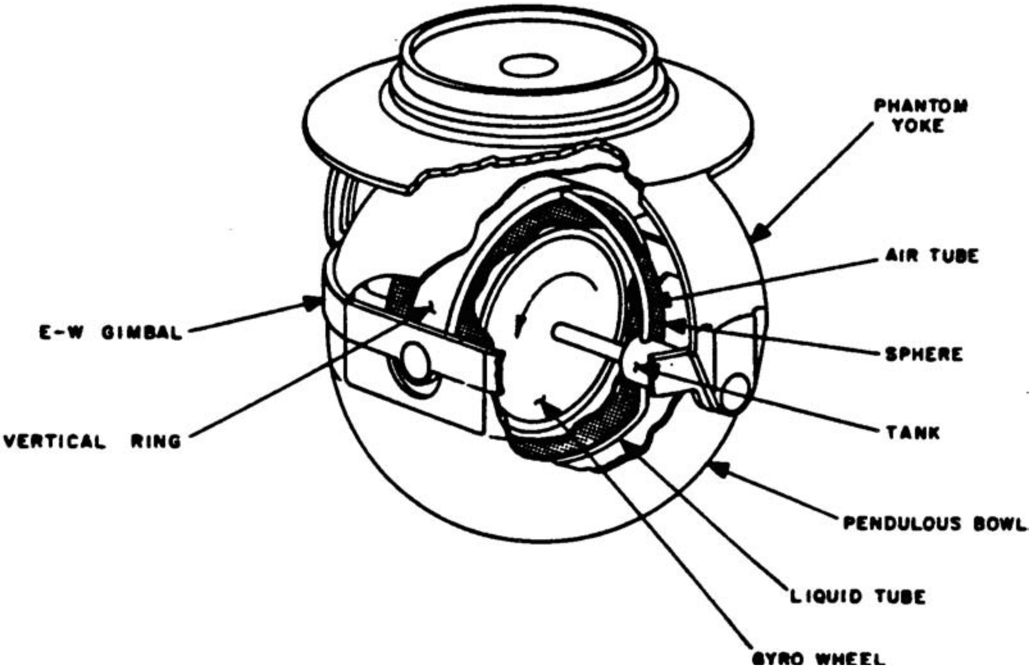


**Figure 8.32** Ballistic system of the Sperry MK 37 VT gyrocompass. (Reproduced courtesy of Litton Marine Systems.)

The liquid ballistic assembly, also known as the control element because it is the component that makes the gyrosphere north-seeking, consists of two interconnected brass tanks partially filled with silicon oil. Small-bore tubing connects the tanks and restricts the free flow of fluid between them. Because the time for fluid to flow from one tank to the other is long compared to the ship's roll period, roll acceleration errors are minimized.

### *Follow-up control*

An azimuth pick-off signal, proportional to the azimuth movement of the vertical ring, is derived from an E-core sensor unit and coupled back to the servo control circuit and then to the azimuth motor mounted on the support plate. When an error signal is detected the azimuth motor drives the azimuth gear to cancel the signal.

Heading data from the synchronous transmitter is coupled to the synchro-to-digital converter (S/D ASSY) where it is converted to a 14-bit word before being applied to the CPU. The synchro heading data, 115 V a.c., 400 Hz reference, 90 V line-to-line format, is uncorrected for ship's speed error and latitude error. Corrections for these errors are performed by the CPU using the data connected by the analogue, digital, isolated serial board (ADIS) from an RS-232 or RS-422 interface.

### *Interface data*

Compass interfacing with external peripheral units is done using NMEA 0183 format along RS-232 and RS-422 lines. Table 8.1 shows data protocols.

**Table 8.1** Sperry MK37 digital gyrocompass I/O protocols. (Reproduced courtesy of Litton Marine Systems)

| | | |
|---|---|---|
| *Inputs* | | |
| Speed: | Pulsed | Automatic. 200 ppnm |
| | Serial | Automatic from digital sources. RS-232/422 in NMEA 0183 format $VBW, $VHW, $VTG |
| | Manual | Manually via the control panel |
| Latitude | | Automatic from the GPS via RS-232/422 in NMEA format $GLL, $GGA<br>Automatic from digital sources via RS-232/422 in NMEA 0183 format $GLL<br>Manually via the control panel |
| *Outputs* | | |
| Rate of Turn | | 50 mV per deg/min (±4.5 VDC full scale = ± 90°/min) NMEA 0183 format $HEROT, X.XXX, A*hh<CR><LF> 1 Hz, 4800 baud |
| Step Repeaters | | Eight 24 VDC step data outputs. (An additional 12-step data output at 35 VDC or 70 VDC from the optional transmission unit)<br>7 – switched, 1 – unswitched |
| Heading Data | | One RS-422, capable of driving up to 10 loads in NMEA 0183 format $HEHDT, XXX.XXX, T*hh<CR><LF><br>Two RS-232, each capable of driving one load in NMEA 0183 format $HEHDT, XXX>XXX, T*hh<CR><LF> 10 Hz, 4800 baud<br>1 – 232 switched, 1 – 232 unswitched, 1 – 422 switched |
| Alarm Outputs | | A relay and a battery-powered circuit activates a fault indicator and audible alarm during a power loss.<br>Compass alarm – NO/NC contacts. Power alarm – NO/NC contacts |
| Course Recorder | | (If fitted) RS–232 to dot matrix printer |
| Synchro Output | | (If fitted) 90 V line-to-line with a 115 VAC 400 Hz reference. Can be switch or unswitched |

### CPU assembly

The heart of the electronic control and processing system, the CPU, is a CMOS architectured arrangement communicating with the Display and Control Panel and producing the required outputs for peripheral equipment. Two step driver boards allow for eight remote heading repeaters to be connected. Output on each channel is a + 24 V d.c. line, a ground line and three data lines D1, D2 and D3. Each three-step data line shows a change in heading, as shown in Table 8.2.

### Scheduled maintenance and troubleshooting

The master compass is completely sealed and requires no internal maintenance. As with all computer-based equipment the Sperry MK 37 VT gyrocompass system possesses a built-in test system (BITE) to enable health checks and first line trouble shooting to be carried out. Figure 8.33 shows the trouble analysis chart for the Sperry MK 37 VT system. In addition to the health check automatically carried out at start-up, various indicators on the control panel warn of a system error or malfunction. Referring to the extensive information contained in the service manual it is possible to locate and in some cases remedy a fault.

**Table 8.2** Step data lines output

| *Step data* | | | *Step fraction* | *Heading* |
|---|---|---|---|---|
| *D3* | *D2* | *D1* | | |
| 0 | 0 | 1 | 0/6 | Decrease |
| 1 | 0 | 1 | 1/6 | ↑ |
| 1 | 0 | 0 | 2/6 | |
| 1 | 1 | 0 | 3/6 | |
| 0 | 1 | 0 | 4/6 | ↓ |
| 0 | 1 | 1 | 5/6 | Increase |

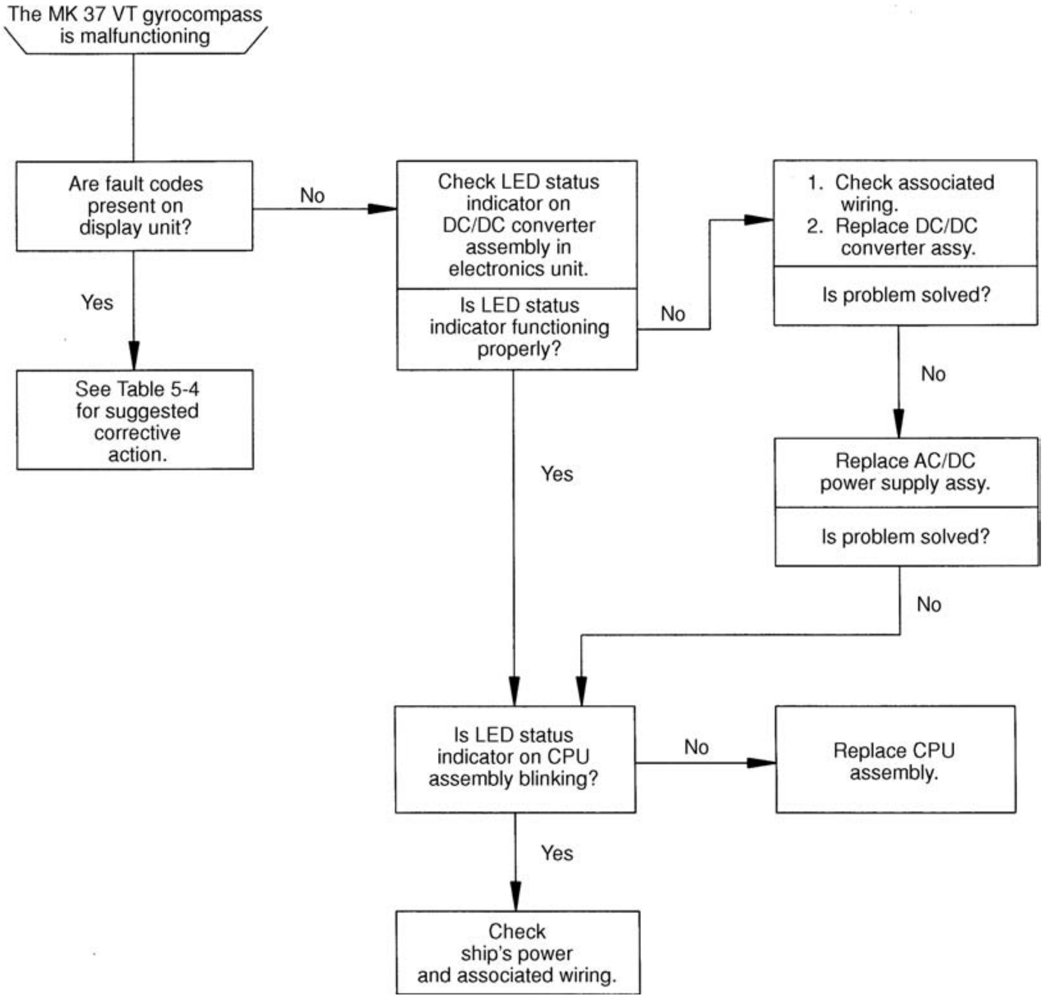


**Figure 8.33** Sperry MK 37 VT digital gyrocompass trouble analysis chart. (Reproduced courtesy of Litton Marine Systems.)

**Table 8.3** Part of a fault location chart for the Sperry MK 37 VT Compass. (Reproduced courtesy of Litton Marine Systems)

| *Symptom* | *Probable cause* | *Remedy* |
|---|---|---|
| Course recorder leaves a blank page every 8–10 inches or has paper feed problems | Printer paper-release lever not in the middle, push-tractor position | Place level in the middle position for push-tractor installation |
| Repeater does not follow MK 37 VT heading | Repeater channel may not be on or not synchronized to the MK 37 VT heading | Check repeater switch on step driver assembly. Make sure repeater is synchronized to the MK 37 VT gyrocompass |
| Speed value does not change | Speed selection may not be in Auto | Verify that speed menu selection is in Auto. Check for faults on serial channel |
| Latitude value does not change | Latitude selection may not be in Auto | Verify that latitude menu selection is in Auto. Check for faults on serial channel |
| Manual transfer (dual system) does not occur | Other system may not be powered, attached, or may have a critical fault. Manual transfer must be initiated from the primary compass only | Verify that other system is powered, attached, and does not have a critical fault |
| Unit makes buzzing sound for at least 15 min after being switched on | If sound persists longer than 15 min, the ac/dc power supply assembly relay is bad | Replace ac/dc power supply assembly |

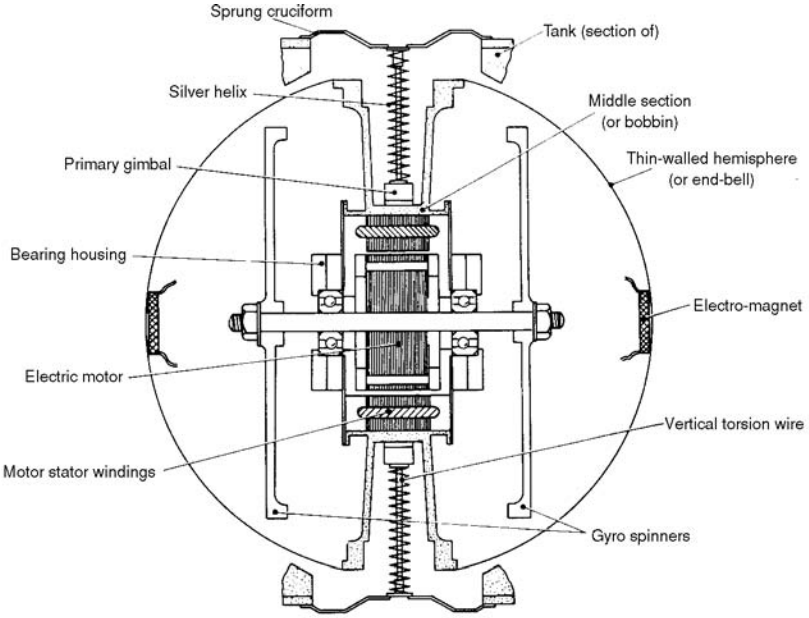
As an example, Table 8.3 shows part of the MK 37 VT gyrocompasses extensive fault diagnosis table. Using this and the data displayed on the main display unit, it is possible to isolate the area of a malfunction.

So far this description has only considered gyrocompass equipment using a top-heavy control mechanism. Many manufacturers prefer to use a bottom-heavy control system. One of the traditional manufacturers, S.G. Brown Ltd, provides some fine examples of bottom-heavy gyroscopic control.

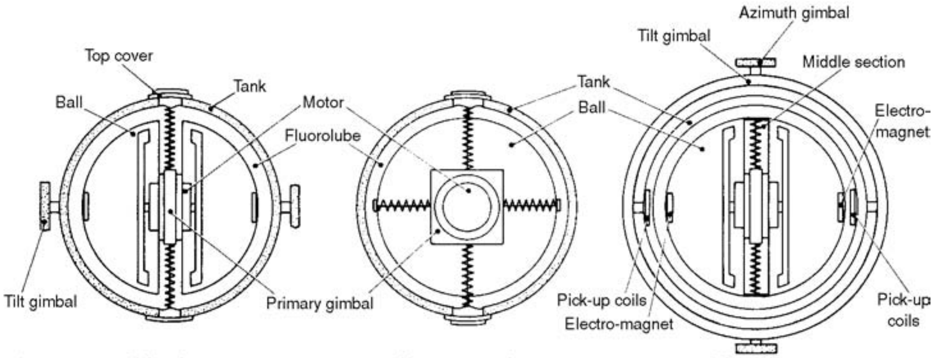
## 8.10 A bottom-heavy control gyrocompass

Modern bottom-heavy controlled gyrocompasses tend to be sealed gyroscopic units with full computer control and electronic interfacing. For the purpose of system description, this early gyrocompass is a good example of bottom-heavy control used to settle and stabilize a compass.

The gyroscopic element, called the sensitive element, is contained within a pair of thin walled aluminium hemispheres joined as shown in Figure 8.34, to form the 'gyroball'. At the heart of this ball is a three-phase induction motor, the rotor of which protrudes through the central bobbin assembly but is able to rotate because of the high quality support bearings. At each end of the rotor shaft, a heavy rimmed gyro spinner is attached to provide the necessary angular momentum for gyroscopic action to be established. Rotational speed of the induction motor is approximately 12 000 rpm.



Vertical section through gyro ball



**Side view**
Showing vertical torsion wires and helixes

**Front/rear view**
Showing horizontal and vertical torsion wires and helixes

**Plan view**
Showing horizontal torsion wires and helixes

**Figure 8.34** Arrangement of the gyroball. (Reproduced courtesy of S.G. Brown Ltd.)

The gyroball is centred within the tank by means of two vertical and two horizontal torsion wires forming virtually friction-free pivots. The torsion wires permit small controlling torques to be applied in both the vertical and the horizontal axes to cause precessions of the axes in both tilt and azimuth. In addition, the torsion wires are used to route electrical supplies to the motor. The gyroball assembly is totally immersed in a viscous fluid called halocarbon wax, the specific gravity of which gives the ball neutral buoyancy, at normal operating temperatures, so that no mass acts on the torsion wires.

The tank containing the gyroball sensitive element is further suspended in a secondary gimbal system, as shown in Figure 8.35, to permit free movement of the spin axis. This axis is now termed the 'free-swing axis' which under normal operating conditions is horizontal and in line with the local meridian. The secondary gimbal system also permits movement about the east–west axis. Each of the

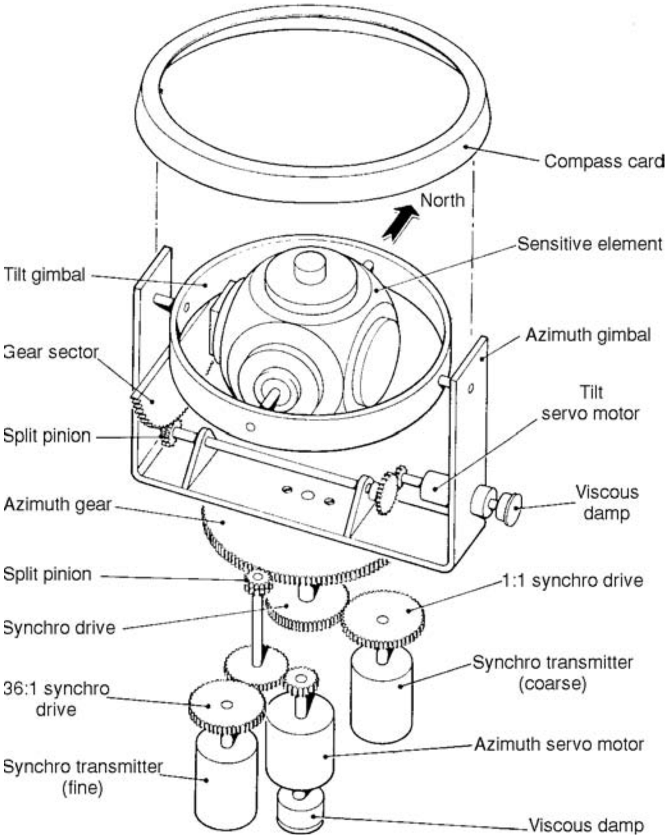


**Figure 8.35** Schematics showing the arrangement of the secondary gimbals.

movable axes in the secondary gimbal system can be controlled by a servomotor, which in turn provides both tilt and azimuth control of the gyroball, via a network of feedback amplifiers.

An electromagnetic pick-up system initiates the signal feedback system maintaining, via the secondary gimbals and servomotors, the gyro free-swing (spin) axis in alignment with the north–south axis of the tank. If there is no twist in the two pairs of torsion wires, and no spurious torques are present about the spin axis, no precession of the gyroball occurs and there will be no movement of the control servomotors. The gyro spin axis is in line with a magnet mounted in each hemisphere of the gyroball.

Pick-up coils are mounted on the north/south ends of the containment tank and are arranged so that when the gyro-ball is in alignment with the tank, no output from the coils is produced. If any misalignment occurs, output voltages are produced that are proportional to the displacement in both tilt and azimuth. These small e.m.f.s are amplified and fed back as control voltages to re-align the axis by precession caused by moving the secondary gimbal system. The tiny voltages are used to drive the secondary gimbal servomotors in a direction to cancel the sensor pick-up voltages and so maintain the correct alignment of the gyroball within the tank.

With a means of tank/gyroball alignment thus established, controlled precessions are produced. Referring to Figure 8.36, to precess the gyroball in azimuth only, an external signal is injected into the tilt amplifier. The null signal condition of the pick-up coils is now unbalanced and an output is produced and fed back to drive the tilt servomotor. This in turn drives the tilt secondary gimbal system

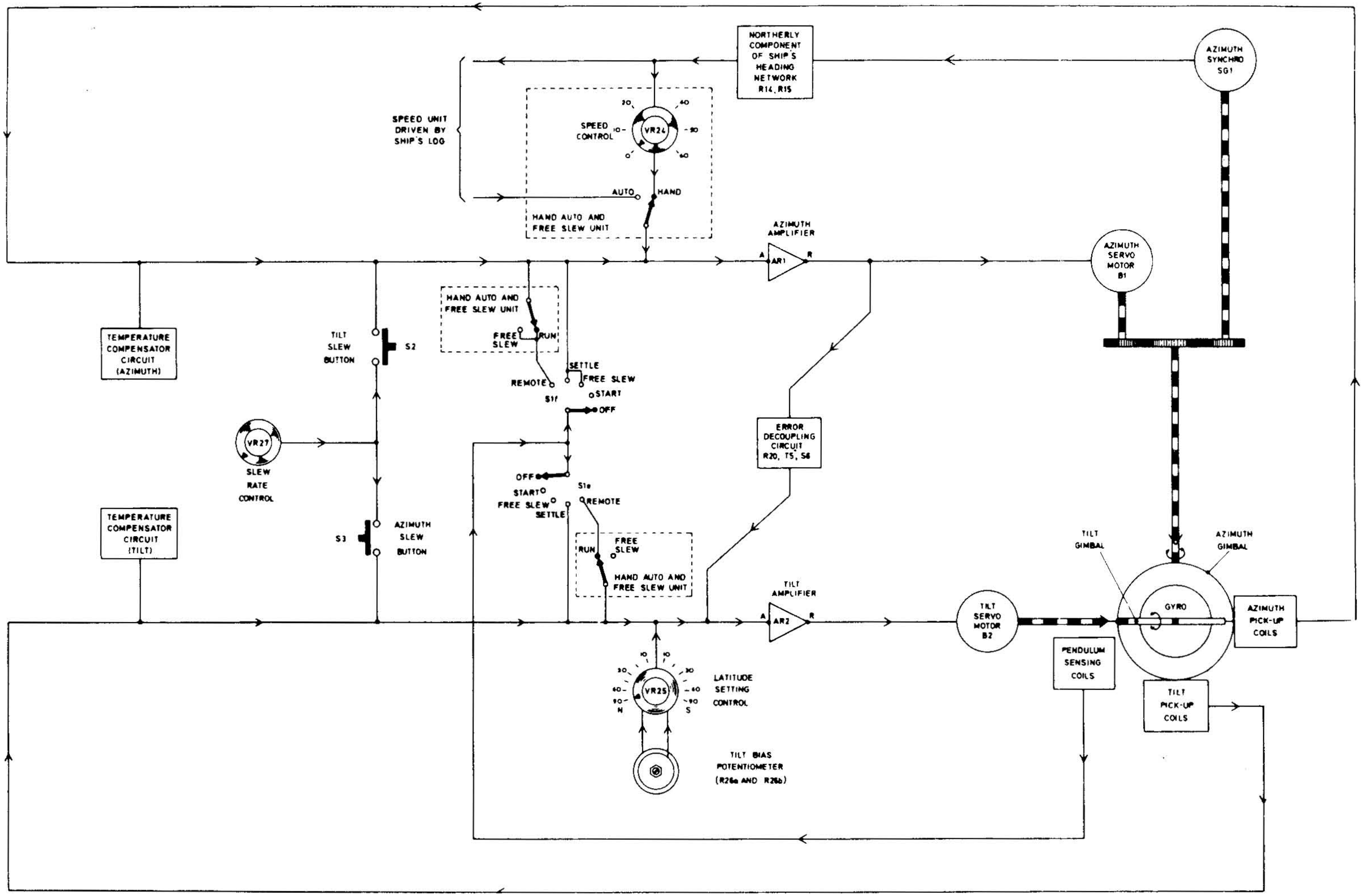


**Figure 8.36** Compass circuits schematic. (Reproduced courtesy of S.G. Brown Ltd.)

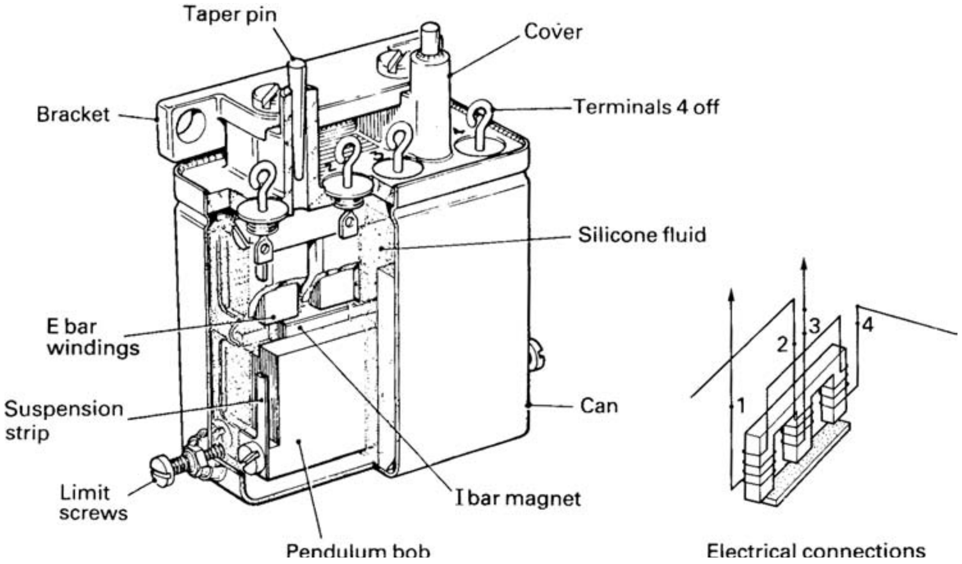
to a position in which the tilt pick-up coil misalignment voltage is equal and opposite to the external voltage applied to the amplifier.

The tilt servo feedback loop is now nulled, but with the tank and gyroball out of alignment in a tilt mode. A twist is thus produced of the horizontal torsion wires, creating a torque about the horizontal axis of the gyroball and causing it to precess in azimuth. As azimuth precession occurs, azimuth misalignment of the tank/gyroball also occurs but this is detected by the azimuth pick-up coils. The azimuth servomotor now drives the secondary gimbal to rotate the tank in azimuth to seek cancellation of the error signal. Since the azimuth secondary gimbal maintains a fixed position relative to the gyro spin axis in azimuth, a direct heading indication is produced on the compass card mounted on this gimbal.

Control of the sensitive element in tilt is done in a similar way. Therefore signals injected into the tilt and azimuth servo loops, having a sign and amplitude that produce the required precessional directions and rates, will achieve total control of the gyrocompass.

It is a relatively simple task to control the gyroball further by the introduction of additional signals because each of the feedback loops is essentially an electrical loop. One such signal is produced by the 'gravity sensor' or 'pendulum unit'. The pendulum unit replaces the liquid ballistic system, favoured by some manufacturers, to produce gravity control of the gyro element to make the compass north-seeking.

To produce a north-seeking action, the gyroscopic unit must detect movement about the east–west (horizontal) axis. The pendulum unit is therefore mounted to the west side of the tank, level with the centre line. It is an electrically-operated system consisting of an 'E'-shaped laminated transformer core, fixed to the case, with a pendulum bob freely suspended by two flexible copper strips from the top of the assembly. The transformer (Figure 8.37) has series opposing wound coils on the outer 'E' sections and a single coil on the centre arm. The pendulum-bob centres on the middle arm of the 'E' core and is just clear of it. The whole assembly is contained in a viscous silicon liquid to damp the short-term horizontal oscillations caused by the vessel rolling.



Electrical connections

**Figure 8.37** The pendulum assembly and its electrical connections. (Reproduced courtesy S. G. Brown Ltd.)

Initially the bob will centre in the middle of the 'E' core, but if the gyro tank tilts, the bob will offset causing the normally equalized magnetic field to be unbalanced and produce a stronger field on the outer arm towards which it is offset. The result is that a tilt signal, of correct sense and amplitude, is produced. This signal is fed to the tilt and azimuth amplifiers as required.

The output signal of the pendulum unit is also used to enable the gyro to settle in the meridian and become 'north settling'. A small carefully calibrated portion of the output signal is applied to the azimuth amplifier to cause azimuth misalignment of the gyro tank and hence a twist of the vertical torsion wires. The result is a tilt of the sensitive element, the direction of which depends on whether the gyro spin axis is north or south end up with respect to the horizontal. The amplitude of the pendulum signal fed to the azimuth amplifier will determine the settling period of the gyro, which for this compass is 40 min.

Loop feedback versatility is again made use of by applying signals in order to achieve the necessary corrections for latitude and speed errors. The injected signals result in the required precessional rates in azimuth, for latitude correction and in tilt, for speed correction.

### 8.10.1 Speed correction

A signal that is proportional to the ship's speed and the cosine of the ship's course, is coupled back to the azimuth amplifier to cause the gyroball to tilt in opposition to the apparent tilt caused by the northerly or southerly component of the ship's speed. The signal will therefore be maximum in amplitude when the course is due north or south, but will be of opposite sense. If the course is due east or west no correction is necessary. The system uses a 1:1 ratio azimuth synchronous transmitter SG1, which is mechanically driven by the azimuth servomotor gearing, and a balanced star connected resistor network as shown in Figure 8.38.

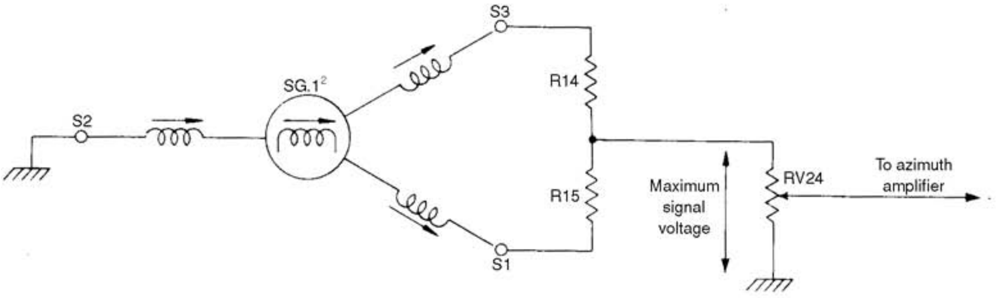
Alternatively an external signal derived from the ship's speed log may be used. In Figure 8.38 the error for a ship sailing due north is maximum and therefore the feedback signal produced across RV24, by the currents flowing through SG1, S1 and S2 coils, will be maximum. A portion of this signal, dependent upon the speed setting of RV24, is fed to the azimuth amplifier to produce a tilt of the gyroball. For a course due south, the signal is again maximum, but is of opposite phase to the northerly signal. This will cause an opposite tilt of the gyroball to be produced. With the ship sailing due east, the synchronous transmitter SG1 is in a position which will produce a zero signal across RV24 and no correction signal is applied to the azimuth amplifier irrespective of the speed setting of RV24. Any intermediate setting of SG1 will produce a corresponding correction signal to be developed across RV24.

### 8.10.2 Latitude correction

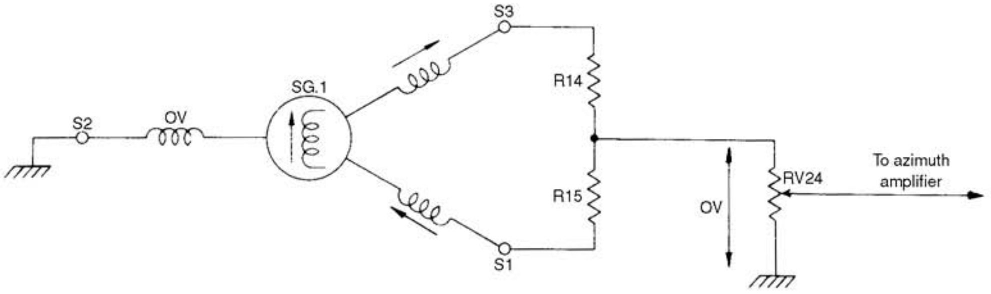
The latitude correction circuit provides a signal, proportional to the sine of the vessel's latitude, to cause the gyroball to precess in azimuth at a rate equal and opposite to the apparent drift caused by the rotation of the earth. This signal will be zero at the equator and maximum at the poles. It must also be of opposite phase for north or south latitudes. VR25 (see Figure 8.36), the latitude potentiometer, derives its signal from the 24V centre-tapped secondary winding of a transformer, and therefore has signals of opposite phase at either end. This control sets the amplitude of the correction signal and is manually adjusted.

### 8.10.3 Temperature compensation

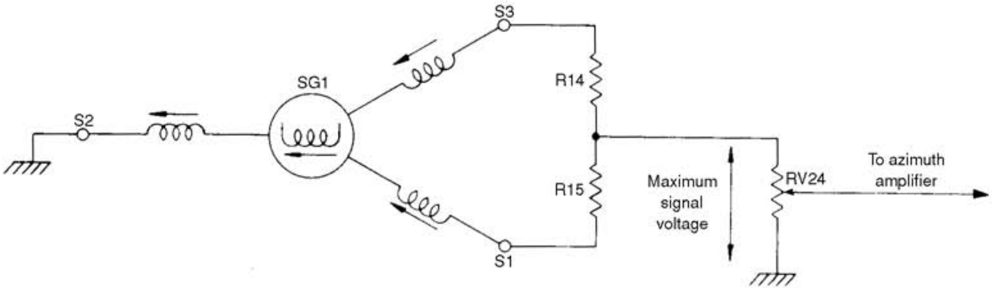
Both the vertical and horizontal torsion wires may twist with a change in ambient temperature. A corrective signal is produced in each of the tilt and azimuth temperature compensation circuits to



(a) Ship sailing north



(b) Ship sailing east



(c) Ship sailing south

Note:-
Arrows denote instantaneous current flow

**Figure 8.38** Signal output of synchro SG1 for different headings. (Reproduced courtesy S.G. Brown Ltd.)

counteract any precession of the gyroball caused by a change in temperature. The corrective signals are produced in the compensation circuits and connected to the tilt and azimuth amplifiers in such a way that both signal amplitude and sense will cause torques to be produced which are equal and opposite to those produced by twisting of the torsion wires. The effect of ambient temperature on the torsion wires is therefore cancelled.

### 8.10.4 Error decoupling circuit

The accuracy of a gyrocompass can be seriously affected by violent movement of the vessel, particularly heavy rolling caused by severe storms and rapid manoeuvring. A carefully calibrated error signal is derived from the output of the azimuth amplifier (which will be present due to misalignment of the tank and gyro spin axis during such conditions) and applied to the tilt amplifier to control the tilt gimbals. The system will provide partial and adequate compensation for errors that arise due to violent rolling conditions. The correction system is more than adequate for fittings on Merchant Navy vessels that are rarely subjected to rapid manoeuvres.

### 8.10.5 Slew rate

The purpose of the slew rate control VR27 (see Figure 8.36) is to rapidly level and orientate the gyro during the start-up procedure. The potentiometer VR27 is connected across the 24 V centre-tapped secondary winding of a transformer and is therefore able to produce an output of opposite phase and varying amplitude. The signal voltage level set by VR27 may be applied to the input of either the azimuth or tilt amplifiers separately by the use of push buttons. The buttons are interconnected in such a way that the signal cannot be applied to both amplifiers at the same time.

If the output of VR27 is firstly applied to the tilt servo amplifier (by pressing the azimuth slew button) the gyro will precess towards the meridian. If the tilt slew button is now pressed, the gyro will be levelled by applying the output of VR27 to the azimuth servomotor. The slew rate control VR27 adjusts the rate at which the gyro precesses and not the extent of precession, which is a function of time. It is essential that this control is centred before either slew button is pressed, otherwise a violent kick of the gyro ball will occur in one direction making compass alignment more difficult to achieve. The selector switch S1 must be in the ‘free slew’ position during this operation.

## 8.11 Starting a gyrocompass

As has been previously stated, from start-up a gyrocompass needs time to settle on the meridian. The time taken depends upon the make, model and the geographic location of the compass, but in general it is between one and several hours. The duration also depends upon whether the gyro wheel is already rotating or not. If the compass has been switched off, it will take much longer to bring the compass into use. Inputting the ship’s heading to reduce the initial error factor can reduce the time period. As an example, the following section considers the start-up procedure for the Sperry MK37 VT Digital Gyrocompass.

At power-up and prior to entering the settle mode, the system performs the automatic ‘bite’ procedure to determine if the equipment is operating within specified parameters. The CPU also initializes the system hardware and communication channels. During this procedure the gyro wheel is checked for movement. If it is stationary, the system ops for a cold start, if it is rotating a hot start is programmed. During a cold start, if no heading data is input to the system when requested, the gyrocompass selects Automatic.

### 8.11.1 Cold starting the compass

After an initial period, during which the bite is active, the following sequence is initiated and the settle indicator lamp will be lit.

- Two bleeps prompt the operator for a heading input. If heading data is not entered within 5 min, the gyro switches to an 'auto level' process.
- Assuming heading data has been input, the yoke will be offset based on this data. It will be slewed from the meridian, either clockwise or anticlockwise.
- The gyrowheel is brought up to speed within 14 min.
- The yoke is slewed back and forth to level the ballistic. This action takes about 4 min.
- Again assuming heading data has been input, the gyrocompass will settle within 1 h and the settle indicator lamp goes out. If no heading data was entered, the compass will automatically settle within 5 h.

Other inputs to the gyrocompass are as follows.

- Heading: in the range 0 to 359°. If the entered heading is in error by more than 20° from the true heading, the compass takes 5 h to settle.
- Initialize and Synchronize Step Repeaters. An operator selects a repeater and when requested uses the keypad's left or right arrow switches to scroll the display to the repeater's current position. After 10 s the system steps the repeater to the compass heading. It is essential to repeat and double check this procedure because there must be no alignment errors in a repeater system.
- Speed Input. Using the left or right arrow keys, an operator inputs a speed in the range 0–70 knots.
- Latitude Input. Using the arrow keys, an operator inputs latitude in degrees north or south of the equator.

## 8.12 Compass repeaters

Remote analogue compass repeaters are simply mechanized compass cards driven either by a stepper motor or a synchro bearing transmission system. Digital heading displays can also be produced by digitizing the stepper 'grey code' waveform before applying it to a suitable decoding system. This section deals with the most popular bearing transmission systems.

### 8.12.1 Stepper systems

Figure 8.39 shows a mechanical switching stepper system which, because its robustness, is still found on many merchant ships for bearing transmission to remote repeaters. The rotor of the transmitter is geared to the azimuth ring gearing of the master compass. The transmitter is a multi-contact rotary switch that completes the circuit for current to flow through the appropriate repeater motor coils. The transmitter rotor has two rotating arms spaced at 165° to each other. Each rotor arm makes contact with copper segments arranged in four groups of three, with each segment being wired to its corresponding number in the other three groups.

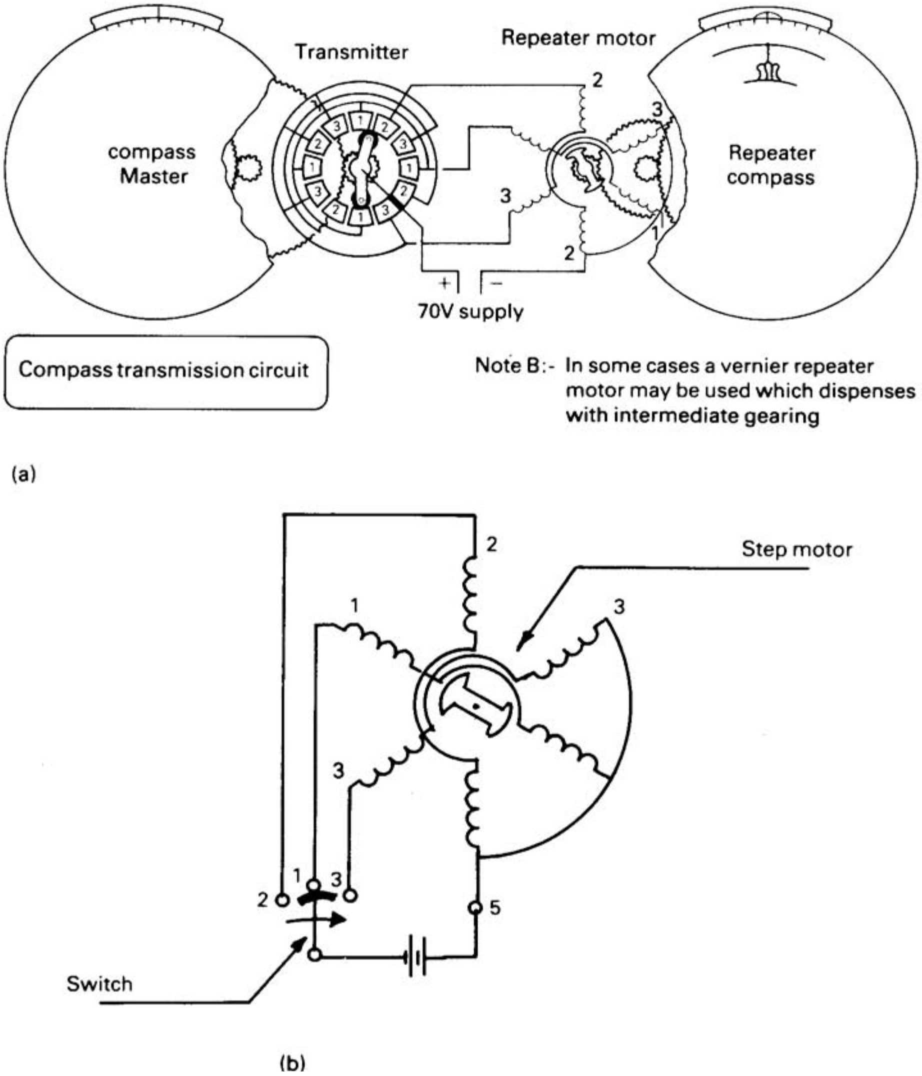


**Figure 8.39** Stepper repeating system. (a) Early mechanical switching system; (b) diagrammatic representation of a simple step motor receiver. (Reproduced courtesy of Sperry Ltd.)

The gear ratio of transmitter rotor to azimuth gear is 180:1. Therefore:

$$\begin{aligned} 180\ \text{rev} &= 360^\circ \\ 1\ \text{rev} &= 2^\circ \\ 12\ \text{seg} &= 2^\circ \\ 1\ \text{seg} &= 2/12^\circ \text{ or } 10\,\text{min of arc} \end{aligned}$$

The rotating arms make 12 steps per revolution. Because of the 180:1 gear reduction, each step therefore corresponds to 1/6th of a degree or 10 min of arc on the compass card.

A simplified step by step receiver is shown in Figure 8.39(b). Three pairs of coils are wound, and located at 60° intervals on the stator assembly of the receiver. The rotor is centrally located and capable of rotating through 360°. With the switch in the position shown, current flows through the series connected coils (1) and, under the influence of the magnetic field produced, the rotor takes up the position shown. As the switch moves to position 3, its make-before-break action causes current to flow through both coils 1 and 3 and the rotor moves to a position midway between the coils, due east–west. The next movement of the switch energizes coil 3 only causing the rotor to line up with this coil. In this way the rotor is caused to rotate one revolution in 12 steps. The construction details of a step motor are given in Figure 8.40.

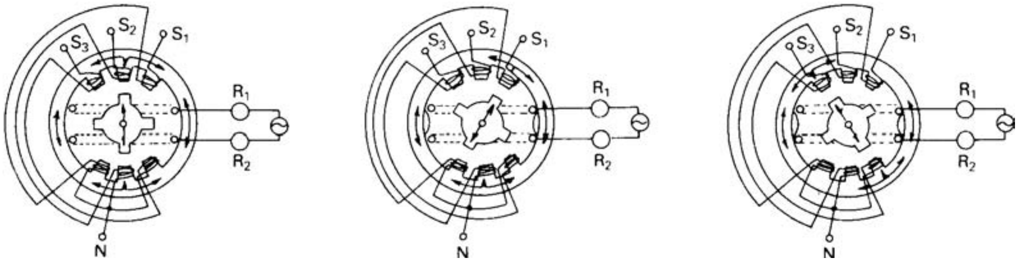


**Figure 8.40** Construction details of a step motor.

A stepper system such as this may also be used as part of a 'direct digital control' (d.d.c.) system in which signals are generated digitally to control movement of the repeater. Such a stepper system uses a cyclic binary code or gray code for its operation. The gray code is easily produced using shaft or disc encoders geared to the compass azimuth gearing.

