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.

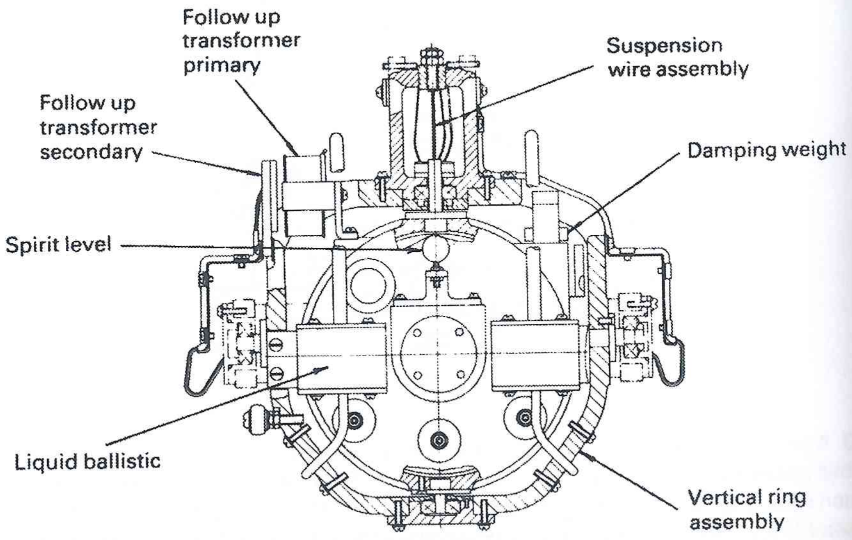


Figure 8.24 The compass sensitive element.

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.

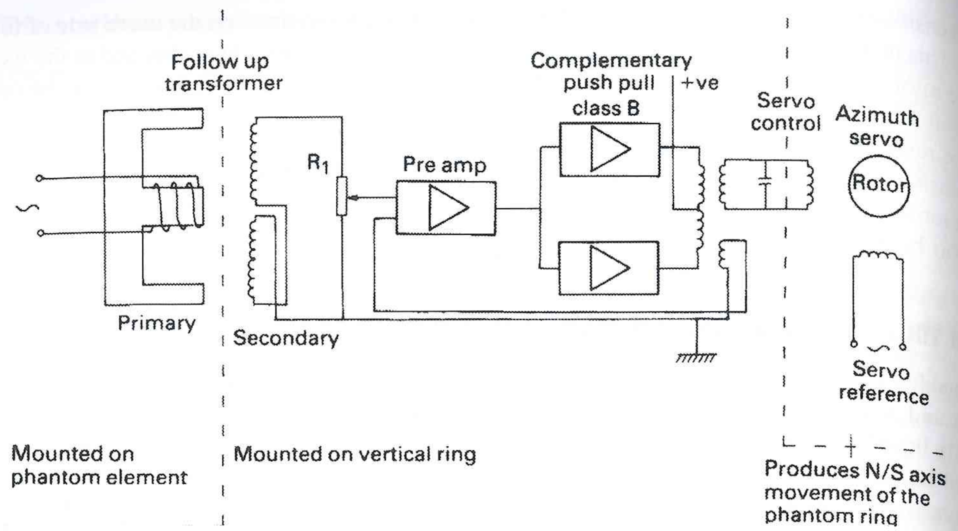


Figure 8.25 The Sperry compass azimuth follow-up circuit.

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 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.

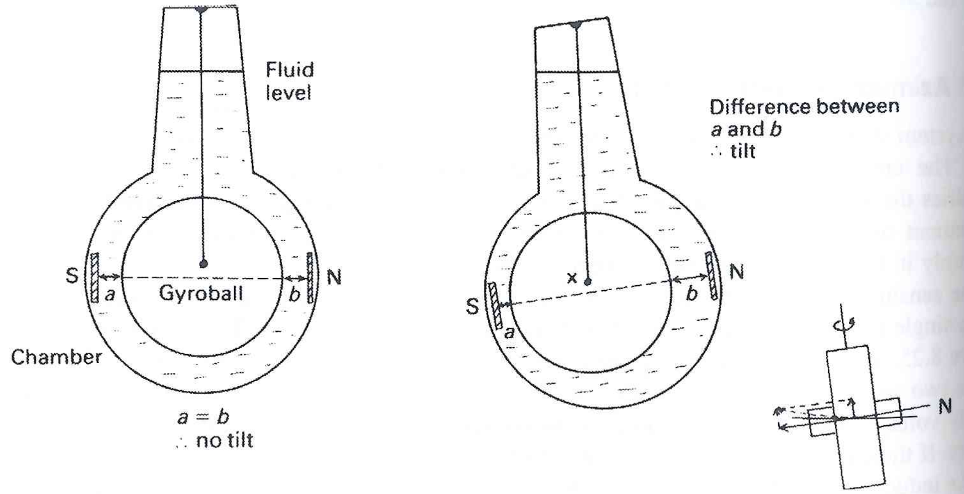


Figure 8.26 Simplified diagrams of the gyroball action in the Sperry SR220 gyrocompass.

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; ϕ = a signal corresponding to vertical movement; and θ = 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

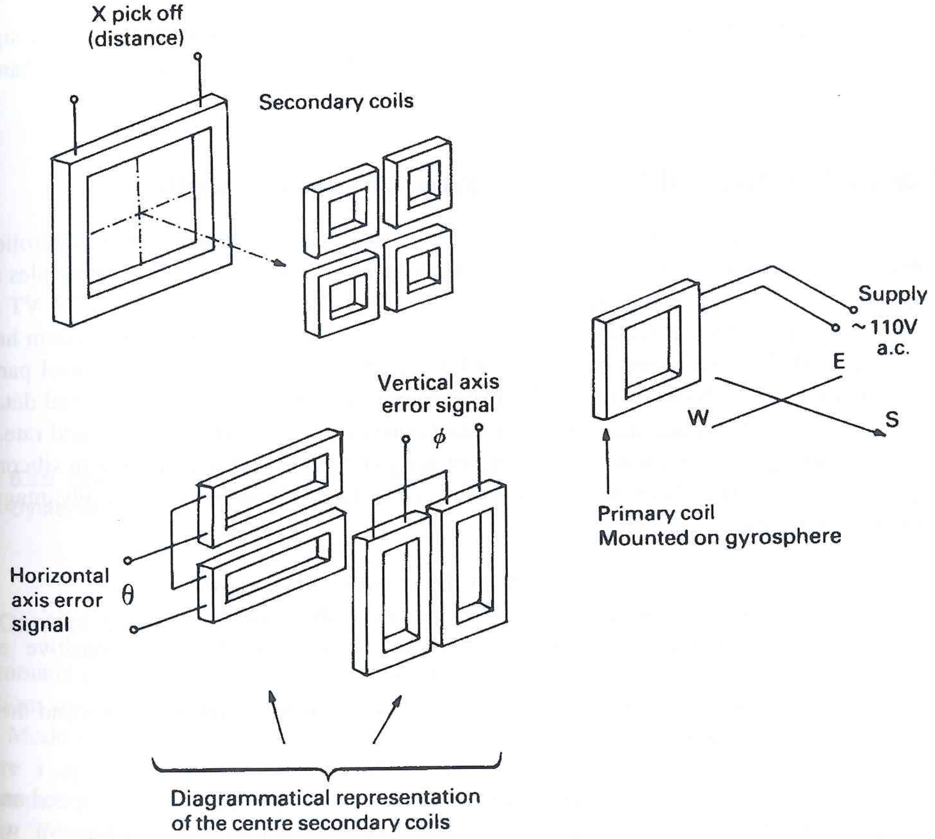


Figure 8.27 Follow-up signal pick-off coils.