### 8.12.2 Synchro systems

A synchro is a device that uses the basic principle of a single-phase transformer with magnetic coupling between a rotating primary (rotor) and a number of secondaries (stators). For the purpose of this description three secondaries are located at 120° intervals on the stator. The rotor may be rotated through 360° within the laminated stator assembly holding the three secondary windings. The primary coil is energized by a low frequency a.c. applied via slip rings located on the main shaft. The magnitude and phase of the secondary induced e.m.f.s is dependent upon the relative position of the rotor in relation to the stator windings.

Figure 8.41 shows a synchro repeater system using the basic 'synchro error detecting' method of operation common to many control applications. The rotor of the synchro transmitter is reduction geared to the azimuth ring of the gyrocompass. A reference low frequency a.c. supply to the transmitter rotor coil couples with the three secondaries to produce e.m.f.s which cause current to flow around the three circuits. Each current flow produces a magnetic field around the corresponding receiver secondary and a resultant error signal is induced in the receiver rotor coil. No error signal is produced if the system is in the synchronous state with the transmitter and the receiver rotors at 90° to each other.

The error signal present, when the rotors are not synchronized, is directly proportional to the error angle ($\psi$) existing between the horizontal and the plane of the rotor. This error signal is amplified to the level required to drive a servo to turn the compass card. Also mechanically coupled to the servo

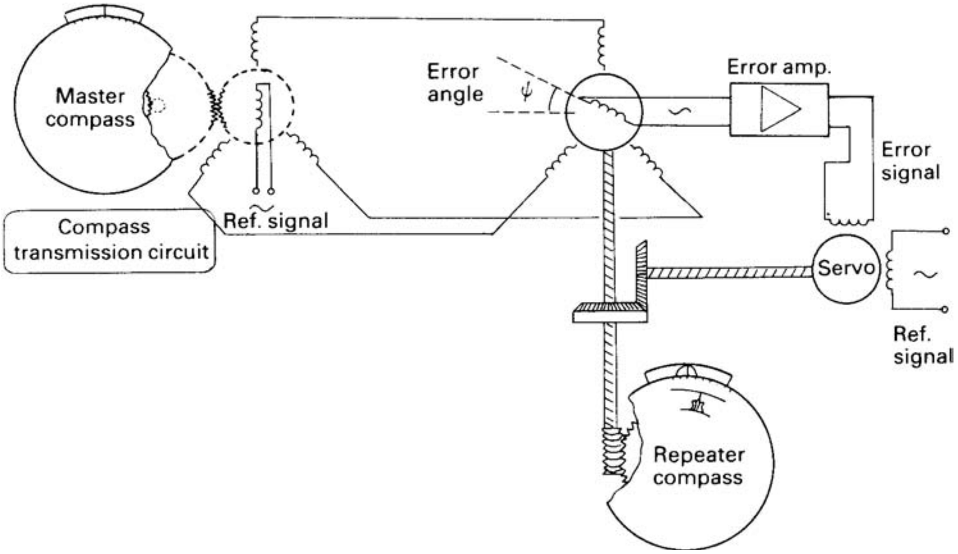


**Figure 8.41** A synchro bearing transmission system .

shaft is the receiver rotor that turns to cancel the error signal as part of a mechanical negative feedback arrangement. The receiver rotor will always therefore line up (at 90°) with the transmitter rotor to produce the synchronous state.

## 8.13 The magnetic repeating compass

Magnetic compasses are still popular with mariners and can easily be converted into a repeating compass with the addition of a flux gate assembly. A flux gate element is effectively a magnetometer that is used to detect both the magnitude and the direction of a magnetic field. Flux gate elements in common use are of the 'second harmonic' type, so called because if excited by a fundamental frequency, $f$, an output voltage will be generated which varies in both phase and amplitude, depending upon its position within the magnetic field, at a frequency of $2f$.

### 8.13.1 Construction

The basic flux gate consists of two thin wires of mumetal or permalloy, each contained in a glass tube around which is wound a coil. Two such assemblies are used. They are mounted side by side and parallel to each other. The two coils are connected in series so that their magnetic fields are in opposition when a low frequency a.c. (typically 2 Hz) is applied. Mumetal is used for the wire cores because of its property of magnetically saturating at very low levels of magnetic flux. (Mumetal magnetically saturates at a field strength of approximately 8 ampere turns per metre compared to 250 000 ampere turns per metre for steel wire.)

A secondary coil, wound around the whole assembly, provides a mutually induced e.m.f. as the output voltage.

Figure 8.42 illustrates the basic construction of a simple flux gate. Note that the primary coils are connected in series. In a practical unit a balancing system would be included to ensure that in the

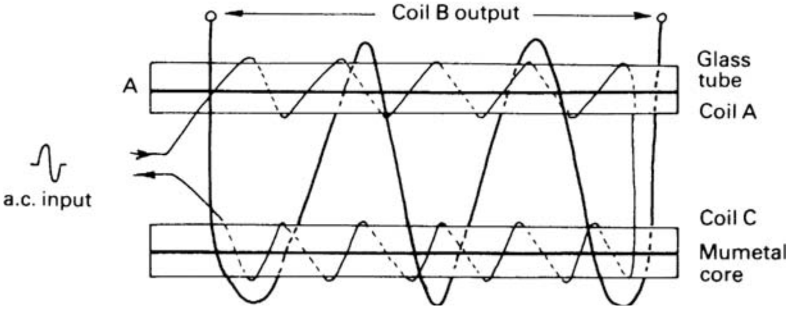


**Figure 8.42** A basic flux gate showing the primary windings of equal turns around tubes A and C and a secondary coil wound around the whole assembly.

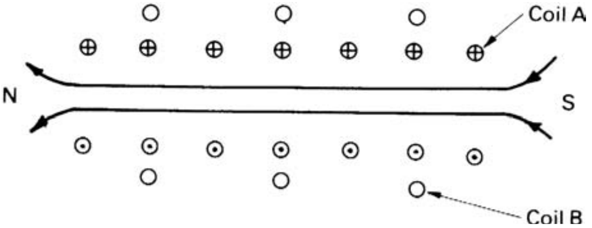


**Figure 8.43** A cross-section of part of a flux gate. Current flowing in coil A is 'into the diagram' on the top half of the winding and 'out' on the bottom.

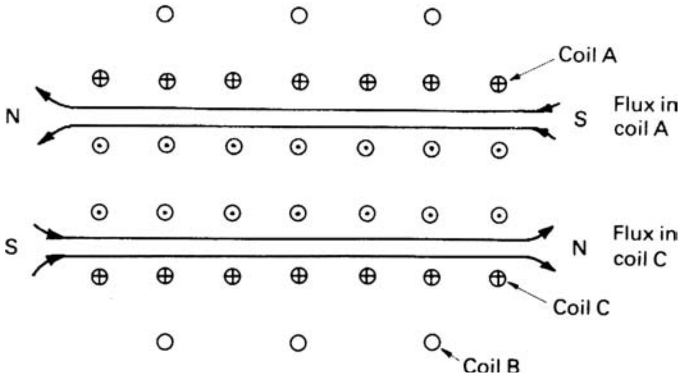


**Figure 8.44** A cross-section of a completed flux gate.

absence of any externally produced magnetic field, the magnetic field produced by the two primary windings will cancel and consequently no output will be generated. If the current in coil A changes (see Figure 8.43), the magnetic flux it causes will correspondingly change either in value or direction. Any change will produce a self-induced e.m.f. across coil A and a mutually-induced e.m.f. across coil B. Figure 8.44 shows a cross-section of a complete flux gate with coils A and C forming the primary function and coil B the secondary output coil.

If the magnetic fluxes produced by both coil A and C are of the same value but of opposite polarity, there will be no mutually induced e.m.f. in coil B. This is because the two magnetic fields linking with

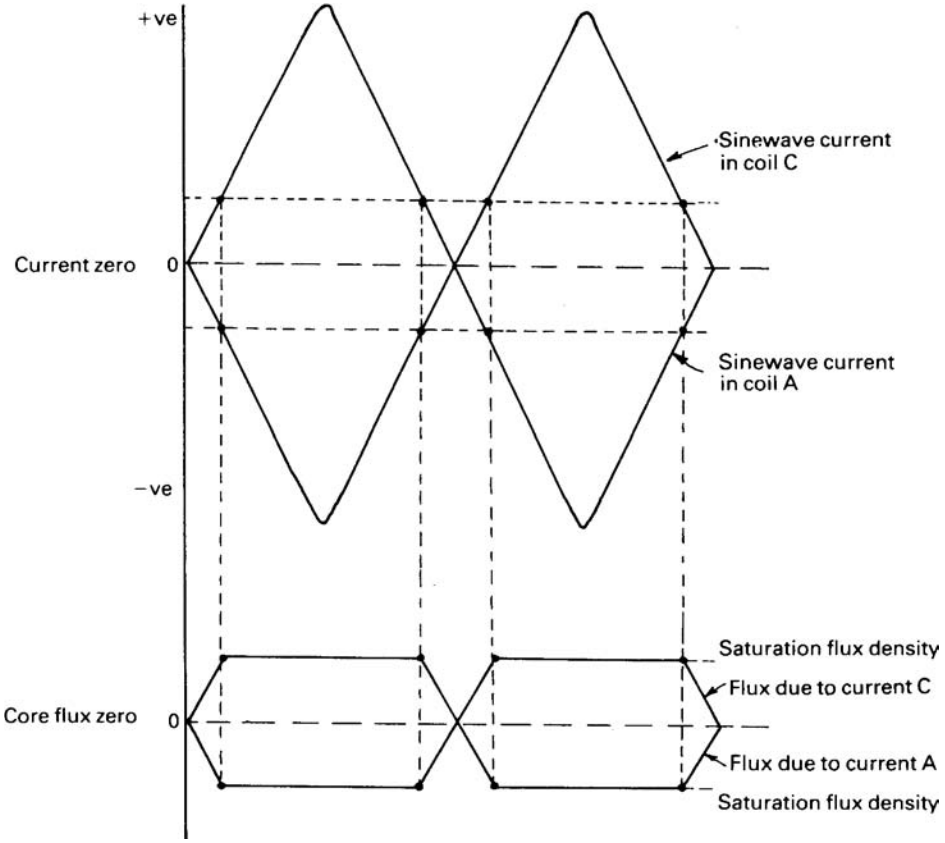


**Figure 8.45** Currents and flux saturation levels.

the turns of coil B will be effectively zero. This state can only exist if the two coils A and C are connected in series causing the current flow through the two coils to be the same value at any instant. When this is the case the system is said to be balanced and the output voltage across coil B will be zero.

Figure 8.45 shows the currents and flux saturation levels for both coil A and coil C when the assembly is balanced.

If a permanent magnet is placed in proximity of the flux gate as shown in Figure 8.46 its magnetic field will produce cancelling fields.

In the parts of the cores that carry flux in the same direction as the magnet, the core will saturate with a lower value of coil current. In the other half of the same core the two fluxes will oppose so that this part of the core does not saturate until a much larger current is flowing. These two effects will therefore not affect the balancing of the core fluxes so there will be no mutually induced e.m.f. across the secondary coil B. If the permanent magnet is now placed parallel to the two cores of the flux gate, as in Figure 8.47, an imbalance occurs.

The flux due to the magnet will now be in the same direction as that due to the coil current in one core but in the opposite direction in the other. The magnet will cause one core to saturate with a lower value of coil current and the other to require a larger value of coil current for saturation to occur.

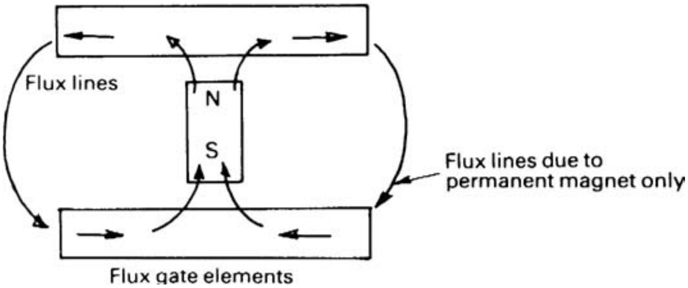


**Figure 8.46** Flux lines due to the addition of a permanent magnet to the flux gate.

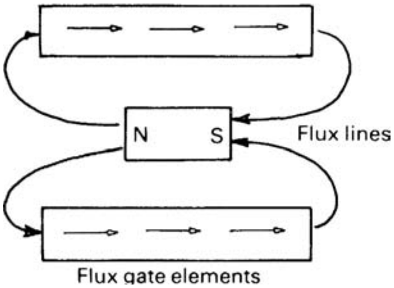


**Figure 8.47** Flux lines with the permanent magnet in line with the flux gate.

Figure 8.48 shows how the permanent magnet flux affects the flux produced in each core by the low frequency a.c. primary current on each half cycle of input voltage.

Figure 8.49 shows that the value of the a.c. induced into coil B is twice the frequency of the energizing supply, but depends upon the amplitude of the permanent magnet field. The output also varies as the cosine of the angle between the line of the magnet and the flux gate. The a.c. output is then amplified and used to drive a servomotor which rotates the gate until the output is zero. This corresponds to the magnet being at an angle of 90° to the gate elements.

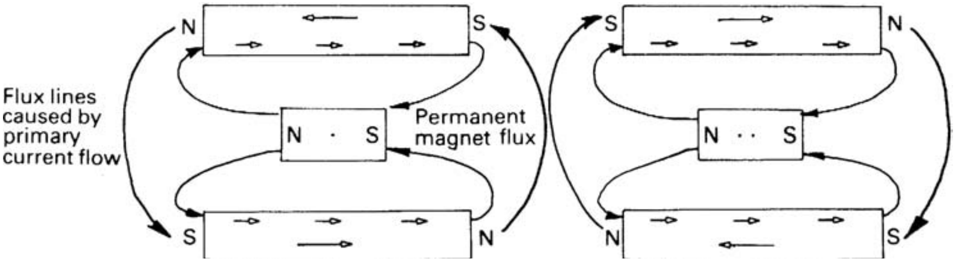


**Figure 8.48** The intensity of the magnetic flux in each core is changed on each half cycle of primary alternating current.

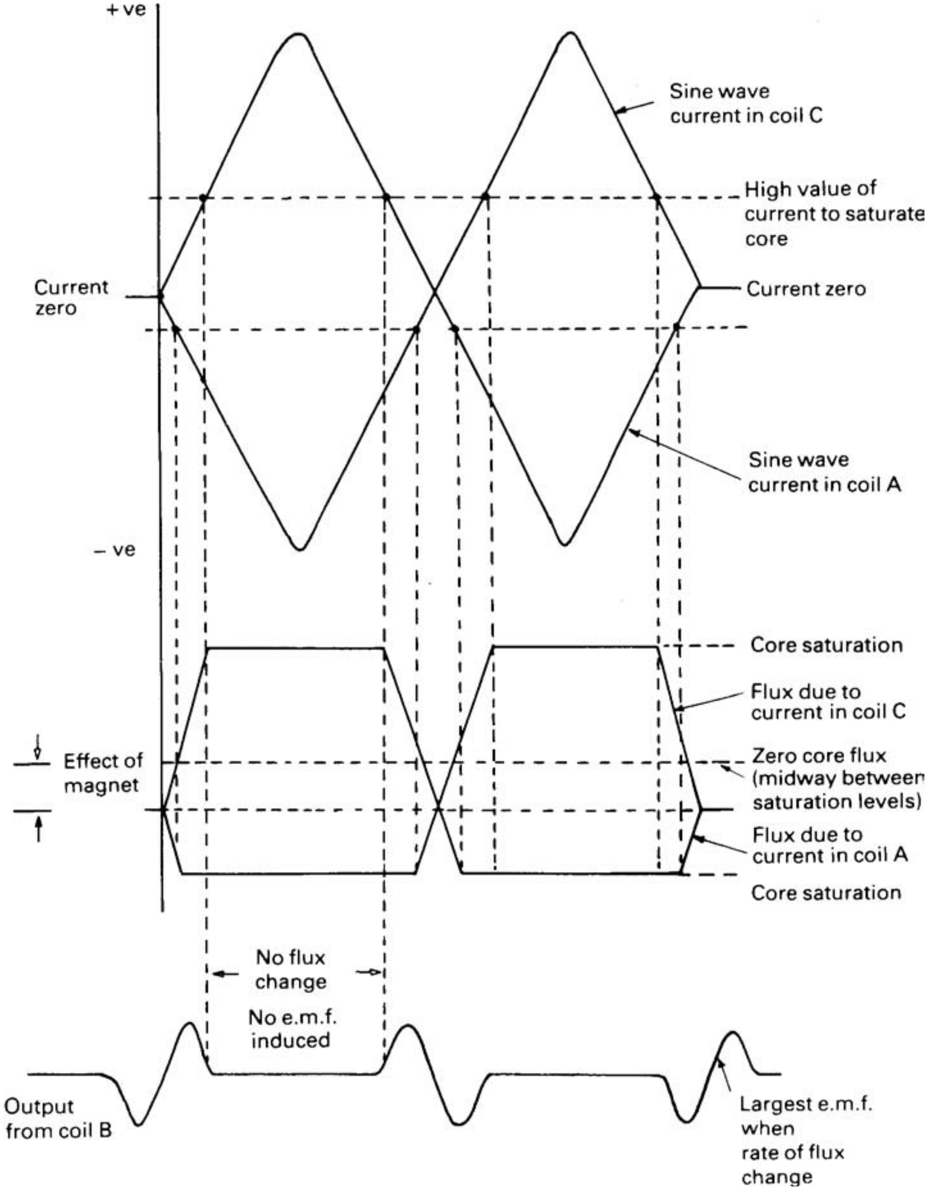


**Figure 8.49** An illustration of the fluxes and output e.m.f. produced by an unbalanced flux gate assembly.

### 8.13.2 Practical flux gate systems

There are currently two main systems of flux gates used in a repeating compass. The simplest of these uses a flux gate in conjunction with an ordinary magnetic compass as shown in Figure 8.50.

The flux gate is mounted on a rotating platform below the compass card of a standard marine magnetic compass and uses the north-seeking property of a permanent magnet. The core elements of the flux gate will therefore come under the influence of the permanent magnetic field produced by the

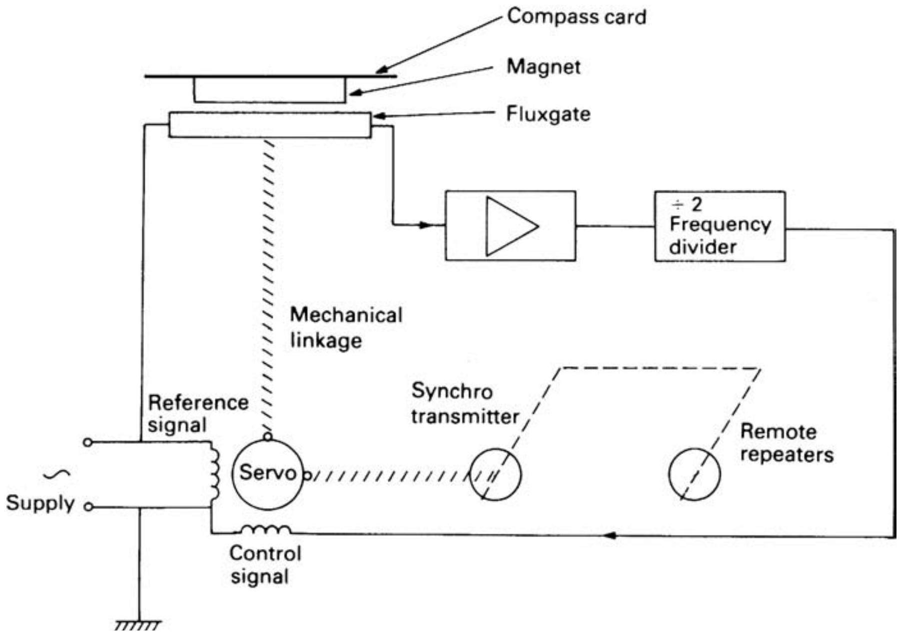


**Figure 8.50** A flux gate system used in conjunction with a magnetic compass.

compass pointer. As previously shown, the magnetization will have maximum effect when the flux gate and the compass magnet are parallel and zero effect when they are at 90° to each other. This point is referred to as the NULL point. The resultant output voltage from the secondary winding of the flux gate varies as the cosine of the angle between magnet and flux gate. Output from the flux gate secondary winding is amplified and its frequency divided by two before being applied to the control winding of a servomotor. This servo, which is mechanically coupled to the flux gate platform, drives the whole assembly towards a null point.

Assuming the flux gate and magnet are not at 90° to each other, an output from the flux gate secondary is produced which, after processing, is fed to the control winding of the servomotor. The reference winding supply is taken directly from the low frequency oscillator. This ensures that correct phasing of the servomotor is achieved and that the flux gate will always be driven towards the correct null point. When the null point is reached, the servo amplifier input falls to zero causing the servo to stop. The flux gate is therefore always kept in correct alignment with the compass magnet.

### 8.13.3 Dual axis magnetometer magnetic compass

As an alternative to using a flux gate in conjunction with a magnetic compass, it is possible to use a dual axis magnetometer to sense the earth's magnetic field to produce an indication of flux direction. The earth's magnetic lines of force are not horizontal to the earth's surface, thus it is necessary that the angle between the lines of force and the earth's surface be resolved into both vertical and horizontal components, as shown in Figure 8.51.

If we assume that a vessel is heading due north as shown in Figure 8.52, the two horizontally-orientated flux gates sense the magnitudes of the earth's horizontal magnetic flux lines diminished by

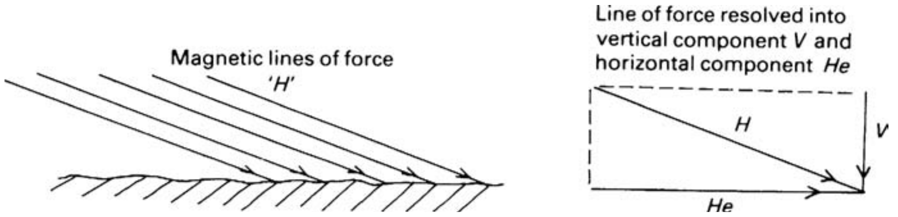


**Figure 8.51** An illustration showing how the lines of force of the earth's magnetic field may be resolved into vertical and horizontal components.

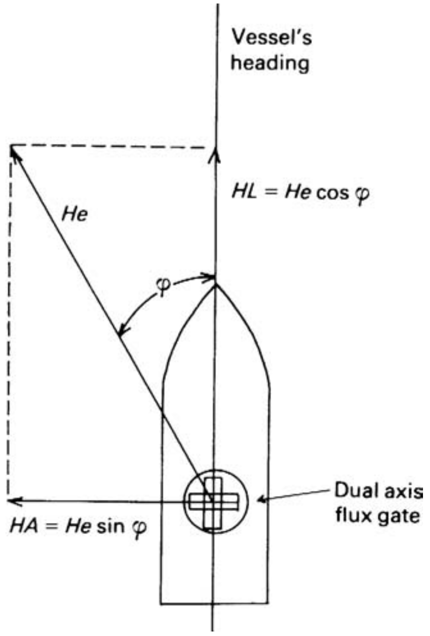


**Figure 8.52** The vessel's course shown as a cosine function of *He* – the direction of the earth's magnetic field.

sine and cosine functions of the heading. The resulting outputs produced, designated *HL* and *HA*, are derived as shown in Figure 8.52.

In Figure 8.53, flux gate 1 is mounted along the fore and aft line of the vessel and flux gate 2 athwartships. The fore and aft line component of the earth's magnetic field causes flux gate 1 to produce an output voltage proportional to the amplitude of this component. Similarly, gate 2 produces an output proportional to the athwartships component. Both signals are coupled to the stator coils of a synchro that produces two magnetic fields proportional to the amplitude of the original fields acting upon the flux gates. The line of the resultant field within the synchro is the same as the direction of the earth's magnetic field, *He*.

Output from the rotor of the synchro is connected, via a servo amplifier, to drive a servomotor which rotates the synchro rotor mechanically until it is at 90° to the resultant field, at which point output from the rotor is zero and the servo stops. The synchro rotor is thus kept in alignment with the resultant direction of the magnetic field within the synchro, which in turn depends upon the direction,

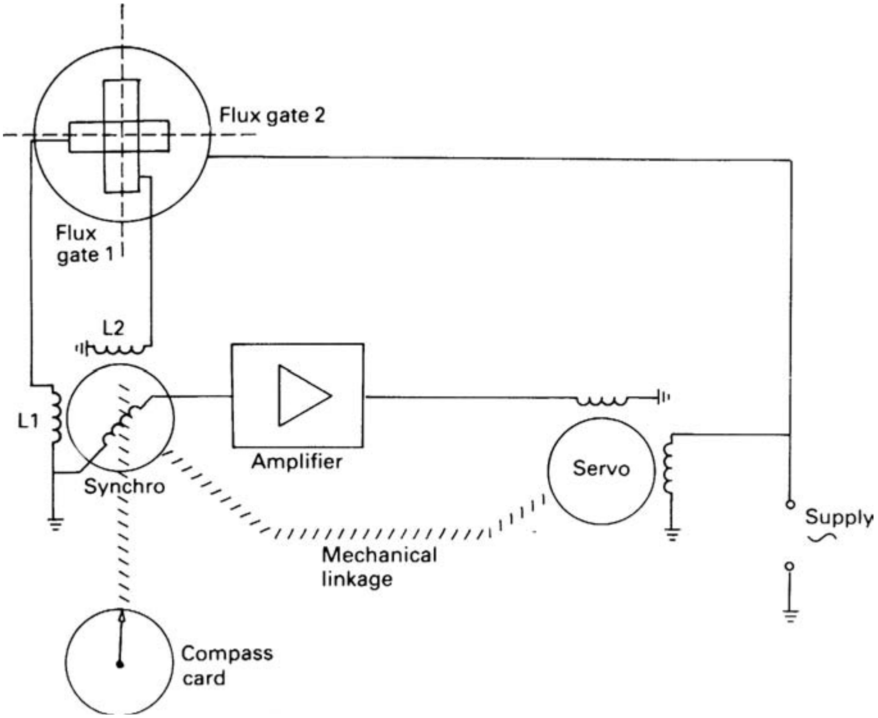


**Figure 8.53** A simplified diagram of a dual axis magnetometer type of magnetic compass.

relative to both flux gates, of the earth's magnetic field. A compass card is directly driven by the rotor of the synchro. Remote repeaters can be fitted, as illustrated in the previous case, by the use of a synchro transmission system. A compass has thus been produced which eliminates the conventional pivoted magnet arrangement to provide an electrical indication of magnetic north.

## 8.14 Glossary

| | |
|---|---|
| **Angular momentum** | In the case of a gyrowheel, this is the product of its linear momentum and the radius of the rotor. |
| **Ballistic pots** | Containers of viscous liquid to add damping to a gyrocompass. |
| **BITE** | Built-in test equipment. Automatic or manually commanding equipment test circuits. |
| **Compass repeaters** | Remote display of compass information. |
| **Controlled gyroscope** | One in which the movement caused by earth rotation is controlled. |
| **Drift** | The apparent movement in azimuth of a gyroscope due to earth rotation. |
| **Dynamic errors** | Errors caused by the angular motion of the vessel during heavy weather or manoeuvring. |
| **Flux gate** | The electrical sensing unit of a magnetic compass. |
| **Free gyroscope** | A gyroscope with a spin axis fixed by inertia to some celestial reference point and not to a terrestrial point. Not suitable as a gyrocompass. |

| | |
|---|---|
| **Follow-up** | A system enabling control of the gyro when it is fitted on board a moving platform. |
| **Gyroscope** | A perfectly balanced wheel that is able to spin at high speed symmetrically about an axis. |
| **Gyroscopic inertia** | A gyroscope rotor maintains the direction of its plane of rotation unless an external force of sufficient amplitude to overcome inertia is applied to alter that direction. |
| **Latitude error** | A constant value error the magnitude of which is directly proportional to earth rotation at any given latitude. |
| **Linear momentum** | The product of mass and velocity. |
| **Manoeuvring error** | An error caused by a vessel's rapid changes of speed and/or heading. |
| **North-seeking gyro** | One which is partly controlled and as a consequence will seek to locate north but will not settle. Further control is required to convert this type of gyro into a compass. |
| **North-settling gyro** | One which is fully controlled and will settle to point north. |
| **Precession** | Movement at 90° from the applied force. If a force is applied to a spinning rotor by moving one end of its axle, the gyroscope is displaced at an angle of 90° from the applied force. |
| **Rolling error** | As the name suggests, this error is caused by a vessel rolling. The error cancels when the ship is steaming north or south and is maximum when following an east/west course. |
| **Settling time** | The period taken for a gyrocompass to settle on the meridian from start-up. |
| **Slew rate control** | A control setting an electrical input to rapidly level and orientate the gyro during start-up. |
| **Stepper systems** | A step motor compass repeater circuit. |
| **Synch. systems** | A synchronous motor compass repeater circuit. |
| **Tilt** | By virtual of precession, the earth's rotation causes the spin axis to tilt upwards to an angle dependent upon its position in latitude. |
| **Transmission error** | An error existing between the master compass and any repeaters. |

## 8.15 Summary

- There are three axes in which a gyroscope is free to move: the spin axis, the horizontal axis and the vertical axis.
- In a free gyroscope none of the three axes is restricted.
- A free gyroscope is subject to the laws of physics, the most important of which, when considering gyrocompass technology, is inertia.
- Precession is the term used to describe the movement of the axle of a gyroscope under the influence of an external force. Movement of the axle will be at 90° to the applied force.
- Tilt is the amount by which the axle tilts because of the gyroscope's position in latitude.
- Azimuth drift is the amount by which the axle drifts due to the earth's rotation.
- A controlled gyroscope is one with its freedoms restricted.
- A north-seeking gyroscope is a controlled gyro that never settles pointing north.
- A north-settling gyroscope is a damped controlled gyro that does settle on the meridian.
- Bottom- and top-heavy controls are methods used for settling a north-seeking gyroscope.

- A gyrocompass fitted on board a ship is affected by dynamic errors. They are rolling error, manoeuvring error, speed and course error and latitude or damping error. All these errors are predictable and controllable.
- When starting from cold, gyrocompasses require time to settle on the meridian. A settling time period of 75 min is typical.
- Stepper systems are transmission devices that relay the bearing on the master compass to remote repeaters.
- Magnetic repeating compasses are based on flux gate technology.
- A flux gate is an electrical device that interprets the compass bearing to produce control functions.

## 8.16 Revision questions

1. Describe what you understand by the term gyroscopic inertia?
2. What do you understand by the term precession when applied to a gyrocompass?
3. Why is a free gyroscope of no use for navigation purposes?
4. How is earth's gravity used to turn a controlled gyroscope into a north-seeking gyroscope?
5. How is a north-seeking gyroscope made to settle on the meridian and indicate north?
6. When first switched on a gyrocompass has a long settling period, in some cases approaching 75 min. Why is this?
7. Explain the terms gyro-tilt and gyro-drift.
8. How is a gyrocompass stabilized in azimuth?
9. What is rolling error and how may its effects be minimized?
10. Why do gyrocompass units incorporate some form of latitude correction adjustment?
11. What effect does an alteration of a ship's course have on a gyrocompass?
12. What are static errors in a gyrocompass system?
13. When would you use the slew rate control on a gyrocompass unit?
14. Why is temperature compensation critical in a gyrocompass?
15. What is a compass follow-up system?
16. What is a compass repeater system?
17. A flux gate is the central element of magnetic repeating compasses. Explain its operation.
18. Flux gate elements are known as 'second harmonic' units. Why is this?
19. What are the advantages of using a dual axis magnetometer in preference to a flux gate?
20. Why is a magnetic repeating compass not influenced by the vessel's position in latitude or by violent manoeuvring?

# Chapter 9

# Automatic steering

## 9.1 Introduction

It has already been implied that a modern merchant vessel must be cost-effective in order to survive the ever-increasing pressure of a financially orientated industry. A good automatic pilot, often called an Autohelm, although a registered trade name, can improve the profit margin of a vessel in two ways. First, it enables a reduction to be made in the number of ships' personnel, and second, a considerable saving in fuel can be achieved if the vessel makes good its course with little deviation. This chapter, dealing with the principles of automatic pilots, enables the reader to understand fully the electronic systems and the entire operator control functions.

Early autopilots were installed in the wheelhouse from where they remotely operated the vessel's helm via a direct drive system as shown in Figure 9.1. This figure gives an excellent indication of system first principles.

Although efficient, the main drawback with the system was the reliance upon a hydraulic telemotor system, which required pressurized tubing between the transmitter, on the ship's bridge, and the receiver unit in the engine room. Any hydraulic system can develop leaks that at best will cause the system to be sluggish, and at worst cause it to fail. To overcome inherent inefficiencies in hydraulic transmission systems, they have been replaced with electrical transmitters, and mechanical course translating systems have been replaced with computer technology.

## 9.2 Automatic steering principles

Whatever type of system is fitted to a ship, the basic principles of operation remain the same. Before considering the electronic aspects of an automatic steering system it is worthwhile considering some of the problems faced by an automatic steering device.

In its simplest form an autopilot compares the course-to-steer data, as set by the helmsman, with the vessel's actual course data derived from a gyro or magnetic repeating compass, and applies rudder correction to compensate for any error detected between the two input signals. Since the vessel's steering characteristics will vary under a variety of conditions, additional facilities must be provided to alter the action of the autopilot parameters in a similar way that a helmsman would alter his actions under the same prevailing conditions.

For a vessel to hold a course as accurately as possible, the helm must be provided with data regarding the vessel's movement relative to the course to steer line. 'Feedback' signals provide this data consisting of three sets of parameters.

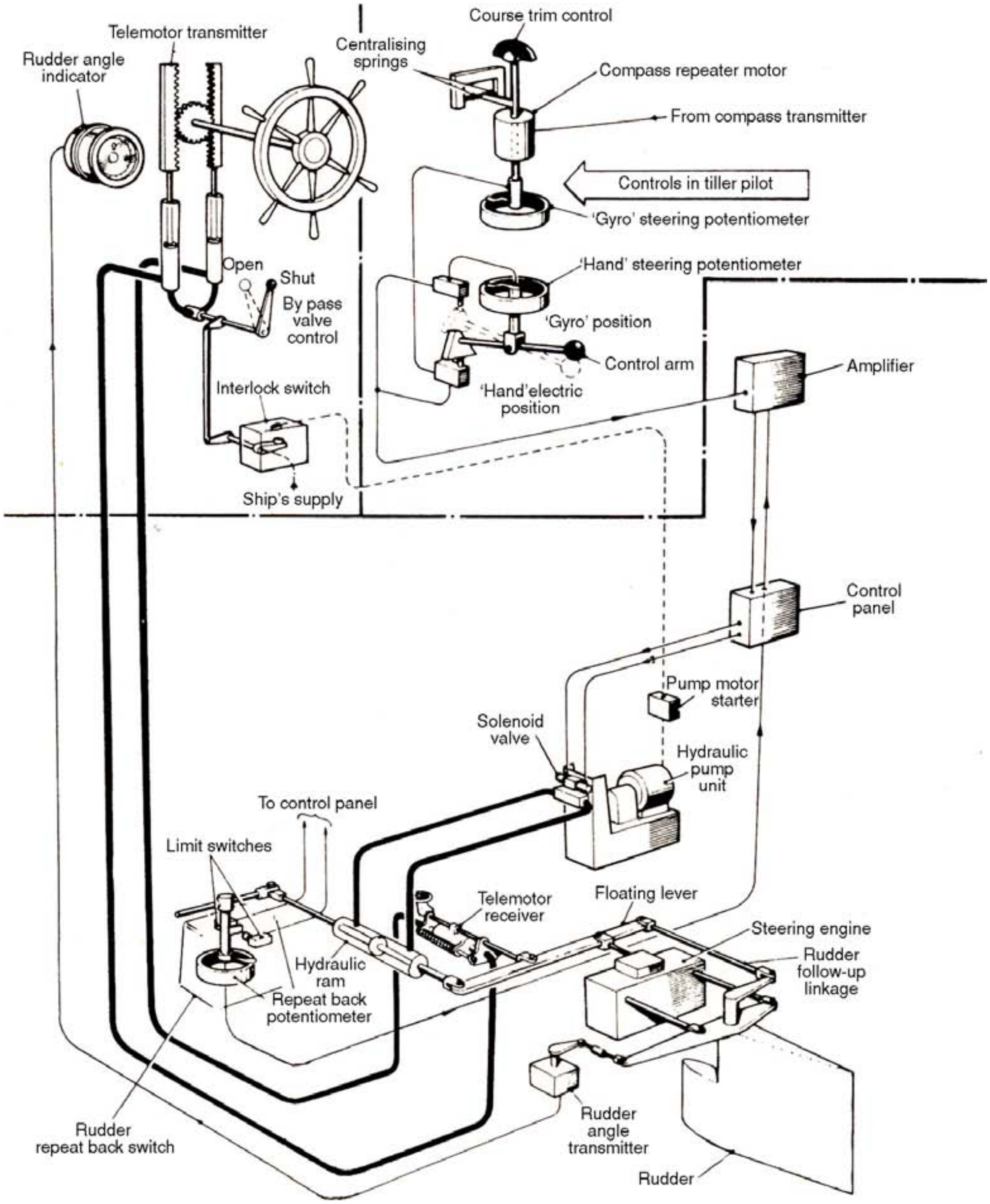


**Figure 9.1** An early electro-mechanical autopilot system using telemotors. (Reproduced courtesy of Sperry Ltd.)

- Position data: information providing positional error from the course line.
- Rate data: rate of change of course data.
- Accumulative error data: data regarding the cumulative build-up of error.

Three main control functions acting under the influence of one or more of the data inputs listed above are: proportional control, derivative control and integral control.

### 9.2.1 Proportional control

This electronic control signal causes the rudder to move by an amount proportional to the positional error deviated from the course line. The effect on steering, when only proportional control is applied, is to cause the vessel to oscillate either side of the required course, as shown in Figure 9.2. The vessel would eventually reach its destination although the erratic course steered would give rise to an increase in fuel expended on the voyage. Efficiency would be downgraded and rudder component wear would be unacceptable.

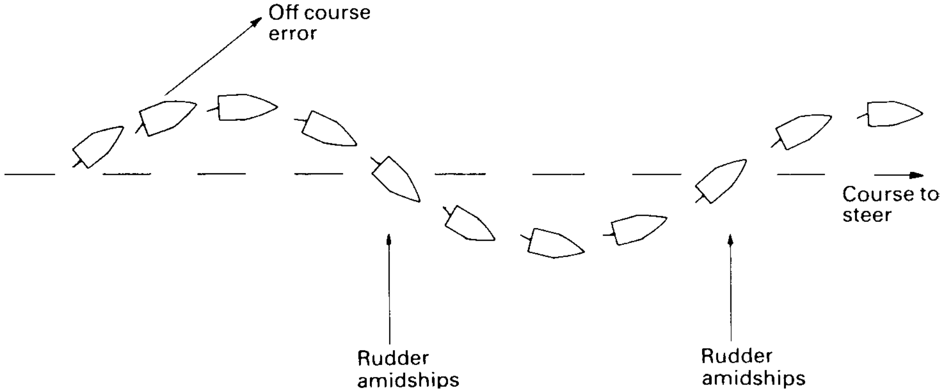


**Figure 9.2** The effect of applying proportional control only. The vessel oscillates about the course to steer.

At the instant an error is detected, full rudder is applied, bringing the vessel to starboard and back towards its course (Figure 9.2). As the vessel returns, the error is reduced and autopilot control is gradually removed. Unfortunately the rudder will be amidships as the vessel approaches its course causing an overshoot resulting in a southerly error. Corrective data is now applied causing a port turn to bring the vessel back onto course. This action again causes an overshoot, producing corrective data to initiate a starboard turn in an attempt to bring the vessel back to its original course. It is not practical to calculate the actual distance of the vessel from the course line at any instant. Therefore, the method of achieving proportional control is by using a signal proportional to the rudder angle as a feedback signal.

### 9.2.2 Derivative control

With this form of control, the rudder is shifted by an amount proportional to the 'rate-of-change' of the vessel's deviation from its course. Derivative control is achieved by electronically differentiating the actual error signal. Its effect on the vessel's course is shown in Figure 9.3.

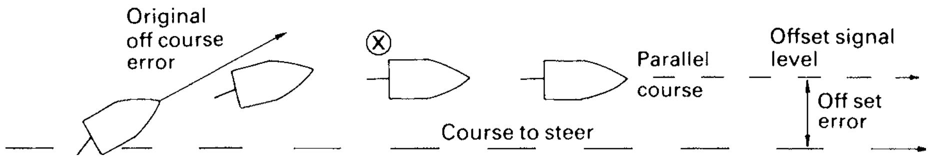


**Figure 9.3** The effect of applying derivative control only.

Any initial change of course error is sensed causing a corrective starboard rudder command to be applied. The rate-of-change decreases with the result that automatic rudder control decreases and, at point X, the rudder returns to the midships position. The vessel is now making good a course parallel to the required heading and will continue to do so until the autopilot is again caused to operate by external forces acting on the vessel.

An ideal combination of both proportional and derivative control produces a more satisfactory return to course, as shown in Figure 9.4.

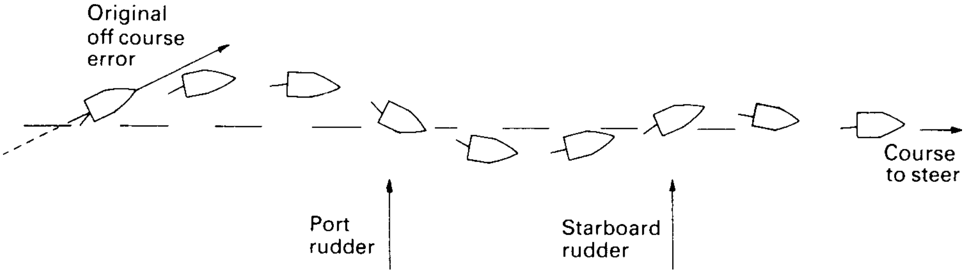


**Figure 9.4** Applying a combination of proportional and derivative control brings the vessel back on track.

The initial change of course causes the rudder to be controlled by a combined signal from both proportional and derivative signals. As the vessel undergoes a starboard turn (caused by proportional control only) there is a change of sign of the rate of change data causing some counter rudder to be applied. When the vessel crosses its original course, the rudder is to port, at some angle, bringing the vessel back to port. The course followed by the vessel is therefore a damped oscillation. The extent of counter rudder control applied is made variable to allow for different vessel characteristics. Correct setting of the counter rudder control should cause the vessel to make good its original course. Counter rudder data must always be applied in conjunction with the output of the manual 'rudder' potentiometer, which varies the amount of rudder control applied per degree of heading error.

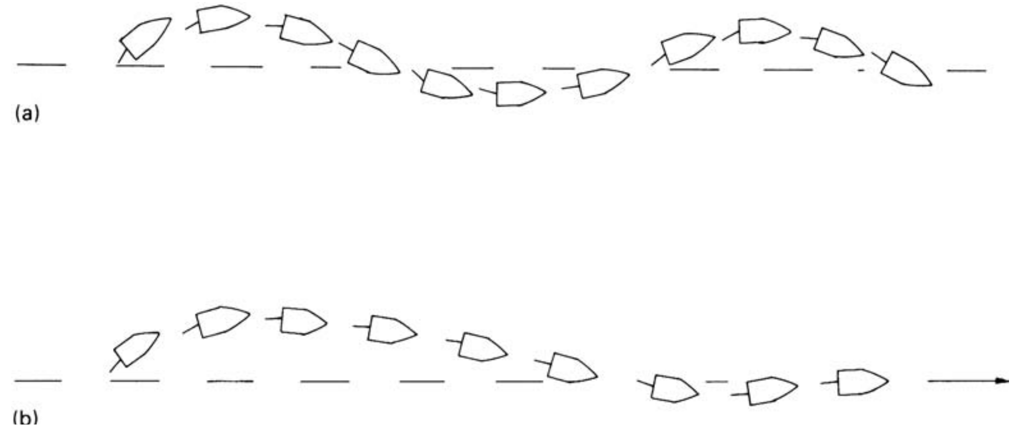


**Figure 9.5** (a) If 'counter rudder' and 'rudder' controls are set too high, severe oscillations are produced before the equipment settles.(b) If 'counter rudder' and 'rudder' controls are set too low, there will be little overshoot and a sluggish return to the course.