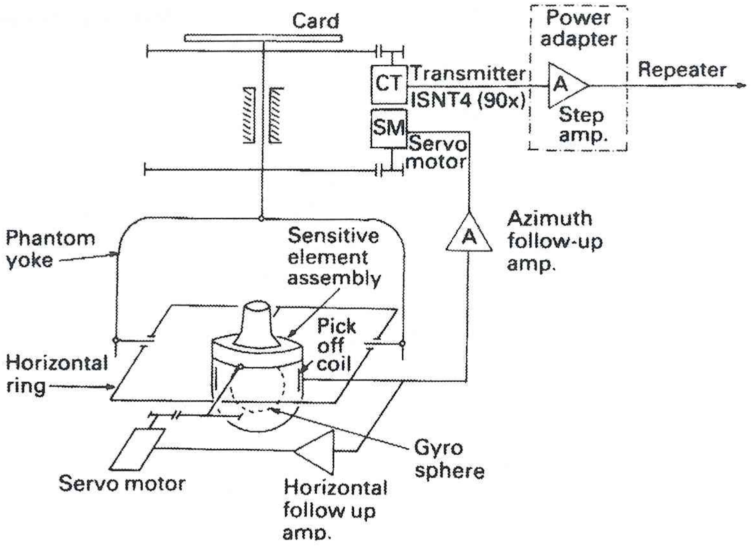


Figure 8.28 The Sperry SR220 follow-up system.

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.

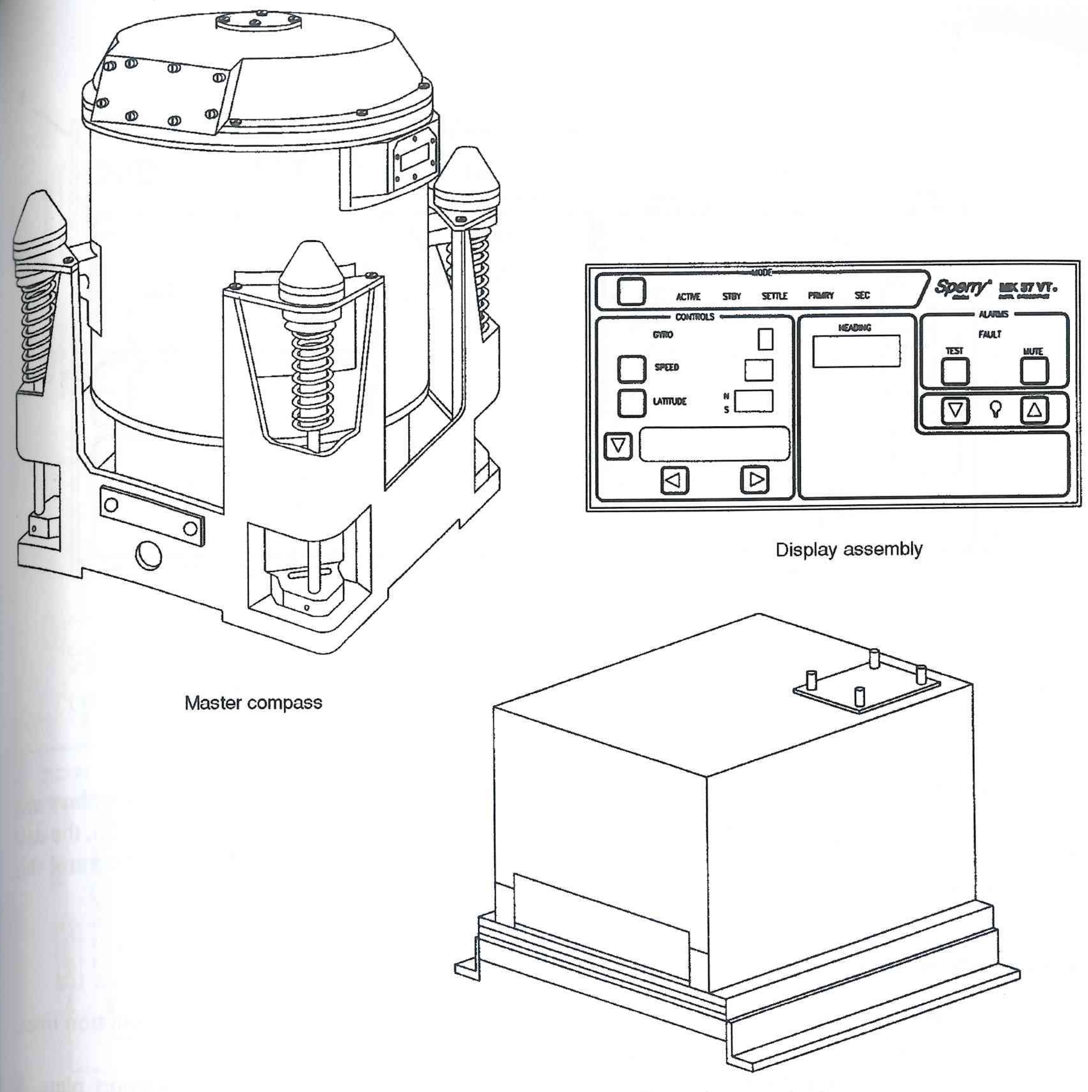
8.9 A digital controlled top-heavy gyrocompass system

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.