Figures 9.5(a) and (b) show the effect on vessel steering when the counter rudder and rudder controls are set too high and too low, respectively.

### 9.2.3 Integral control

Data for integral control is derived by electronically integrating the heading error. The action of this data offsets the effect of a vessel being moved continuously off course. Data signals are produced by continuously sensing the heading error over a period of time and applying an appropriate degree of permanent helm.

In addition to proportional control, derivative control and integral control, autopilots normally have the yaw, trim, draft, rudder limit, and weather controls, which will be dealt with in more detail later in this chapter.

## 9.3 A basic autopilot system

The simplest form of autopilot is that shown in Figure 9.6. An output from a gyro or magnetic repeating compass is coupled to a differential amplifier along with a signal derived from a manual course-setting control. If no difference exists between the two signals, no output will be produced by the amplifier and no movement of the rudder occurs. When a difference is detected between the two sources of data, an output error signal, proportional in magnitude to the size of the difference, is applied to the heading error amplifier. Output of this amplifier is coupled to the rudder actuator circuit, which causes the rudder to move in the direction determined by the sign of the output voltage. The error signal between compass and selected course inputs produces an output voltage from the differential amplifier that is proportional to the off-course error. This type of control, therefore, is termed ‘proportional’ control. As has previously been shown, the use of proportional control only, causes the vessel to oscillate either side of its intended course due to inertia producing overshooting.

With a Proportional, Integral and Derivative steering control system, the oscillation is minimized by modifying the error signal ($\psi$) produced as the difference between the selected heading and the

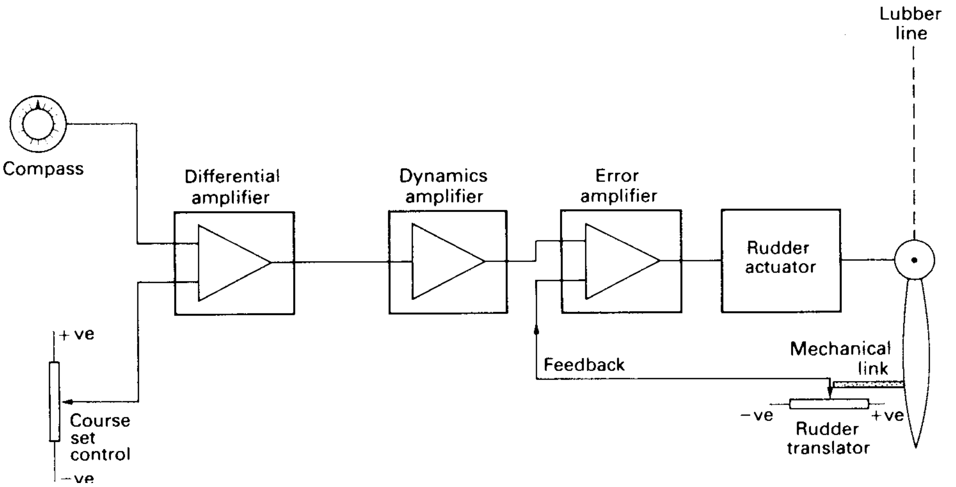


**Figure 9.6** A simple autopilot system.

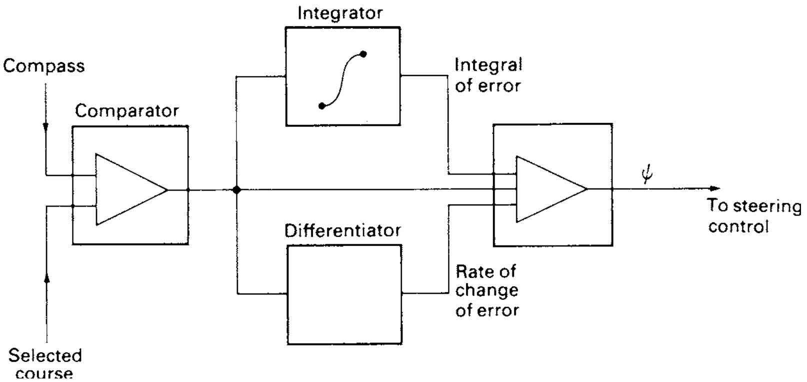


**Figure 9.7** Error signal summing circuit.

compass heading. Figure 9.7 shows that a three-input summing-amplifier is used, called a dynamics amplifier, to produce a resultant output signal equal to the sum of one or more of the input signals.

The demanded rudder error signal ($\psi$) is inspected by both the differentiator and the integrator. The differentiator determines the rate of change of heading as the vessel returns to the selected course. This sensed rate of change, as a voltage, is compared with a fixed electrical time constant and, if necessary, a counter rudder signal is produced. The magnitude of this signal slows the rate of change of course and thus damps the off-course oscillation. Obviously the time constant of the differentiation circuit is critical if oscillations are to be fully damped. Time constant parameters depend upon the design characteristics of the vessel and are normally calculated and set when the vessel undergoes initial trials. In addition, a 'counter rudder' control is fitted in order that the magnitude of the counter rudder signal may be varied to suit prevailing conditions.

Permanent disturbances of the course due to design parameters of the vessel must also be corrected. These long-term errors, typically the shape of the hull or the effect of the screw action of a single propeller driving the ship to starboard, may be compensated for by the use of an integrator. The integral term thus produced is inserted into the control loop offsetting the rudder. This permits proportional corrections to be applied about the mean offset course (the parallel course shown in Figure 9.3). The offset signal amplitude causes a permanent offset-error angle of the rudder. The output of the dynamics amplifier is now the total modified error signal ($\psi$) which is regulated by the 'rudder' control to determine the amount of rudder correction per degree of heading error to be applied.

An overall simplified diagram of an autopilot is shown in Figure 9.8.

The rudder error amplifier is provided with variable sensitivity from the 'weather' control, which in effect varies the gain of the amplifier by varying the feedback portion of the gain-determining components. In this way the magnitude of the heading error signal required, before the output from this amplifier causes the rudder to operate, may be varied. Using this control a delay in rudder operation may be imposed if weather conditions cause the vessel to yaw due to a heavy swell aft of the beam.

Under certain conditions, mainly draft and trim of the vessel, a degree of permanent rudder may be required. The 'permanent helm' control provides an input to the rudder error amplifier that may be positive or negative depending on whether the rudder needs to be to starboard or to port. Since the effect of rudder movement does not influence the setting of this control, the rudder will remain

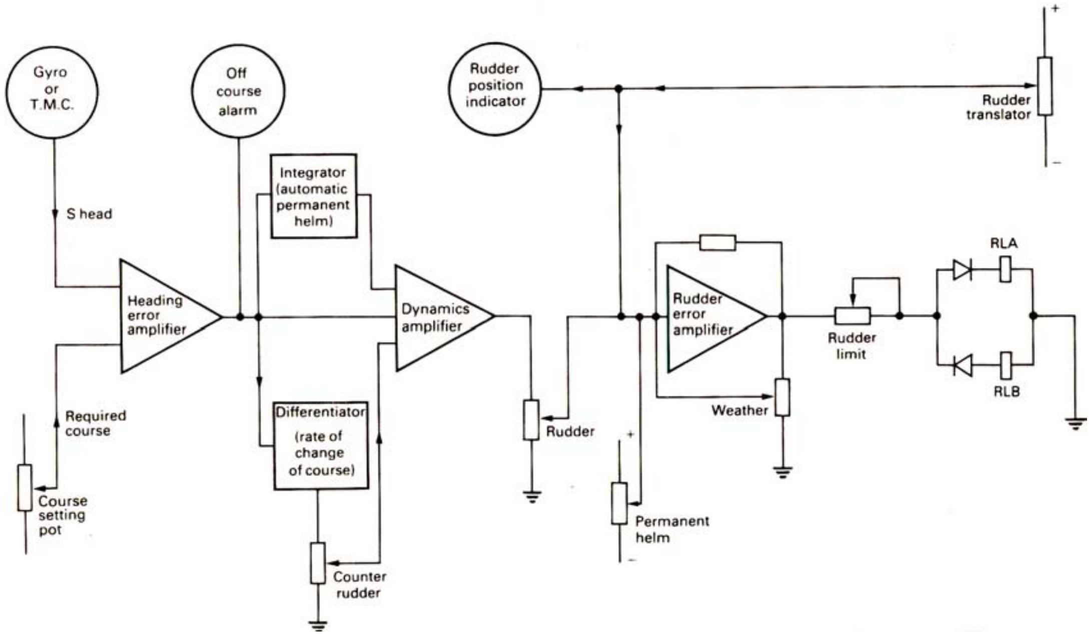


**Figure 9.8** A simplified diagram of an autopilot system.

permanently in the position set by the control (assuming no other control signals are produced). Permanent helm will also be applied automatically by sensing the build-up of heading error in the integrator circuit.

In the system described control relays RLA and RLB are used to switch power to the steering gear contactors, which in turn supply power of the correct amplitude and polarity to the prime rudder mover. As the rudder moves, a mechanical linkage drives the slider of a potentiometer to produce the rudder feedback signal. Output from this ‘rudder translator’ potentiometer is normally used to indicate the instantaneous rudder angle. Excursions of the rudder may be limited by the manually operated ‘rudder limit’ control which fixes the maximum amount by which the rudder may move from the midships position.

An off-course alarm circuit senses the error signal at the output of the heading error amplifier and causes an audible alarm to be sounded when a signal amplitude outside pre-determined limits is detected. A manual off-course limit control (not shown) is provided to enable an operator to select the point at which the alarm will sound.

## 9.4 Manual operator controls

### 9.4.1 Permanent helm

This control is intended for use when the vessel is being driven unilaterally off-course by a crosswind. Its function is to apply sufficient permanent rudder angle to offset the drift caused by the wind, thus holding the vessel on the required heading. Permanent helm is also applied automatically when the steering system is in the automatic mode of operation.

Automatic application of permanent helm makes no use of the permanent helm control. The degree of rudder offset required for course holding is now electronically computed and applied automatically.

Since the computing process involves the charging of a capacitor, the required degree of permanent helm is built-up gradually over a period of minutes. This period may be changed by altering the charging time of the capacitor.

### 9.4.2 Rudder

Rudder limit control sets a finite limit on the rudder angle obtained irrespective of the angle commanded by the automatic control circuitry. Obviously if the rudder was permitted to exceed design parameters severe damage may be caused.

The rudder potentiometer enables the ship's steering characteristics to be modified in accordance with the changing requirements caused by loading and speed factors. This control determines the absolute degree of rudder command obtained for every degree of steady-rate heading error. For example, if this control is set to '2', the rudder will move through 2° for every degree of heading error.

The counter rudder control determines the degree of opposite helm to be applied if it is demanded by the control circuit. The control permits daily adjustments to be made as dictated by loading conditions.

### 9.4.3 Weather

The effect of weather and sea conditions can be effectively counteracted by the use of this control. The circuits controlled by this switch progressively desensitize the control amplifier, which in turn causes an increase in the deadband width. The control also imposes an increasing time delay on the rudder command signal in order that the ship will recover naturally when under the influence of repetitive yaw. This means that the steering gear is not subjected to continual port/starboard commands. Thus the higher the setting of the weather control, the wider will be the deadband. This increases the amplitude of yaw that can be tolerated before the steering gear is enabled.

### 9.4.4 Non-follow-up mode (NFU)

The rudder is manually controlled by means of two position port/starboard lever switches. These switches energize the directional valves on the hydraulic power unit directly, thus removing the rudder feedback control. In this mode the normal autopilot control with repeat back is by-passed and the rudder is said to be under 'open loop' control. There is no feedback from the rudder to close the loop. The helmsman closes the loop by observing the rudder angle indicator and operating the NFU control as appropriate.

### 9.4.5 Follow-up mode (FU)

In this mode the FU tiller control voltage is applied to the error amplifier (Figure 9.9) along with the rudder feedback voltage. Rudder action is now under the influence of a single closed loop control.

## 9.5 Deadband

Deadband is the manually set bandwidth in which the rudder prime movers do not operate. If the deadband is set too wide, the vessel's course is hardly affected by rudder commands. With the control set narrow, the vessel is subjected to almost continuous rudder action causing excessive drag.

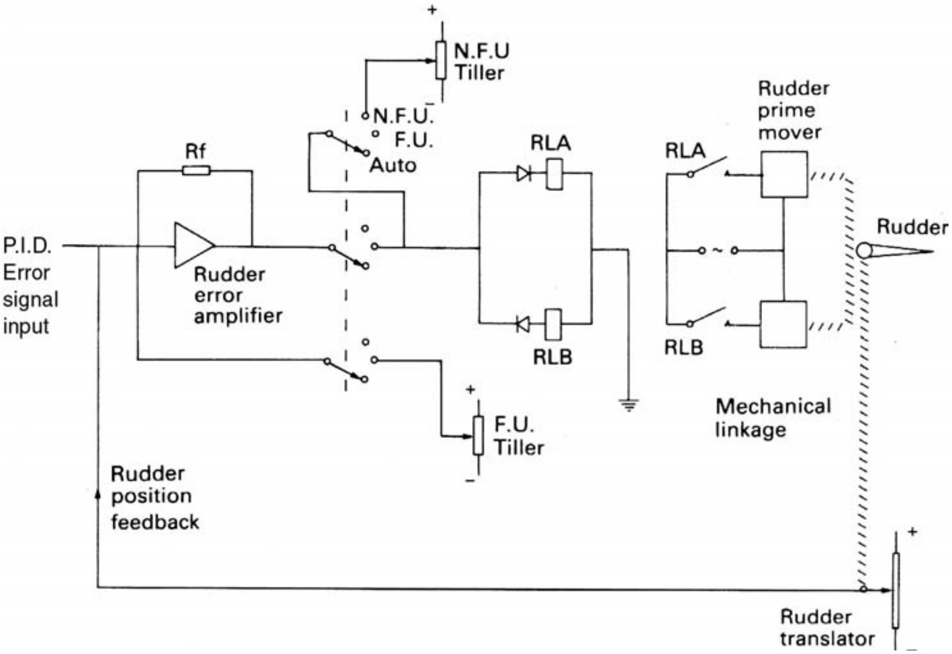


**Figure 9.9** FU and NFU control of tiller operation. (Reproduced courtesy of Racal Marine Controls.)

### 9.5.1 Overshoot

For optimum course-keeping performance it is imperative that an autopilot operates with as narrow a deadband as possible. All steering systems suffer a degree of inherent overshoot. The effect of this overshoot on the stability of the rudder positioning system can be graphically represented as shown in Figure 9.10.

Two scales are plotted on the vertical axis, the first shows the rudder angle in degrees with respect to the midships position and the second, the voltage corresponding to that angle produced by the rudder translator.

It is assumed that a starboard rudder command is applied to the autopilot at time $t = 0$ s, and as a result the starboard rudder controller pulls in to cause the rudder to move to starboard. Since the mechanical linkage of most autopilot systems take a finite time to develop full stroke, the rudder does not reach its terminal velocity until $t = 2$ s. At time $t = 9$ s, the position feedback signal ($Vp$) crosses the release threshold of the starboard relay. Prime power is now removed from the steering gear pump. Because of inherent overshoot, caused by inertia, the rudder will continue to move to starboard as shown by the solid line. If the overshoot is of sufficient magnitude, it will cause the position feedback signal to cross the operating threshold of the port relay ($t = 12.5$s), and thus set the rudder moving towards the midships position. When, at $t = 15.25$s, $Vp$ crosses the release threshold of the port relay, power is again removed from the steering gear. Overshoot now carries the $Vp$ signal back through the operating threshold of the starboard relay and the rudder once again moves to starboard. The control system is now described as unstable and the rudder is caused to oscillate or hunt.

The dotted curve in Figure 9.10 illustrates the operational characteristics of a stable system. Here, overshoot does not cause the port relay to be activated and thus the rudder arrives at the commanded position in one continuous movement.

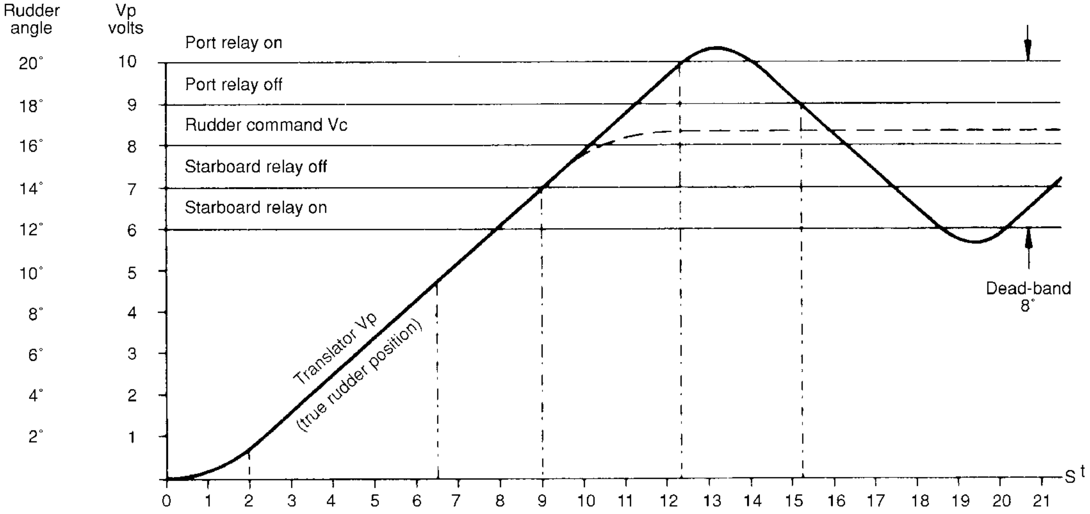


**Figure 9.10** Effect of overshoot on control system stability. (Reproduced courtesy of Racal Marine Controls.)

One method of stabilizing an unstable system is to decrease the sensitivity of the rudder amplifier. This solution is not satisfactory because it has the effect of increasing the distances between the 'operate and release' thresholds of the steering relays thus producing a wider deadband and a degradation of the steering performance and efficiency.

A better solution is to remove power from the steering gear at some determinate time before *Vp* crosses the release threshold of the starboard relay. The extent of this pre-determined release time must be dependent upon individual steering gear overshoot characteristics. In Figure 9.10, if power was removed from the steering gear at $t = 6.5$ s (a time advance of 2.5 s), the inherent overshoot would not carry *Vp* through the operating threshold of the port relay and rudder movement will follow the dotted line illustrating a stable system. This principle is an outline of a system known as phantom rudder.

## 9.6 Phantom rudder

Dependent upon the setting of the 'phantom rudder speed' control, a determinate d.c. voltage is applied to an integrator input resistance with the result that the circuit starts to generate the positive going ramp voltage *Vp* defined by the solid line in Figure 9.11.

It should be noted that the polarity of the integrator output is the reverse of that of the translator output *Vt*, hence the provision of separate voltage scales on the y-axis of the graph. It is arranged so that the slope of *Vp* and *Vt* are equal. On the assumption that the steering gear takes 1 s to run up to speed, the phantom output establishes a lead of approximately 0.75 V (1.5°) during this period. At time $t = 2.4$ s, the phantom output, functioning as a position feedback signal, arrives at the release threshold of the starboard relay, one contact of which removes the input from the integrator causing the output to halt at +3 V. It is arranged that at this time a second input is applied to the phantom rudder circuit integrator, which now produces a negative going ramp. The slope of this ramp is made to be gradual by limiting the amplitude of the signal applied to the integrator.

At time $t = 3$ s, the phantom (*Vp*) and translator (*Vt*) outputs will be equal and of opposite polarity causing the output from the integrator (*Vp*) to stop increasing. This condition is not stable because as

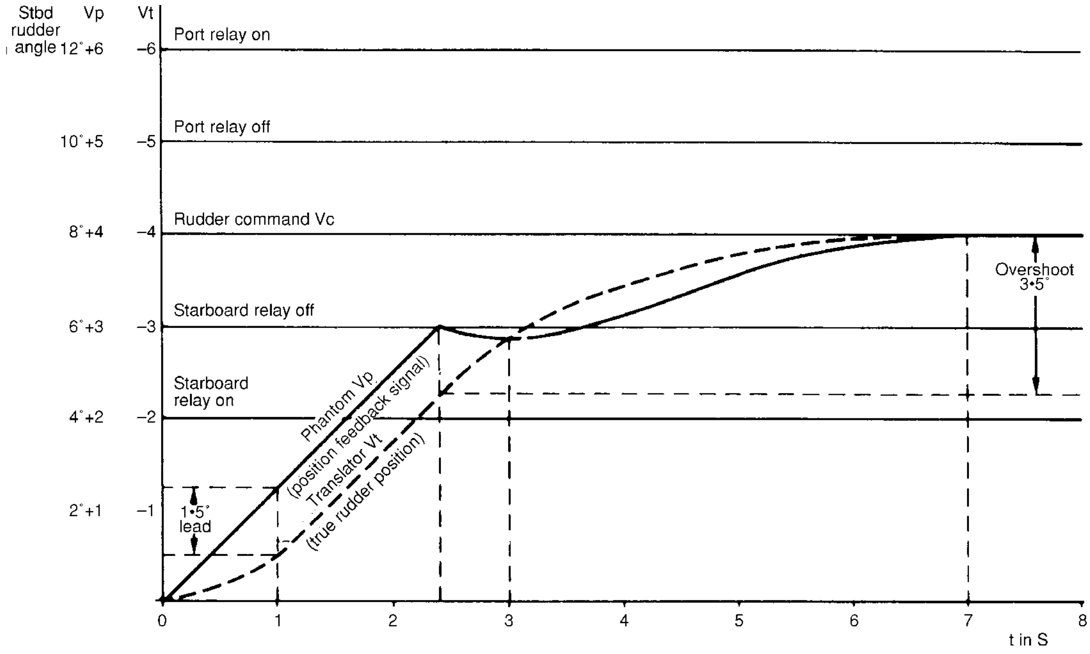


**Figure 9.11** Operational principle of a phantom rudder. (Reproduced courtesy of Racal Marine Controls.)

*Vt* is carried progressively more negative by rudder overshoot, the integrator generates a positive going ramp with a low incline. Output from the integrator will now continue to rise, and the slope will gradually decrease as the positive potential of *Vp* approaches parity with the negative potential of *Vt*. Ultimately at $t = 7$ s, *Vp* will be equal to *Vt*. Since no input is now applied to the integrator, its output *Vp* will be held at the attained level, and the hypothetical position of the phantom rudder will be the same as that of the true rudder.

In the foregoing example, the lead of the phantom rudder on the true rudder was obtained purely as a result of the slow take-off of the steering gear. In practice, it is desirable that the phantom rudder speed output be set 20% higher than that of the true rudder. Since, with this arrangement, the phantom rudder output will continue to increase its lead on the translator output so long as the steering gear is energized, some means has to be provided to limit the lead that the phantom output is permitted to build up. This function is performed by the ‘steering gear overshoot’, effectively limiting the rise time of the integrator causing *Vp* to level off in stages as illustrated in Figure 9.12.

## 9.7 The adaptive autopilot

Autopilot systems so far described have operated under various command functions, the origins of which have been small signals produced by feedback loops. The rudder command-loop signals have been further modified by the proportional, integral and derivative terms to form the nucleus of the PID autopilot systems. The adjustment of operator controls on the PID autopilot requires considerable expertise if the system is to operate efficiently. It is not feasible to continually reset

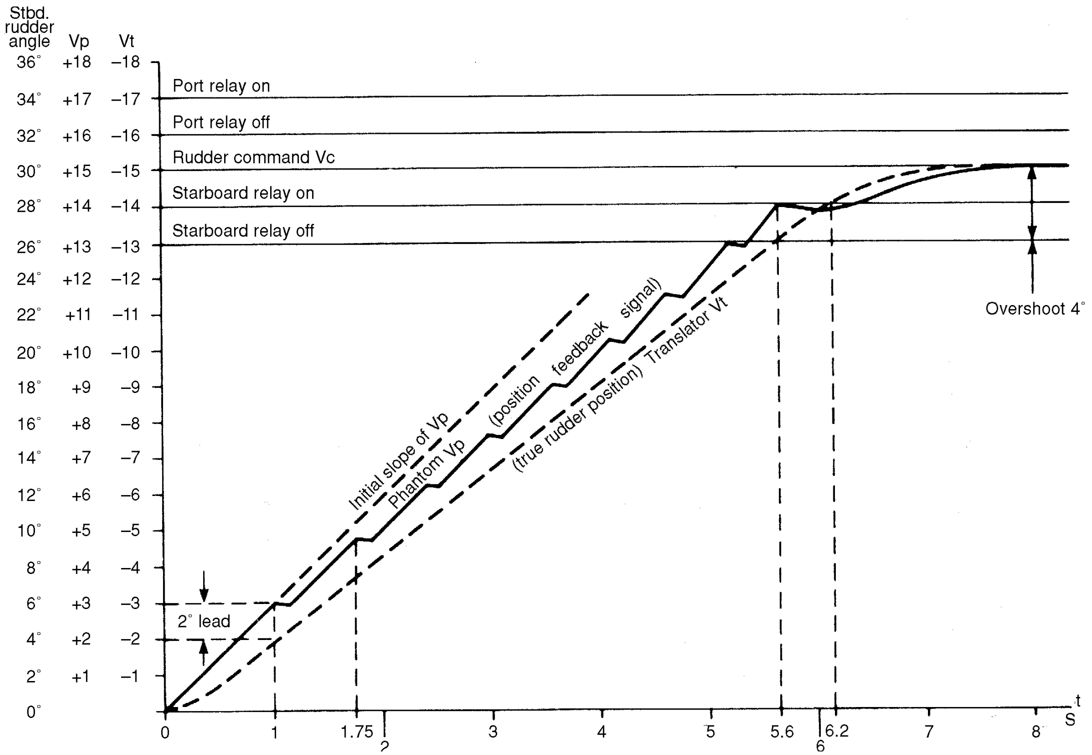


**Figure 9.12** Characteristics of a practical application of phantom rudder. (Reproduced courtesy of Racal Marine Controls.)

the potentiometers during constantly varying weather conditions; thus the system cannot be absolutely efficient.

The PID autopilot was developed in an effort to enable a vessel to follow a course as accurately as possible by reducing drag caused by excessive rudder angles whilst limiting rudder excursions to a low level in order to minimize wear on the steering gear. Considerable research has been undertaken into the effects of the ship's natural yaw action in relation to the course to be steered, and it has been found that a straight course is not necessarily the most economical and that the ship's natural yaw action should not be smoothed out.

Operating parameters for modern adaptive autopilots (AAPs) have been developed by a number of notable design engineers over the past three decades. Probably the most influential of these is N. H. Norrbin who, in the early 1970s, derived a performance index relating to added resistance due to imperfect steering control. This he produced in the measurable term of 'the square of the average heading error'. Most modern AAP controllers use this index as the fundamental control term. In addition to the fact that a straight course is not the most economical course it was decided that steering control should always be optimized with respect to the prevailing environmental conditions and a low bandwidth should be used to minimize losses. There are, therefore, two main factors that affect the steering control.

- The complex characteristics of the vessel. Handling parameters will be different for each vessel, even of the same type, and will change with the loading factor.

- Environmental influences, namely wind and tide which will be constantly shifting and introducing instantaneous variable course errors.

As has been standard practice for many years, ship handing characteristics can be programmed into a standard autopilot system and their effects counteracted. However, environmental effects are a different matter. They can, to some extent, be counteracted by the helmsman. But to nullify their effects totally would require the skills of an extraordinary person, one with the ability to instantaneously predict all ship and environmental effects before applying corrective rudder. Such a helmsman would be a treasure indeed. It is more logical to replace the helmsman with a computer that is able to react more quickly to the constantly changing parameters that are input from various sensors.

The AAP is, in its simplest form, a good quality autopilot system with the addition of a digital control system (microcomputer) producing the final rudder command signal. Contained in the microcomputer are data relating to the dynamics of a 'virtual ship' which may be analysed in order that rudder commands for the actual ship can be predicted. Obviously the dynamics of this 'virtual ship' are critical to the AAP operation and in practice will be accurately set for the vessel on which the AAP is fitted.

### 9.7.1 The 'virtual ship' principle

Most adaptive autopilot equipment is designed around the 'virtual ship' principle, a computer-generated model vessel, and the following criteria.

- The ship's operating envelope, including the vessel's speed, load factor and external environmental conditions.
- Precise dynamics of the vessel that relate directly to its steering control.
- The dynamics of the ship's steering system.
- The dynamics of the gyrocompass.
- The dynamics of the seaway.

It is then necessary to define the principal modes of operation that require specific performance criteria. The most used of these modes is open sea course keeping where optimized steering can lead to potentially large savings in fuel oil.

### 9.7.2 Open sea course keeping

Fuel consumption, which is of major importance for the economic operation of a vessel, is affected by a number of factors, such as engine performance, trim, and the condition of the hull below the waterline. These factors are, however, predictable and counteractive data is easily obtained and input to the AAP. It is essential that the central processor is able to distinguish between ship/engine loss parameters and rudder movements, and apply corrective rudder only when course keeping is affected by environmental conditions and not by the natural yaw of the vessel. Various mathematical formulae have been developed to analyse the AAP integral term to optimize rudder performance. Thus the AAP system automatically minimizes propulsion losses and is termed an adaptive control system. The term adaptive is used because the mathematical parameters of the model ship have been 'adapted' to match those of the actual vessel.

The performance criterion, when reduced to a form suitable for online evaluation on board ship, may be represented as

$$J = (\lambda\psi^2 + \delta^2)dt$$

where $\psi$ = ship's heading error,
$\delta$ = rudder angle, and
$\lambda$ = weighting factor derived from analytical expressions of drag forces due to steering.

Obviously the adaptive autopilot must be able to detect that a course change has been commanded. This is the function of the course changing control circuitry.

### 9.7.3 The course changing controller

When changing course it is standard practice to consider three phases of the manoeuvre:

- the start of the turn
- the period of steady turn
- the end of the turn.

The measure of rudder applied determines the rate-of-turn and also the peak roll-angle. In practice therefore the maximum roll-angle is determined by the maximum permissible rudder limit. Proportional and rate gains can be obtained for each vessel and its loaded condition as a function of speed. In an AAP, gains are chosen based on the optimized results of the virtual ship during a controlled turn. The primary concern of the AAP whilst manoeuvring in confined waters must be safety.

### 9.7.4 Confined waters mode

When manoeuvring in confined waters, it is essential that cross-track error be minimized. Since the central processor cannot determine cross-track data, an alternative mathematical concept is used. Balancing the heading error against the rudder rate derives cross-track data.

$$J = (\lambda\psi^2_e + \delta^2)dt$$

The main difference between the open sea course keeping controller and the confined waters controller is that the gain of the latter is varied only as a function of the ship's speed.

## 9.8 An adaptive digital steering control system

Sperry Marine Inc., now part of the Litton Marine Systems group, is a traditional manufacturer of compass and control equipment, and their ADG 3000VT Adaptive Digital Gyropilot® Steering Control system is a good example of an up-to-date autopilot using many of the principles described in this chapter (see Figures 9.13 and 9.14).

At the heart of the autopilot is a sophisticated microcomputer and electronic circuitry providing control signal outputs to the steering gear pump controllers. The microelectronic circuitry is programmed (calibration/configuration CALCON) at installation to set controller gains and time

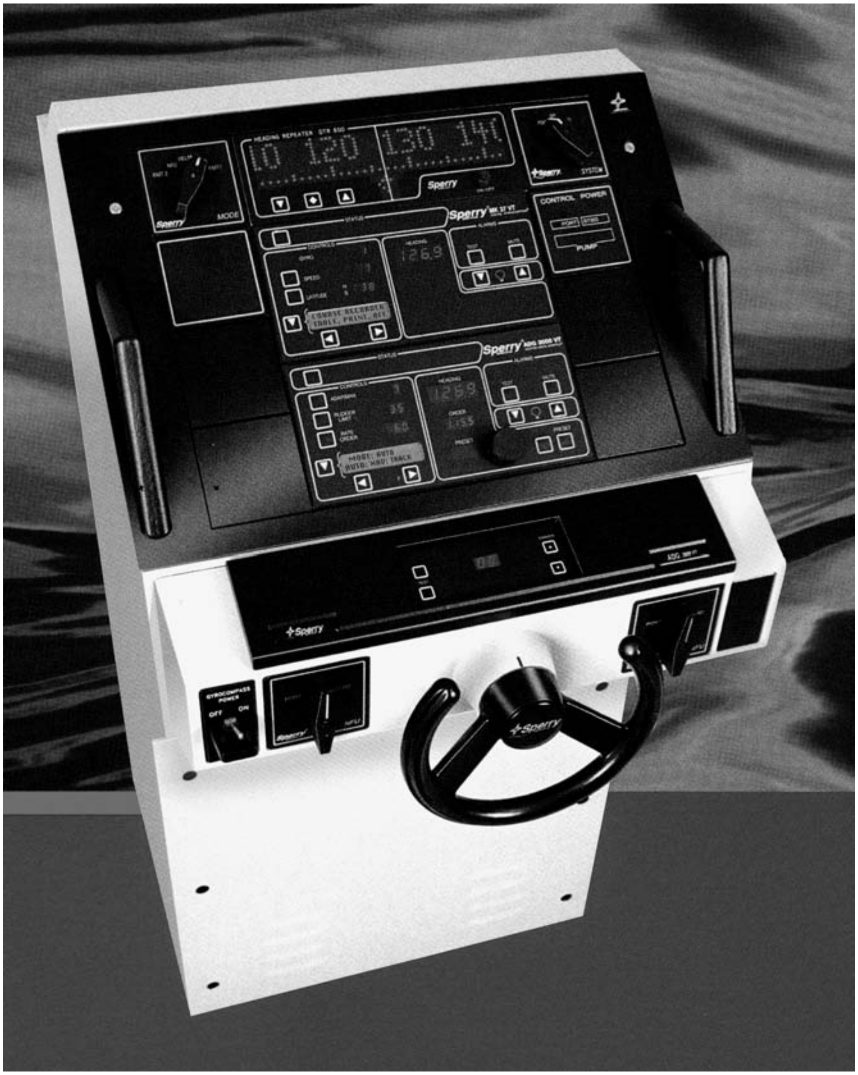


**Figure 9.13** The Sperry Adaptive Digital Gyropilot“ Steering Control Console. (Reproduced courtesy Litton Marine Systems.)

constants specific to the ship's design affecting heading keeping and manoeuvring. Inputs of speed and heading data are provided from a speed log and a gyrocompass for automatic operation, and manual control is from the primary helm unit, primary NFU controller or remote FFU controllers. Rudder angle information feedback data is interfaced with the main electronics unit. System control is from the main control panel, although a serial I/O data line enables automatic course order entry from an integrated bridge unit, such as the Sperry Marine Inc's Voyage Management System VMS.

Manual steering and input commands are under the control of the helmsman whereas automatic steering can be performed from three different automatic steering modes.

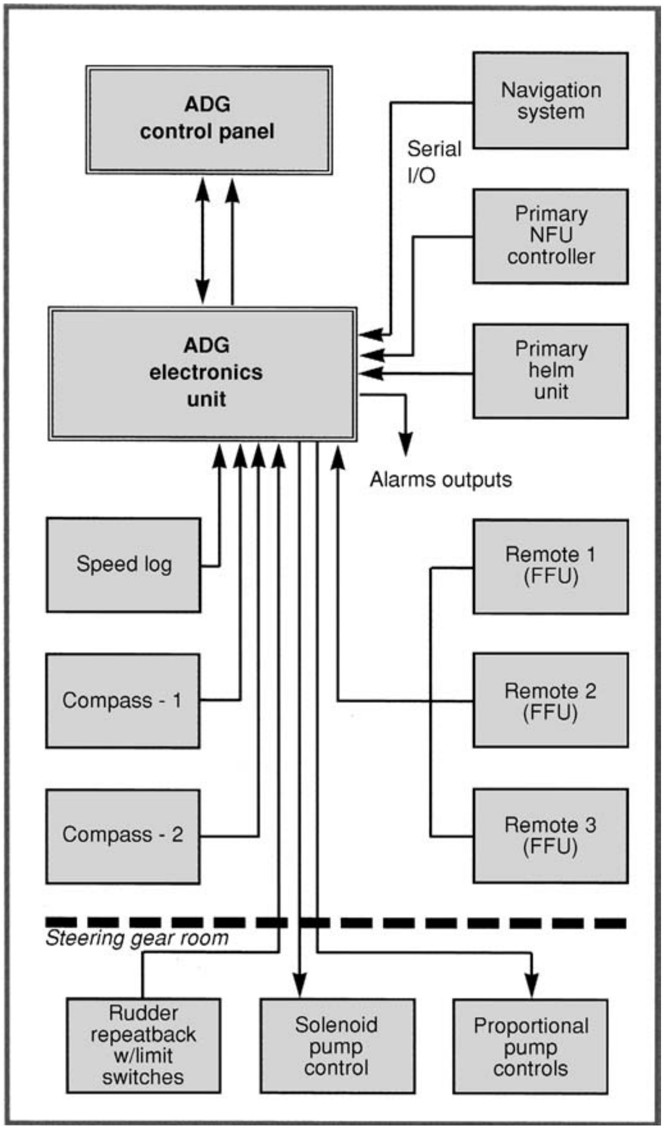


**Figure 9.14** Sperry ADG 3000VT autopilot bridge configuration. (Reproduced courtesy of Litton Marine Systems).

- AUTO mode. The primary mode, used for automatic heading keeping with data input from a gyrocompass and the helmsman's Order setting.
- NAV mode. An optional steering mode performing automatic heading keeping using inputs from an external management system to steer the vessel to pre-determined waypoints. Only available if installed by Calcon during installation.
- TRACK mode. An optional steering mode using inputs from an external navigator, which is corrected for cross-track error by the autopilot to steer the ship to a waypoint over a designated ground track.

## 9.8.1 Operation

As previously stated, the autopilot can be operated in one of three modes; Auto, Nav and Track mode. This section considers the autopilot when using the Auto mode with interfaced data from a gyrocompass and a helmsman inputting data via the keypad.

Referring to Figure 9.15, the number 1 is the Status selector switch and numbers 2–6 are the indicators. In this case Auto will be selected. Number 7 is the Adap/Man display and number 20 is the control. An 'A' is displayed indicating that adaptive (automatic) gain selection has been chosen by the helmsman to compensate for sea conditions. If a fixed gain setting is selected, the display shows a number 1–7, with the highest number indicating the lowest gain and therefore the lowest number providing tighter heading keeping. This choice will depend upon sea conditions. In this case Adaptive has been selected permitting the autopilot to determine automatically the gain, based upon heading error and rudder activity.

The number 8 (displaying 15 in the diagram) is the rudder set-limit display, set by the helmsman with switch 17. This may be set to any value between 1° and the ship's maximum permitted rudder angle. It is the 'effective' rudder limit based on 'weather helm' and may differ from the true rudder angle. The indicator 18 lights to show that the rudder order output is equal to the selected rudder limit.

Display number 9 (showing 014) is the rate order display. This is set by control 15, which selects the turn rate to be followed during turn manoeuvres.

Display number 10 is the status and control display showing autopilot information. Menu scroll buttons, 11 and 12, and switch 13 select different display data. As an example, the display is currently showing Turn Radius Order data but it may be switched to one of many other functions affecting the operation of the autopilot and not immediately obvious on the control panel display. These include deadband, turn radius, speed selection, load selection, and rudder order bias amongst other functions.

Number 22 indicates the vessel's heading display derived from the master compass input data. The display numbered 40 shows the current heading order (in 1/10th of a degree) and display number 38 displays the heading order entered via the control knob 37 or from the pre-set values, controls 34 and 35.

Indicator number 24 warns of an off-course situation, indicator 27 warns of a malfunction in the system, and indicator number 28 warns of errors or failure of the compass input data.

After a successful power-up, during which the autpilot performs a self-diagnostic function, the following sequence enables automatic operation.

- Adjust the autopilot controls to the desired settings.
- Verify that the steering control system is selecting the autopilot.
- Press the Status switch to select Auto.
- Rotate the Order knob until the required heading-to-steer appears in the Order display.

## 9.8.2 System description

The heart of the autopilot is the central processing unit holding a processor, various I/O ports and buffers, and ROM/RAM memory. A block diagram of the overall system is given in Figure 9.16.