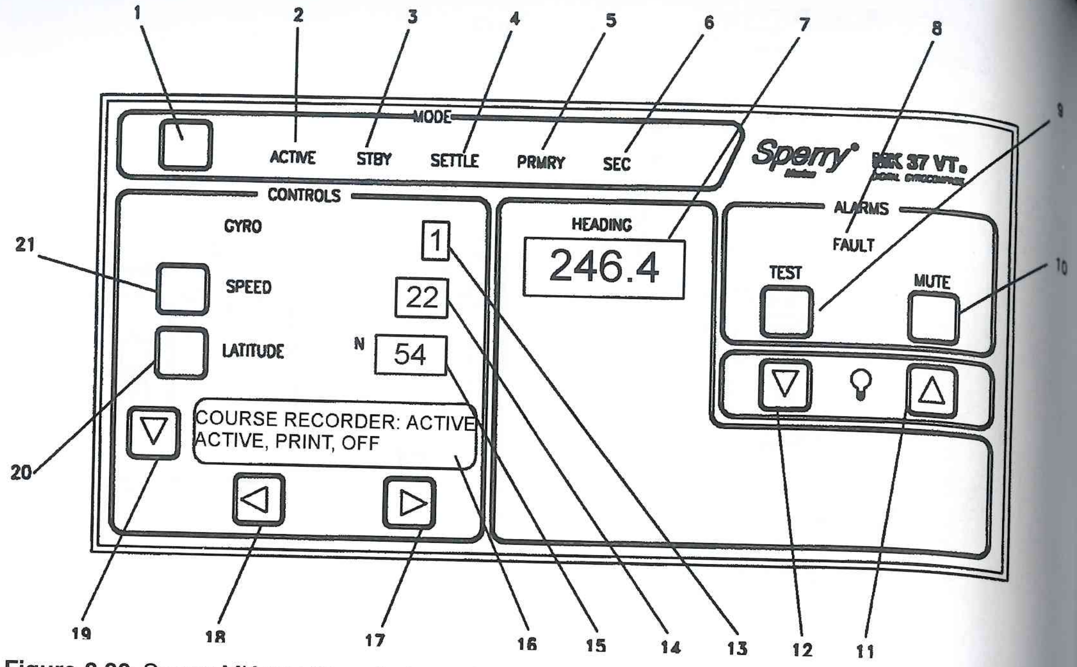


Figure 8.30 Sperry MK 37 VT control panel. (Reproduced courtesy of Litton Marine Systems.)

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.

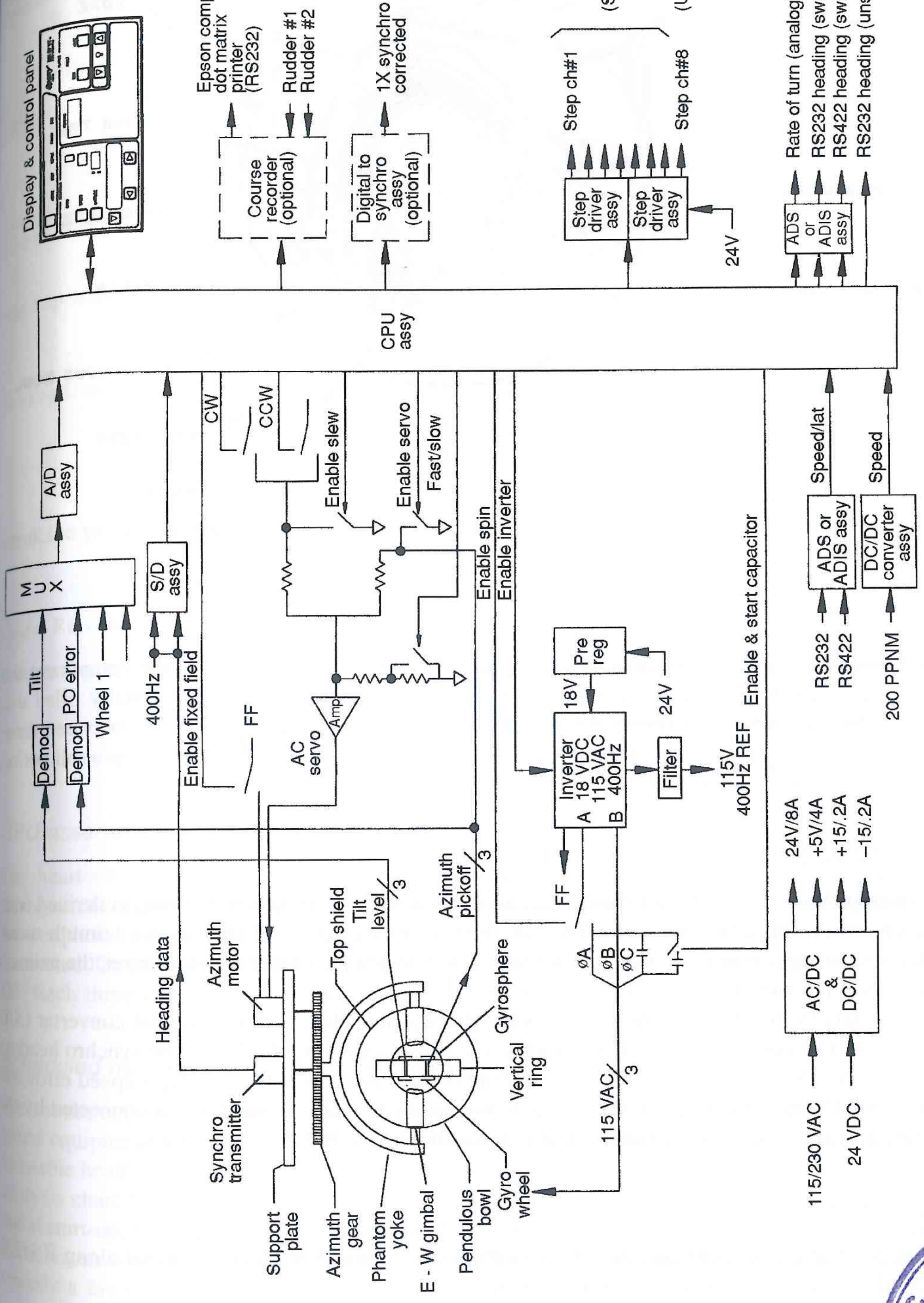


Figure 8.31 Overall functional block diagram. (Reproduced courtesy of Litton Marine Systems.)

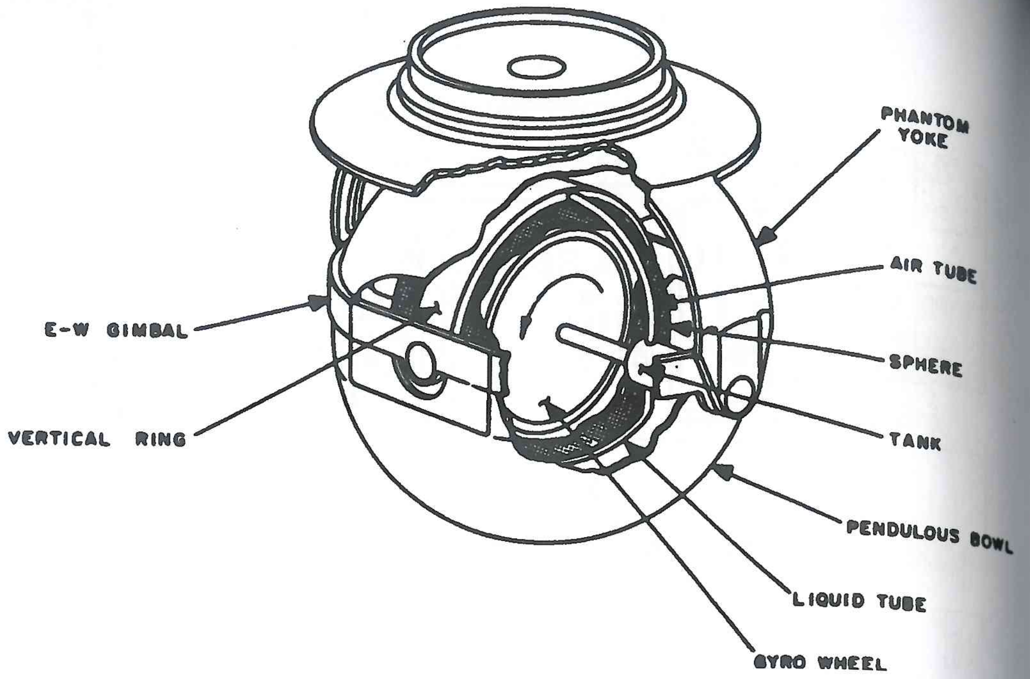


Figure 8.32 Ballistic system of the Sperry MK 37 VT gyrocompass. (Reproduced courtesy of Litton Marine Systems.)

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 Automatic from digital sources via RS-232/422 in NMEA 0183 format $GLL 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) 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> Two RS-232, each capable of driving one load in NMEA 0183 format $HEHDT, XXX>XXX, T*hh<CR><LF> 10 Hz, 4800 baud 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. 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

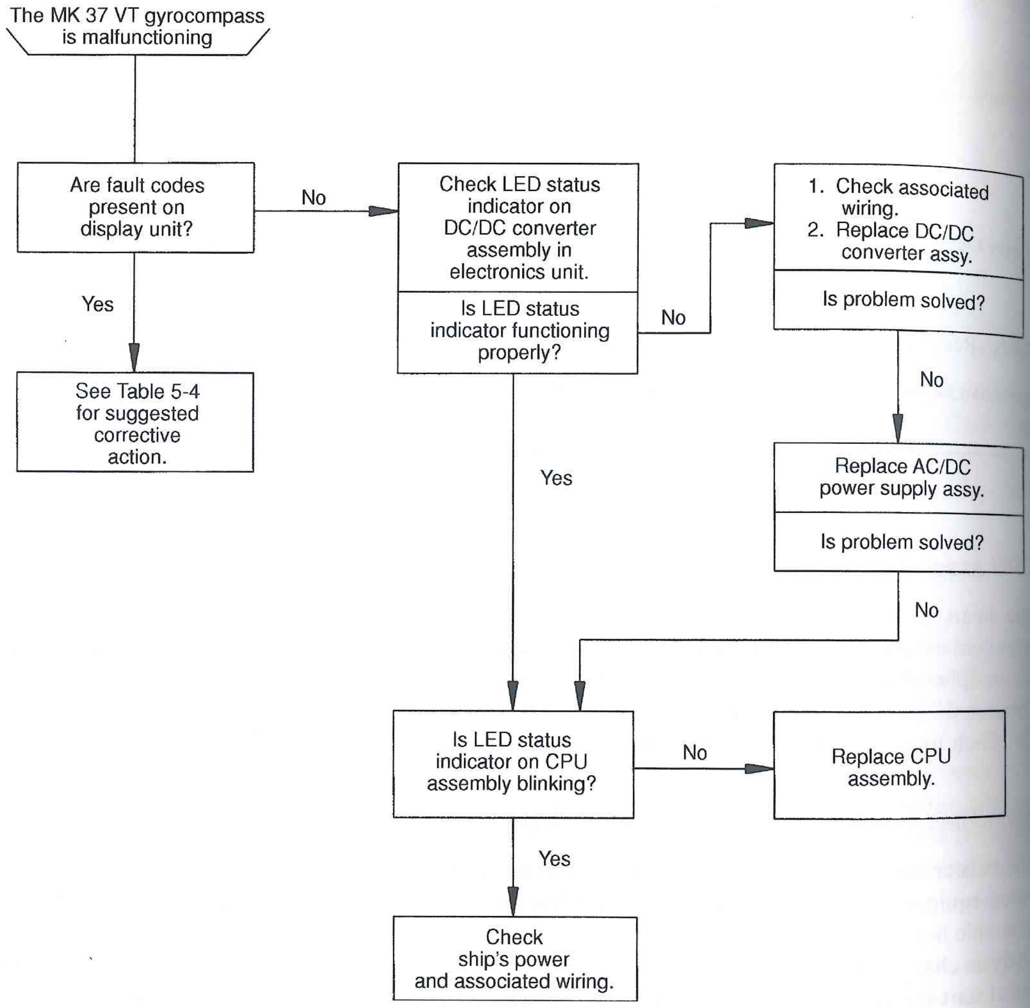


Figure 8.33 Sperry MK 37 VT digital gyrocompass trouble analysis chart. (Reproduced courtesy of Litton Marine Systems.)

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

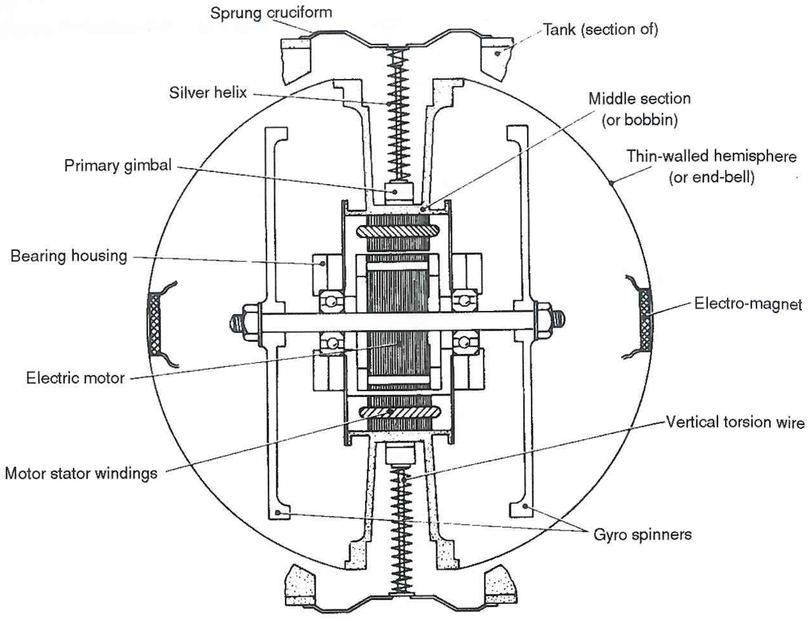
As an example, Table 8.3 shows part of the MK 37 VT gyrocompasses extensive fault diagnosis table. Using this and the data displayed on the main display unit, it is possible to isolate the area of a malfunction.