Rudder commands are output to U6, a dual channel 12-bit digital-to-analogue converter (DAC) on the Analogue Digital Serial board (ADS), giving an analogue output in the range ±5 V to the rudder servo-amplifier. Ultimately this circuitry provides a dual proportional rudder order (RO) analogue

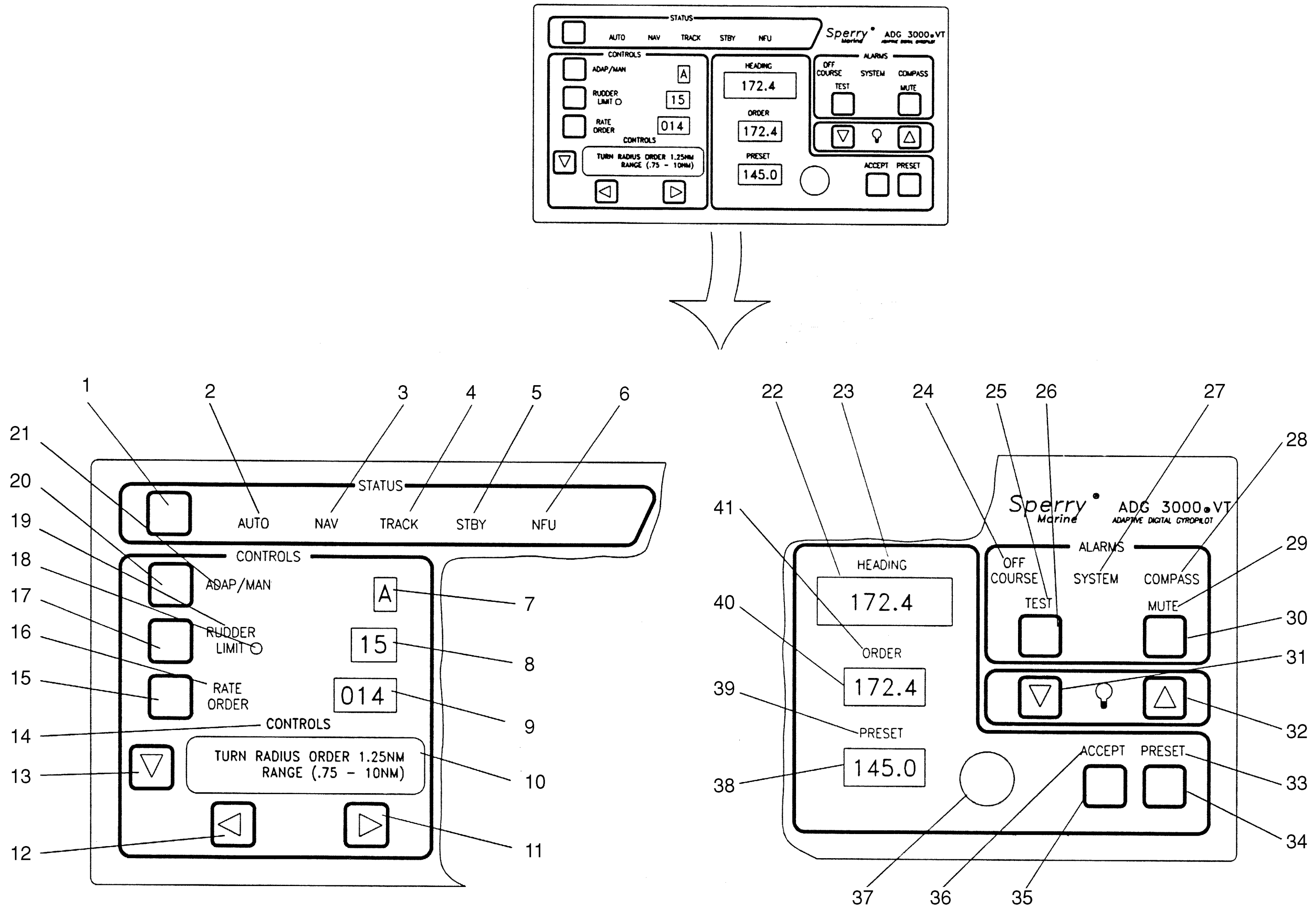


**Figure 9.15** Display unit. (Reproduced courtesy Litton Marine Systems.)

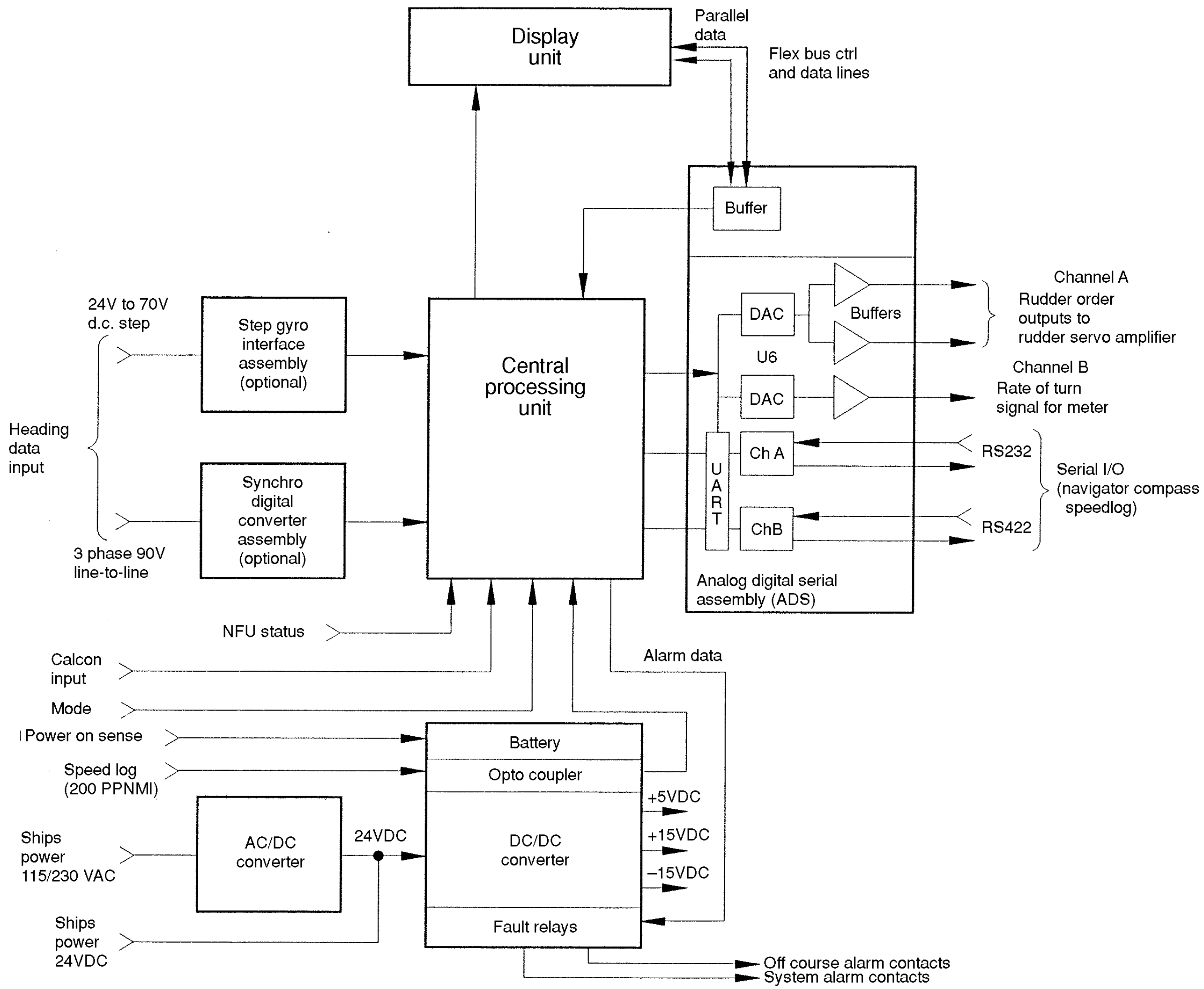


**Figure 9.16** Overall system block diagram. (Reproduced courtesy of Litton Marine Systems.)

voltage capable of driving rudder servo-amplifiers. The output voltage is in the range –11.25 to +11.25 V, corresponding to 45° left and right rudder orders, or 0.25 V/°. The output of DAC-A is routed through an analogue switch that prevents any output at power-up and thus produces 0 V to the rudder servo-amplifier. This ensures there are no rudder commands until the central processor is ready. The output of DAC-B is gated to provide rate-of-turn signals.

The ADS circuitry also communicates with the display unit via an 8-bit bi-directional parallel data bus interface (flex bus). This is the main command path between the CPU and the display unit. The ADS board also has two serial I/O channels to communicate with an electronic navigator, compass or speed log equipment: channel A is configured to be an RS-232 interface and channel B an RS-422 line. Speed information may be input on the RS-422 interface line in the NMEA 0183 format or via the power board optocoupler in 200 pulses per nautical mile (200 PPNMI). See Table 9.1.

**Table 9.1** Interface signals. (Reproduced courtesy of Litton Marine Systems)

| | |
|---|---|
| *Inputs* | |
| Speed log input | |
| Pulsed | 200 pulse/nautical mile (PPNMI) format (contact closure) |
| Serial | RS-232 (channel A or C) or RS-422 (channel B) communications in NMEA 0183 format, $VBW, $VHW |
| Navigator (vessel management system) input | Serial data for heading order, rate order, and cross track error information in RS-232 or RS-422 communication on channel A, B or C. in NMEA format $APB, $HSC, $HTR, $HTC or $XTE. |
| Compass | |
| Step data | Positive or negative step data (24 or 70 V) |
| Synchro | 1X, 90X or 360X |
| Data | $HDT (on channels A, B or C) |
| Serial data | |
| Mode switch sense contacts | External switched opened or closed to inform autopilot to change from Standby mode to an automatic mode |
| NFU sense contacts | External contacts to indicate when the NFU controller is active |
| Power failure circuits | Closed contacts on external power switch to activate power failure alarm |
| *Outputs* | |
| Interface to external rudder servo control amplifiers | Bipolar analogue voltage proportional to the rudder order. ± 11.25 V (maximum limit) equal to ± 45° or rudder |
| Rate of turn interface | Bipolar analogue voltage proportional to a turn rate indicator.<br>± 4.5 V (max) equal to ±90° turn/min. Resolution equal to 0.5°/min. |

Gyrocompass input data is coupled via either a synchro digital converter assembly (SDC) or a step gyro interface assembly to the CPU board. Both boards have the same function, i.e. to convert the azimuth data from a gyrocompass to a suitable data input for the CPU. The SDC board accepts a synchro azimuth input as three-phase 90 V line-to-line signal to a resolver circuit. A built-in test (BITE) circuit detects any errors or failure in the azimuth data at this point. Output from the resolver is a 16-bit data line to the CPU.

All the external operator command functions are requested through the display unit. The CPU scans the X select lines of an X–Y matrix and monitors each of the Y lines sequentially searching for a keypad command. When a switch is pressed the X select gets transferred to a particular Y line and the command is initiated.

### 9.8.3 NMEA 0183 interface format

Communication with other shipboard navigation equipment is via the RS-232 and RS-422 ports. Message format and field definitions are outlined below using the speed serial interface as an example. The heading, heading order, and speed messages follow the NMEA 0183 format with extensions for status and tenths resolution.

Incoming messages are required to begin with the string, $ttsss, where: tt = (upper case characters) talker identifier; and sss = (upper case characters) sentence identifier. Incoming messages may omit the '*cc' checksum field.

**Table 9.2** Sperry ADG 3000 VT Autopilot NMEA 0183 input message styles. (Reproduced courtesy of Litton Marine Systems)

*(a) Input message styles*

| *Sentence* | *Data* | *Expected rate (Hz)* | *Time delta without message before alarm(s)* |
|---|---|---|---|
| HDT | Heading, true | 8 | 1 |
| VBW | Velocity, bottom and water | 1 | 4 |
| HSC | Heading order command | $\frac{1}{2}$ | 15 |
| HTC | Heading of course to next waypoint | $\frac{1}{2}$ | 15 |
| HTR | True rate order | $\frac{1}{2}$ | 15 |
| XTE | Cross track error | $\frac{1}{2}$ | 15 |
| APB | Alternate order command and cross track error | $\frac{1}{2}$ | 15 |
| VHW | Alternate water speed | 1 | 4 |

*(b) Output message style*

| *Sentence* | *Data* | *Output rate (Hz)* |
|---|---|---|
| FLT | Faults | 1 |
| HSC | Heading to steer | 1 |
| HTC | Heading to waypoint | 1 |
| ROR | Autopilot rudder order | 1 |
| STA | Autpilot status and commands | 1 |
| STB | Autopilot controls | 1 |
| VHW | Heading and water speed | 1 |

### *Serial speed interface*

If the autopilot is configured for serial speed input, the CPU reads fore/aft speed data from the configured channel. In an automatic mode, the CPU expects one message per second. If it does not receive at least one message within 4 s, the system alarm is set and the autopilot defaults to the last known speed input.

If water speed is supplied but marked invalid, the processor uses it for steering; if water speed is unavailable, the processor uses bottom speed. In either case, a system alarm is set for misformatted messages. If speed is constantly less than 1 knot, the processor sets a system fault and uses the normal service speed instead. The processor reports speed system faults only in automatic steering modes.

The NMEA 0183 input speed message format is:

$ttVBW,sww.w,sx.xx,a,syy.y,szz.z,a*cc<cr><lf>

where tt = talker ID;
s = negative for aft/port speeds, omitted for fore/starboard speeds;
ww.w = alongship water speed in knots;
xx.x = athwartship water speed in knots;
yy.y = alongship bottom speed in knot;
zz.z = athwartship bottom speed in knots;
a = status sign: A, if valid speed data is available; V, if not;
cc = ASCII hex 8-bit XOR characters after '$' through the letter before the '*';
<cr><lf> = carriage return and line feed end-of-sentence markers.

Examples are:

$ttVBW,20.0,,A,,,V = Valid water speed with trailing zeroes omitted.

$ttVBW,,,V,18.2,,A = Bottom speed.

### 9.8.4 Troubleshooting

The system possesses an extensive fault identification system that enables system malfunctions to be isolated to a circuit board or major subassembly level. Extensive use is made of the system's BITE function to identify types of malfunctions by means of pre-programmed diagnostics. It is also possible for an operator to diagnose certain types of faults that are undetectable by the processor-dependent BITE functions.

If a System or Compass alarm occurs, the operator presses and holds the Mute switch for 4 s or longer. During this time, the CPU will search for a malfunction and, if it is one of the listed faults in Table 9.3, it will display error information and a code corresponding to that condition. For instance, if there is an error in the input data from a speed log, the autopilot display may show Speed Log Error 40 Enter Manual Speed. Referring to the fault code and corrective action chart, part of which is reproduced as Table 9.3, it is possible to locate and/or replace a faulty assembly.

The fault logic diagram, shown in Figure 9.17, shows the procedure to be followed if no fault code is present on the control unit display.

**Table 9.3** Sperry ADG 3000 VT fault codes and corrective action chart. (Reproduced courtesy of Litton Marine Systems)

| *Fault message (20 spaces per line)* | | *Description* | *Corrective action* |
|---|---|---|---|
| (DEVICE B) LOST<br>(See Above) | 14 | Loss of receiver interrupts for 15 s (when NMEA 0183 device installed on RS-422 port (Channel B)) | a. Check source.<br>b. Check wire connection.<br>c. Replace ADS Assembly.<br>d. Replace CPU Assembly. |
| (DEVICE B) ERROR<br>(See Above) | 15 | Framing error; invalid message bit format (on RS-422) | a. Check serial channel wire protocol of source.<br>b. Replace ADS Assembly. |
| (DEVICE B) ERROR<br>(See Above) | 16 | Overrun error (on RS-422) | a. Check serial channel wire protocol of source.<br>b. Replace ADS Assembly. |
| (DEVICE B) ERROR<br>(See Above) | 17 | Loss of transmitter interrupts for 1 s after character sent (on RS-422) | Replace ADS Assembly. |
| (DEVICE C) ERROR<br>(See Above) | 18 | Overrun error; input too fast (on RS-232) | a. Check serial channel wire protocol of source.<br>b. Replace ADS Assembly. |
| CALCON I/O ERROR<br>Frame/Overrun/Noise | 18 | Framing, overrun, or noise error (if CALCON connected) | a. Check CALCOM cable connection.<br>b. Try running CALCON from a different PC.<br>c. Replace CPU Assembly. |
| (DEVICE C) ERROR<br>(See Above) | 19 | Loss of transmitter interrupts for 1 s after character sent (on RS232) | Replace ADS Assembly. |
| CALCON I/O ERROR<br>Loss of interrupts | 19 | Loss of transmitter interrupts for 1 second after character sent | a. Check CALCON cable connection.<br>b. Try running CALCON from a different PC.<br>c. Replace CPU Assembly. |
| SPEED LOG ERROR<br>Enter Manual Speed | 40 | No VBW/VHW message received for 4 s | a. Check source.<br>b. Check wire connection.<br>c. Replace ADS Assembly.<br>d. Replace CPU Assembly. |
| SPEED LOG ERROR<br>Enter Manual Speed | 41 | Invalid format in VBW/VHW sentence | a. Check message string output by the source.<br>b. Check connection.<br>c. Replace ADS Assembly.<br>d. Replace CPU Assembly. |
| SPEED LOG ERROR<br>Enter Manual Speed | 42 | Data out of range 1 … XX kts per CALCOM (low speed detected in AUTO/NAV/TRACK modes; others detected always) | a. Check speed log source (log data strings from source).<br>b. Check wire connection speed log.<br>c. Replace DC/DC Assembly.<br>d. Replace CPU Assembly for pulse log. |
| SPEED LOG ERROR<br>Enter Manual Speed | 43 | Speed data null or marked invalid | Check speed log source (log data strings from source). |
| RADIUS DISABLED<br>Log Speed Required | 44 | Speed setting changed to a manual entry while in RADIUS control | Operator misuse. |

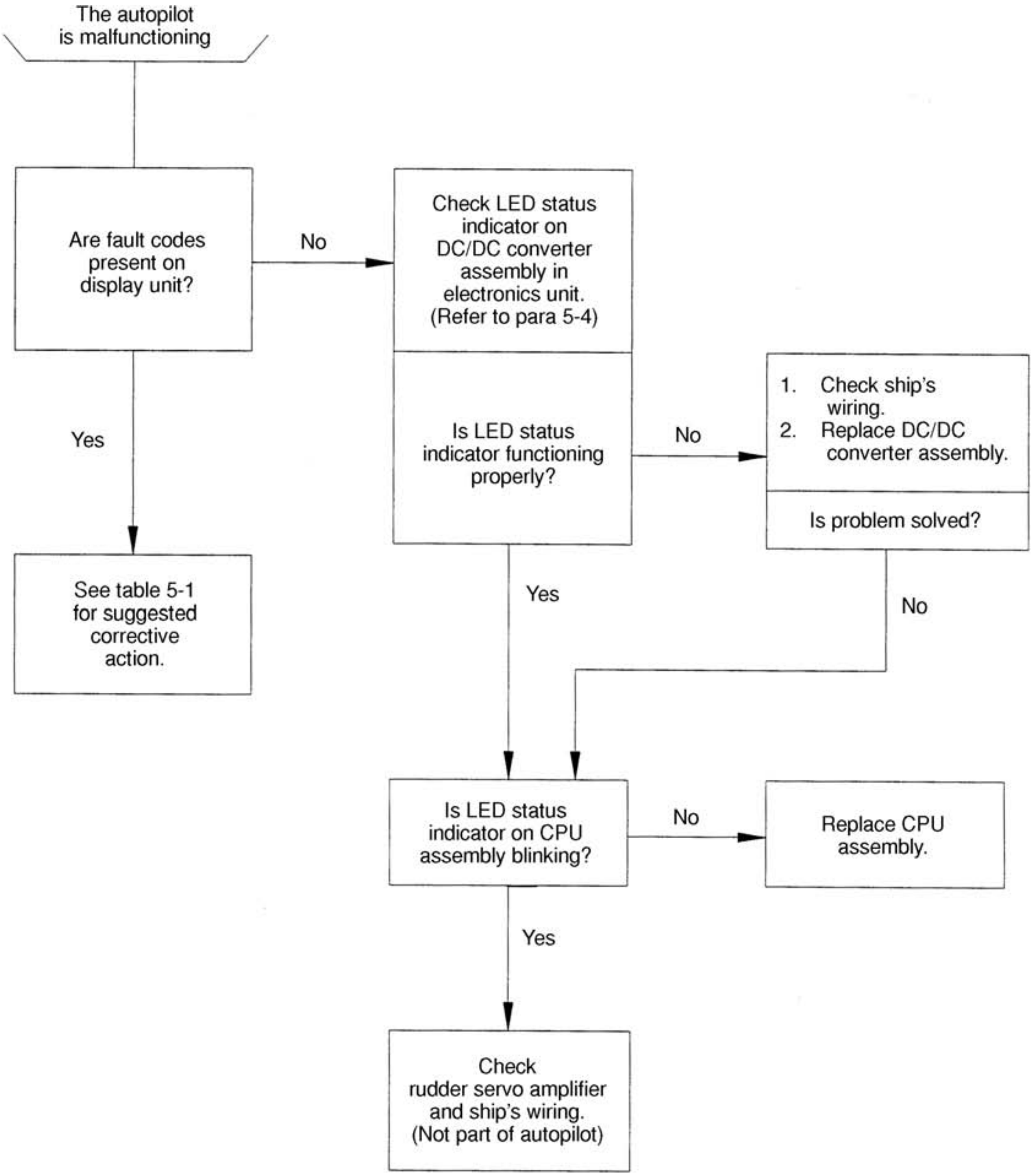


**Figure 9.17** Sperry ADG 3000 VT fault logic diagram. (Reproduced courtesy of Litton Marine Systems.)

## 9.9 Glossary

| | |
|---|---|
| **Adaptive autopilot** | One in which all the control signals are adapted to suit the vessel and environmental requirements. |
| **Counter rudder control** | Determines the degree of opposite helm control which may be applied. |
| **Deadband** | A manually set bandwidth in which the rudder movers do not operate. |
| **Derivative control** | A control signal proportional to the rate of change of the vessel's deviation from its intended course. It is produced by electronically differentiating the actual error signal. |
| **Follow-up mode (FU)** | Full autopilot functions apply. |
| **Integral control** | A control signal created by electronically integrating the heading error. This signal offsets the effect of a vessel being moved continuously off course. |
| **Non-follow-up mode (NFU)** | The autopilot is switched off. Steering is by manual control. |
| **Permanent helm** | Used when the vessel is being driven unilaterally off course by a cross wind. |
| **Phantom rudder** | An electronic signal modelling rudder control. |
| **PID controller** | A system using all three of the feedback signals: Proportional, Integral and Derivative. |
| **Proportional control** | A control signal the extent of which is proportional to the positional error deviation of a vessel from its intended course line. |
| **Rudder limit control** | Sets a finite limit on the maximum rudder angle that can be commanded. |
| **Virtual ship** | A computer-generated model vessel used during the design of autopilot systems. |
| **Weather control** | Used to slow down the effects of severe weather affecting the vessel's course-to-steer by effectively damping the response of the electronic feedback circuitry. |

## 9.10 Summary

- A simple autopilot compares the course-to-steer data, as set by a helmsman, with the vessel's actual course, derived from a master compass, and applies rudder correction to compensate for any error existing between the two input signals.
- To be effective an autopilot requires the following inputs: information about the positional data from the course line, rate of change of course data, and data specifying the cumulative build-up of error.
- PID control systems use proportional, integral and derivative feedback signals.
- A number of operator controls permit rudder control to counteract the static parameters of a vessel and the dynamic effects of the environment. As an example, the weather control dampens the electronic feedback circuitry to reduce the effects of heavy weather on rudder demand thus reducing the likelihood of damage to the rudder mechanism.
- An adaptive autopilot system is one in which all the feedback signals used for rudder control are adapted to the existing requirements of the vessel and the environment.
- A virtual ship is a software model used during the computer design of an autopilot system for a specific vessel and ultimately used on the ship for control.
- Rudder commands vary between open sea course keeping, course changing and confined waters requirements. All are catered for in modern autopilot systems.

## 9.11 Revision questions

1 What are the three main control functions known as PID in an autopilot system?
2 What are the three main feedback parameters required by an autopilot system?
3 What parameter determines the extent of proportional control fed back into the system?
4 Why is the rate of change of the ship's deviation from its course important in autopilot systems?
5 What is an error summing system?
6 What is the function of the weather control?
7 The permanent helm control provides a degree of bias in the system. Why is this?
8 What is the main difference between FU and NFU rudder control?
9 What is deadband and why should it be kept as narrow as possible?
10 What is phantom rudder and why are its characteristics important to course keeping?

# Chapter 10

# Radio direction finding

## 10.1 Introduction

With the advent of the GPS and the massive leaps forward in microelectronic technology, the system of radio direction finding (RDF) looks distinctly aged. It is, of course, the oldest of the position fixing systems having been around in one form or another since the First World War. RDF systems used throughout the last century owed their existence to Sir R. A. Watson-Watt who invented the original concept and to Adcock who designed the non-rotating antenna system that eliminated the earlier troublesome mechanical rotating antenna. To this day, RDF system principles remain unchanged, it is the signal processing and computing functions offered by modern microelectronics that has propelled RDF into the 21st century.

Once the mainstay of maritime position fixing the medium frequency RDF receivers and the large loop antenna that once dominated a ship's superstructure, have now been assigned to the scrap heap. But RDF is still alive and modern vessels do carry VHF RDF equipment. It is still an efficient system for localized position fixing and remains the only method for finding the bearing of a transmitter in an unknown location. If the relative bearings taken by two suitably equipped ships are laid-out on a chart, the two bearing lines will intersect at the position of the unknown transmitting station. Such a station need not be a radio beacon. It could be a vessel in distress and thus the two receiving ships are able, by triangulation, to pinpoint the distress position at the intersection of vectors drawn on a chart from their two known locations. Naturally, the same holds true for two land-based RDF stations.

Because the use of RDF at sea has diminished over the years, its description in this book has been simplified. Whilst the system principles remain the same, the standard of the receiving equipment has dramatically improved and automatic direction finders now dominate the field. The nature of radio waves and the antenna system is of prime importance in understanding the system and Chapter 1 should be read before continuing with this chapter.

## 10.2 Radio waves

Radio direction finders work efficiently when using the properties of ground waves or space waves travelling parallel to the earth's surface. Sky waves reflected from the ionosphere seriously affect system accuracy and should be disregarded.

A propagated radio wave shown in Figure 10.1 possesses both electrostatic and electromagnetic fields of energy. It is the plane of the electrostatic field that is used to denote the polarization characteristic of the wave. A radio wave possessing a vertical electrostatic field therefore indicates a vertically polarized transmission. An electromagnetic field lies in quadrature to the vertical electrostatic field. Maritime direction finders use the properties of this horizontally polarized field transmitted from an omnidirectional antenna system.

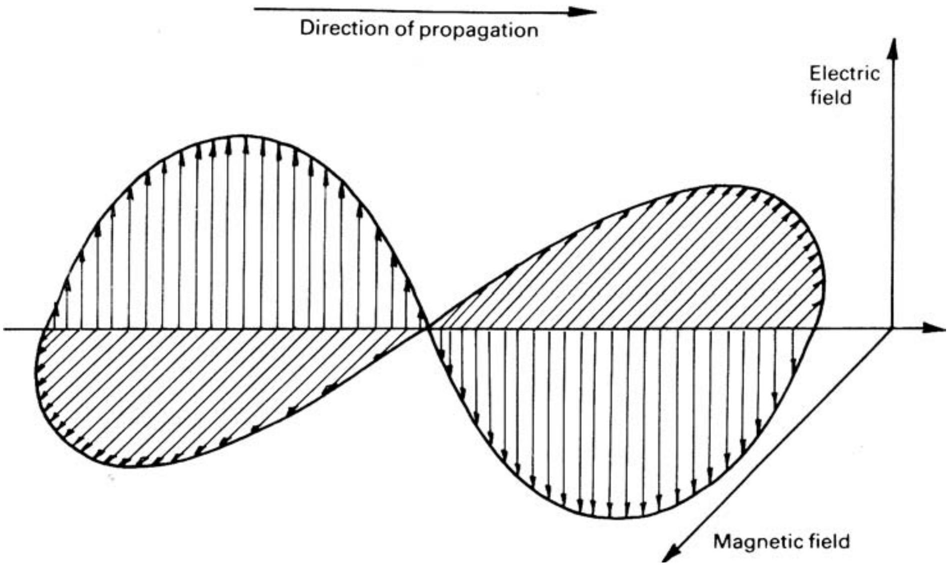


**Figure 10.1** A propagated radio wave illustrating the relationship between both fields of energy and the direction of propagation.

## 10.3 Receiving antennae

This is without doubt the critical component in a RDF system. The electromagnetic component of the radio wave induces tiny voltages termed electromotive forces (e.m.f.) into any vertical conductor (antenna) in its path. If the conductor is a single vertical wire (an omnidirectional antenna), a tiny current will be caused to flow along its length under the influence of the induced e.m.f. The amplitude of the current flow, when applied to the input of a receiver, depends upon a number of factors, but for a given transmitter with a constant power output, it is effectively governed by the distance between the transmitter and the receiver. The frequency of the induced e.m.f. will, of course, be the same as the transmitted frequency.

### 10.3.1 A dipole antenna

A vertical dipole antenna possesses the ability to transmit or receive equally well in all directions and is therefore termed omnidirectional. If a transmitter is arranged to follow a circle at a constant distance from an omnidirectional antenna, the induced e.m.f., at the receiver input, will be constant for all vectors. The pattern thus produced is called the azimuth gain plot (AGP), or sometimes the polar diagram, and illustrates the receptive properties of a vertical antenna as shown in Figure 10.2.

By measuring the induced e.m.f. for all receiving vectors, it is a simple matter to produce an AGP for any antenna. The length of the radial vectors corresponds to amplitude and therefore, in this case, the strength of the signal produced at the receiver input will be constant throughout 360°. This antenna has been designed to be omnidirectional and is used in RDF systems as a ‘sense’ antenna to eliminate bearing ambiguity.

Other antennae are carefully designed to be highly directional. A simple example of this is a Yagi antenna, which is commonly used to receive television pictures and sound. In fact it is possible to use a Yagi antenna and its maximum strength signal indication, to determine the bearing of the

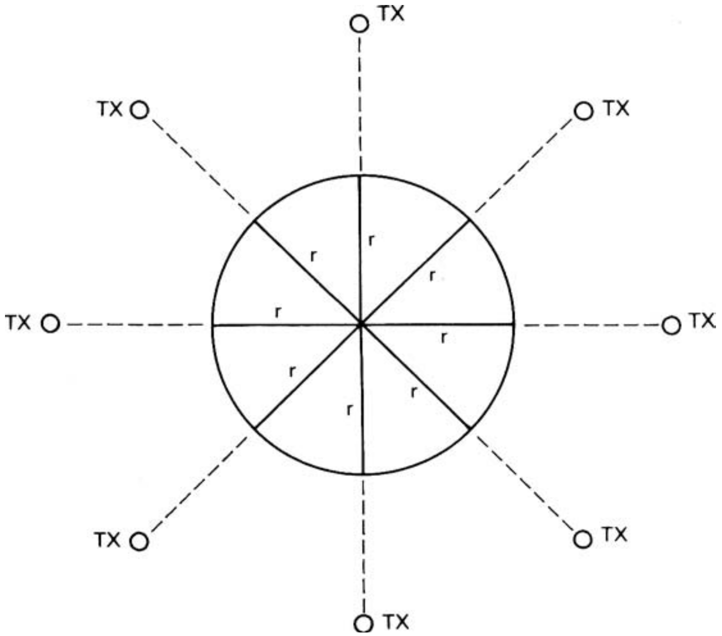


**Figure 10.2** The AGP reception plot for a vertical antenna. The antenna is at the centre of the circle.

transmitting station. Maritime RDF systems, however, use the properties of a simple loop antenna or an Adcock array, and produce a relative bearing indication from a zero or null signal strength.

### 10.3.2 A loop antenna

A simple loop antenna consists of two vertical conductors closed at the top and base to permit current to flow. If the effect of sky waves is ignored (see Polarization error), the shape of the loop is unimportant and for convenience it is often circular. Figure 10.3 shows two vertical antenna joined at the top and at the base via a coil to enable the antenna to be coupled to the input of a receiver.

To be effective the distance between the vertical conductors must be less than one wavelength of the received frequency. For this description, if we assume the distance between the vertical arms to be half of one wavelength and the direction of propagation as shown in the diagram, then maximum e.m.f.s will be induced in both arms AB and CD. The e.m.f.s will cause current to flow through the coil under the influence of an e.m.f. that is the product of the two vertical portion e.m.f.s.

$$\text{Resultant e.m.f.} = (\text{e.m.f. AB}) + (\text{e.m.f. CD})$$

If the direction of the received wave is in the plane shown, or 180° away from it, the resultant current flowing through the pick-up coil will be at its greatest and a maximum signal input to the receiver will result. The single electromagnetic wavelength shown will be at 90° in relation to the vertical antenna arms.

With the transmitter at any angular position from the loop, e.m.f.s will be induced in both vertical arms. The relationship between the plane of the loop and the wavefront will determine the polarity of the induced e.m.f.s, which in turn determines the direction and amplitude of the resultant current

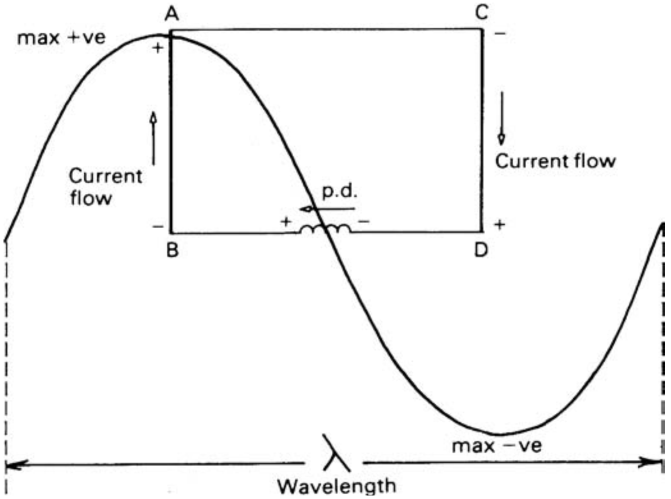


**Figure 10.3** Signal currents induced in the vertical arms of a loop antenna produce a resultant potential difference across the input coil to a receiver.

flowing through the inductor. For convenience we shall consider the plan view of the loop and the wavefront of the propagated signal.

Figure 10.4(a) shows that when the wavefront is parallel to the plane of the loop the e.m.f.s induced in both arms will be of equal amplitude and the same polarity. The two will therefore cancel producing no resultant current flow in the inductor and hence no input to the receiver. This is called a null position because at this point the audio output from a receiver drops to zero. Clearly there will be a second null position, 180° away from the first.

If the loop is turned so that its plane is now 90° with respect to the wavefront, two e.m.f.s will again be induced in both vertical arms, but they will be of equal amplitude but opposite polarity. This causes a maximum circulating current to flow through the coil and a maximum output from the receiver (Figure 10.4(b)). This situation corresponds to a maximum input to the receiver. Once again there will be a second maximum 180° away from the first, the only difference being that the resultant current will flow in the opposite direction through the coupling coil. The AGP produced by such a rotating antenna is shown in Figure 10.5 and for obvious reasons is called a ‘figure-of-eight’ diagram.

A transmitter bearing north or south produces a resultant null output. A transmitter bearing east or west produces a resultant maximum output.

## 10.4 A fixed loop antenna system

At the heart of this system are two permanently fixed loop antennae, mounted on the same mast or base at 90° to each other, one on the fore-and-aft line and the other on the port-and-starboard line of a vessel. An early manual RDF input system is shown in Figure 10.6 to illustrate the principle.

In this case each precisely mounted loop antenna is connected to a pair of precisely aligned fixed coils in a goniometer, a tiny transformer arrangement recreating the electromagnetic fields of the loop antennas. A search coil, able to rotate through 360° inside the fixed coils is tuned to the incoming frequency by the tuning capacitor, C. The resultant circulating current flows through the primary winding of T2 to provide the input to the receiver. The vertical antenna is coupled to the circuit via

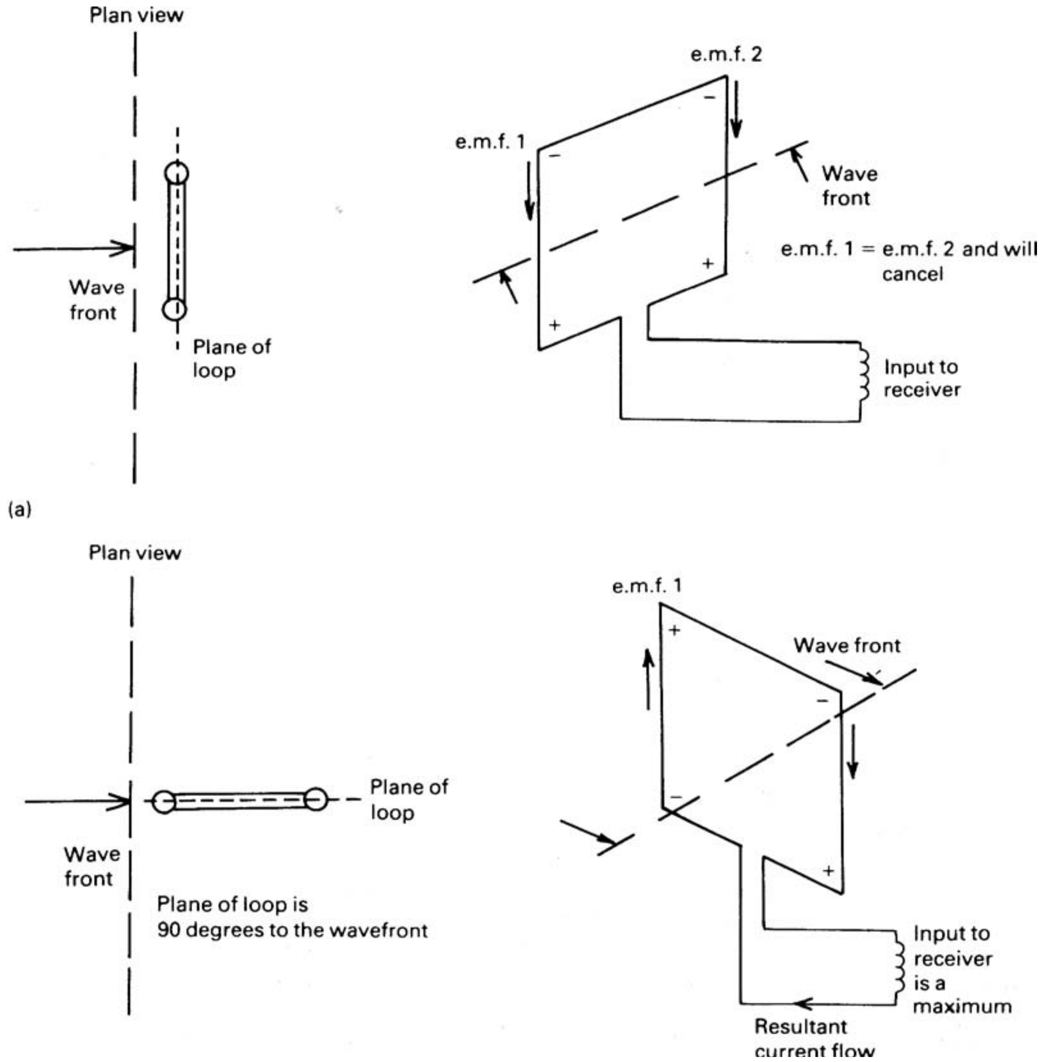


**Figure 10.4** (a)The resultant input to a receiver is zero if the plane of the loop is parallel with the travelling wavefront.(b)The input is a maximum if the loop plane is at 90° to the received signal.

T1. In effect, the goniometer has created a miniaturized version of the rotating loop antenna system without its mechanical disadvantages.

Induced currents in each loop are caused to flow through corresponding fixed field coils in the goniometer. The amplitude and phase relationship of each of the currents will depend upon the relationship between the plane of each fixed loop and the wavefront of the received signal. Current flows will create a magnetic field around the fore-and-aft, and port-and-starboard field coils of the goniometer. A fully rotatable search coil is inductively coupled to each of the field coils. In this way the mutual inductance between the search coil and the field coils follows a true cosine law for any angular position of the search coil to the field coils through 360° of rotation. If the search coil is rotated fully the input to the receiver will consist of a varying signal producing two maxima and two

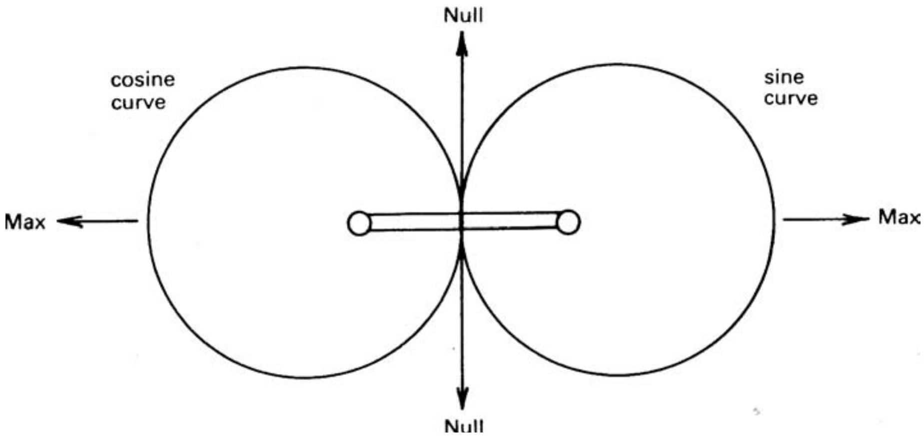


**Figure 10.5** The figure-of-eight azimuth gain plot for a loop antenna.

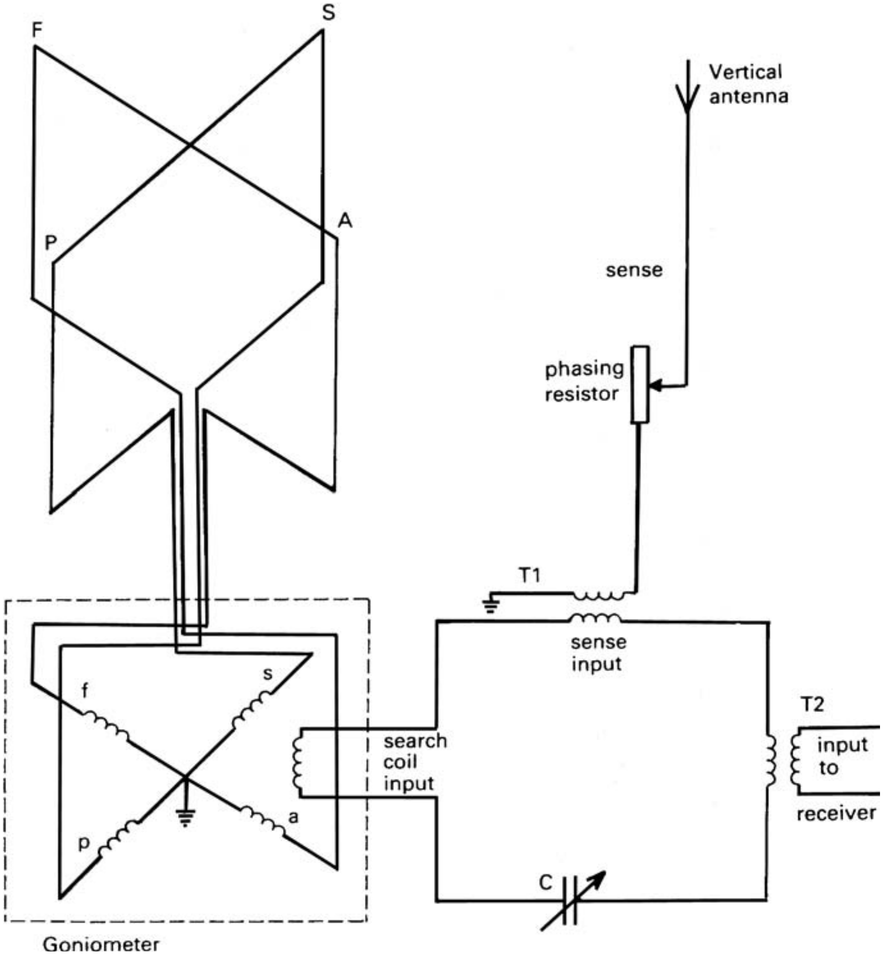


**Figure 10.6** A simple receiver input circuitry for a fixed loops system.

minima positions. A figure-of-eight polar diagram will be created artificially in the confined environment of the goniometer.

Obviously the construction of the goniometer is critical. Early automatic RDF equipment used a tiny servomotor to rotate the search coil but modern equipment dispenses with the mechanical interface and uses software processing to eliminate the reciprocal bearing and produce a true indication.

### 10.4.1 The Adcock antenna

Adcock arrays are capable of covering wide frequency ranges, but for maritime VHF use, the bandspread is relatively small and simple antennas can be used. An Adcock element pair is constructed using two omnidirectional antennas spaced apart by a fraction of the received frequency wavelength in the horizontal plane. Such an arrangement produces an AGP as shown in Figure 10.5. In practice two Adcock pairs are mounted at right angles to each other forming an array.

As in the loop system, Adcock elements are spaced at a fraction of a wavelength apart, often in the region of one-eighth to one-third of the received carrier wavelength. In practice Adcock arrays produce more sharply defined figure-of-eight plots if the spacing between active elements (*d*) is small. Taking the marine VHF communications band at approximately 150 MHz (Channel 16 is 156.8 MHz), one half a wavelength is approximately 1 m and one-eighth wavelength is 25 cm or 10 inches. In Figure 10.7(b), the Adcock array is mounted on a ground conducting base plate, called a ground plane, and the active elements are insulated from it. Distance *d* between the active elements is a constant.

Figure 10.7(c) shows the electrical equivalent of an Adcock array. Induced signal currents *i*1 and *i*2 produce a resultant difference current in the receiver input circuitry. The magnitude of this current is proportional to the element spacing *d* and the length *L* of the elements. Currents induced into the horizontal portions of the array, shown dotted in the diagram, are of equal magnitude and direction and will cancel. Like the loop antenna, the resultant azimuth gain plot is a double figure-of-eight with maximum gain being achieved in line with each pair of dipoles (see Figure 10.8). The length of the active elements *L* is also related to wavelength and because each arm is effectively a dipole antenna, *L* is likely to be one-quarter wavelength or a further subdivision of one wavelength.

On the arrangement shown in Figure 10.7(b), the central element is a sense antenna, the output from which is used to eliminate bearing ambiguity.

#### *Eliminating the reciprocal bearing indication*

The minima or null positions of the figure-of-eight AGP have been chosen to indicate the direction of the bearing because the human ear (used extensively for determining bearings in early systems) is more responsive to a reducing signal than to one that is increasing. For a single Adcock array or loop antenna, there are two null positions, one that indicates the relative (wanted) bearing and the other the reciprocal. Dual antenna arrays create quadruple null indications.

In many cases, reciprocal null indications pose no problem because the relative bearing will be the one that lies within the expected bearing quadrant from a known receiver. However, when taking the bearing of an unknown vessel, for triangulation plotting, it is not known in which quadrant the bearing will lie and therefore a second input to the receiver is required in order that the other null positions can be eliminated. To simplify the explanation, AGPs for a single loop antenna and a vertical antenna have been used. The result of adding the vertical antenna signal, sometimes called a ‘sense’ input, to the resultant loop signal for a single loop is yet another AGP which for obvious reasons is called a cardioid and is shown in Figure 10.9.