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.

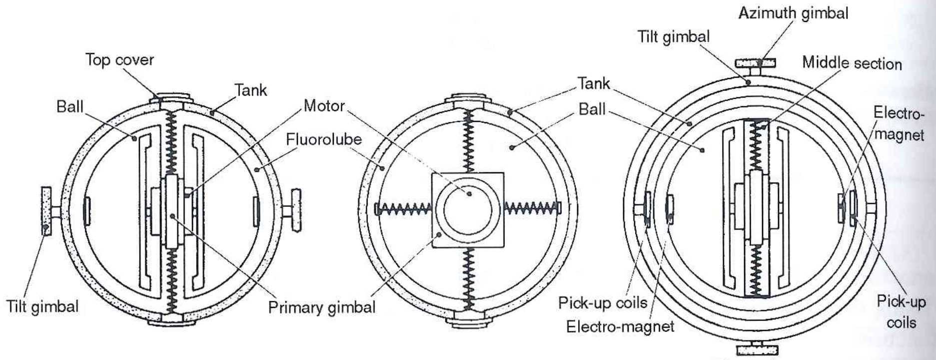
8.10 A bottom-heavy control gyrocompass

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 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.

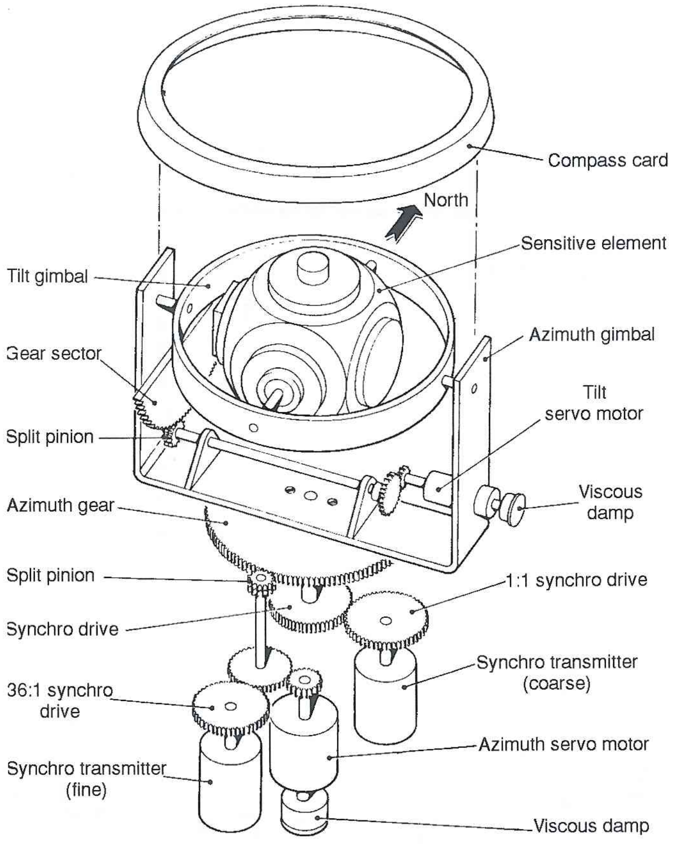


Figure 8.35 Schematics showing the arrangement of the secondary gimbals.

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

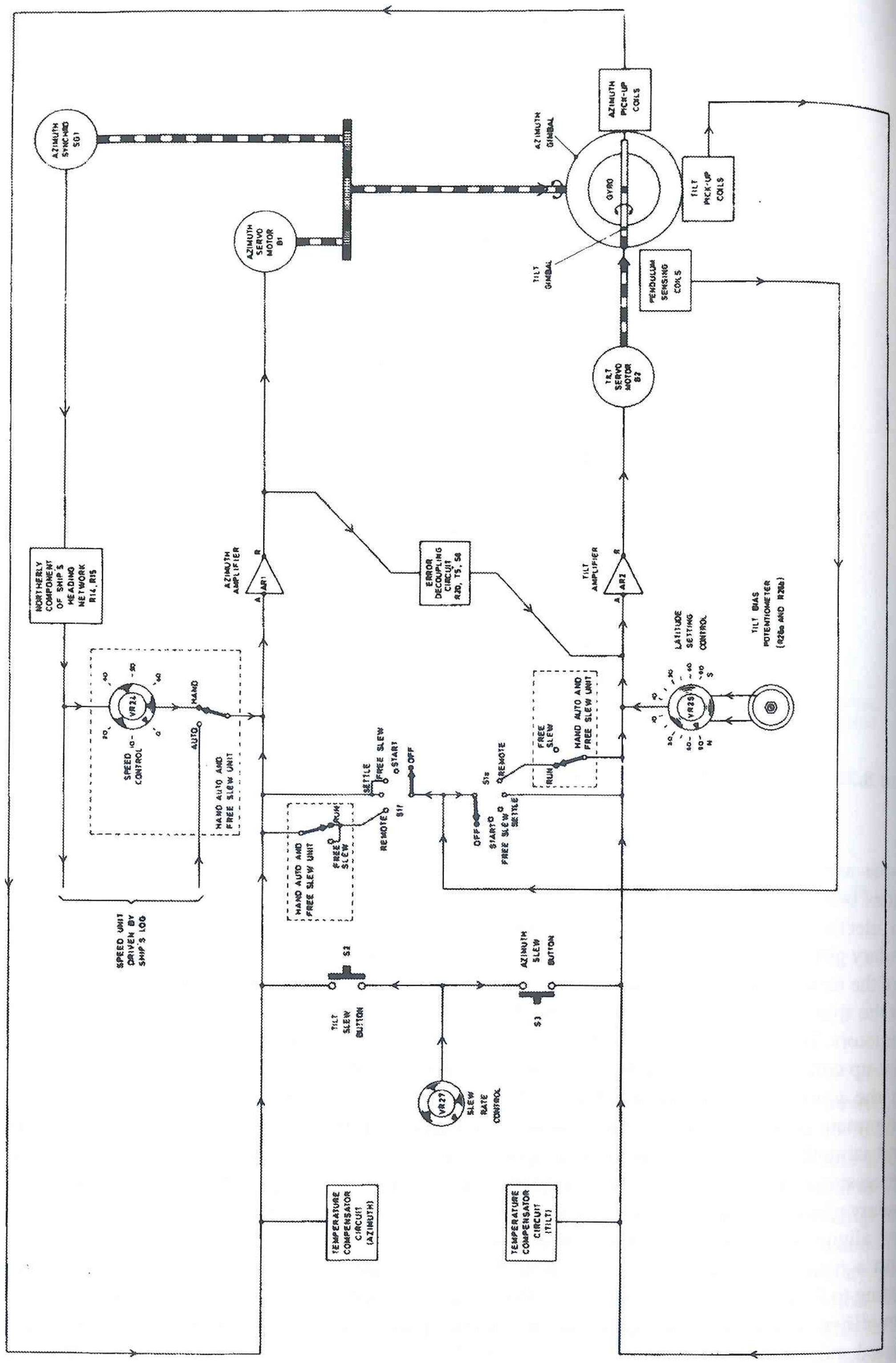


Figure 8.36 Compass circuits schematic. (Reproduced courtesy of S.G. Brown Ltd.)

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.

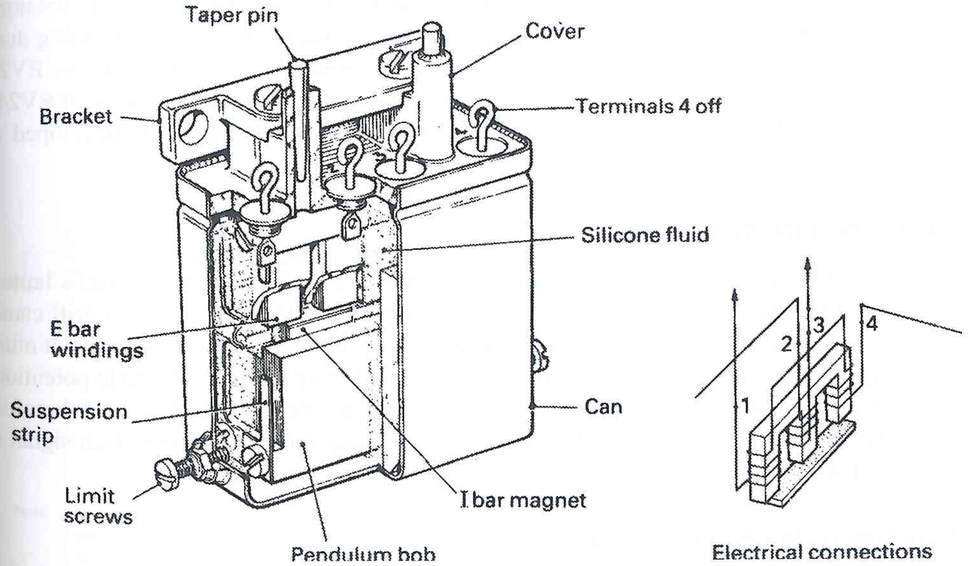


Figure 8.37 The pendulum assembly and its electrical connections. (Reproduced courtesy S. G. Brown Ltd.)

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.

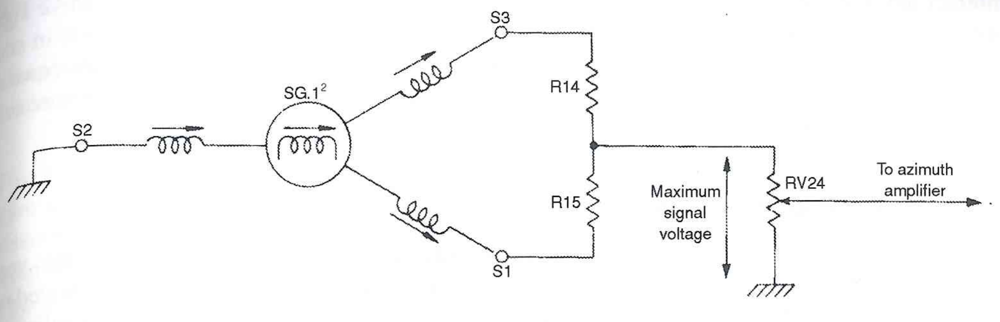
Alternatively an external signal derived from the ship's speed log may be used. In Figure 8.38 the error for a ship sailing due north is maximum and therefore the feedback signal produced across RV24, by the currents flowing through SG1, S1 and S2 coils, will be maximum. A portion of this signal, dependent 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

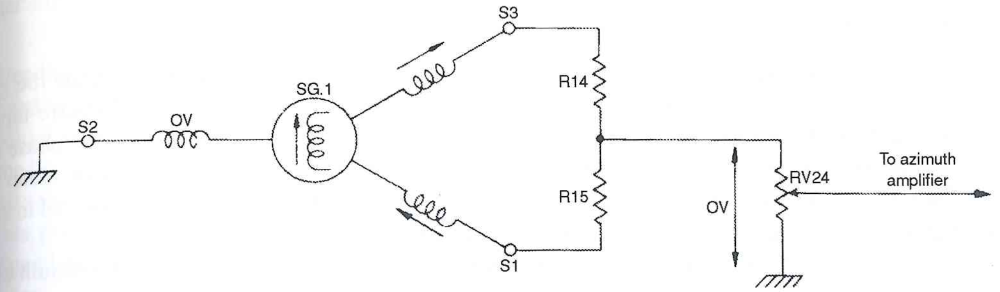
The latitude correction circuit provides a signal, proportional to the sine of the vessel's latitude, to cause the gyroball to precess in azimuth at a rate equal and opposite to the apparent drift caused by the rotation of the earth. This signal will be zero at the equator and maximum at the poles. It 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

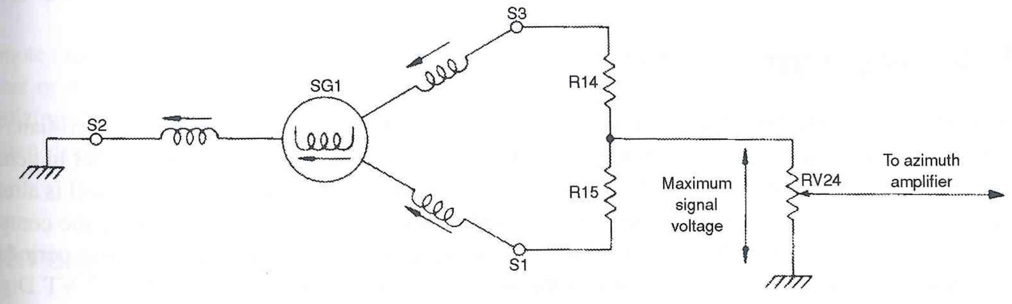
Both the vertical and horizontal torsion wires may twist with a change in ambient temperature. A corrective signal is produced in each of the tilt and azimuth temperature compensation circuits to



(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.