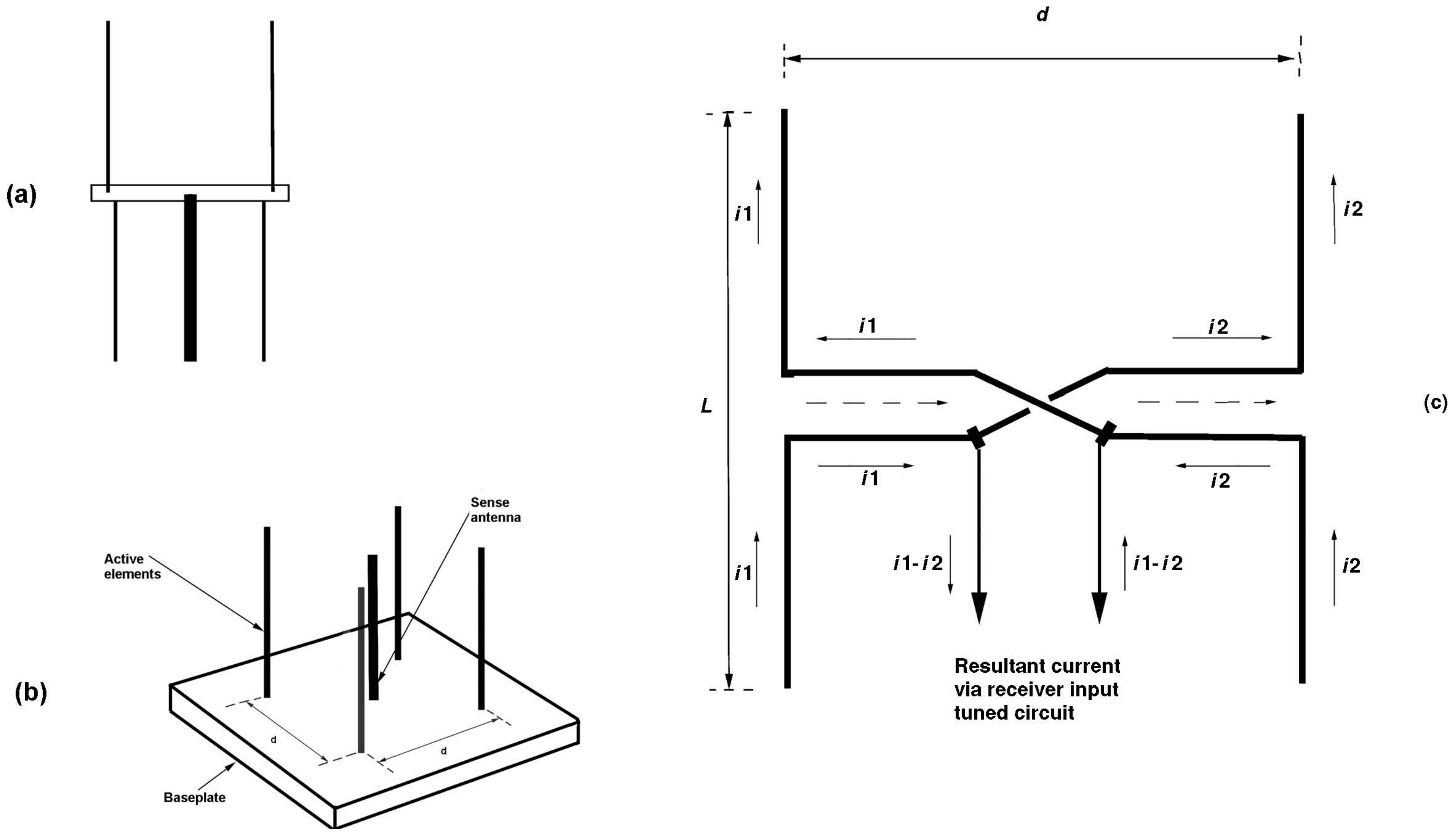


**Figure 10.7** (a) A pole-mounted Adcock antenna and (c) its electrical equivalent. (b) A base plate-mounted Adcock array.

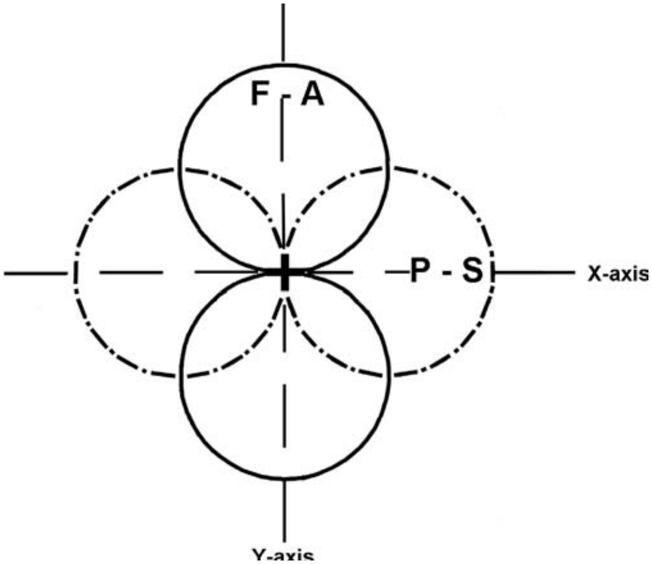


**Figure 10.8** AGP diagram for an Adcock (or a crossed loop) pair.

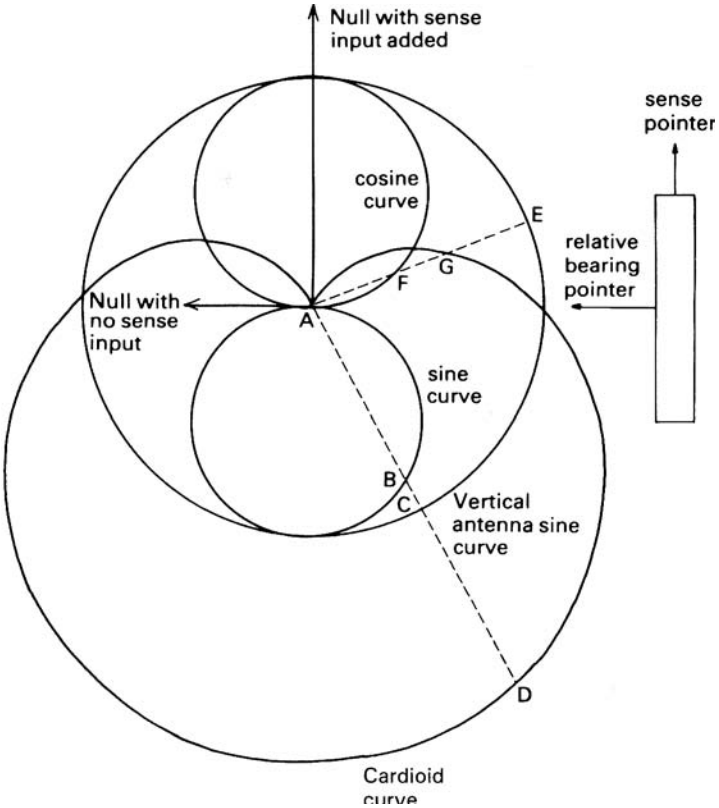


**Figure 10.9** The resultant cardioid AGP produced by the addition of the figure-of-eight and circular plots.

The signal produced by the sense antenna is an omnidirectional sine curve whereas that of the loop figure-of-eight curve possesses both sine and cosine properties. The resultant cardioid is created by radially adding and subtracting the two signal levels. For the sine portion of the loop diagram, AB + AC = AD and for the cosine portion, AE – AF = AG. Unfortunately, although a new single null position has been produced, it has been shifted by 90°. This error is compensated for in the receiver bearing processing circuitry.

The result of adding a sense signal input to a dual loop or Adcock array is to produce a double cardioid and the further bearing ambiguity thus produced is again eliminated during computing. In fact it is possible for modern RDF receivers to produce a relative bearing without a sense antenna input. The microelectronic circuitry computes a virtual sense input for every position in azimuth.

## 10.5 Errors

Although RDF systems are subject to errors, caused mainly by environmental effects, if a fixed loop or Adcock RDF system is correctly installed and accurately calibrated the errors can be reduced to virtually zero. As with any electronic system, it is important to appreciate the error causes and cures. The major error factors affecting RDF systems installed on merchant ships are listed below. Some of these have minimal effect at VHF but they have been included here for reference.

### 10.5.1 Quadrantal error

This error is zero at the compass cardinal points rising to a maximum at 045, 135, 225 and 315°. Each maximum error vector falls into a quadrant and hence the error is termed quadrantal. The cause of the error is a re-radiated signal produced, mainly along the fore-and-aft line of the vessel, by the ship's superstructure receiving and re-radiating the electromagnetic component of the signal. All metallic structures in the path of an electromagnetic wave will cause energy to be received and then re-radiated. In this case the re-radiated signal is in phase with the received wave. The two signals arriving at the RDF antenna will be of the same frequency and phase and will therefore add vectorially causing the relative bearing to be displaced towards the fore-and-aft line of the vessel, as shown in Figure 10.10.

The new bearing is a vector sum of the received and re-radiated signals. The magnitude of the error depends mainly upon the vessel's freeboard and the position of the loop antenna along the fore-and-aft line. For a loop mounted in the after-quarter of the vessel, the effect will be greatest in the two forward quadrants, and vice versa for a loop antenna mounted in the forward quarter. Fortunately the error, for a given mounting position, is constant and is able to be eliminated. For a fixed crossed loop system, the fore-and-aft loop antenna, which is under greater influence from the unwanted signal than the port-and-starboard loop antenna, is made smaller. Also quadrantal error correction is more accurately achieved by placing a quadrantal error variable corrector coil in parallel with the fore-and-aft loop coil.

The effect of varying the inductance of such a coil during calibration is to reduce the signal pick-up along the fore-and-aft line of the vessel. Modern equipment also includes a smaller compensation coil across the port-and-starboard loop circuit. Correct alignment of these coils reduces the effect of quadrantal error.

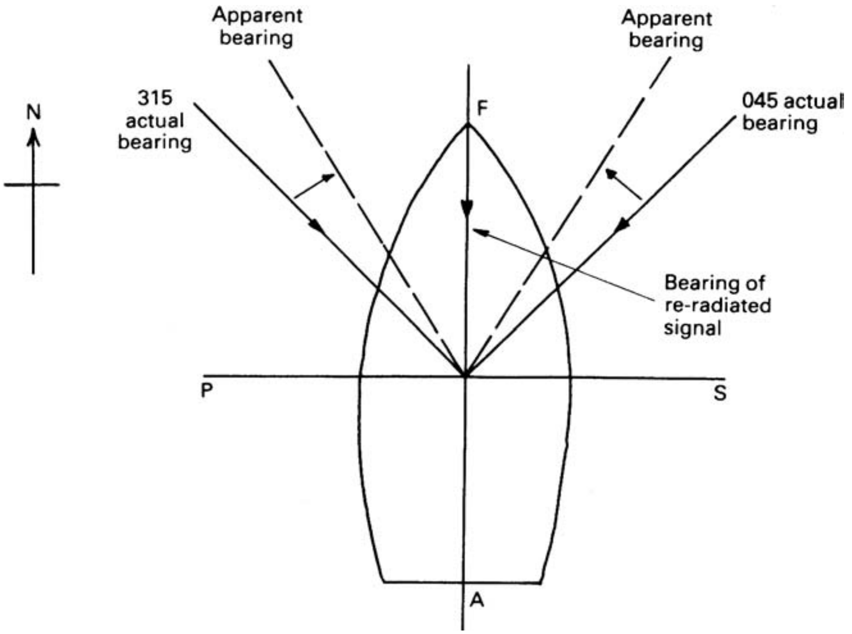


**Figure 10.10** The effects of quadrantal error are to pull the bearing indication towards the vessel's lubber line.

### 10.5.2 Semicircular error

As with quadrantal error, semicircular error is caused by a re-radiated signal arriving at the loop antenna along with the received radio wave. In this case the re-radiated signal is produced by vertical conductors in the vicinity of the loop antenna. This re-radiated signal from such conductors is out of phase with the primary signal and will therefore cause an error that rises to a maximum in two semicircles. Conductors that produce an out of phase re-radiated signal possess a resonant length that is close to the half a wavelength of the received signal.

The most obvious of these conductors are the vessel's various antennae, but wire stays will also have the same effect. For re-radiation to occur, induced current must be able to flow in the conductor. To prevent current flow, wire stays may be isolated by inserting electrical insulators along their length.

### 10.5.3 Polarization error or night effect

A RDF system works on the principle that the electromagnetic component of a propagated space wave parallel to the earth's surface will cause small e.m.f.s to be induced in the vertical arms of an antenna. Under some conditions propagated radio waves are refracted by the ionosphere and will return to earth some distance away from the transmitter. The 'skip distance', the surface range between the transmitter and the receiver, in which radio waves may be returned from the ionosphere, depends upon a number of factors. Two of these are

- the frequency of the propagated wave
- the density of the ionosphere.

The frequency of the radio wave is a constant, but the density of the ionosphere is far from constant as it varies with the radiation it receives from the sun. If two radio waves from the same transmitter are received at a RDF antenna, one directly and the other as a skip from the ionosphere, e.m.f.s will be induced in both the vertical and the horizontal portions of the antenna. Under such conditions it may not be possible to determine the direction of the transmitting station by rotating the loop or search coil because the angular position of the horizontal portions of the loop with respect to the sky wave cannot be changed. The relationship between the ground wave and the sky wave will be constantly changing in phase, amplitude and polarization, which in turn will cause considerable fading and null position shifting to occur when attempting to take a bearing.

Although there is no cure for night effect, using an Adcock array with no horizontal limbs effectively eliminates pick-up from sky waves. However, because the effect is most prevalent 1 h either side of the time of sunrise and sunset, when the ionosphere is most turbulent, if using a loop antenna, it is advisable to treat bearings taken at this time with suspicion.

### 10.5.4 Vertical effect

The error known as vertical effect has been virtually eliminated by the careful construction of a loop antenna. The error was caused by unequal capacitances between the unscreened vertical arms of the loop antenna and the ship's superstructure. Depending upon the shape of a vessel's superstructure, the effect produced an imbalance in the loop antenna symmetry, which in turn produced errors that varied in each quadrant. Mounting the loop conductors inside an electrostatic tubular screen eliminates this error.

As shown in Figure 10.11 the loop conductors are mounted precisely in the centre of the tube, which has the effect of swamping the imbalance of the external capacitance. The loop screening tube is earthed at its centre and is supported at the pedestal by two insulation blocks. The blocks effectively prevent the electrostatic screen from becoming an electromagnetic screen that would block the passage of electromagnetic waves and cause the input to the receiver to fall to zero.

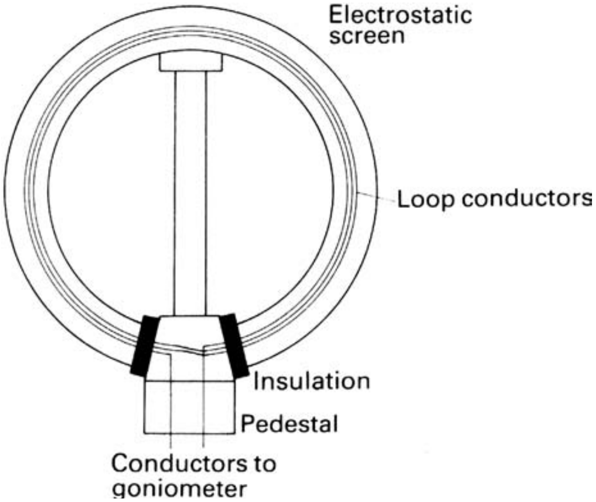


**Figure 10.11** Electrostatic screening of a single loop to minimize vertical error.

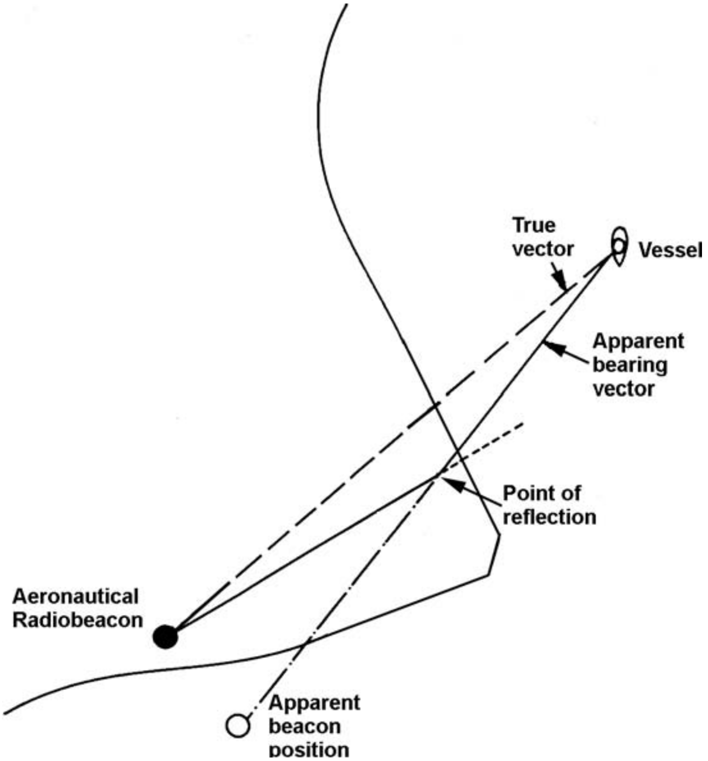


**Figure 10.12** Error introduced by a reflected VHF radio wave.

### 10.5.5 Reflected bearings

Originally, maritime RDF systems relied on the reception of medium frequency ground waves, the velocity of which is influenced by the conductivity of the surface over which the wave is travelling. This factor gave rise to an effect known as ‘coastal refraction’ when bearings were taken from a beacon inland and the radio wave crossed from land to water.

Although VHF space waves do not suffer from velocity changes caused by ground absorption, they do suffer from reflection and it is possible for a RDF bearing to be in error if it is taken from a reflected wave. This can happen when bearings are taken from inland beacons, such as aeronautical VHF beacons, that may be close to high rise buildings or objects (see Figure 10.12). Unless there is published documentation advising of errors, it is advisable to treat bearings taken from aeronautical beacons with suspicion.

## 10.6 RDF receiving equipment

In the early days of radio direction finding, receivers were almost always manually operated. Today however, all RDF equipment is automatic. The first automatic receivers depended upon the use of a servomotor to physically drive the RDF compass card to indicate the relative bearing.

### 10.6.1 An automatic system using a servomotor

This type of RDF has at its heart a low power two-phase servo that, via a mechanical drive mechanism, rotates the goniometer search coil and bearing pointer. This type of system was popular because the bearing is displayed on a compass-like card that revolves to indicate the relative bearing.

First, it is necessary to generate the servomotor signal requirements. A low frequency oscillator generates the necessary two signals, one phase shifted by 90°, to drive the servo. Figure 10.13 illustrates the operational characteristics of the two-phase induction servo used in this type of system.

Two signals, one a reference signal and the other a 90° phase-shifted control signal, are applied, via power amplifiers, to the two stator windings of the servo. Current flows through each of the coils producing magnetic fields along the two axes shown. Each magnetic field causes small e.m.f.s to be induced in the squirrel cage rotor causing it to rotate under their influence. The relative bearing

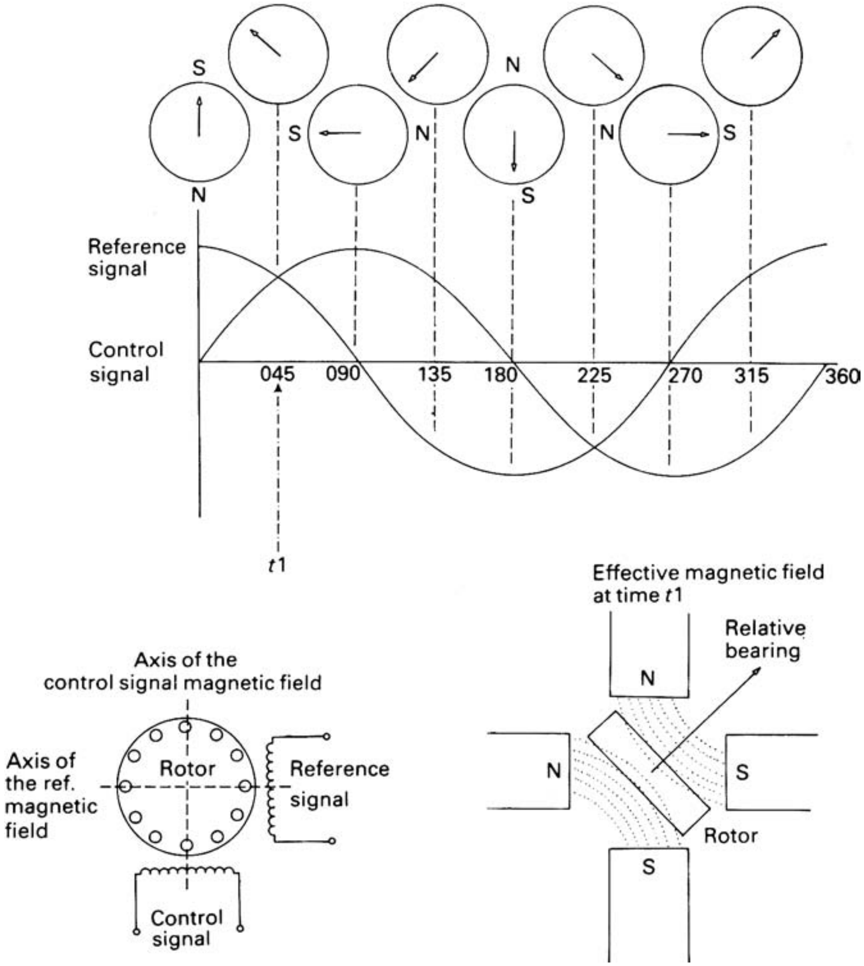


**Figure 10.13** The rotating magnetic field produced in the stator windings of a two-phase induction system.

pointers shown above the two phase-related signals indicate the instantaneous position of the rotor at each of the 45° positions of one cycle of input. The resultant magnetic field produced by the two alternating currents will be continually changing and will create a rotating magnetic field turning the rotor and the search coil in the goniometer via the mechanical linkage.

The search coil continues to rotate as long as the two servo windings are under the influence of the phase quadrature signals. If one signal (the control) disappears the rotor will stop. If the phase relationship between the two signals changes the servo will again stop, unless the change is 180° when the servo rotor will rotate in the opposite direction. This characteristic is exploited in the automatic RDF where the control signal is coupled via the receiver circuits to the control winding of the servo. The control signal is therefore under the influence of the received resultant loop signal amplitude.

Once the electrical signals have been generated and the reference signal is applied to the servo, it is necessary to modulate the control signal with the received bearing signal. This is done by a modulator that is placed between the antenna signal line and the input to the receiver as shown in Figure 10.14.

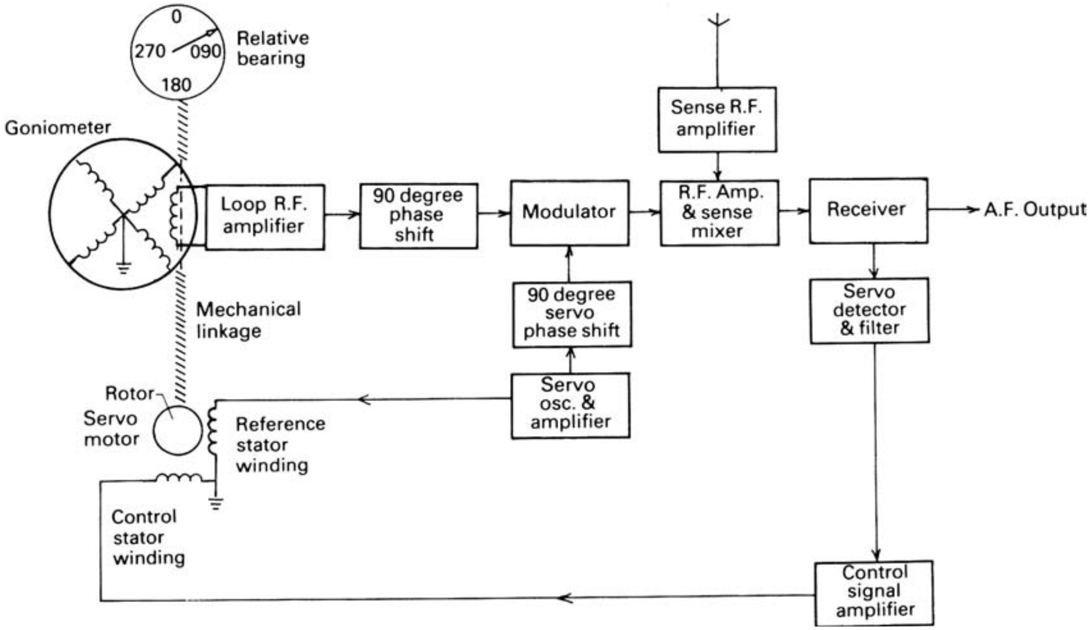


**Figure 10.14** System diagram of a servo-controlled automatic RDF system.

Assuming that the search coil is stationary and sitting 90° away from a relative bearing position, a maximum signal output from the search coil to the loop amplifier results. This signal is then phase shifted by 90° to eliminate the error that will occur when the permanently connected sense input is applied at a later stage.

The control signal is now applied to a Cowan modulator where it is both amplitude- and phase-modulated. The output waveform from the modulator is an alternately 180° phase-shifted signal as shown in Figure 10.15.

In the next radio frequency amplifier, the vertical sense antenna signal is added to the output of the modulator causing the loop signal to be returned to its original phase. This signal is now an amplitude-modulated radio frequency and is processed by the superhet receiver in the normal way. Chopping the

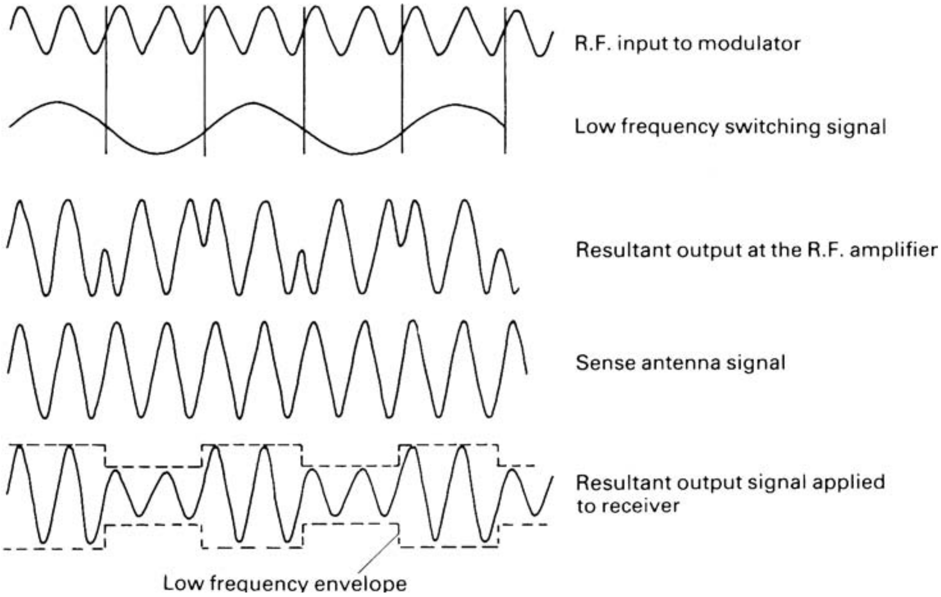


**Figure 10.15** Illustration of the waveform mixing process to produce the servo control signal envelope.

loop signal in the Cowan modulator and then re-constituting it with the sense input signal ensures that the servo cannot rotate if the sense input fails. Thus a failsafe system has been introduced to eliminate the possibility that the servo would stop the search coil on the reciprocal null position of the relative bearing if the sense antenna failed.

The servo detector circuit now detects the amplitude variation of the intermediate frequency and couples the resultant signal through a series resonant filter to the control winding of the servomotor. The filter ensures that only the low frequency servo signal is amplified to become the servo control signal. The rotor now rotates moving the search coil of the goniometer towards a bearing. This in turn will causes the loop signal to the radio frequency amplifier to reduce in amplitude. The output from the modulator reduces causing the output from the servo detector to fall. As the control signal amplitude falls, the magnetic field created around the control stator winding reduces and the rotor slows down. Eventually a null position will be reached where the loop signal falls to zero, no modulation takes place and the servo stops.

Theoretically it is possible for the servo to stop on the reciprocal null position. In practice, however, the reciprocal null position is very unstable due to noise and thus the system will only remain steady in the relative bearing position. To prevent null position overshoot, which may be produced by the torque of the servo as it swings rapidly towards a null, an opposing magnetic field is created within the servo, by a d.c that is introduced when the rotor has moved within prescribed limits of the relative bearing position.

### 10.6.2. A computer-controlled RDF system

A computer-controlled RDF system is shown in Figure 10.16. The description of the system is based upon the discrete logic circuitry of an early RDF receiver manufactured by the STC International Marine Company. It has been used here because of its clarity of operation.

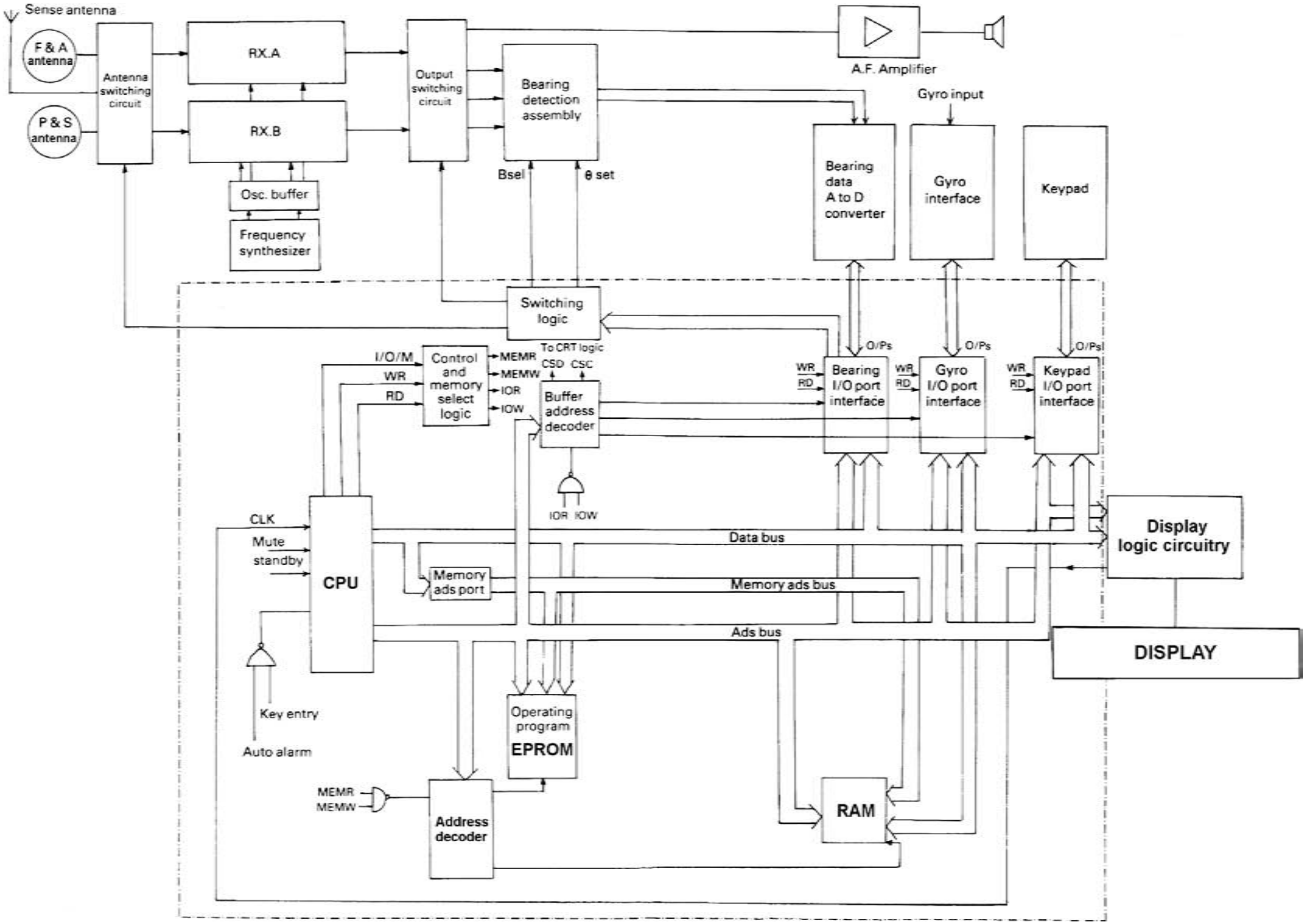


**Figure 10.16** A system diagram for a computer-controlled RDF.

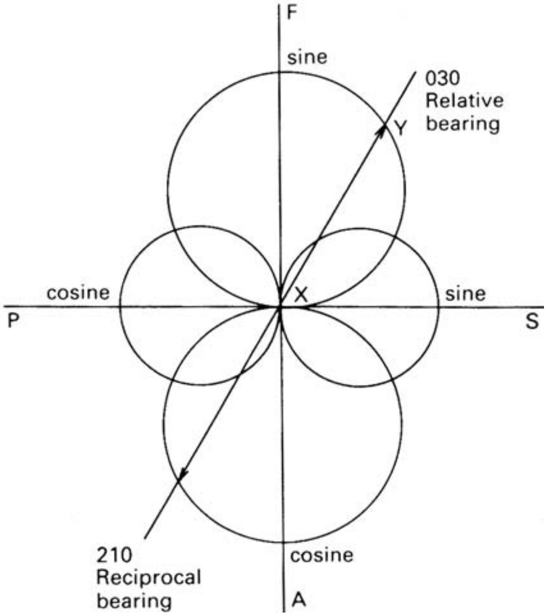


**Figure 10.17** AGP plots of the input to the receiver.

The output signal amplitude of both the port-and-starboard antenna and the fore-and-aft antenna will vary with the azimuth angle of the received radio wave relative to the ship's heading. Figure 10.17 illustrates the resultant polar diagrams produced by the two antennae for a transmission received on a relative bearing of 030°. In this case the output from the fore-and-aft antenna is greater in amplitude than that obtained from the port-and-starboard antenna. The vector XY is an indication of the resultant signal amplitude corresponding to the relative bearing.

Antenna signals are switched to independent receivers where their corresponding amplitudes are compared. The strongest signal, in this case the one from the fore-and-aft antenna, is then switched to the primary receiver. Obviously the fore-and-aft antenna polar diagram also indicates a reciprocal bearing null at 210°. To remove this ambiguity the sense antenna is now connected to receiver B. The phase relationship between the fore-and-aft signal and the sense antenna signal is now compared in the bearing detection assembly board to determine the relative bearing. This process is extremely complex. It is controlled by the θ-set (phase comparison initiation pulse) and the B-sel line (bearing select) both of which originate in the microprocessor. Basically the decoded phase relationship is used to clock an up/down logic counter under command of the B-sel line input. The output from the counters is then connected via an analogue-to-digital converter to the interface circuits of the computer.

Bearing computation is software commanded by a dedicated program held in an EPROM. Central control is from a CPU that, via data and address bus lines, commands all functions. I/O/M, WR (write), and RD (read) control lines are gated to provide four memory and port control lines MEMR (memory read), MEMW (memory write), IOR (input/output port read) and IOW (input/output port write). MEMR and MEMW are further gated to command both the EPROM and RAM memory capacity. Lines IOR and IOW, via the buffer address decoder, control the three data input/output ports: bearing data, gyro data and keypad data.

### *Operation in bearing mode*

Keypad commands are read onto the data bus from the I/O port that has been enabled by the RD line. The line 02 output from the buffer address decoder is also be enabled. The CPU commands receiver and bearing detection assembly functions to produce bearing data at I/O port IC3. Using the RAM as storage, and EPROM software, the CPU inputs bearing and gyro data to complete the computation and produce the bearing data to command the display logic.

### *Bearing presentation*

A RDF bearing display can be as simple as a three-digit numerical readout or as complex as that of an integrated navigation system, but many navigators prefer to see the relative bearing displayed in real-time polar format. In common with all data displays, the relative bearing displayed should be unambiguous and clearly visible. It should also be capable of being displayed in a north-up or ships-head-up mode, depending upon requirements. Other data indications are signal strength, bearing quality, receiver frequency and own ship's heading.

In general there are two outputs from a modern bearing processor to feed the deflection system of a display. They are the vertical or y-axis produced from the fore-and-aft co-ordinates and the horizontal or x-axis produced from the port-and-starboard co-ordinates. Equipment using a cathode ray tube for bearing display uses the two outputs to vary the electrostatic fields generated by x-axis and y-axis deflection plates to deflect the electron beam in the direction of the relative bearing. For instance, equal amplitude positive voltages fed to both the x and y deflection plates will cause the spot

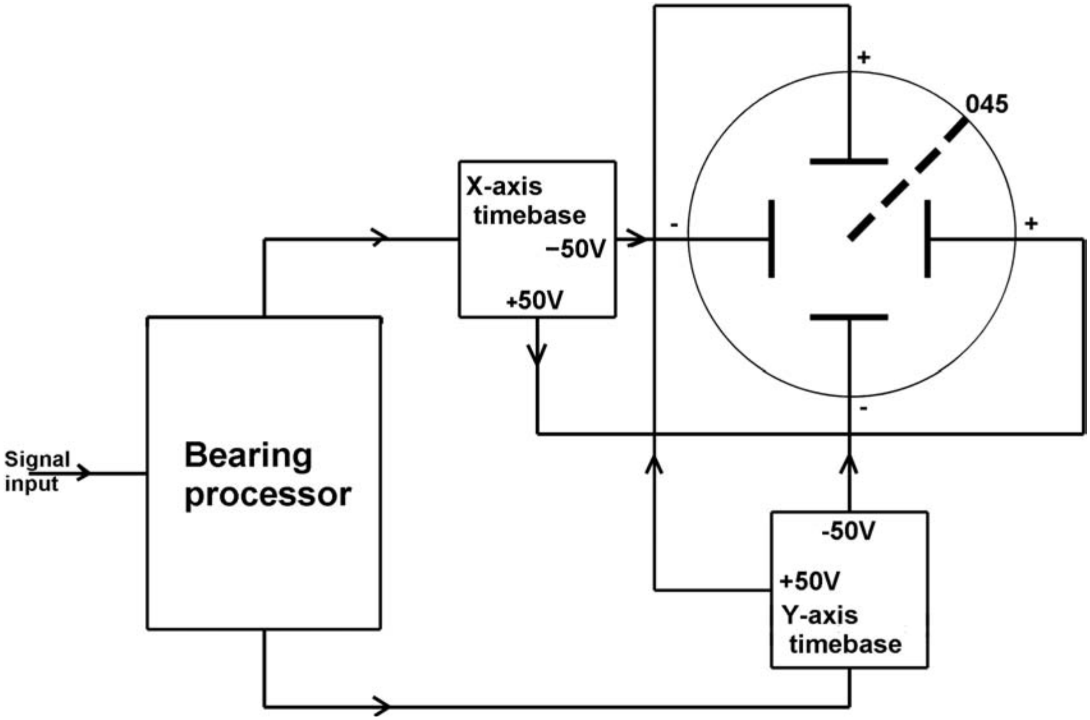


**Figure 10.18** The magnitude of each x and y voltage determines both the azimuth indication and the strength of the signal as shown by the length of the vector.

to deflect to 045° (Figure 10.18). If both voltages are of equal lesser amplitude, the bearing remains the same but the trace length reduces to indicate a weaker signal. If for instance the x deflection voltage is a maximum and the y deflection voltage drops to zero, the displayed bearing will be 090°. This is a simple explanation of the principle. In practice the timebases are more complex.

Modern equipment using flat screen technology uses complex matrix technology but the principle is the same. The relative bearing may be displayed as polar diagram representation, in the form of a bar chart, or it may be in numeric form.

### 10.6.3 VHF scanning RDF equipment

Whilst the carriage of a radio direction finder is not a mandatory requirement on merchant vessels there is no doubt that it is a useful piece of equipment. Since the maritime medium frequency RDF system ceased to function, the number of companies manufacturing and selling maritime RDF equipment has fallen to a mere handful. One traditional marine equipment supplier, Koden, produces a range of RDF equipment designed to operate as stand-alone systems or to be interfaced with an existing VHF communications receiver. One of their models the KS538 is at the forefront of technology in this area (Figure 10.19).

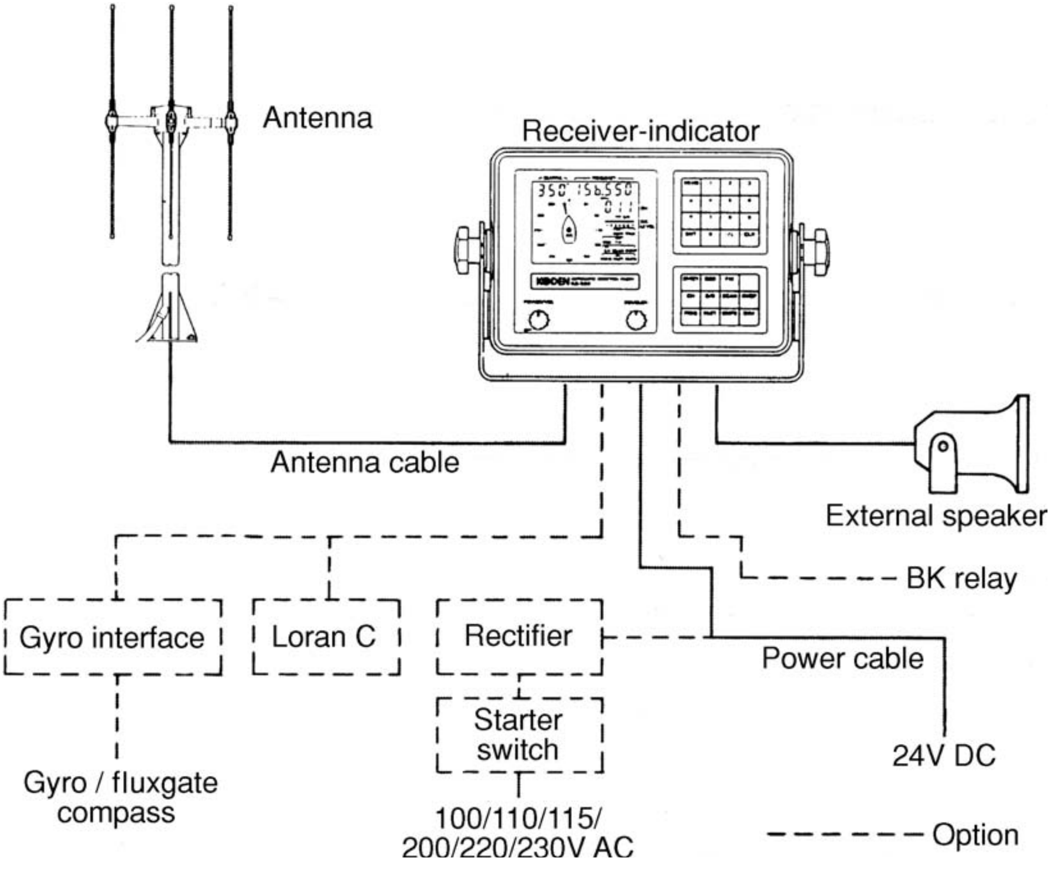


**Figure 10.19** A modern RDF installation showing interface details. (Reproduced courtesy of Koden Electronics Co. Ltd.)

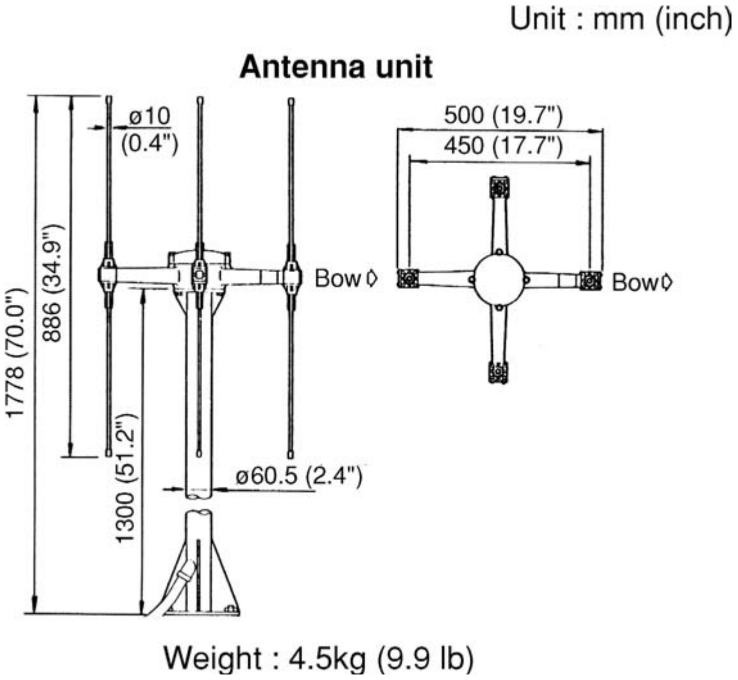


**Figure 10.20** Construction detail of the Adcock antenna unit. (Reproduced courtesy of Koden Electronics Co. Ltd.)

As has previously been stated, a RDF is basically a high quality communications receiver with the addition of a specialized antenna and a suitable visual display. Central to the Koden unit, shown in Figure 10.19, is a fully synthesized VHF receiver able to receive frequencies in the range 110–179·999 MHz in 1 kHz steps. All VHF channels are held in memory including 55 international channels, four US weather channels, three Scandinavian fishing channels, two pleasure craft channels, and the international distress channels. In addition, 99 other channels are operator programmable. Each channel is selected via an alphanumeric keypad and all channels can be automatically scanned.

The system uses a four-element Adcock array antenna for bearing location (see Figure 10.20). Element spacing is approximately 450 mm and the length is 886 mm, which as a subdivision of the short VHF wavelength puts the receptive properties well within the required band.

In common with most modern manufacturers, Koden makes good use of the large backlit LCD display (Figure 10.21). Bearings are presented in the preferred polar form as well as digitally. The dominant feature of the display is the representation of a compass card that clearly shows the relative bearing. It is displayed as a large black triangle, in this case 247° relative. If compass data is interfaced with the unit, a second indication showing the vessel's course appears and bearing data may be shown as a three-digit true bearing for laying-off on charts during a triangulation exercise.

Other display data includes the received frequency and channel number, the signal strength, relative (bow) or true bearing indication, signal modulation, channels and sweep rate, and the period of data presentation.

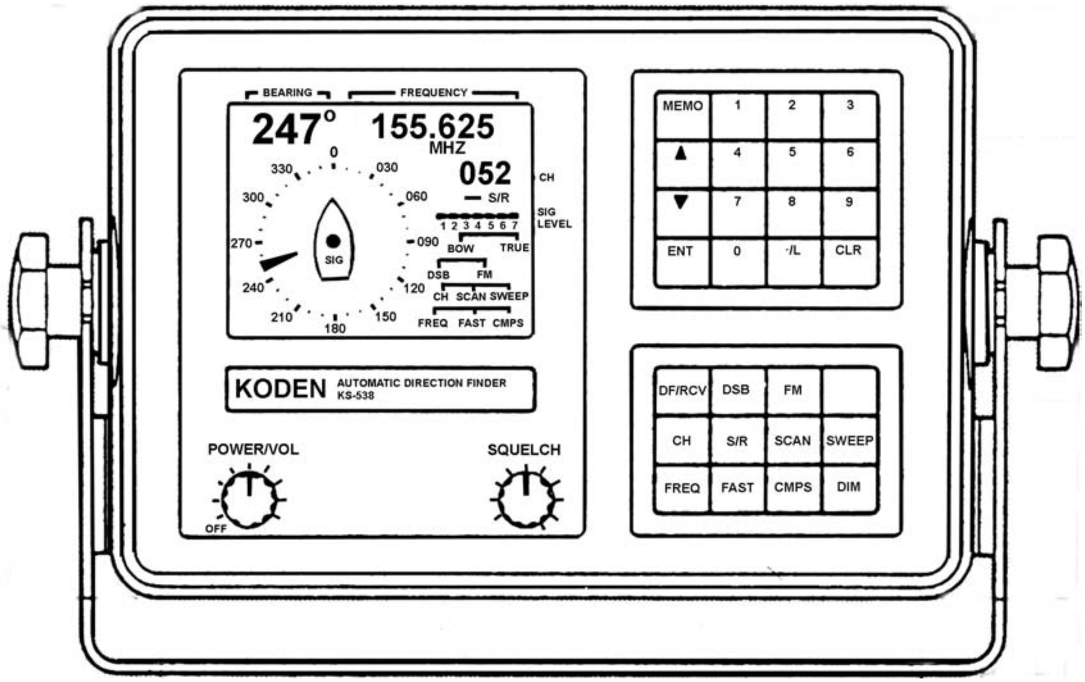


**Figure 10.21** The Koden KS-538 RDF system display is a good indication of the information presented to the user of a modern equipment. (Reproduced courtesy of Koden Electronics Co. Ltd.)

## 10.7 Glossary

| | |
|---|---|
| **Adcock antenna** | A directional antenna constructed from a number of dipole pairs. |
| **Azimuth gain plot (AGP)** | The radiation or reception pattern of an antenna when drawn in azimuth. Occasionally called a polar diagram. |
| **Dipole antenna** | A vertical antenna with the ability to receive equally from all directions. |
| **Loop antenna** | A directional antenna constructed from a coil of wire. Need not be circular; square or triangular shapes are also popular. |
| **Null** | The zero signal condition that indicates the true bearing in manual RDF systems. |
| **Polar diagram** | See azimuth gain plot. |
| **Polarization error (night effect)** | Caused by receiving signal refracted from the ionosphere. |
| **Quadrantal error** | An error existing in all azimuth quadrants of a RDF system. |
| **Reciprocal bearing** | The opposite bearing to the true bearing. |
| **Semicircular error** | Caused by out-of-phase re-radiated signals from structures in the vicinity of the receiving antenna. |
| **Sense antenna** | An omnidirectional antenna providing an input signal to eliminate the reciprocal (unwanted) bearing. |

## 10.8 Summary

- RDF systems operate by receiving ground or space radio waves, not sky waves.
- By triangulating RDF azimuth bearings on a chart it is possible to locate a transmitter at an unknown location.
- Early systems used rotating antenna but modern equipment is automatic and uses fixed receiving antenna.
- A loop antenna is highly directional and two fixed at 90° to each other are used to determine the direction of a transmitter in azimuth.
- An Adock antenna system possesses the same properties as a loop antenna and is often used in RDF systems.
- The input from a dipole antenna, called a sense input, is used to eliminate the reciprocal (unwanted) bearing.
- A number of errors affect system accuracy but they are mostly predictable and are eliminated.
- Modern RDF systems use frequencies in the VHF band and consequently small antenna may be used. Maritime VHF channels are held in memory in modern equipment.
- Modern RDF equipment may be a stand-alone unit or it may be an addition to the bridge VHF equipment fitted on all commercial vessels.

## 10.9 Revision questions

1 How are two RDF-equipped vessels able to triangulate the position of an unknown vessel?
2 How is it possible to produce a null or zero signal at the input to a receiver merely by rotating an antenna?
3 A single loop or Adcock antenna produces an AGP with two nulls. How may the reciprocal (unwanted) null be eliminated?
4 How do sky waves affect the accuracy of an RDF system?
5 How do reflected radio waves affect the accuracy of the indicated bearing?

# Chapter 11

# Global Maritime Distress and Safety System

## 11.1 Introduction

It may seem a little strange to include a chapter about distress communications in a book dedicated to radio navigation, but the Global Maritime Distress and Safety System (GMDSS) is of prime importance to all maritime personnel. The system has been developed to provide mariners with a global communications and locating network, elements of which are capable of being operated by an individual with minimum communications knowledge and yet enable alerting and Search and Rescue (SAR) to be reliably achieved and controlled. A simplified description of the GMDSS and its navigational elements follows. For a full and detailed description of the system, refer to our book *Understanding GMDSS*.

## 11.2 The system

After a lengthy implementation period, the GMDSS became fully operational on 1 February 1999. The basic concept, shown in Figure 11.1, shows that a ship in distress is effectively inside a highly efficient radio net. If the casualty is correctly fitted with GMDSS equipment it will be in a position to alert and communicate with a wide range of ship- and shore-based radio stations and through them initiate a co-ordinated SAR operation based on a rescue co-ordination centre (RCC).

GMDSS relies heavily on digital selective calling (DSC), an electronic system enabling automatic 24-h watchkeeping on specific frequency channels ensuring that a distress call is received and acknowledged.

Two-way global communications with shore stations is via the International Maritime Satellite Organization's (INMARSAT) geostationary satellites or on the HF terrestrial bands. One-way distress alerting may be achieved via the polar orbiting COSPAS/SARSAT satellites.

Navigation elements of the GMDSS include NAVTEX, providing on-board navigation data and meteorological warnings and the new Inmarsat-3 satellites encompassing navigation payloads designed to enhance the accuracy, integrity and availability of both the GPS and the GLONASS systems.

### 11.2.1 Carriage requirements

Whilst the GMDSS is a global system it is not necessary for all ships to carry a full range of communications equipment. Vessels trading solely in coastal water, for instance, may carry less equipment than ocean-going ships. The equipment to be carried is determined by the declared area of

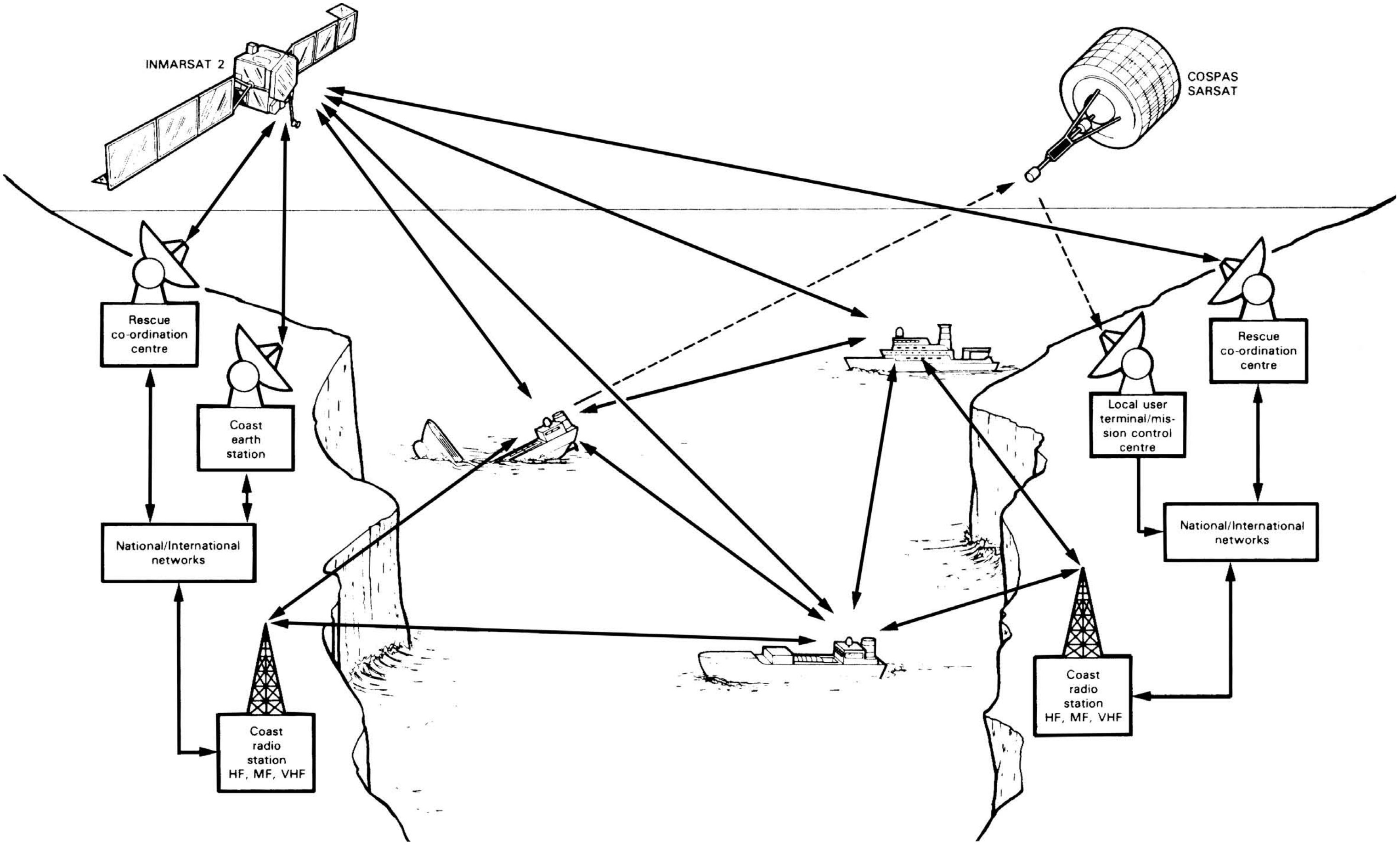


Figure 11.1 General concept of the GMDSS. (Reproduced courtesy of the IMO.)

operation of a vessel within the regions of the GMDSS radio net. The designated areas are as follows.

- Area Al is within the radio range of shore-based VHF coastal radio stations. Typically 20–30 nautical miles, although many countries do not provide sufficient radio stations to guarantee total radio coverage around their coastline. For this reason some countries have not declared an A1 designated area. For example, the UK has declined to do so and vessels trading in UK waters must be fitted with radio equipment to satisfy area A2 requirements.
- Area A2 is within the radio range of shore-based MF coastal radio stations, typically 100–150 nautical miles.
- Area A3 is within the coverage area of Inmarsat satellites, generally defined as the temperate regions of the world between the limits 70° North and 70° South.
- Area A4 is designated as all other remaining areas or defined as full global coverage for those ships not fitted with satellite communications equipment. This assumes that a terrestrial HF communications system is fitted.

In the event of an emergency the first concern of any radio communications operator is that of alerting, which must take precedence over all other communications. Under GMDSS regulations all vessels must be provided with two totally independent methods of distress alerting (see Figure 11.2). Of course when alerting in a distress situation any method or available equipment may be used to attract attention.

If time permits, a GMDSS alert is normally initiated and acknowledged manually using the primary communications system. Such an alert is easily initiated by using the DSC equipment or simply by pressing the distress alarm button on an Inmarsat mobile earth station (MES) terminal. In the event that a disaster overwhelms a vessel before the DSC system can be used or a manual alert sent, a float free satellite emergency position-indicating radio beacon (EPIRB) is automatically released and activated. The alert message is then received by a COSPAS/SARSAT satellite, the position of the casualty calculated and the data transmitted to earth when the satellite next passes within range of a download station.

Once the RCC for an ocean region has been advised of a distress position it will use either terrestrial or satellite communications to alert other vessels in the area of the casualty. This again implies the use of DSC.

Because DSC forms such an integral part of GMDSS a description follows. Readers should remember that DSC is a highly complex electronic calling system and only a relatively brief organizational description can be provided here.

### 11.2.2 Digital Selective Calling (DSC)

GMDSS distress alerting relies heavily on the automated DSC system fitted in shore-based radio stations and carried on all GMDSS equipped ships. DSC effectively enables a 24-h radio watch to be maintained on specific terrestrial frequency channels. For ships at sea, DSC radio watch must be maintained on the following frequencies.

- VHF channel 70
- MF 2187.5 kHz – in A1 and A2 areas
- HF 8414.5 kHz and at least one other HF DSC frequency appropriate to the time of day and the location of the ship in a A4 or/and A3 area for those ships not fitted with an Inmarsat MES.

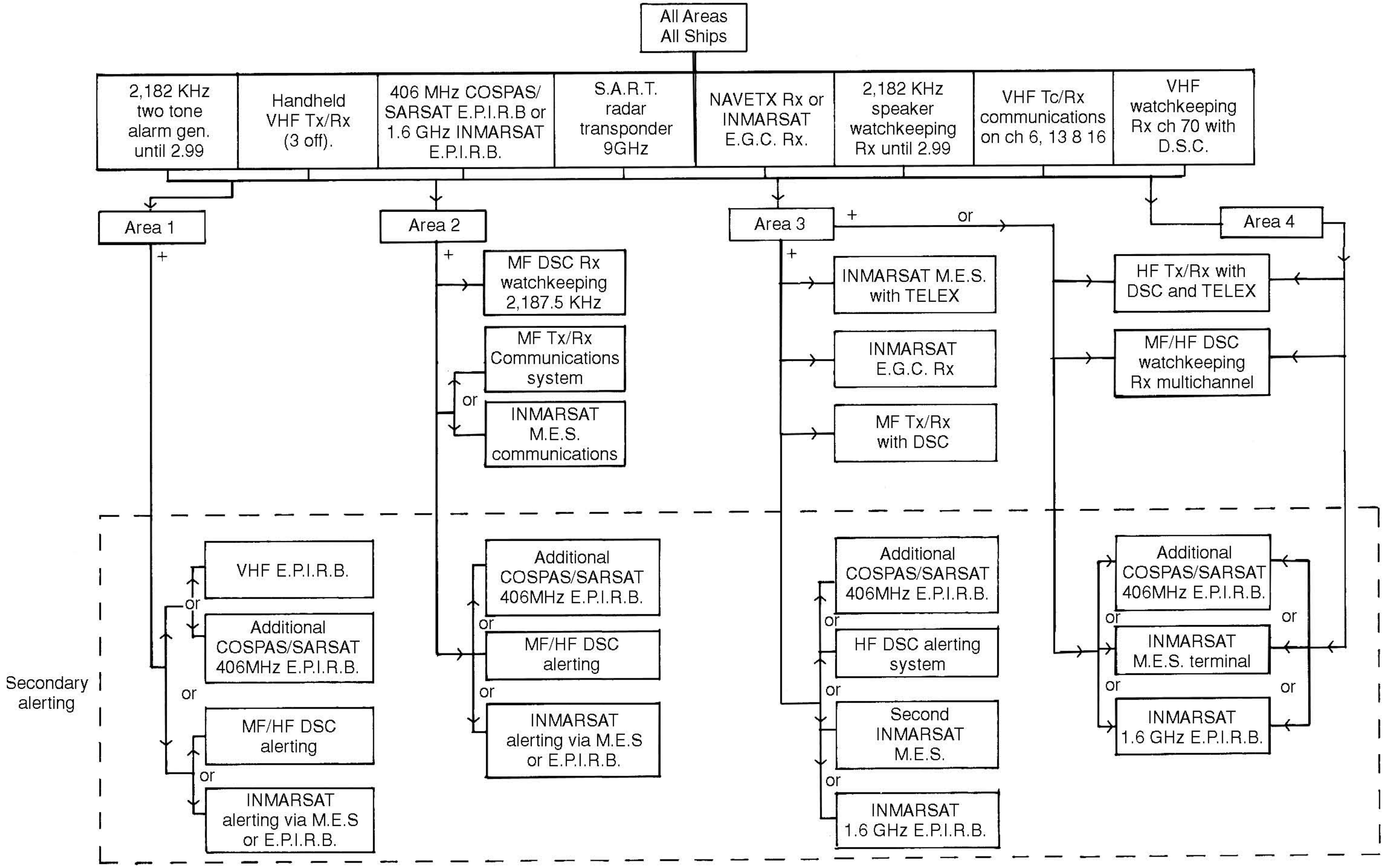


**Figure 11.2** GMDSS equipment carriage requirement table.

DSC terrestrial distress alerts should be transmitted on one or more of the following terrestrial frequency channels depending upon the time of day and position of the vessel: VHF channel 70, MF 2187.5 kHz, HF 4207.5 kHz, HF 6312 kHz, HF 8414.5 kHz, HF 12577 kHz or/and HF 16804.5 kHz. For further information on the selection of a frequency band for terrestrial communications refer to Chapter 1.

GMDSS distress alerting and communications may also be carried out using the MES if the vessel is fitted with satellite communications equipment. Under international regulations all transmitting stations must identify themselves and consequently each station is provided with a selected code. For DSC this is a group of nine digits unique to a single vessel.

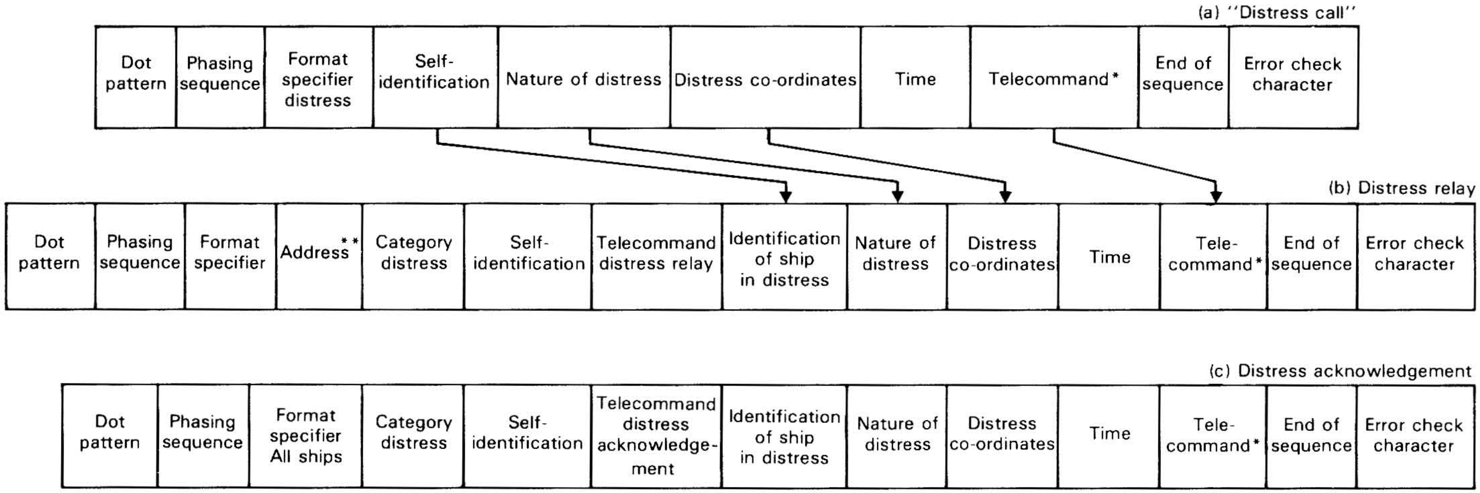
By using the ship's call number, a coastal radio station is able to call a selected vessel. Collective calls may also be made to all ships, or ships belonging to one company or trading in one area of the world. Selective calling is, depending upon the propagational characteristics of the radio wave, a reliable way of automatically calling ships. Although the ship's call number is shown in decimal format, DSC uses a sequence of seven unit binary combinations. Whilst DSC calls are of primary importance for distress alerting and acknowledgement, the system is also capable of handling other more routine communications.

Figure 11.3 shows the sequence and content of the data blocks used for distress alerting, relay and acknowledgement. A distress call is initiated simply by pressing the distress button on the DSC equipment. An incoming distress call will trigger the printer along with audio and visual alarms.

With all methods of automatic digital transmission it is necessary to include error correction coding in the transmission. This is necessary to provide the receiving apparatus with a means of identifying, and in some cases, correcting errors. A DSC sequence transmits each single character twice and uses an overall message check at the end. The transmission speed of a DSC call varies depending upon the frequency band used. On MF and HF it is fairly slow at 100 bauds, but on VHF it is 1200 bauds. A single call on MF or HF therefore varies between 6.2 and 7.2 s, whereas on the faster baud rate of VHF it is between 0.45 and 0.63 s depending upon message content. To increase further the chances of a DSC call or alert being received it is automatically transmitted for five consecutive attempts. Additionally when a DSC alert is made on MF or HF it is transmitted up to six times over any or all of the frequencies available (one on the MF band and five on HF).

Once the DSC distress button has been activated the automatic transmission format shown in Figure 11.3 is transmitted. The first two blocks permit the receiving DSC unit to synchronize with other equipment and then, as an example, the following data is sent signifying a distress alert.

- Format specifier. A distress code will automatically be sent.
- Self-identification. The unique nine digit number (in binary form) identifying the vessel in distress.
- Nature of distress. This is selected by the operator from one of nine codes, i.e. fire or explosion, flooding, collision etc. In the absence of a front panel input, the system defaults to 'undesignated distress'.
- Distress co-ordinates. Automatically included from the interfaced satellite navigation data or defaults to 'no position' information.
- Time. The time at which the distress co-ordinates were valid.
- Telecommand. Indicates whether subsequent distress communication will be by radiotelephony or Narrow Band Direct Printing (NBDP) telegraphy. The system defaults to radiotelephony. When a valid alert is received and acknowledged by a regional RCC, SAR operations are immediately initiated.



*Type of subsequent communication (radiotelephony or teleprinter)
**Address is not included if the format specifier is ''all ships''
(courtesy I.M.O.)

**Figure 11.3** DSC sequence of (a) a distress call, (b) a distress relay call and (c) a distress acknowledgement. (Reproduced courtesy of the IMO.)

On-scene SAR communications are, by definition, short range and will normally take place on VHF between the casualty and other ships or aircraft. Locating a casualty may be done in a number of ways.

- At long range by precise latitude and longitude co-ordinates sent in the alerting message or by using COSPAS/SARSAT fix co-ordinates.
- At short range using VHF radio direction finding triangulation (not a required part of the GMDSS) or, if the casualty has activated a search-and-rescue radar transponder (SART), by using the assisting vessel's radar. A SART generates a series of signals that are easily identified by a 9 GHz shipboard or aircraft radar. The radar display shows a line of 20 blips extending outwards for 8 nautical miles along the bearing line of the SART.

### 11.2.3 The space segment

Satellite communications play a crucial role in the operation of GMDSS. Using satellites, suitably equipped vessels are able to send a distress alert and receive an acknowledgement instantly and reliably from virtually anywhere in the world.

To ensure full global coverage for alerting purposes, two satellite segments, the Inmarsat system and the COSPAS/ SARSAT system, are in operation. The Inmarsat system uses geostationary equatorial orbiting satellites whereas COSPAS/SARSAT uses polar orbiting satellites. Communication via the Inmarsat system is instantaneous and two-way whereas the COSPAS/SARSAT system is outward from the ship only.

#### *COSPAS/SARSAT*

COSPAS/SARSAT (Space system for search of distress vessels/Search and Rescue Satellite-aided Tracking) is an international satellite-aided search and rescue system established and operated by Canada, France, the USA and the USSR.

COSPAS/SARSAT satellites receive digital signals on 406 MHz (for maritime GMDSS EPIRBS) from a casualty, electronically process them and then transmit the data back to a Mission Control Centre (MCC). Various parameters including Doppler frequency shift are used to determine the position of an alert transmitted from a maritime EPIRB, an aeronautical ELT (Emergency Locating Transmitter) or a PLB (Personal Locator Beacon). When a COSPAS/SARSAT satellite passes over an MCC, the data are transmitted to earth for onward transmission to an RCC where the distress position is computed. Depending upon the relative position of a satellite with respect to the casualty there may be some delay in downloading the information but this is insignificant when one considers that the system allows for truly global distress alerting. See Figure 11.4 for the basic concept of this alerting system.

#### *INMARSAT*

To give the reader an understanding of Inmarsat's involvement in the GMDSS, a brief outline of the satellite communication system follows. A full description of satellite communications and Inmarsat can be found in the book *Understanding GMDSS.*

Over 40 countries are signatory members of Inmarsat and each one appoints an organization to represent its investment and interests in the system. Inmarsat signatories are responsible for the establishment and operation of the land earth stations (LES) that are the downlink stations communicating with Inmarsat satellites. Mobile users, on the other hand, purchase, install and operate Mobile Earth Station (MES) equipment that has been constructed to Inmarsat-approved standards by approved suppliers.

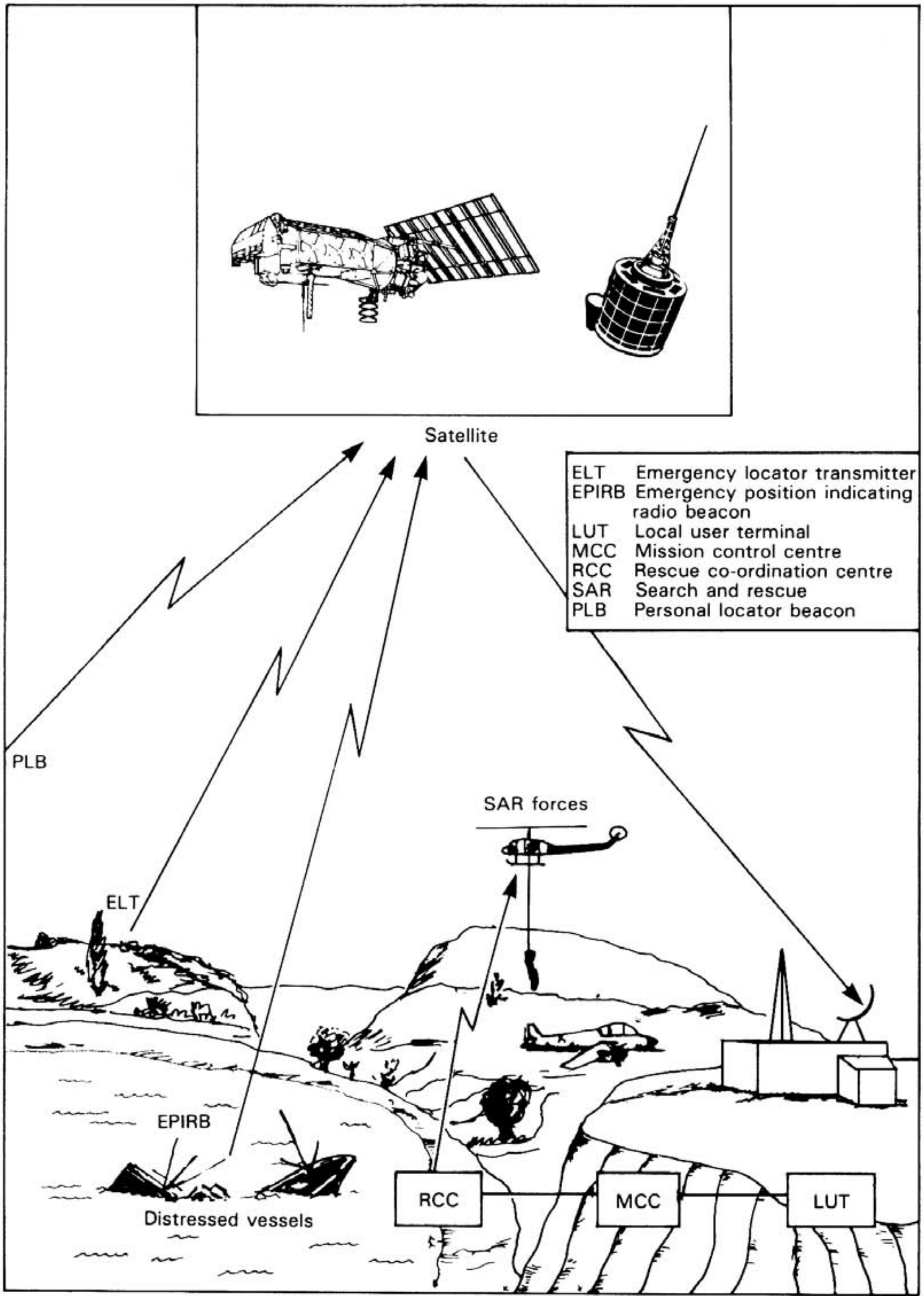


**Figure 11.4** Basic concept of the COSPAS/SARSAT alerting system. (Reproduced courtesy of the IMO.)

Inmarsat's operations control centre (OCC) forms the nucleus of the system's control. It is located on the outskirts of London, England from where technical operators monitor the network for all three ocean regions. Each of the ocean regions, Atlantic (AOR E/W), Indian (IOR) and Pacific (POR), is served by one or more satellites in geostationary orbit approximately 36 000 km above the equator. Currently there are four satellites in each region, some are active and others are available for standby, producing coverage 'footprints' as shown in Figures 11.5 and 11.6.

There are several classes of MES and equipment available in the Inmarsat system of interest to mariners.

- Inmarsat-A. This is physically the largest and oldest of the four and, although the technology has been improved upon in the new digital Inmarsat-B MES, it still provides a good service for the many ships that carry it. It is an analogue system providing two-way direct-dial phone, fax, telex, electronic mail and data communications at 9.6 kbit $s^{-1}$, although a high speed data (HSD) option is sometimes fitted giving rates up to 64 kbit $s^{-1}$. The above-decks parabolic antenna is easily recognized on a ship by the large radome in which it is enclosed. Certified for use within the GMDSS.

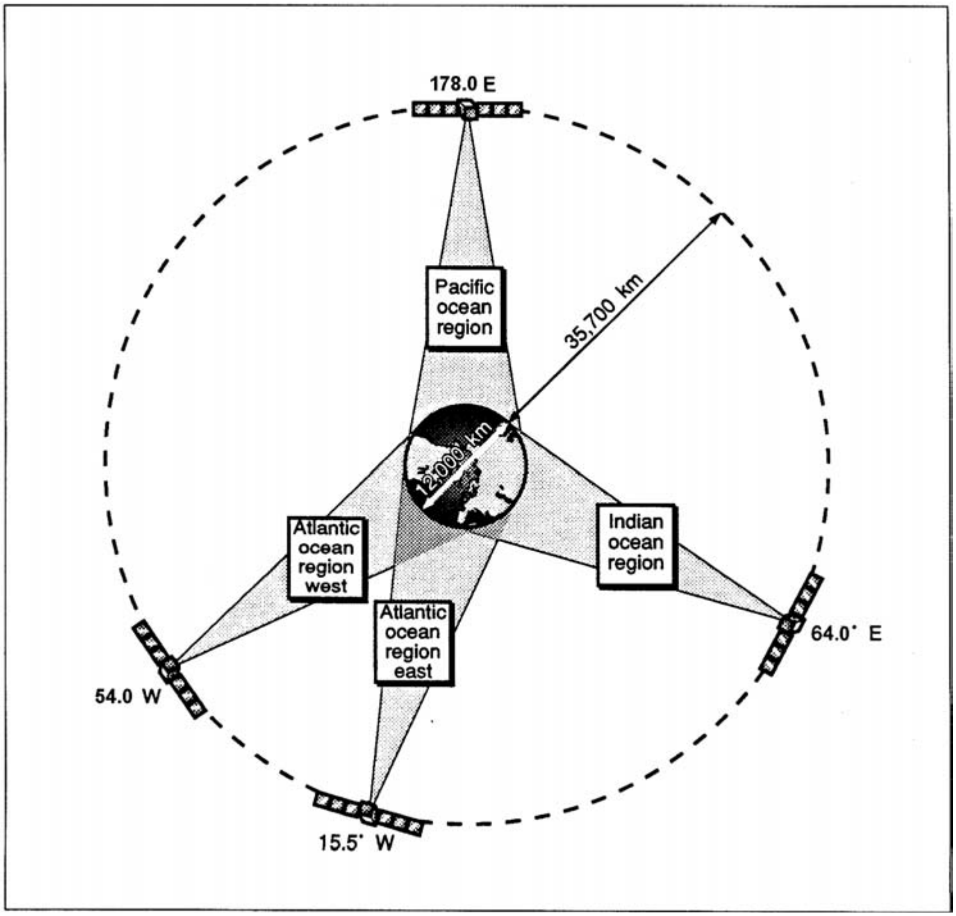


**Figure 11.5** Earth total coverage from Inmarsat's four geostationary satellite configuration. (Reproduced courtesy of Inmarsat.)

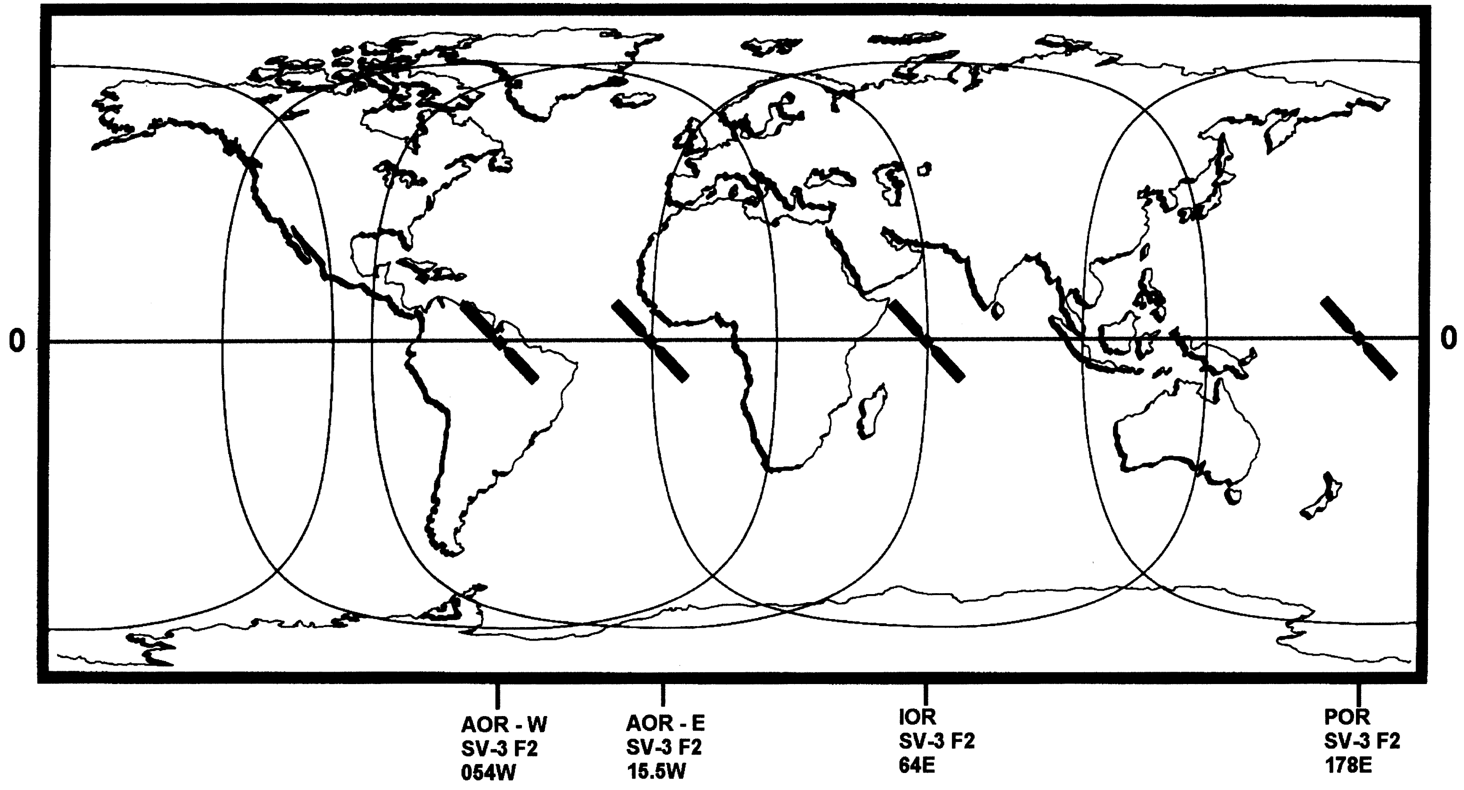


**Figure 11.6** Footprint coverage of the earth's surface from Inmarsat-3 four geostationary satellites. (Reproduced courtesy of Inmarsat.)

- Inmarsat-B. This MES is a smaller and more compact digital version of Inmarsat-A and may eventually replace the older analogue system. Because of its use of digital technology, an Inmarsat-B MES is able to communicate more efficiently and at much faster rates than an Inmarsat-A MES. Its services include, two-way direct-dial high-quality phone, Group 3 facsimile, telex, and 64 kbit $s^{-1}$ and 56 kbit $s^{-1}$ high sped data. Enhanced terminals are also able to offer multiple channel access and other high speed networks. Certified for use within the GMDSS.
- Inmarsat-C. A smaller and cheaper MES providing two-way data communications at 600 kbit $s^{-1}$. It does not handle voice but provides two-way communications via telex or computer data services. The electronics unit can be very small, similar in size to a laptop computer, and uses a small omnidirectional antenna. Inmarsat-C has been approved for use within the GMDSS and supports Enhanced Group Calling (EGC), the SafetyNET and FleetNET services. Other services include, two-way messaging, data reporting and polling, position reporting, safety/emergency alerting and Internet email. Certified for use within the GMDSS.
- Inmarsat-D and D+. Using equipment as small as a personal hi-fi system, Inmarsat-D offers two-way data communications within the full coverage of Inmarsat satellites. It is a data-only system that is able to store and display up to 40 messages of up to 128 characters each and is used for personal paging and group calling as well as two-way communications. When a unit is integrated with a GPS receiver, then labelled Inmarsat-D+, it is able to transmit position information for tracking and tracing services.
- Inmarsat-E. In the GMDSS system, the Inmarsat-E system provides global alerting, via Inmarsat satellites, from Emergency Position Indicating Radio Beacons (EPIRBs). A float-free EPIRB may also incorporate a GPS receiver that is interfaced with the transmitter to provide location data.
- Inmarsat mini-M. Designed to use the spot beam power of Inmarsat-3 satellites, Inmarsat mini-M equipment offers two-way digital phone, voice, fax and data services. Inmarsat mini-M equipment is small and cheap to operate but it is not certified for use within the GMDSS service.

Inmarsat provides the following services as part of the GMDSS radio net.

### *Ship-to-shore distress alerting*

The Inmarsat system provides instant priority access to shore in emergency situations. A maritime operator is provided with a distress button which when activated instantly sends a distress alert. The message is recognized at a LES and a priority channel is allocated. The system is entirely automatic and once activated will connect a ship's operator directly with an RCC. Because the MES is interfaced with the vessel's satellite navigation equipment, the geographical location of the distress will also be automatically transmitted.

### *Shore-to-ship distress alerting*

This may take one of three forms.

- An All Ships Call made to vessels in one ocean region.
- A Geographical Area Call made to vessels in a specific area. Areas are based on the IMO NAVAREA scheme. A MES will automatically recognize and accept a geographical area call only if it carries a specific code.
- A Group Call to Selected Ships alerting ships in any global area again providing specific codes have been input to the MES. Calls are made using the Enhanced Group Calling (EGC) network.

### *Enhanced Group Calling*

The EGC system has been designed by Inmarsat to provide a fully automated service capable of addressing messages to individual vessels, pre-determined groups of ships, or all ships in specified geographical areas. EGC calls may be addressed to groups of ships designated by fleet, flag or geographical area. A geographical area may be further defined as a standard weather forecast area, a NAVAREA, or other pre-determined location. This means that in addition to efficient GMDSS shore-to-ship alerting, the system is also able to provide automated urgency and safety information, as well as fleet calls made by the owner.

## 11.3 The NAVTEX system

### 11.3.1 Introduction

NAVTEX is not a position fixing system, it is an information network. The service forms an integral part of both the Global Maritime Distress and Safety System (GMDSS) and the World Wide Navigational Warning Service (WWNWS) operated by the International Maritime Organization (IMO). These broadcast systems are designed to provide the navigator with up-to-date navigational warnings in English and, using the EGC SafetyNET message service, provide a means of shore-to-ship alerting announcing distress and urgency traffic (Figure 11.7).

NAVTEX services are based on the IMO's 16 global NAVAREAS chart shown in Figure 11.8. Each NAVAREA is subdivided and covered by a number of transmission stations, A to Z. This geographical spread of transmitters minimizes the risk of interference between transmitting stations in adjoining areas.