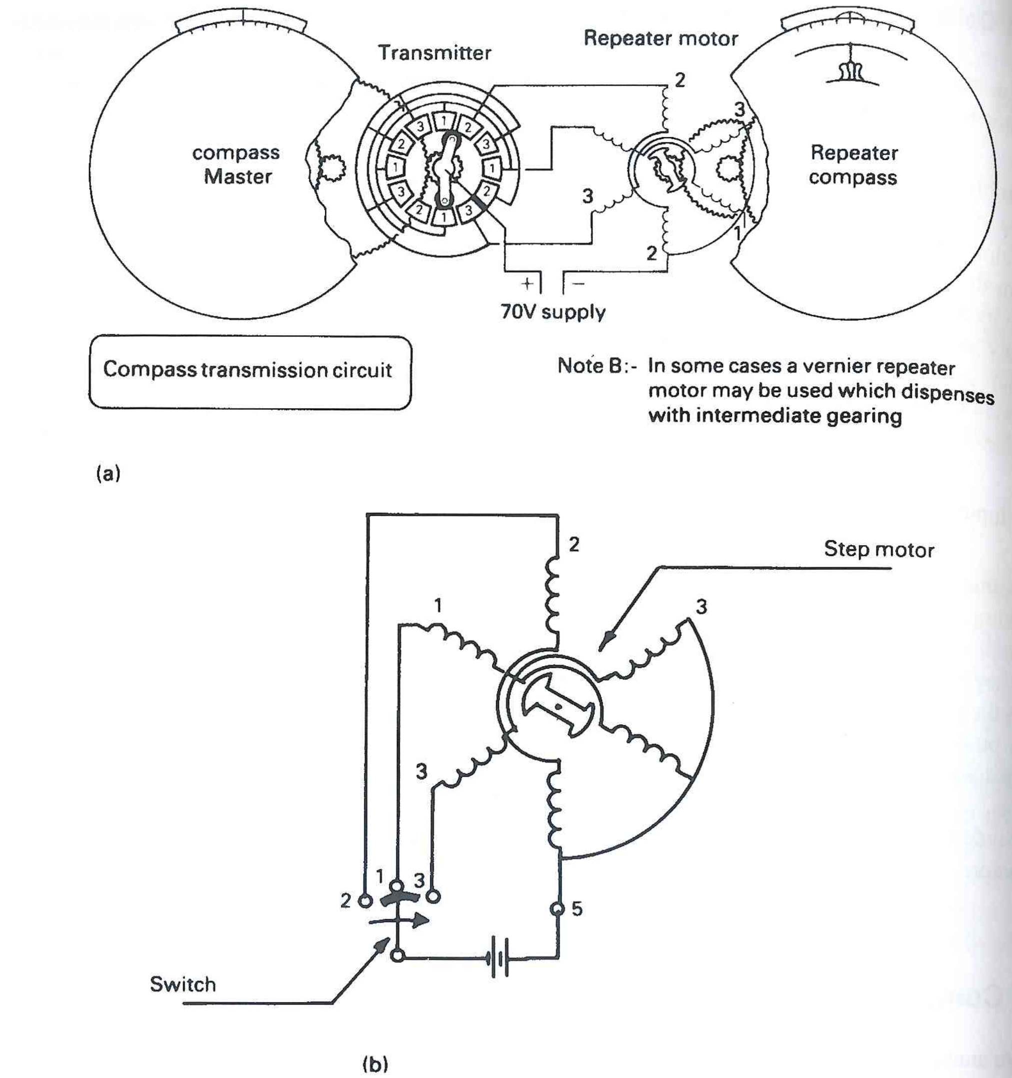
- Heading: in the range 0 to 359°. If the entered heading is in error by more than 20° from the true heading, the compass takes 5 h to settle.
- Initialize and Synchronize Step Repeaters. An operator selects a repeater and when requested uses the keypad's left or right arrow switches to scroll the display 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.



Note B:- In some cases a vernier repeater motor may be used which dispenses with intermediate gearing

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.

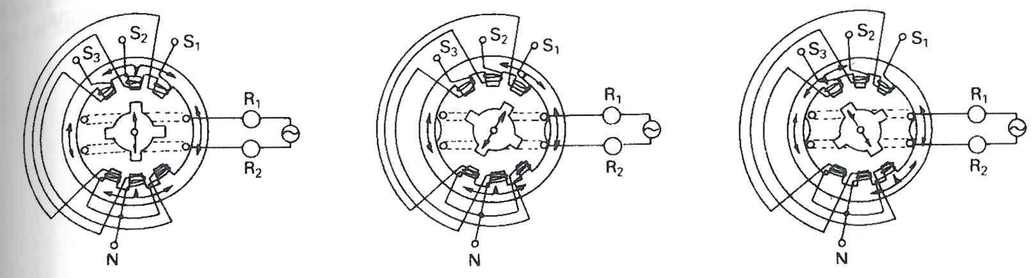


Figure 8.40 Construction details of a step motor.

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 (ψ) 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

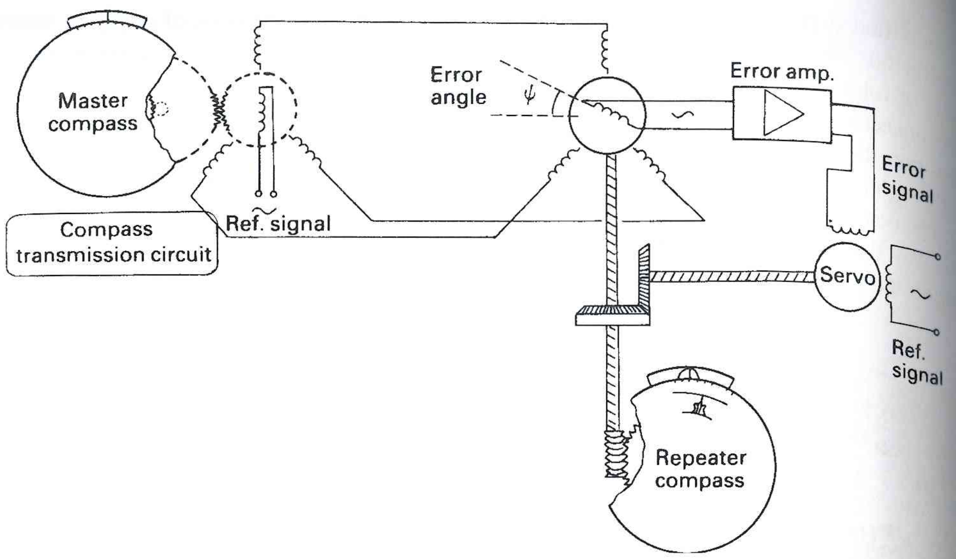


Figure 8.41 A synchro bearing transmission system .

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 2*f*.

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

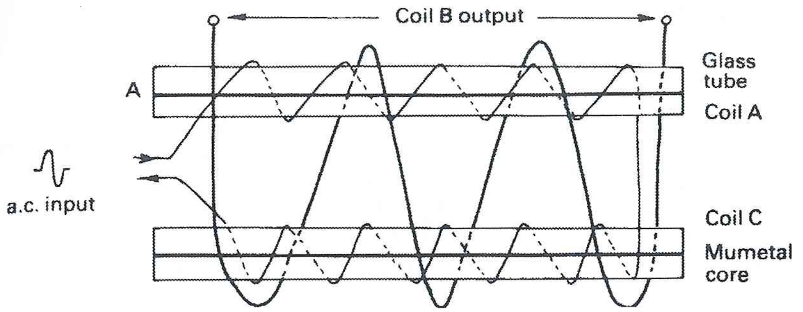


Figure 8.42 A basic flux gate showing the primary windings of equal turns around tubes A and C and a secondary coil wound around the whole assembly.