The transmission schedule for NAVAREA1, Western Europe, is shown in Table 11.1 and the transmitting station locations and coverage areas in Figure 11.9. Similar station groupings occur in other parts of the world.

### 11.3.2 System parameters

Messages are transmitted on a frequency of 518 kHz using narrow band direct printing (NBDP) techniques. Modulation is by FM, F1B designation, using a 7-unit forward error correcting (FEC or Mode B) at 100-bauds frequency shift keying (FSK) with a carrier shift of 170 Hz. The centre frequency of the audio spectrum is 1700 Hz and the receiver bandwidth 270–340 kHz (at 6 dB).

**Table 11.1** European TDM schedule for NAVTEX transmissions

| *Code* | *Name* | *Times of transmission* | | | | | |
|---|---|---|---|---|---|---|---|
| H | Harnosand | 0000 | 0400 | 0800 | 1200 | 1600 | 2000 |
| S | Niton | 0018 | 0418 | 0818 | 1218 | 1618 | 2018 |
| U | Tallin | 0030 | 0430 | 0820 | 1230 | 1630 | 2030 |
| G | Cullercoats | 0048 | 0448 | 0848 | 1248 | 1648 | 2048 |
| F | Brest-le-Conquet | 0118 | 0518 | 0918 | 1318 | 1718 | 2118 |
| O | Portpatrick | 0130 | 0530 | 0930 | 1330 | 1730 | 2130 |
| L | Rogaland | 0148 | 0548 | 0948 | 1348 | 1748 | 2148 |
| T | Oostende | 0248 | 0648 | 1048 | 1448 | 1848 | 2248 |
| R | Reykjavik | 0318 | 0718 | 1118 | 1518 | 1918 | 2318 |
| J | Stockholm | 0330 | 0730 | 1130 | 1530 | 1930 | 2330 |
| P | Scheveningen | 0348 | 0748 | 1148 | 1548 | 1948 | 2348 |
| B | Bodo | 0018 | 0418 | 0900 | 1218 | 1618 | 2100 |

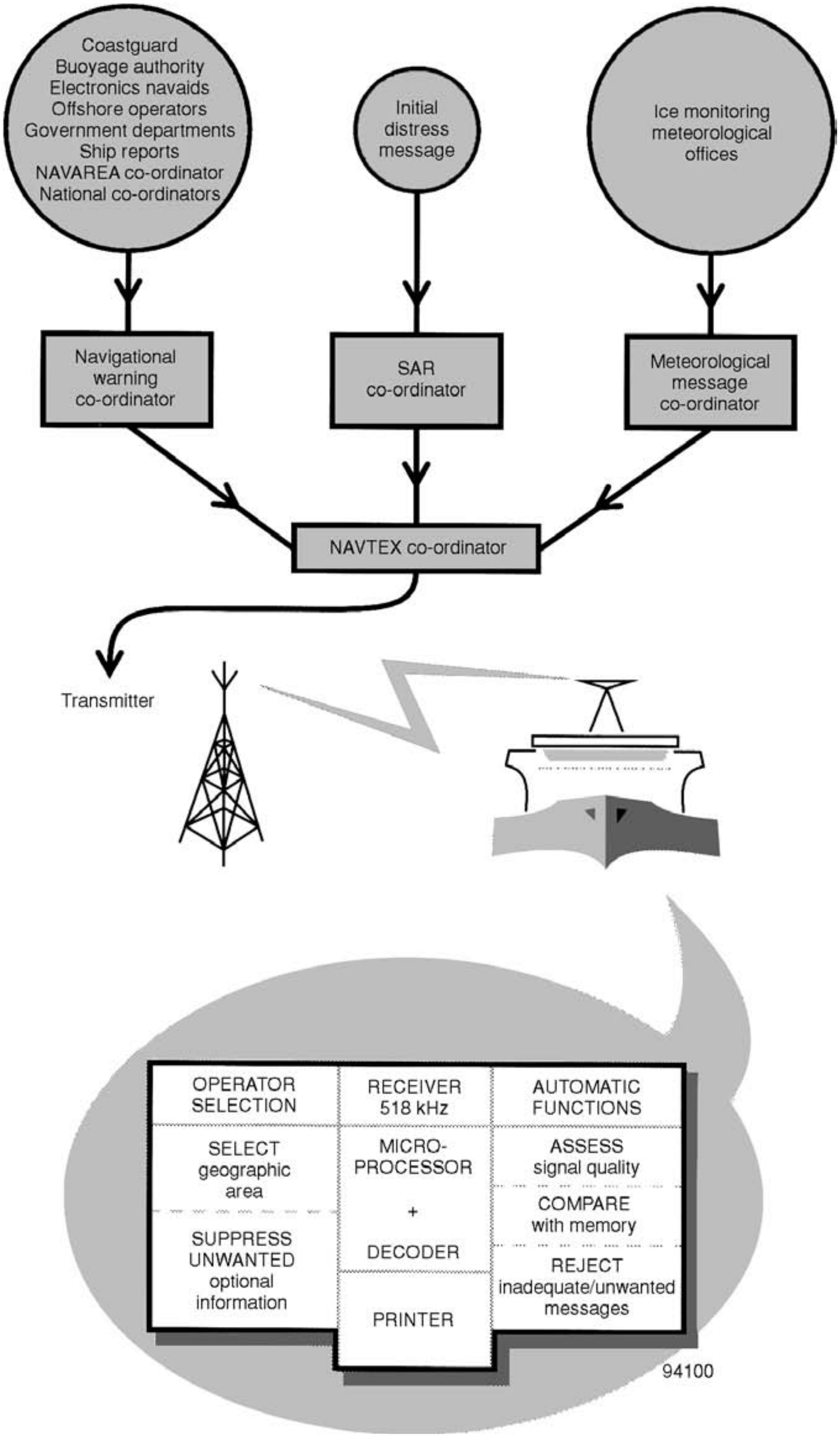


**Figure 11.7** Structure of the NAVTEX service. (Reproduced courtesy of the IMO.)

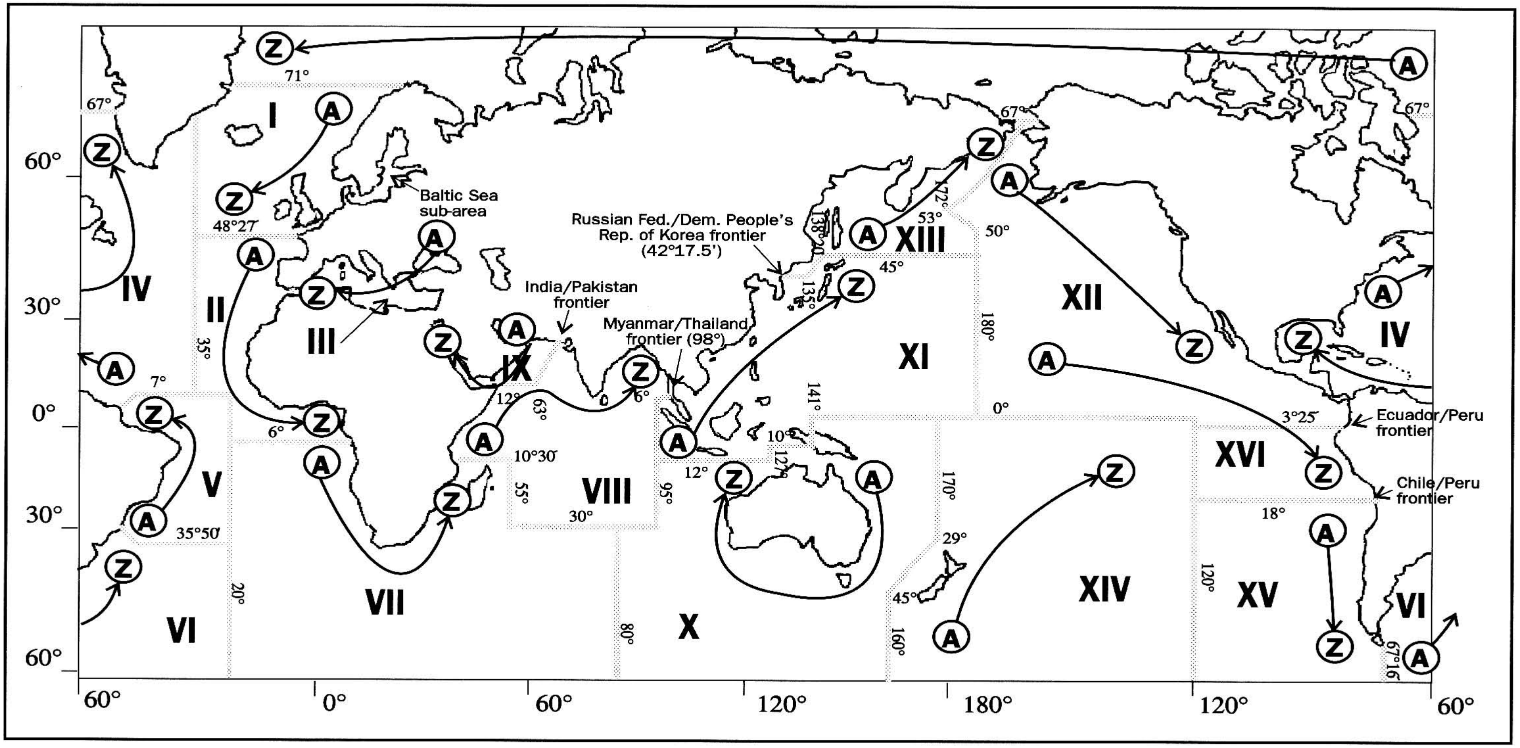


**Figure 11.8** NAVAREAS of the World Wide Navigational Warnings Service (WWNWS) showing the basic scheme for allocation of transmitter identification characteristics. (Reproduced courtesy of the IMO.)

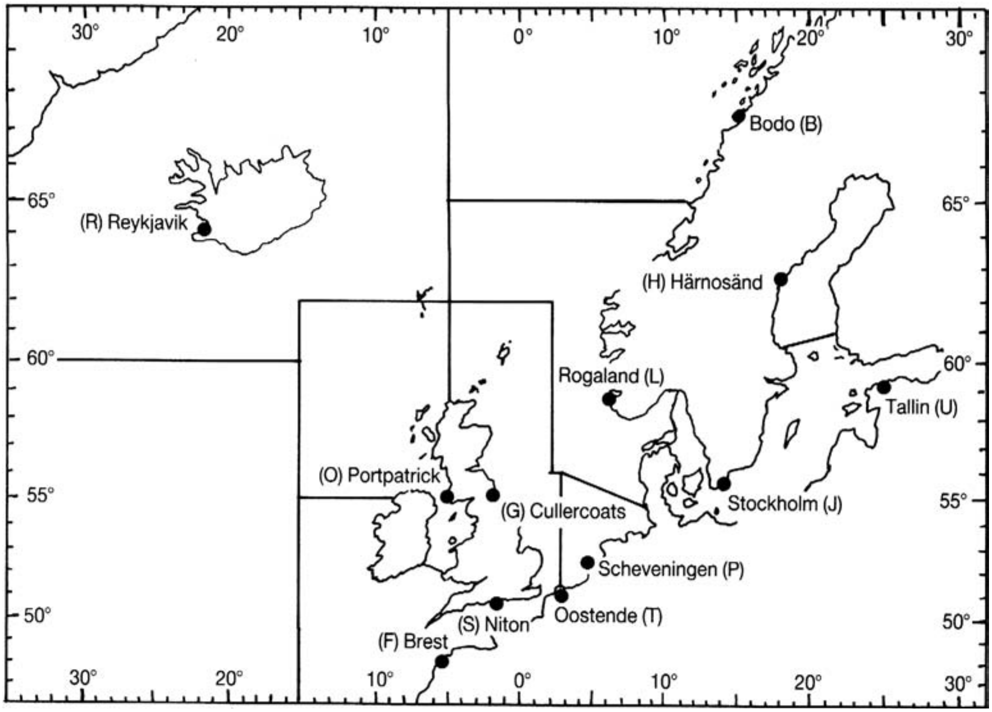


**Figure 11.9** NAVTEX coverage areas within NAVAREA1.

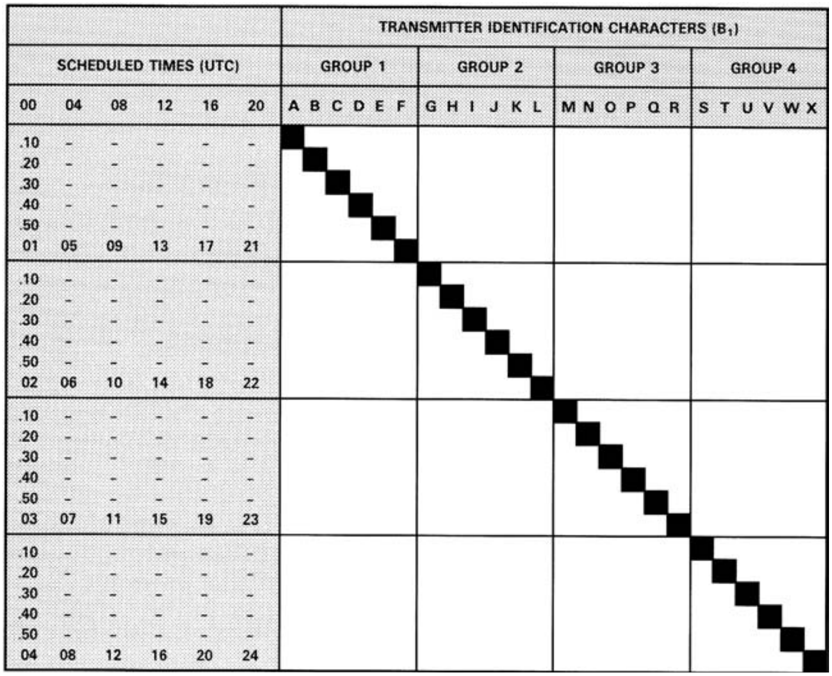
Marine safety information (MSI) is also transmitted by NBDP with FEC on 490 kHz, in tropical areas and there are future plans to use 4209.5 kHz to extend the service.

The NAVTEX primary frequency 518 kHz propagates mainly by surface wave and, if all other factors remain constant, its range is determined by carrier power at the transmitter. NAVTEX transmitters are designed to have an effective range of 400 nautical miles. This figure has been based upon a transmitter carrier power of 1 kW and a receiver input sensitivity better than 1 μV and a 10 dB signal-to-noise ratio. The accepted range for reception of NAVTEX broadcasts may be greatly increased when the sky wave is returned from the ionosphere. Naturally the system is not designed for sky wave reception and messages received via that route may be unreliable. In addition to limiting range by capping the transmitted power, time division multiplex (TDM) of the carrier frequency is also used to limit the chance of interference from neighbouring stations. A simple organizational transmission matrix is used as shown in Figure 11.10.

NAVAREAs are subdivided into four groups each containing six transmitters each with a 10-min allocated transmission slots every 4 h. It should be noted that the matrix is designed for the broadcasting of routine navigational information and that a large volume of data can be transmitted in 10 min at a rate of 100 bauds. It is unlikely that all time slots will be allocated within one frame in any one NAVAREA. Distress and vital warnings are transmitted upon receipt.

### 11.3.3 Signalling codes

Every NAVTEX message is preceded by a four-character header $B_1$, $B_2$, $B_3$, $B_4$ and every NAVTEX receiver is able to read the codes and take action accordingly.



| SCHEDULED TIMES (UTC) | | | | | | TRANSMITTER IDENTIFICATION CHARACTERS ($B_1$) | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GROUP 1 | | | | | | GROUP 2 | | | | | | GROUP 3 | | | | | | GROUP 4 | | | | | |
| 00 | 04 | 08 | 12 | 16 | 20 | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X |
| .10 | - | - | - | - | - | ■ | | | | | | | | | | | | | | | | | | | | | | | |
| .20 | - | - | - | - | - | | ■ | | | | | | | | | | | | | | | | | | | | | | |
| .30 | - | - | - | - | - | | | ■ | | | | | | | | | | | | | | | | | | | | | |
| .40 | - | - | - | - | - | | | | ■ | | | | | | | | | | | | | | | | | | | | |
| .50 | - | - | - | - | - | | | | | ■ | | | | | | | | | | | | | | | | | | | |
| 01 | 05 | 09 | 13 | 17 | 21 | | | | | | ■ | | | | | | | | | | | | | | | | | | |
| .10 | - | - | - | - | - | | | | | | | ■ | | | | | | | | | | | | | | | | | |
| .20 | - | - | - | - | - | | | | | | | | ■ | | | | | | | | | | | | | | | | |
| .30 | - | - | - | - | - | | | | | | | | | ■ | | | | | | | | | | | | | | | |
| .40 | - | - | - | - | - | | | | | | | | | | ■ | | | | | | | | | | | | | | |
| .50 | - | - | - | - | - | | | | | | | | | | | ■ | | | | | | | | | | | | | |
| 02 | 06 | 10 | 14 | 18 | 22 | | | | | | | | | | | | ■ | | | | | | | | | | | | |
| .10 | - | - | - | - | - | | | | | | | | | | | | | ■ | | | | | | | | | | | |
| .20 | - | - | - | - | - | | | | | | | | | | | | | | ■ | | | | | | | | | | |
| .30 | - | - | - | - | - | | | | | | | | | | | | | | | ■ | | | | | | | | | |
| .40 | - | - | - | - | - | | | | | | | | | | | | | | | | ■ | | | | | | | | |
| .50 | - | - | - | - | - | | | | | | | | | | | | | | | | | ■ | | | | | | | |
| 03 | 07 | 11 | 15 | 19 | 23 | | | | | | | | | | | | | | | | | | ■ | | | | | | |
| .10 | - | - | - | - | - | | | | | | | | | | | | | | | | | | | ■ | | | | | |
| .20 | - | - | - | - | - | | | | | | | | | | | | | | | | | | | | ■ | | | | |
| .30 | - | - | - | - | - | | | | | | | | | | | | | | | | | | | | | ■ | | | |
| .40 | - | - | - | - | - | | | | | | | | | | | | | | | | | | | | | | ■ | | |
| .50 | - | - | - | - | - | | | | | | | | | | | | | | | | | | | | | | | ■ | |
| 04 | 08 | 12 | 16 | 20 | 24 | | | | | | | | | | | | | | | | | | | | | | | | ■ |

**Figure 11.10** Scheme for the allocation of transmission schedules. (Reproduced courtesy of the IMO.)

- $B_1$ is an alpha character identifying a specific transmitting station that is used by a receiver to determine messages to be accepted or rejected. In order to prevent erroneous reception by a receiver that happens to be in a position to receive two transmissions using the same $B_1$ code, each code's allocation is based on the NAVAREAS shown in Figure 11.8. Transmitters are allocated, according to an IMO-adopted strategy, an alphabetical listing in sequence through each NAVAREA with no two transmitters, in ground wave range of each other, bearing the same alphabetical character.
- $B_2$, another alpha character, identifies the different classes of message available (Table 11.2). The $B_2$ code is used by the receiver to reject unwanted messages.
- Subject indicators $B_3$ and $B_4$ indicate the numbering of the messages transmitted commencing with 00 and ending at 99. The use of the number 00 indicates a message that will be printed by all receivers. This number is reserved for distress alerting.

### 11.3.4 Message format

A NAVTEX transmission data frame is shown in Figure 11.11. A 10-s synchronizing frame is followed by the sequence ZCZC indicating the end of the phasing period. The B code characters indicate coverage area, message type and numbering. Carriage return and line feed are included for NBDP control. The message follows and is concluded with NNNN. More printer control signals follow before the entire sequence is repeated.

### 11.3.5 Signal characteristics

FSK modulation is used to encode message data onto the 518 kHz carrier frequency. The FSK modulator shifts the carrier frequency either side of 518 kHz by ±85 Hz. Thus to encode a logic 0, the

**Table 11.2** NAVTEX subject indicator characters for code $B_2$

| *Code* | *Meaning* |
|---|---|
| A | Navigational warnings* |
| B | Meteorological warnings* |
| C | Ice reports |
| D | Search and rescue information and pirate warnings* |
| E | Meteorological forecasts |
| F | Pilot service messages |
| G | Formerly DECCA messages (This service is no longer in use) |
| H | LORAN-C messages |
| I | Formerly OMEGA messages (This service is no longer in use) |
| J | SATNAV messages – GPS and GLONASS |
| K | Other electronic navaid messages |
| L | Navigational warnings additional to letter A* |
| V | Notices to fishermen (USA only) |
| W | Environmental messages (USA only) |
| Z | No messages to hand |

* Messages that cannot be rejected by a receiver.
Note: Subject indicator letters B, F and G are not normally used in United States waters because the US National Weather Service includes weather warnings as part of a forecast. NAVTEX meteorological warnings are broadcast under the subject character E. Indicators V, W, X and Y are allocated by the NAVTEX Panel for special services.

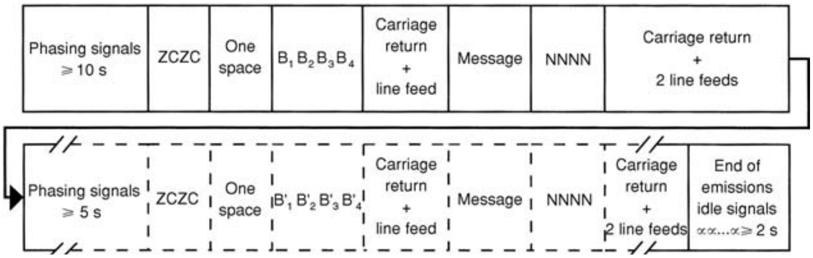


**Figure 11.11** Data format of NAVTEX transmissions. (Reproduced courtesy of the IMO.)

carrier is retarded to 517.915 kHz and for a logic 1, it is advanced to 518.085 kHz conforming to CCIR recommendation 540. In the receiver the 517.915 kHz signal is demodulated to an audio frequency of 1615 Hz representing logic 0 and the 518.085 kHz signal is demodulated to a logic 1 of 1785 Hz.

Each alphanumeric character is serially encoded as a 7 data-bit word (7-unit SITOR code) with a data rate of 100 bauds.

Table 11.3 shows the complete NAVTEX coding standard that conforms to CCIR recommendation 476. There are, however, only 35 possible combinations using this code and consequently each data string represents two possible characters. For instance, data string 0010111 may represent a T or a 5. To eliminate this error, each 7-bit data character is preceded by the letter or figure shift codes.

To eliminate errors caused by noise in the transmission path the system employs the same transmission protocol as that used by marine radiotelex services, i.e. forward error correction (FEC). Each symbol is transmitted twice, the first time known as DX (direct) and the second as RX (repeat).

**Table 11.3** NAVTEX coding standard

| *Data input* | *Hex* | *Meaning* | |
|---|---|---|---|
| | | *Letters* | *Figures* |
| 0001111 | OF | Carriage return | |
| 0010111 | 17 | T | 5 |
| 0100111 | 27 | 8 | ? |
| 1100011 | 47 | 0 | 9 |
| 0011011 | 13 | Line feed | |
| 0101011 | 28 | No perforation | |
| 1001011 | 48 | H | |
| 0110011 | 33 | Phasing signal q | |
| 1010011 | 53 | L | > |
| 1100011 | 63 | Z | + |
| 0011101 | 1D | Space | |
| 0101101 | 2D | Letter shift | |
| 1001101 | 4D | N | ‘ |
| 0110101 | 35 | E | 3 |
| 1010101 | 55 | R | 4 |
| 1100101 | 65 | D | $ |
| 0111001 | 39 | U | 7 |
| 1011001 | 59 | I | 8 |
| 1101001 | 69 | S | |
| 1110001 | 71 | A | – |
| 0011110 | 1E | V | = |
| 0101110 | 2E | X | / |
| 1001110 | 4E | M | · |
| 0110110 | 36 | Figure shift | |
| 1010110 | 56 | G | @ |
| 1100110 | 66 | Phasing signal b | |
| 0111010 | 3A | Q | 1 |
| 1011010 | 5A | P | 0 |
| 1101010 | 6A | Y | 6 |
| 1110010 | 72 | W | 2 |
| 0111100 | 3C | K | ( |
| 1011100 | 5C | C | : |
| 1101100 | 6C | F | % |
| 1110100 | 74 | J | BEL |
| 1111000 | 78 | Phasing signal a | |

By referring to the coding standard it can be seen that all the 7-bit codes possess four logic 1s and three logic 0s. This enables the demodulator to identify and correct a single bit error in the received signal. If either the DX or RX words are corrupted, the processor will print the other as the correct character. If both are corrupted, an ‘*’ is printed to indicate that the character is unreliable.

### 11.3.6 Messages

A NAVTEX receiver is designed with the ability to select the messages to be printed. However, various messages including distress alerts cannot be excluded. The message printed is determined by the four-character header code that appears in all message preambles or alternatively may be selected by an operator. An example of a routine message printed by a NAVTEX receiver may be as follows.

ZCZC SB03 (phasing and identity information)

041402 UTC APR 02 (date and time)

NAVAREA 1 156 (Series identity and consecutive number)

Dover Wight SW winds expected

storm force ten imminent.

NNNN (end of message)

where:

ZCZC = phasing sequence
S = the transmitting station (Niton Radio)
B = category of message (meteorological warning)
= message number
041402 = 04 (date) 14 (hour) 02 (minutes)
UTC = Universal Time Co-ordinated
APR = month
= year (2002)
NAVAREA1 = series identity
= consecutive number (identifies the source of the report. Not the same as the NAVTEX serial number $B_3$ $B_4$)

Message text

NNNN = end.

Full and complete details of the NAVTEX system can be found in the International Maritime Organization's NAVTEX Manual available from their office. See the web site www.imo.org

**Table 11.4** Definition symbols for classes of modulation

| | |
|---|---|
| A3E | Double sideband (DSB) |
| H3E | Single sideband (SSB) full amplitude carrier |
| R3E | Single sideband (SSB) reduced carrier amplitude |
| J3E | Single sideband (SSB) fully suppressed carrier |
| J2E | SSB suppressed carrier NBDP and DSC |
| G2E | Phase modulation (PM) DSC channel 70 VHF |
| G3E | PM radio telephony VHF |
| F1B | FM direct printing telegraphy DSC |

## 11.4 Glossary

| | |
|---|---|
| **AORE** | Atlantic Ocean Region East satellite. |
| **AORW** | Atlantic Ocean Region West satellite. |
| **DSC** | Digital selective calling. A NBDP transmission system used for priority alerting. |
| **EGC** | Enhanced group call. A group calling system using Inmarsat-C terminals. |
| **EPIRB** | Emergency position indicating radio beacon. An automatic beacon released from a ship in distress to alert a shore station via the COSPAS/SARSAT network of satellites. |
| **FEC** | Forward error correction. An encoding system providing the ability to detect errors in a digital transmission system. Used in maritime text equipment. |
| **FleetNET** | Inmarsat EGC-based broadcast system permitting shipowners to transmit to some, or all of their fleet. |
| **FM** | Frequency modulation. A voice modulation system of a carrier wave. |
| **FSK** | Frequency shift keying modulation used in the NAVTEX service. |
| **IMO** | The International Maritime Organization. |
| **INMARSAT** | The International Maritime Satellite Organization. |
| **IOR** | Indian Ocean Region satellite. |
| **ITU** | The International Telecommunications Union. |
| **MCC** | Mission Control Centre. |
| **MES** | Inmarsat mobile earth station. The satellite communications equipment fitted on board a ship. |
| **MRCC** | Maritime Rescue Co-ordination Centre. |
| **MSI** | Maritime safety information. A broadcast service providing information for navigators. |
| **NAVAREA** | IMO designated global navigation area. |
| **NAVTEX** | NBDP broadcast system transmitting navigational information on 518 kHz. |
| **NBDP** | Narrow band direct printing. A narrow band transmission system used for teletype text messages. |
| **NCC** | Network Co-ordination Centre. |
| **Priority-3** | Inmarsat designation for distress calls via satellite. |
| **RCC** | Rescue Co-ordination Centre. |
| **SafetyNET** | Inmarsat EGC system for the transmission of maritime safety notices. |
| **SAR** | Search and rescue. |
| **SARSAT** | Search and rescue satellite-aided tracking. |
| **SART** | Search and rescue radar transponder. A radar beacon that indicates its position in response to surface or airborne radar signals. |
| **SES** | Inmarsat ship earth station. |
| **SOLAS** | Safety of Life at Sea convention. |
| **TOR** | Telex over radio. |
| **UTC** | Co-ordinated universal time. |
| **WARC** | World Radio Administrative Conference. A sub-group of the ITU producing the regulations governing the use of radio frequencies. |

## 11.5 Summary

- The GMDSS is effectively a world radio net in which vessels may communicate a distress situation either via terrestrial or satellite communications.

- A network of Maritime Rescue Co-ordination Centres (MRCCs), one to each global designated area, process the distress communication and co-ordinate SAR units.
- Two-way satellite communication is via Inmarsat satellites located above the Atlantic, the Indian and the Pacific Oceans.
- Ship-to-shore alerting may be done via the orbiting COSPAS/SARSAT satellites on 406 MHz. Distress alerting may also be achieved via the COSPAS/SARSAT system from a float-free EPIRB.
- On-board ship carriage requirements depend upon the GMDSS area in which the vessel is trading. Areas are designated A1–A4.
- Digital selective calling (DSC) is used extensively in the GMDSS for distress alerting and communication. DSC operates on a range of transmission frequencies from MF to VHF.
- Enhanced group calling (EGC), FleetNET and SafetyNET are all services operating within the GMDSS.
- NAVTEX is a broadcast service offering navigational and safety information.

## 11.6 Revision questions

1. State the four designed areas of the GMDSS radio net and explain the difference between areas A3 and A4.
2. What are the major differences between the Inmarsat and COSPAS/SARSAT satellite systems?
3. All vessels must carry two independent methods of distress alerting. Explain the alternative systems that are available for a vessel trading in area A3.
4. What information should the initial distress alert message contain?
5. If a disaster overwhelms a vessel before a manual distress alert can be transmitted, how is an automatic alert activated?
6. How may this alert message be acknowledged by a shore-based station?
7. What is a SART and how does it provide position information to rescue vessels?
8. How may vessels in a specific ocean region be alerted of a casualty by a shore station?
9. NAVTEX provides navigational and other information for shipping. Over what range would you expect to receive NAVTEX signals?
10. Which of the NAVTEX broadcast signals using subject indicators $B_3$ and $B_4$ cannot be rejected by an operator?

# Appendices

# A1

# Computer functions

## Introduction

The function of a computer is to perform operations on data (usually arithmetic or logical) according to a set of specified instructions. The specified set of instructions is a computer program and is known as software. The physical aspects of the computer system, such as the circuitry, monitor, keyboard, printer, cabling etc., is known as the hardware. Computers can be categorized according to the functions for which they are designed.

- Supercomputer. Mainly used for research and capable of ‘number crunching’ on a massive scale with extremely rapid calculations.
- Mainframe. Mainly used in large commercial concerns such as banking and large automated plants where large amounts of data have to be processed on a daily basis.
- Minicomputer. Smaller version of the mainframe and suited to smaller scale businesses and research establishments.
- Microcomputer/Workstation. Less complex than the others, although still powerful, they tend to be operated by a single user. Workstations tend to be a dedicated version of the microcomputer and could well operate faster and contain more memory.

This appendix will concentrate on the last of the computer types since it is the one most likely to be used on board ship.

The heart of a computer is its central processing unit (CPU) which, for a microcomputer, is a microprocessor. More detail on the microprocessor is included later. It is sufficient for the moment to say that it is a circuit available as a single integrated circuit (IC) ‘chip’ which, when connected with other IC chips, can produce the microcomputer. A basic system is discussed below.

## Basic system

Essentially, the microcomputer consists of three elements as shown in Figure A1.1.

In addition to the three hardware elements, there are three sets of connections, known as buses, that interconnect the chips. Details of each bus and its function are as follows.

- **Data bus.** Provides a path for the data which is to be processed. The data is usually in ‘words’ which can be anything from 4 bits to 32 bits in length. A ‘bit’ is a contraction of ‘binary digit’ and can have the value of 1 or 0; thus a combination of 1s and 0s in a word can represent specific data. It can be shown that for a 4-bit word there are $2^4$ or 16 possible combinations ranging from 0000 to 1111. Obviously with 8, 16 or 32-bit words the number of combinations will be increased. A

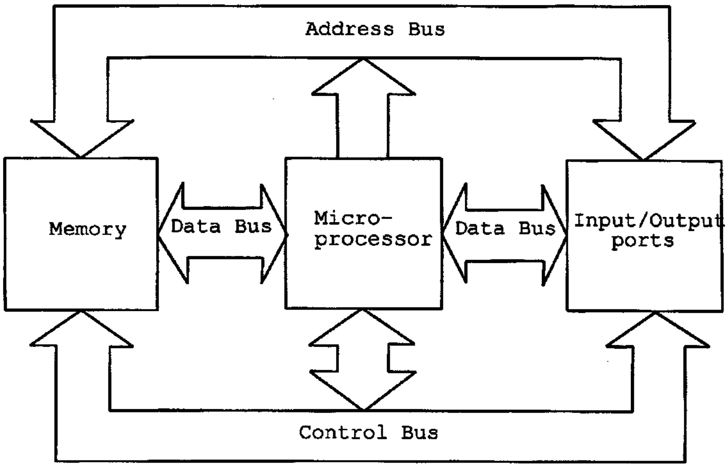


**Figure A1.1** A basic microcomputer system.

group of 8 bits is known as a byte while 4 bits is a nibble. Thus two nibbles make a byte. A group of 16 bits is said to be made up of 2 bytes, etc.

- **Address bus.** The memory device will consist of a number of memory cells which can be uniquely identified by an address. The memory cells can contain data or program instructions and each cell could contain several bits.

  As shown in the basic system diagram of Figure A1.1 the input/output (I/O) chip is also accessed via the address bus. This arrangement is known as memory-mapped I/O. An alternative arrangement allows the microprocessor to be connected to the I/O with a dedicated bus structure giving what is known as dedicated or port addressed I/O.

  The size of the address bus can vary; for an 8-bit system the address bus would be typically 16 bits wide giving $2^{16}$ or about 64 000 (64 kbyte) address locations. For a 16-bit system the address bus is typically 20 bits wide giving 1 Mbyte (one million) addressable locations. When an address is accessed by the microprocessor, all other address locations are disabled so that the microprocessor communicates with only one address location at a time.

- **Control bus.** This bus carries the signals required to synchronize the operations of the system. For example, if the microprocessor needs to read data from (or write data into) a memory location, the control bus carries the necessary signal. The signal in this case is the Read/Write (R/$\overline{\text{W}}$) signal, which is sent from the microprocessor to allow the necessary data movement to be carried out. The microprocessor would send a logic 1 via the control bus if a read operation were to be performed from the memory location whose address was currently on the address bus. For a write operation the signal would be a logic 0, as indicated by the bar over the letter W, i.e., $\overline{\text{W}}$, indicates an operation carried out with a signal that is active low.

  Some I/O elements can send signals to the microprocessor via the control bus; such signals include interrupts, where the system is designed to respond to an external event, and Reset, where the system could be reset to a specified start condition.

  The most important signal carried by the control bus is the system clock which, operating at frequencies up to 1 GHz, provides the necessary synchronization for the system to operate. The clock is crystal controlled and, although not shown on the system diagram above, is an integral part of the microprocessor block.

## Microprocessor

As stated earlier this is a device responsible for executing arithmetic and logical operations and for controlling the timing and sequence of operations. A basic block diagram is shown in Figure A1.2.

The ALU is the arithmetic and logic unit while the control unit undertakes the timing and sequence functions. Additionally there are registers which can hold data while data manipulation takes place. Registers also assist in the role of program execution. A register is simply a store which can contain a set of logic states, i.e. logic 1 and logic 0.

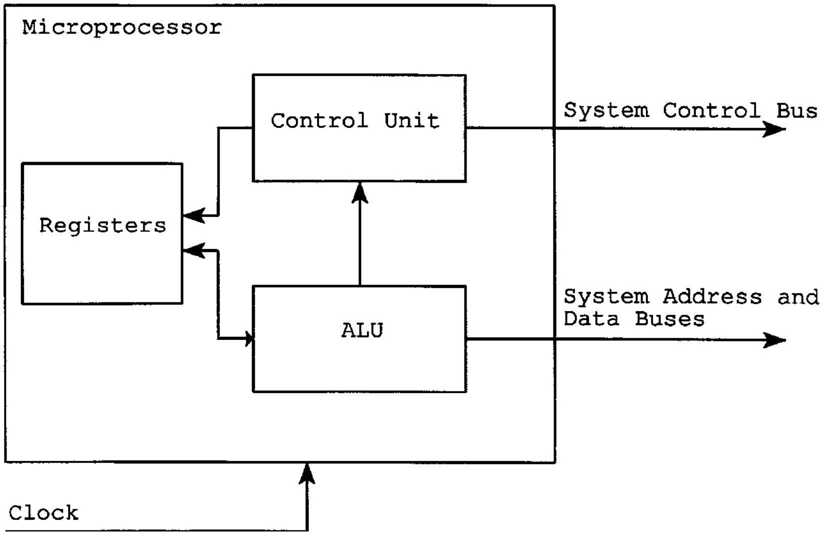


**Figure A1.2** Block diagram of a microprocessor.

The ALU performs arithmetic manipulations, such as binary addition, subtraction and, possibly, multiplication and division. Also the logical functions such as AND, OR, NOT and Exclusive-OR can be implemented. The ALU consists of gates which are organized to receive binary inputs and provide binary outputs according to the instruction code in force, i.e. for the addition process the gates are arranged as an adder while for the AND process the gates are arranged as an AND gate etc.

The control unit provides the essential timing of operations within the system including the process of 'fetch and execute' whereby an instruction is fetched from memory and caused to be executed. This is known as the instruction cycle.

The register group contains the data that the processor needs while performing the task of executing a program. Information held by the registers includes the program counter (PC) which allows the processor to keep track of its position within the program. Other registers include the accumulator and the stack pointer. There are many types of microprocessor available, and the registers and the names given to them may vary from device to device.

## Memory devices

Memory is necessary to store the program instruction codes, the data used in computation and the results of the computation. The memory devices can consist of one or more ICs which can be interconnected to provide the necessary unique location addresses required by the system. Devices fall into two basic categories: random access memory (RAM), perhaps better described as read/write memory, and read only memory (ROM).

For RAM there is a matrix assembly of flip-flops each forming a memory cell and, for this type of memory, it is possible to determine the contents of any cell by a read operation or to change the contents of a cell by a write operation. The read operation is non-destructive since reading will not alter the contents of a cell. However RAM is volatile since removing the power supply from the memory will destroy the contents; restoring the power supply will allow the cells to once again have particular values (logic 1s and logic 0s) that will not be the same as before the removal of the power.

RAM can be static (SRAM) or dynamic (DRAM); in the latter case use is made of stored charge on a capacitor and since such charge can leak away in time, the cells need to be constantly refreshed to maintain the state of charge. ROM also has a matrix array of cells but is non-volatile in that the contents, written by the manufacturer or the customer, will not vary and can be read to give the value of its contents.

In the normal way ROM cannot be written into since its purpose is to provide a pre-determined fixed value for its contents. However, some ROMs are capable of having their contents altered. By a process known as field programming, users may purchase a ROM containing all 1s, or all 0s, in each cell, and by an electrical process cause certain cells to change value to obtain the required contents. Such a memory is a programmable read only memory (PROM). A PROM once programmed has its memory fixed for good. An erasable PROM, or EPROM, can have its contents removed, by using electrical means or UV light, and new contents put in place.

Typically RAM is available in units of 8, 16, 32, 64 and 128 Mbyte and combinations can be used to produce the desired system memory capacity. Secondary, or auxiliary, memory storage devices are available in the form of magnetic tape and disks which are non-volatile. Hard disk drives are available in excess of 30 Gbyte capacity while floppy disk drives exist in 3.5- and 5.25-inch format. Programs are available on floppy disks and CD-ROMs which can be loaded into computer systems for storage on the hard disk if desired. CD-ROMs are also available as memory storage devices that can be written to once (CD-R) or written to, erased and rewritten to (CD-RW).

## Memory organization

A complete computer system needs both RAM and ROM. A memory map will show how the memory locations are divided between the different types of memory. For a system operating with 8 bits of data and a 16-bit address bus, the memory map would extend from location 0000 to location FFFF. This representation of the memory location is known as hexadecimal and provides a simple way of identifying a location without the need to specify all 16 digits of the actual address. There are 16 total variations represented by 4 bits (i.e. $2^4 = 16$), but only 0 (0000) to 9 (1001) can be represented by decimal numbers and the remaining six combinations (1010) to (1111) are represented by alphabetic figures A to F, respectively. Thus 16 bits can be represented by four hexadecimal alphanumeric figures. Thus the first memory location is 0000 0000 0000 0000, represented in hexadecimal form by 0000, while the last is 1111 1111 1111 1111, represented by FFFF. A typical memory map for an 8-bit system is shown in Figure A1.3.

The operating system sits at the top of memory. This system ensures that the facilities of the machine are co-ordinated and also contains information regarding the start address for routines which are executed at reset or external interrupt.

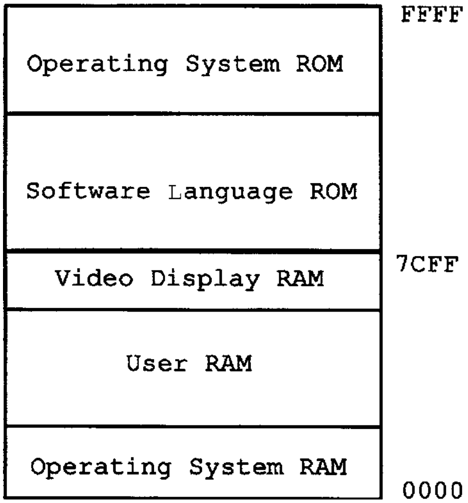


**Figure A1.3** Memory map for a system with a 16-bit address bus.

The visual display unit (VDU) screen supported by the system can have any screen location identified by a particular memory address in RAM. The size of the video RAM depends on the system used and its graphics resolution requirements. The operating system also requires some RAM which should not be used by the system user.

Larger systems with 16 data bits and 20 address lines would have 1 Mbyte of addressable memory with a typical memory map as shown in Figure A1.4.

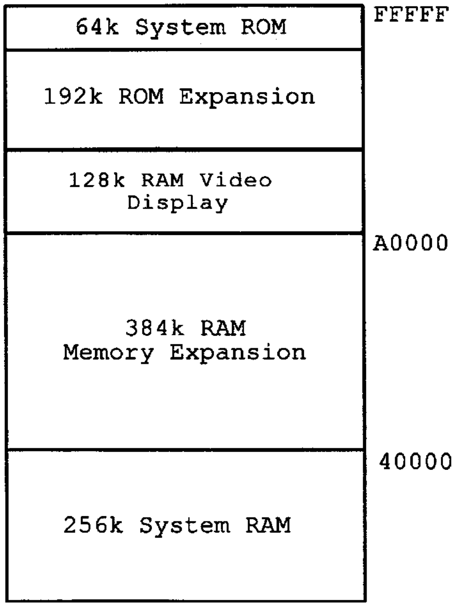


**Figure A1.4** Memory map for a system with a 20-bit address bus.

## Input/output (I/O)

The system will need an interface with the ‘outside world’. The I/O interface allows the connection of input data via, say, a keyboard and sensors which can transpose information such as movement, pressure, temperature etc., into electrical signals. For output data there could be, say, a monitor to display instructions/data and outputs that can feed external devices, such as relays, solenoids, LEDs etc.

As mentioned earlier it is possible to ‘memory map’ the I/O interface so that data read from the interface comes directly from the external device while data transferred to the interface is data fed directly to the external device. The I/O interfaces are usually referred to as I/O ports. Most microcomputer systems have ICs which perform the function of I/O ports and some are programmable which means that the operating mode may be changed to suit the particular system requirements.

If the microcomputer is used to monitor and control an external quantity (such as pressure, temperature, displacement etc.) The signals produced by the transducer are likely to be analogue in form. Such a signal would need translating into a digital signal using an analogue-to-digital converter (ADC) before the input can be fed to the input port of the microcomputer. Once the computer has evaluated the received data, it is likely to send a control signal back to the transducer to maintain or amend the quantity being measured. The control signal is digital in form and must be translated into analogue form using a digital-to-analogue converter (DAC) before being applied to the transducer. A possible arrangement is shown in Figure A1.5.

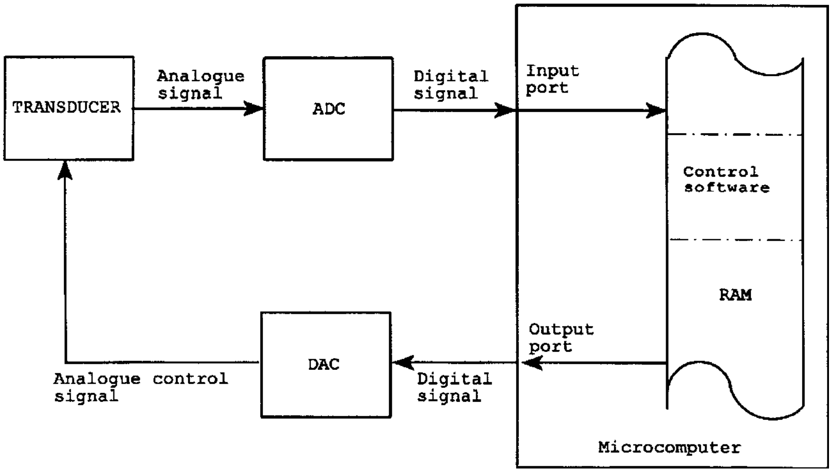


**Figure A1.5** Control of an external transducer parameter by a microcomputer.

The ports may be serial, for moving data a bit at a time, or parallel, where data is moved in a block, with the rate of transfer determined by the system clock.

## Local Area Networks (LANs)

It is often required to interconnect microcomputers/workstations to form a network which, if distances involved are small, is known as a Local Area Network or LAN. A LAN is typically used to share data and peripherals and to allow communication between stations. It is possible, for instance, to use several outlets to display data, for example ECDIS chart information, to areas other than just a single station.

## Interrupts

A microcomputer system may be operated in such a way that the processor is able to respond to a servicing request from an external device as required. Such requests are usually asynchronous in that they are transmitted in an arbitrary manner without being synchronized by an external clock. Two techniques are polling and interrupt.

- Polling is a technique whereby the external devices are interrogated in turn on a priority basis to determine which device made the servicing request. The servicing request will cause the processor to stop its normal program and move to a polling subroutine. After the servicing is carried out the return from subroutine will restore the processor to the task it was engaged in prior to the servicing request being made.
- Interrupt is a technique that causes an interrupt signal to be sent to the processor itself. Such a signal causes the processor to suspend its current operations and transfer to a servicing routine for the device requesting the interrupt. The routine is similar to a subroutine except that it is initiated by hardware instead of software.

## Software

The computer will operate according to a program that contains a set of sequential instructions. The categories of programming are:

1 machine code
2 assembly language
3 high-level language.

1 and 2 above provide what is known as ‘low-level’ programming while 3 provides ‘high-level’ programming. Machine code is in the form of a set of logic 1s and 0s on which the processor operates. Assembly language uses mnemonics which represent the required machine code, with the required transition from assembly language into machine code being performed by an assembler. A computer operator controlling the processor via a keyboard will use a high-level language to produce what is known as source code which is translated into machine code by instructions stored in ROM.

A similar translation process is necessary for changing the machine code results of programming back to a form which can be understood by the operator. The translation process with a high-level language is undertaken by a compiler or interpreter. The difference in level between assembly and high-level languages is that:

- in assembly language, each symbolic code instruction is related to one machine code instruction
- in high-level language, the compiler can deal with complex symbolic instructions each of which would convert to several machine code instructions.

The language BASIC (beginner’s all-purpose symbolic instruction code) is an interpreted high-level language. The source code consists of line-numbered instructions and during program execution each line is converted into machine code prior to execution. No complete machine code translation of the complete source code is produced and program execution is slow but has the advantage that it is easily changed.

Languages such as ADA, C, FORTRAN and PASCAL are compiled with the source code first produced using an editor according to the language rules. The compiler then translates the source programme into machine code form which is termed the object code. The main advantage of compiled high-level languages is that of transportability of the codes between microcomputers that use the same family of microprocessors.

# A2

# Glossary of microprocessor and digital terms

This appendix is not intended to be a complete listing of terms relating to microprocessor and digital systems. The aim is to give a brief outline description of those terms found in the various chapters so that each section can be understood without the need to refer to specialist texts. Should the reader wish to go further than this then obviously textbooks dealing with these topics can be used.

## Using the glossary

Many terms are referred to by an abbreviated form, or acronym, and where applicable the definition appears under this heading. The heading under the full version of the term will direct the reader to the acronym version.

Certain terms are included more than once, although under different headings, with cross-references to link the headings. Cross-references are only used when it is felt necessary, for easier understanding, to expand a particular definition.

| | |
|---|---|
| **ADC** | Analogue-to-digital converter. A device that samples an analogue signal and converts the observed analogue level to digital form. The digital form is made up from several binary digits, or *bits*. |
| **Active** | A signal may be described as active high or active low to indicate which of the two logical levels (logic 1 or logic 0) causes the digital circuit to be enabled. |
| **Address** | A coded instruction that specifies the location in memory of stored data. |
| **Algorithm** | A set of rules laid out in a logical sequence to define a method of solving a particular problem. |
| **Alphanumeric** | A system where the required information is in a combination of alphabetic characters and numbers. |
| **Analogue** | A system where the signal can be considered to vary continuously with time. A digital system on the other hand may be considered to consist of a finite number of discrete levels. The number of levels may only be two as in the case of a binary system. |
| **Analogue-to-digital converter** | See ADC |
| **AND gate** | For a description of a *gate* see under that heading. An AND gate is an electronic circuit of two or more inputs which will only generate an output at logic 1 if all the inputs are at logic 1. All other combinations of input signals will give a logic 0 output. The performance of an AND gate may be defined in terms of a truth |

table which lists the output level for all possible input combinations. The truth table for a two input AND gate is:

| A | B | F |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

where A and B are the inputs and F is the output.

| | |
|---|---|
| **ASCII** | American Standard Code for Information Interchange. This is a common code which gives a 7-bit word to define letters, numbers and control characters. |
| **Basic** | Beginner's all-purpose symbolic instruction code. This is a high-level language that enables the computer user to program the system using an easily understood set of instructions. Within the computer memory there is a 'translator' which converts the BASIC language into the binary signals, or machine code, which the machine understands. |
| **BCD** | Binary Coded Decimal. A system of representing the numbers 0 to 9 inclusive by a binary equivalent. The relationship is as shown: |

| Decimal number | Binary coded value |
|---|---|
| 0 | 0000 |
| 1 | 0001 |
| 2 | 0010 |
| 3 | 0011 |
| 4 | 0100 |
| 5 | 0101 |
| 6 | 0110 |
| 7 | 0111 |
| 8 | 1000 |
| 9 | 1001 |

| | |
|---|---|
| **Binary** | A system of numbers using a base 2, whereas the decimal system uses base 10. The binary system only requires two symbols, i.e. 1 and 0. |
| **Bit** | Contraction of binary digit. A single bit may be a logic 1 or logic 0 and is usually represented by the presence or absence, respectively, of a voltage level. |
| **Buffer** | An electronic circuit connected between other circuit elements to prevent interactions between those elements. The buffer may also provide extra drive capability. A buffer may be used also as a temporary storage device to hold data that may be required at a later time while the computer is engaged on other tasks. |
| **Bus** | A collection of conductors used to transmit binary information in parallel around the system. For microprocessor applications there would be an Address bus used by the central processing unit (CPU) to identify storage locations and a Data bus used for the transmission of data around the system. |

| | |
|---|---|
| **Byte** | A collection of 8 bits. In a microprocessor system using 8-bit data buses and a 16-bit address bus, then the data can be contained in one byte while the address needs two bytes to define it. |
| **Character** | The letters A–Z, numbers 0–9 and other special symbols used by a computer or microprocessor system and coded for use by the system. |
| **Character generator** | The electronic circuitry required in order to prepare a character for display purposes. Such generators possess memory where the binary-coded characters can be stored. |
| **Chip select** | An input to an integrated circuit which, when active, allows the integrated circuit to be operative. If the input is not active then the integrated circuit is inactive. This control signal is sometimes called a 'chip enable' input. |
| **Clock** | A periodic timing signal used to control a system. |
| **Code** | A set of rules allocated to groups of bits. The combination of the bits in a group gives a unique meaning based on following the rules. |
| **Coincidence gate** | An electronic circuit used to indicate, by means of a certain output level, when two inputs are identical. When the inputs are binary in form then all bits of one input should be coincident with the corresponding bits of the other input before the required output level is generated. An Exclusive-NOR gate could be used for this purpose. |
| **Command** | A signal, or group of signals, used to begin or end an operation. |
| **Computer** | In the case of a digital computer the basic system consists of a central processing unit (CPU), memory, input and output units and a control unit. The computer is able to perform such tasks as: manipulate data, perform arithmetic and logical operations on data and store data. |
| **Computer language** | A set of conventions, rules and representations used to communicate with the computer system. The language may be low-level, such as assembler (uses mnemonics), or high-level, using user-orientated language like BASIC. |
| **Converter** | See under the headings of analogue-to-digital converter (ADC) or digital-to-analogue converter (DAC). |
| **Counter** | A circuit used to count the number of pulses received. The counter may be arranged to start from zero and count from there in increments of one (up-counter) or to start from the counter maximum capacity and decrement from that value one pulse at a time (down-counter). |
| **CPU** | Central processing unit. Part of a computer system which contains the main storage (registers), arithmetic and logic unit (ALU) and control circuitry. Sometimes referred to simply as the processor. |
| **Data** | Information or signals, usually in binary form. |
| **D-type flip-flops** | An electronic circuit which on receipt of a clock pulse will give an output logic level the same as that present at the input terminal prior to the arrival of the clock pulse. It is widely used as a data latching buffer element. |
| **Decoder** | An electronic circuit which has several parallel inputs and the ability to recognize one or more of the possible input combinations and output a signal when these combinations are received. All signal levels are binary. |
| **Dedicated** | A dedicated system is one designed to perform a specific operation, i.e. a dedicated microprocessor system is programmed to perform only one specific task. |
| **Demultiplexer** | A device used to direct a time-shared input signal to several outputs in order to separate the channels. |

| | |
|---|---|
| **Digital** | Information in discrete or quantized form, i.e. not continuous as in the case of an analogue signal. |
| **DAC** | Digital-to-analogue converter. An electronic device for converting discrete signal levels into continuous form. |
| **Disable** | A control signal that prevents a circuit or device from receiving or sending information. |
| **Display** | A means of presenting information required by a user in visual form. Includes the use of CRT (cathode ray tube), LED (light emitting diode), liquid crystal, gas discharge and filament devices. |
| **Driver** | An electronic circuit that provides the input for another circuit or device. |
| **Enable** | A control signal that allows a circuit or device to receive or transmit information. |
| **Encoder** | This is the inverse process of decoding. An encoder has several inputs but only one is in the logic 1 state. A binary code output is generated depending on which of the inputs has the logic 1 level. |
| **EPROM** | Erasable and programmable read-only memory. A memory circuit with stored data which can be read at random. The data are capable of being erased and the chip reprogrammed with new data. |
| **Exclusive-OR gate** | A circuit with two inputs and an output which can be at logic level 1 when either of the two inputs is at logic 1 and logic 0 if neither or both the inputs are at logic level 1. |
| **Filament display** | A 7-segment filament wired element whereby an alphanumeric character may be displayed when certain of the filaments are caused to be lit. |
| **Flag** | A flip-flop that can be set or reset to inform of an event that has occurred or a condition that exists within a system. |
| **Flip-flop** | An electronic circuit having two stable states that can be used to store one bit. The circuit uses two gates, the output from each being cross-coupled as an input to the other. The output from one gate is usually referred to as the Q output while the output from the other gate, being the complement of the first output, is called $\overline{Q}$. |
| **Gate** | This is a circuit with two or more inputs and an output which allows a logic level 1 to exist at the output, or not, as the case may be, when certain defined criteria are met. |
| **Hard copy** | Printed or graphical output produced on paper by a computer system thus allowing a record to be kept. |
| **Input/output ports** | These circuits allow external circuits to be connected to the computer internal bus system. |
| **Integrated circuit (IC)** | A small ‘chip’ of silicon processed to form several elements directly interconnected to perform a given unique function. |
| **Interface** | A common boundary between systems to allow them to interact. |
| **Interrupt** | A computer input that temporarily suspends the main program and transfers control to a separate interrupt routine. Interrupt inputs to the microprocessor systems discussed in the main text are usually referred to by acronyms such as $\overline{\text{IRQ}}$ (interrupt request) and $\overline{\text{INT}}$. |
| **Interrupt masking** | A technique that allows the computer to specify if an interrupt will be accepted. $\overline{\text{IRQ}}$ and $\overline{\text{INT}}$ are maskable interrupt inputs whereas NMI (non-maskable interrupt) is not. |

| | |
|---|---|
| **Keypad (or Keyboard)** | A unit which forms part of an input device. This may have a full QWERTY type key layout or be a simplified arrangement to suit the needs of the system. |
| **Language** | See computer language. |
| **Latch** | A temporary storage element, usually a flip-flop. |
| **Logic** | Electronic circuits which control the flow of information through the system according to certain rules. These circuits are known as gates since the ‘gates’ are opened and closed by the sequence of events at the inputs. |
| **Logic level** | Using binary notation the levels may be logic 1 or logic 0. According to the rules mentioned in the definition of logic, level 1 is taken to mean a logical statement is ‘true’ while level 0 means the logical statement is ‘false’. |
| **Magnetic tape** | A flexible, standard width, magnetic powder coated tape which can be used to store, and retrieve, binary-based data. |
| **Mask bit** | With reference to an interrupt request, an internal flip-flop in the MPU can be set to disable an interrupt (interrupt masked) or reset to allow the interrupt to be accepted. |
| **Memory** | In a digital system, it is that part of the system where information is stored. |
| **Microprocessor** | See MPU. |
| **Monostable** | An electronic circuit which has only one stable state. The circuit is normally in the stable state and is triggered into the unstable state where it remains for a period of time determined by a CR time constant value of external components. After this period of time the circuit returns to the stable state. |
| **MPU** | An IC that can be programmed with stored instructions to perform a wide variety of functions, consisting of at least a controller, some registers and an ALU (arithmetic and logic unit). Thus the MPU contains the basic parts of a simple CPU. |
| **Multiplexing** | A method of selecting one of several inputs and placing its value on a time-shared output. |
| **NOT gate** | An inverter. A circuit whose output is high if the input is low and vice versa. |
| **Octal latch** | An integrated circuit package that offers eight separate flip-flop (or latch) circuits. |
| **OR gate** | For a description of a *gate* see under that heading. An OR gate is an electronic circuit of two or more inputs which will generate an output at logic 1 if any one or all of the inputs are at logic 1. Only when all inputs are at logic 0 will the output be at logic 0. The performance of an OR gate may be defined in terms of a truth table which lists the output level for all possible input combinations. The truth table for a two-input OR gate is: |

| A | B | F |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |

where A and B are the inputs and F is the output.

| | |
|---|---|
| **Port** | Terminals (input and output) which allows access to or from a system. |
| **Printer** | The output peripheral of a computer system which allows a hard copy to be obtained. |
| **Program** | A sequence of instructions logically ordered to perform a particular task. |

| | |
|---|---|
| **PROM** | Programmable Read Only Memory. This is the type of memory used to hold microprocessing instructions. It is a form of ROM that can be programmed by the supplier and not the user. |
| **Pulses** | Those signals used to energize a circuit digitally. There is a transition in signal level between discrete values and each level is maintained for a period of time. |
| **Quad gate** | An integrated circuit package which offers four separate gate circuits of a particular type. |
| **RAM** | Random Access Memory. A memory that can be read from and written into. |
| **Readout** | A presentation of input information from a computer. It can be displayed on a screen, stored on tape or disk or be a hard copy when it is usually referred to as a printout. |
| **Read/Write** | Can refer to the type of memory element (RAM is sometimes referred to as read/write memory) or to the signal input line to a RAM chip, the logic level on which determines whether the memory is read or overwritten. |
| **Register** | A group of memory cells used to store groups of binary data in a microprocessor. |
| **Reset** | This could be an input to a flip-flop to bring the Q output to a logic 0 state or that facility which allows a microprocessor to be returned to a pre-determined state. Where the point of return is situated in memory depends on the system. |
| **ROM** | A memory element containing information which cannot be altered under computer operation. The data can only be read by the computer. |
| **Self-test** | In some equipment this is a facility by which when power is first applied the system is checked by running through a special software routine. If a faulty area is found some indication is made to the user. |
| **Sensor** | A device, possibly a transducer, which converts physical data into electrical signal form. If digital in form the electrical signal can be processed by the computer directly while if analogue in form, it requires analogue-to-digital conversion (ADC) before being applied to the computer. |
| **Seven-segment display** | That form of display element comprising seven segments where each segment can be individually energized. The element is thus able to display a variety of alphanumeric characters depending on which segments are energized. |
| **Shift register** | A register in which the stored data can be shifted, a bit at a time, to the left or right. |
| **Signal** | An electrical variation, either continuously variable or variable between discrete levels, which can be interpreted as information. |
| **Software** | A program which can be loaded into a computer system and resides in RAM. Such programs can be loaded and changed at will. |
| **Storage** | A term used to describe any device capable of storing data. Memory elements are storage devices. |
| **Subroutine** | Part of a master program which can be entered frequently from the master program. Used to save programming space where a part of a program is repetitive. |
| **Tape** | The media, either paper or magnetic, used to store binary coded data for a computer system. |
| **Test** | The routine for establishing that a device or system is responding as it was designed to do. |
| **VDU** | Visual display unit. An input/output peripheral, which has a keyboard for data input and a monitor screen for viewing both the input data and any outputted data. The system usually includes buffer storage facilities so that data may be loaded off-line. Often used to communicate directly with the computer in real time. |

# A3

# Serial data communication

With a wide variety of electronic devices available to perform specific functions there is a need to interconnect the devices so that efficient error-free communication can occur. This appendix will look at the RS-232, RS-422, and RS-485 standards as well as the NMEA 0183 interfacing protocol since they are the ones that are most often used in the marine environment.

## Serial communication

Data in digital form has voltage levels that define a logic 1 and logic 0 level; a binary digit, often abbreviated to the term ‘bit’, will be either 1 or 0. A byte of data is made up of 8 bits while two bytes would comprise 16 bits, etc. If we assume that data comprises a single byte then the data may be sent through a parallel port of a device where all 8 bits would be transmitted simultaneously. The data is transmitted quickly but the required cable is bulky because eight separate wires are required, one for each bit. If a serial port is used then each bit of the data is sent in turn over a single wire. The time taken to transmit the data is eight times longer than the time taken using a parallel port but fewer wires are needed. In fact full duplex (simultaneous transmission in both directions) is possible with just three wires, one for sending, one for reception and a common signal ground. RS-232 is a good example of a full duplex arrangement.

Half duplex devices can transmit data in both directions but not at the same time, i.e. one device transmits while a second device receives; at some other time the direction of transmission can be reversed. RS-485 is an example of a half duplex arrangement while RS-422 can operate in either full duplex or half duplex as required. Simplex transmission (i.e. in one direction only), where one device always transmits and the connected device always receives, would require just two wires.

Serial data can be transmitted in two ways.

- Synchronous. In this arrangement the interconnected devices initially synchronize with each other and continually send characters to keep synchronization even when data is not being transmitted.
- Asynchronous. In this arrangement the sending and receiving devices are not synchronized and each byte of data sent must be identified by a start bit inserted before the data and a stop bit at the end of the data. The extra bits involved means that this type of transmission is slower than the synchronous form although it does not need to transmit idle characters to maintain synchronization.

When transmitting a data byte it is possible to insert an extra bit, known as a parity bit, alongside the data. The logic value of the parity bit can be changed so that the number of data bits sent can be identified as an even or odd number. As an example if even parity is used and the data byte is, say, 00101100 then the parity bit sent would be 1; if the data byte is, say, 01100110 then the parity bit sent would be 0. The converse would be true if odd parity were chosen. The use of a parity bit allows a degree of error-check on received data. Suppose a data byte 00001100 is received together with a parity bit 1, it follows that, using even parity, one of the data bits received is incorrect although it is not specific as to which one. Also if two data bits are incorrect the parity bit would not show any error at all.

Data transmission rates are quoted in bauds, a unit named after the Frenchman Jean Baudot who is credited with devising an original 5-bit code for alphabetical characters in the latter part of the 19th century. The baud rate defines the number of times per second that a line changes state. The baud rate may be the same as the bit rate (i.e., number of bit $s^{-1}$ transmitted) but there may be circumstances where bit rate and baud rate are not the same.

## RS-232 Serial Interface

The original Recommended Standard-232C was approved in 1969 by the Electronic Industries Association (EIA) for interconnecting serial devices. In 1987 the EIA produced a new version of the standard which became the EIA-232D. By 1991 the EIA had joined forces with the Telecommunications Industry Association (TIA) and the standard became known as the EIA/TIA-232E. However, the increasing length of the title was too much for most users and the standard is still commonly known as the RS-232C or as simply the RS-232.

The RS-232 standard specifies the physical interface, together with associated electrical signalling, between serial transmitting/receiving Data Communication Equipment (DCE) and Data Terminal Equipment (DTE). A computer, for example, is a DTE device, as are printers and terminal equipment, while other, remote, devices such as a modem are DCE devices. A typical arrangement is shown in Figure A3.1.

The type of signal is known as single-ended unbalanced because each signal line has a voltage level that is set with respect to signal ground. RS-232 drivers (transmitters) are specified with an output voltage more negative than –5 V for a logic 1 level and more positive than +5 V for a logic 0 level. The defined maximum output voltage of a driver stage on open-circuit is ± 25 V. The RS-232 receivers will interpret a voltage level more negative than –3 V as logic 1 while a voltage level more positive than +3 V is a logic 0. This permits a noise immunity of 2 V for the transmission. If a parallel port is used then a Universal Asynchronous Receiver Transmitter (UART) must be placed between the transmitter (and/or receiver) and the RS-232 interface.

The maximum transmission rate for the standard is defined as 20 kbaud with a cable length not exceeding 15 m; the cable length can be increased for lower baud rates and if shielded cable is used.

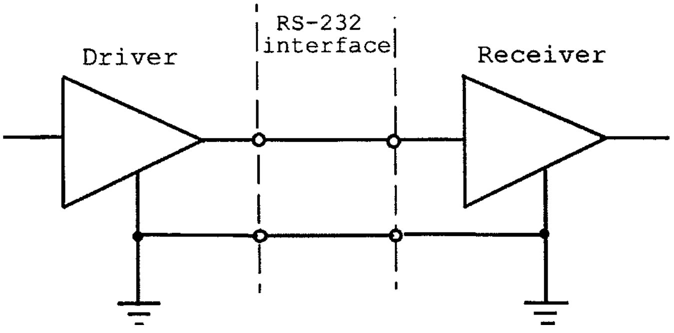


**Figure A3.1** Driver and receiver circuit connected via an RS-232 interface.

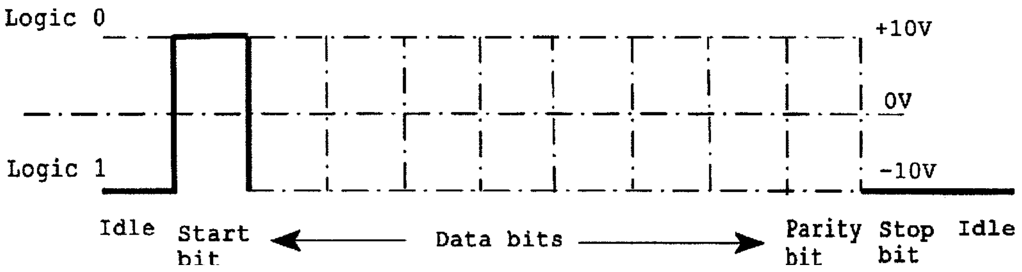


**Figure A3.2** Typical arrangement for the transmission of an ASCII character using the RS-232 standard.

Voltage levels could be in the range ±5 to ±15 V for the loaded driver stage. If a voltage level of ±10 V is assumed and with the data transmitted as an 8-bit group consisting of 7 data bits and a parity bit, the arrangement would be as shown in Figure A3.2. The 8-bit group is framed by a start bit at logic 0 and a stop bit at logic 1. If the group represents an ASCII character then the use of 7 bits can only allow ASCII values up to 127.

The RS-232 standard supports two types of connectors, a 25-pin D-type connector (DB-25) and a 9-pin D-type connector (DB-9). The pin assignments for a DB-25 connector is shown in Table A3.1

**Table A3.1** DB-25 pin assignment

| *Pin* | *Signal* | *Source* | *Key* |
|---|---|---|---|
| 1 | - | - | Frame ground |
| 2 | TD | DTE | Transmitted data |
| 3 | RD | DCE | Received data |
| 4 | RTS | DTE | Request to send |
| 5 | CTS | DCE | Clear to send |
| 6 | DSR | DCE | Data set ready |
| 7 | SG | - | Signal ground |
| 8 | DCD | DCE | Data carrier signal |
| 9 | - | - | Positive voltage |
| 10 | - | - | Negative voltage |
| 11 | - | - | Unassigned |
| 12 | SDCD | DCE | Secondary DCD |
| 13 | SCTS | DCE | Secondary CTS |
| 14 | STD | DTE | Secondary TD |
| 15 | TC | DCE | Transmit clock |
| 16 | SRD | DCE | Secondary RD |
| 17 | RC | DCE | Receive clock |
| 18 | - | - | Unassigned |
| 19 | SRTS | DTE | Secondary RTS |
| 20 | DTR | DTE | Data terminal ready |
| 21 | SQ | DCE | Signal quality detector |
| 22 | RI | DCE | Ring indicator |
| 23 | DRS | DTE/DCE | Data rate selector |
| 24 | SCTE | DTE | Clock transmit external |
| 25 | - | - | Busy |

**Table A3.2** DB-9 pin assignment

| *Pin* | *Signal* | *Key* |
|---|---|---|
| 1 | DCD | Data carrier detect |
| 2 | RD | Received data |
| 3 | TD | Transmitted data |
| 4 | DTR | Data terminal ready |
| 5 | SG | Signal ground |
| 6 | DSR | Data set ready |
| 7 | RTS | Request to send |
| 8 | CTS | Clear to send |
| 9 | RI | Ring indicator |

Typically in many applications only nine of the DB-25 pins are important and the DB-9 connector reflects this as shown in Table A3.2.

Considering the DB-25 connector, signals are carried as single voltages referred to a common earth point SG (pin 7). The TD (pin 2) connection allows data to be transmitted from a DTE device to a DCE device; the line is kept in a mark state by the DTE device when it is idle. The RD (pin 3) connection is the one where data is received by a DTE device; the line is kept in a mark state by the DCE device when idle.

Pins 4 and 5 are the RTS and CTS connections, respectively, and provide handshaking signals. The DTE device puts the RTS line in a mark state when ready to receive data from the DCE; if unable to receive data the DTE puts the line in a space state. For CTS the DCE device puts the line in a mark state to inform the DTE device it is ready to receive data; a space on the line indicates the DCE is unable to receive data.

The DSR/DTR connections (pins 6/20, respectively) are used to provide an indication that the devices are connected and turned on. DCD (pin 8) is used to indicate that the carrier for the transmit data is on. The DCD and RI (pin 22) are only used in connections to a modem. The state of the RI line is toggled by the modem when an incoming call rings the user's telephone.

If the RS-232 link is used to connect devices operating with transistor–transistor logic (TTL) levels then interface integrated circuits (ICs) must be used to convert the TTL logic levels to the RS-232 standard and vice versa.

## RS-422

The use of RS-232 is universal and popular but it does have its limitations. The use of a single line to carry the signal does make it susceptible to noise. Screening the cable can mitigate external noise but will do nothing to stop internal noise. An improved standard introduced by the EIA is the RS-422, which uses a balanced line interface. A pair of lines (Line A and Line B) are used to carry each signal and data is encoded/decoded as a differential voltage between the two lines. See Figure A3.3.

Voltage levels at the driver stage output are typically between 2 and 6 V across the A and B terminals while at the input to the receiver stage the voltage levels are in the range 0.2–6V. The lower threshold voltage is to allow for signal attenuation on the line.

Logically, a '1' ('Mark' or 'off' state) is a voltage on line A which is negative with respect to line B, while a '0' ('Space' or 'on' state) is a voltage on line A which is positive with respect to line B. Using RS-422, up to 10 receivers may be connected to one driver stage.

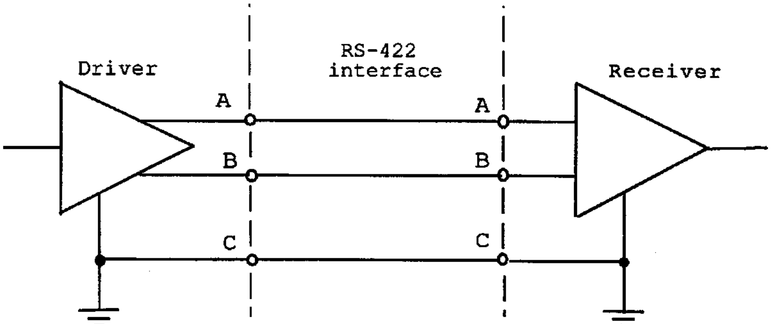


**Figure A3.3** Driver and receiver circuit connected via an RS-422 interface.

Because the voltage is differential, the interface is less likely to be affected by differences in ground voltage between transmitter and receiver. Also if the lines are twisted together the effect of external noise will be the same in each line and hence eliminated. This is known as common-mode rejection. Common-mode signals are defined as the average value of the sum of the voltages on the A and B lines. RS-422 can withstand a common mode voltage of ±7 V.

The use of RS-422 allows higher data rates to be transmitted over longer distances. A maximum length of 1300 m is recommended at 100 kbaud, while for distances up to 13 m it can deliver signals at 10 Mbaud.

## RS-485

This is also a balanced arrangement similar in detail to RS-422. The RS-485 standard allows up to 32 devices to communicate at half duplex on a single pair of wires, with devices up to 1300 m apart at 120 kbaud, in what is known as a multidrop network. Figure A3.4 shows the arrangement.

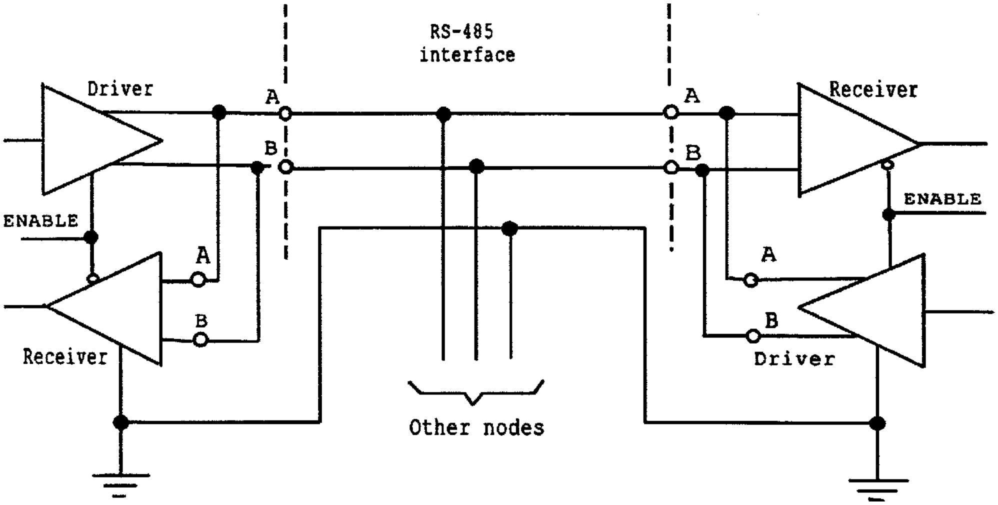


**Figure A3.4** Typical arrangement for an RS-485 two-wire multidrop network.

It can be seen from Figure A3.4 that each device has an 'enable' input. Since only one driver stage can be connected to the line at any time, an 'on' signal on the enable input will connect that driver to the line while all other drivers have an 'off' signal on their enable line. This puts their outputs to the line in a high impedance state, effectively disconnecting them from the line. At the same time the associated receivers will have an on signal on their enable line allowing them to be connected to the line and receive a transmission from the connected drive stage. This change in signalling on the enable line can be achieved using hardware or software techniques. The range of common mode voltage levels that the system can tolerate is increased to +12 V to –7 V. Since the driver can be disconnected from the line it must be able to withstand this common mode voltage level while in the high impedance state.

An alternative wiring arrangement allows full duplex operation by having one 'master' port with the driver connected to each of the 'slave' receivers using one twisted pair. In turn each slave driver is connected to the master receiver using a second twisted pair.

All the above descriptions are of the hardware requirements for particular RS connections. There is also a software requirement that has not been discussed because such a requirement depends on the particular application.

## NMEA interfacing protocols

The National Marine Electronics Association (NMEA) has established standards to be employed by the manufacturers of marine electronic equipment to ensure compatibility when different equipment is fitted together on a ship. The NMEA Standard 0180 was published in late 1980, NMEA 0182 in early 1982, followed by NMEA 0183 which has had several revisions, the latest of which is version 2.30, issued in March 1998. There are differences in transmission parameters between the various NMEA standards which means that NMEA 0183 is not directly compatible with its predecessors.

NMEA 0180 and 0182 standards are concerned with connections between Loran-C receiver and an autopilot using a simple or complex data format. The former consists of a single data byte transmitted at intervals of between 0.8 and 5 s at 1200 baud using a parity bit and bit 7 always set to zero. The complex data format uses a block of data of 37 bytes of ASCII characters transmitted at intervals of 2–8 s with bit 7 always set to one.

## NMEA 0183

This NMEA standard specifies the signal parameters, data communication protocol and timing together with sentence formats for serial data bus transmission rates of 4800 baud. The serial data communication between equipments is unidirectional with one 'talker' and possibly many 'listeners'. The data uses ASCII format and typically a message might contain between 11 and 79 characters in length and require transmission at a rate no greater than once every second.

The arrangement for interconnecting the 'talker' to the many 'listeners' requires just two wires (classified as signal lines 'A' and 'B') and a shield. The 'A' line of the talker should be connected in parallel to the 'A' lines of every listener, and similarly each listener 'B' line is connected in parallel to the talker 'B' line. The listener shield connections should be made to the talker chassis but not to each other.

The talker signal is required to be similar in form to that shown in Figure A3.2 but there are eight data bits and no parity bit. The talker device must have its drive capability defined in order to establish the possible number of listener devices it can drive. Each listener device should contain an optoisolator and protective circuit which limits current, reverse bias and power dissipation at the point of optical coupling.

The standard defines the logic 1 state in the range –15 V to +0.5 V while the logic 0 state is in the range +4–15 V, while sourcing is not more than 15 mA. The receiver circuit should have a minimum differential input voltage of 2.0 V and should not draw more than 2.0 mA from the line under those conditions. The voltage conditions on the data bus should be in accordance with the RS-422 specification.

As described for Figure A3.2, the data bits use the 7-bit ASCII format and for this standard the data bits d0–d6 will contain the ASCII code, while data bit d7 is always set to 0. The ASCII character set consists of all printable ASCII characters in the range 20h–7Eh except for those characters reserved for specific formatting purposes. The individual characters define units of measure, indicate the type of data field, type of sentence etc. A sentence always starts with the character '$' followed by an address field, a number of data fields, a checksum, and finishes with carriage return/line feed.

A field consists of a string of valid characters located between two appropriate delimiter characters. An address field is the first field in a sentence and follows the $ delimiter. The types of address field include the following.

- Approved address field. This consists of five digits and upper-case letter characters. The first two characters are the talker identifier. The following three characters are used to define the format and type of data.
- Query address field. This consists of five characters and is used to request transmission of a specific sentence on a separate bus from an identified talker. The first two characters represent the talker identifier of the device requesting data , the next two characters represent the talker identifier of the device being addressed, while the final character is the query character Q.
- Propriety address field. This consists of the character 'P' followed by a three-character manufacturer's mnemonic code, used to identify the talker issuing a propriety sentence.

Other fields include the following.

- Data fields. These are contained within the field delimiters ','. Data field may be alpha, numeric, alphanumeric, variable or fixed length or constant, with a value determined by a specific sentence definition.
- Null fields. This is a field where no characters are transmitted and is used where the value is unavailable or unreliable.
- Checksum field. This will always be sent and is the last field in a sentence and follows the checksum delimiter character '*'. The checksum is the 8-bit Exclusive-OR (XOR) of all characters in the sentence including the '$' and '*' delimiters. The hexadecimal value of the most significant and least significant 4 bits of the result is converted to two ASCII characters (0–9, A–F(upper case)) for transmission with the most significant character transmitted first.

Sentences may have a maximum number of 82 characters which consists of the maximum 79 characters between the starting delimiter '$' and the terminating <CR><LF>. The minimum number of fields in a sentence is one. The first field shall be the address field, which identifies the talker and the sentence formatter, which specifies the number of data fields in the sentence, the type of data within them and the order in which they are sent. The maximum number of fields in a sentence is limited only by the maximum length of 82 characters. Null fields may be present in a sentence and should always be used if data for that field is unavailable. A talker sentence contains the following elements in the order shown:

$aaccc,df1,df2,df3*hh<CR><LF>

where

| | |
|---|---|
| $ | is the start of the sentence, |
| aa | are alphanumeric characters which identify the talker, |
| ccc | are alphanumeric characters identifying the sentence formatter which gives the data type and string format of following fields |
| , | is the field delimiter which is present at the start of all fields except the address and checksum fields. The field delimiter will still be present even if a null field is transmitted, |
| df1/2/3 | represent the data fields which contain all data to be transmitted. The data field sequence is fixed and is identified by the ‘ccc’ characters in the address field. Data fields may be of variable length, |
| * | is the checksum delimiter which follows the last data field. The two characters following represent the hex value of the checksum, |
| hh | is the checksum field, |
| <CR><LF> | is the end of the sentence. |

An example of a talker sentence is given for a rudder order output message:

$AGROR,uxx.x*hh<CR><LF>

where:

| | |
|---|---|
| AG | is a general autopilot, |
| ROR | is autopilot rudder order, |
| u | is sign, negative for left order, omitted for right or zero order, |
| xx.x | is automatic rudder order up to 45.0°, empty if unavailable. The field here is for a variable number and the use of a decimal point gives a value to one decimal place, |
| hh | is ASCII hex 8-bit XOR of characters after $ through to the letter before ‘*’, |
| <CR><LF> | is the end of sentence marker. |

Hence, if sentence reads:

$AGROR,-10.2*hh

it indicates an automatic rudder order of 10.2° left.

A ‘query’ sentence is used when a listener device requests information from a talker. As an example a query message could be transmitted to a GPS receiver to request ‘distance to waypoint’ data to be sent. The general form of a query sentence is:

$aaaaQ,ccc*hh<CR><LF>

where the first two characters after the ‘$’ start symbol represent the talker identifier of the request. The next two characters represent the talker identifier of the device from which data is requested. ‘Q’ identifies that the message is a query and ‘ccc’ contains the approved sentence formatter for data being requested. An example could be:

$CCGPQ,GGA*hh

where the computer (CC) is requesting the GPS receiver (GP) to send data using the mnemonic GGA which represents global positioning system fix data. Such data would then be transmitted at 1 s intervals.

A ‘proprietary’ sentence may be used by a manufacturer to transfer data which, although using the sentence structure of the standard, does not come within the scope of approved sentences. The general form of the proprietary sentence is:

$Paaa,df1,df2*hh<CR><LF>

where ‘P’ indicates a proprietary message and ‘aaa’ is the manufacturers code, i.e. FUR for Furuno, SMI for Sperry Marine Inc. etc. ‘df1,df2’ represents manufacturer’s data fields that must still conform to the valid character set of the standard.

Details of characters used for data content, talker identifier mnemonics, approved sentence formatters for data fields, field types and manufacturer’s mnemonic code identifiers are too numerous to list here. Some of the detail can be found in those chapters relating to equipment where the NMEA standard is used. Also manufacturer’s manuals should contain references where applicable.

## NMEA 2000

The NMEA has established a working group to develop a new standard for data communication between shipborne electronic equipment. The working group will liaise with the International Standards Organization (ISO), the International Electrotechnical Commission (IEC) and the International Maritime Organization (IMO) to develop a new standard, NMEA 2000, to meet the needs of ships in the 21st century.

NMEA 2000 is expected to be a bi-directional, multi-transmitter, multi-receiver serial data network with the ability to share commands, status and other data with compatible equipment over a single channel link. The capacity of the new system is expected to be much greater than the current NMEA 0183 standard and testing has already begun with a few manufacturers participating in trials. It is anticipated that NMEA 2000 should be available by the middle of 2001.

# A4

# The United States Coast Guard Navigation Center (NAVCEN)

NAVCEN provides quality navigation services that promote safe transportation, support the commerce of the United States and directly benefit worldwide international trade. As a centre of excellence, NAVCEN is proud to be at the forefront of US transportation and navigation initiatives, leading the nation and the international maritime communities into the 21st century.

## Radionavigation and information services

NAVCEN controls and manages Coast Guard radionavigation systems from two sites: Alexandria, Virginia, and Petaluma, California. NAVCEN provides worldwide users with reliable navigation signals, timely operational status, general navigation and other information.

## GPS

NAVCEN gives access to a massive amount of information on GPS. The NAVCEN website lists the following GPS data files.

- Press releases
- Status messages
- Active Nanus
- YUMA Almanacs
- SEM Almanacs

## DGPS

NAVCEN operates the DGPS service, consisting of two control centres and more than 50 remote broadcast sites. The DGPS service broadcasts correction signals on marine radiobeacon frequencies to improve the accuracy and integrity of the GPS (see Chapter 5).

## LORAN-C

Atlantic and Pacific LORAN-C user notifications and system health information is listed on the NAVCEN site (see Chapter 4].

## Other services

Other files of interest to navigators on the NAVCEN site are:

- RNAV radio frequency spectrum issues
- local Notices to Mariners
- maritime telecommunications
- Federal Radio navigation plan.

## Contact

The easiest way to contact NAVCEN is via the web. The primary site is http://www.navcen.uscg.mil. If you do not have access to the net, NAVCEN's mailing address is: The Commanding Officer, USCG NAVCEN, 7323 Telegraph Road, Alexandria VA22315.

Below is a full list of services and contact numbers.

**Table A.4.1**

| *Service* | *Availability* | *Info type* | *Contact no.* |
|---|---|---|---|
| NIS watchstander | 24 hours a day | User inquires | Phone (703) 313–5900 Fax (703) 313–5920 |
| Internet | 24 hours a day | Status Fore/Hist/ Outrages/NGS Data/ Omega/FRP and Misc.info | http://www.navcen.uscg.mil<br>ftp://ftppp. navcen.uscg.mil |
| Internet Mirror Site | 24 hours a day | Status GPS/DGPS Outrages/ | http://www.nis-mirror.com |
| NIS Voice Tape Recording | 24 hours a day | Status forecasts historic | (703) 313–5907–GPS |
| WWV | Minutes 14 & 15 | Status forecasts | 2.5, 5, 10,15, and 20 MHz |
| WWVH | Minutes 43 & 44 | Status forecasts | 2.5, 5, 10, and 15 MHz |
| USCG MIB | When broadcast | Status forecasts | VHF Radio marine band |
| NIMA Broadcast Warnings | When broadcast received | Status forecasts | |
| NIMA Weekly Notice to Mariners | Published & mailed weekly | Status forecast outrages | (301) 227–3126 |
| NAVTEX Data Broadcast | All Stations Broadcast 6 times daily at alternating times | Status forecast outrages | 518 kHz |

# Index