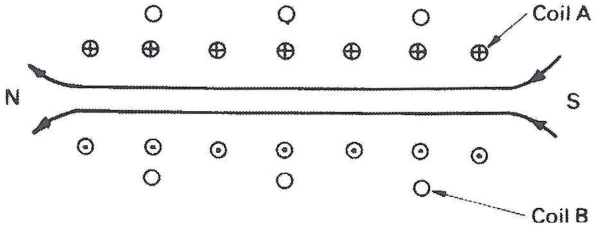


Figure 8.43 A cross-section of part of a flux gate. Current flowing in coil A is 'into the diagram' on the top half of the winding and 'out' on the bottom.

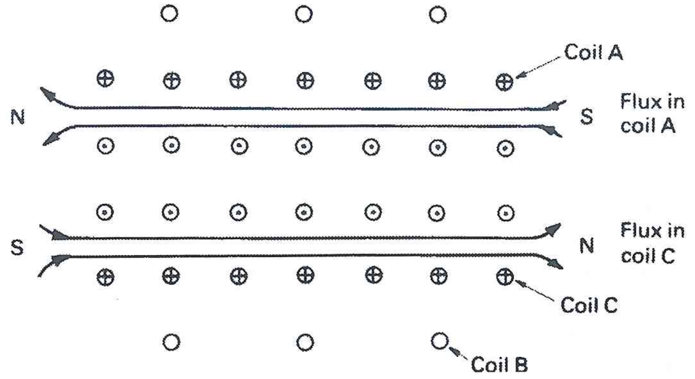


Figure 8.44 A cross-section of a completed flux gate.

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

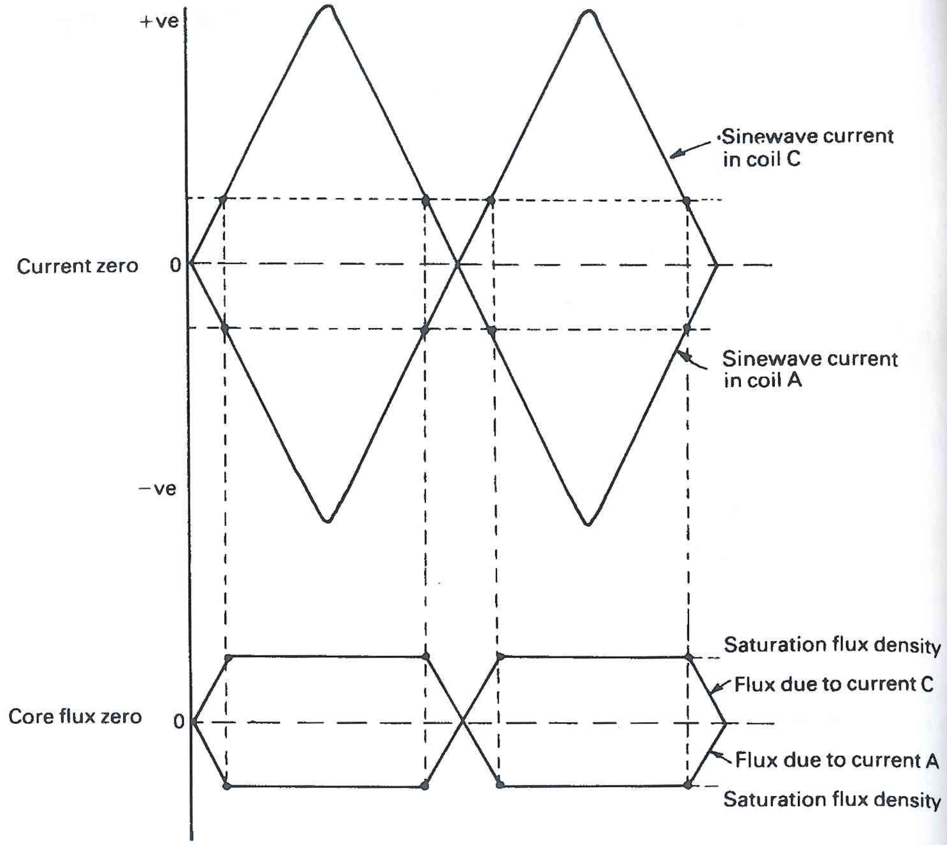


Figure 8.45 Currents and flux saturation levels.

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.

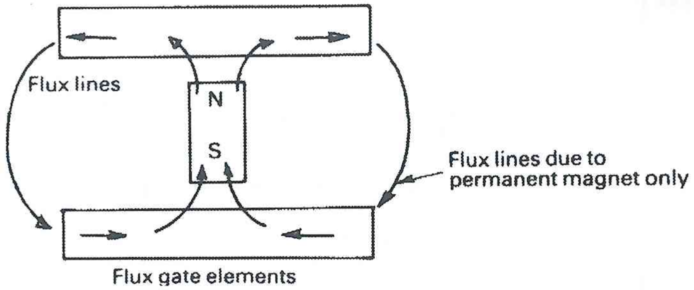


Figure 8.46 Flux lines due to the addition of a permanent magnet to the flux gate.

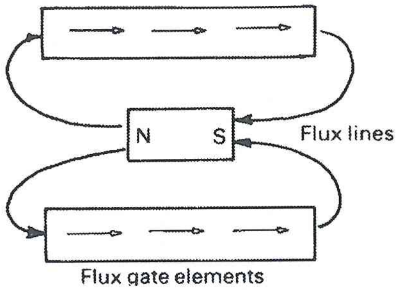


Figure 8.47 Flux lines with the permanent magnet in line with the flux gate.

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.

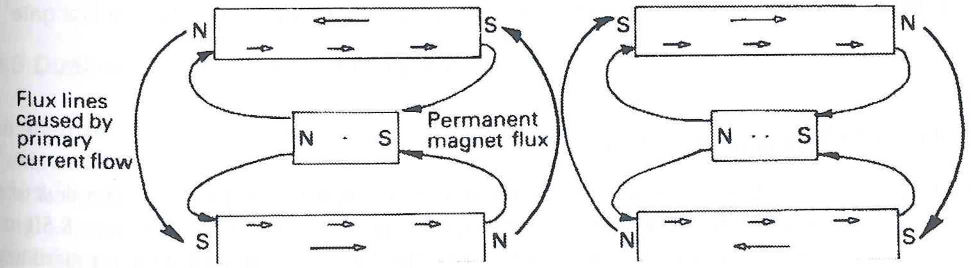


Figure 8.48 The intensity of the magnetic flux in each core is changed on each half cycle of primary alternating current.

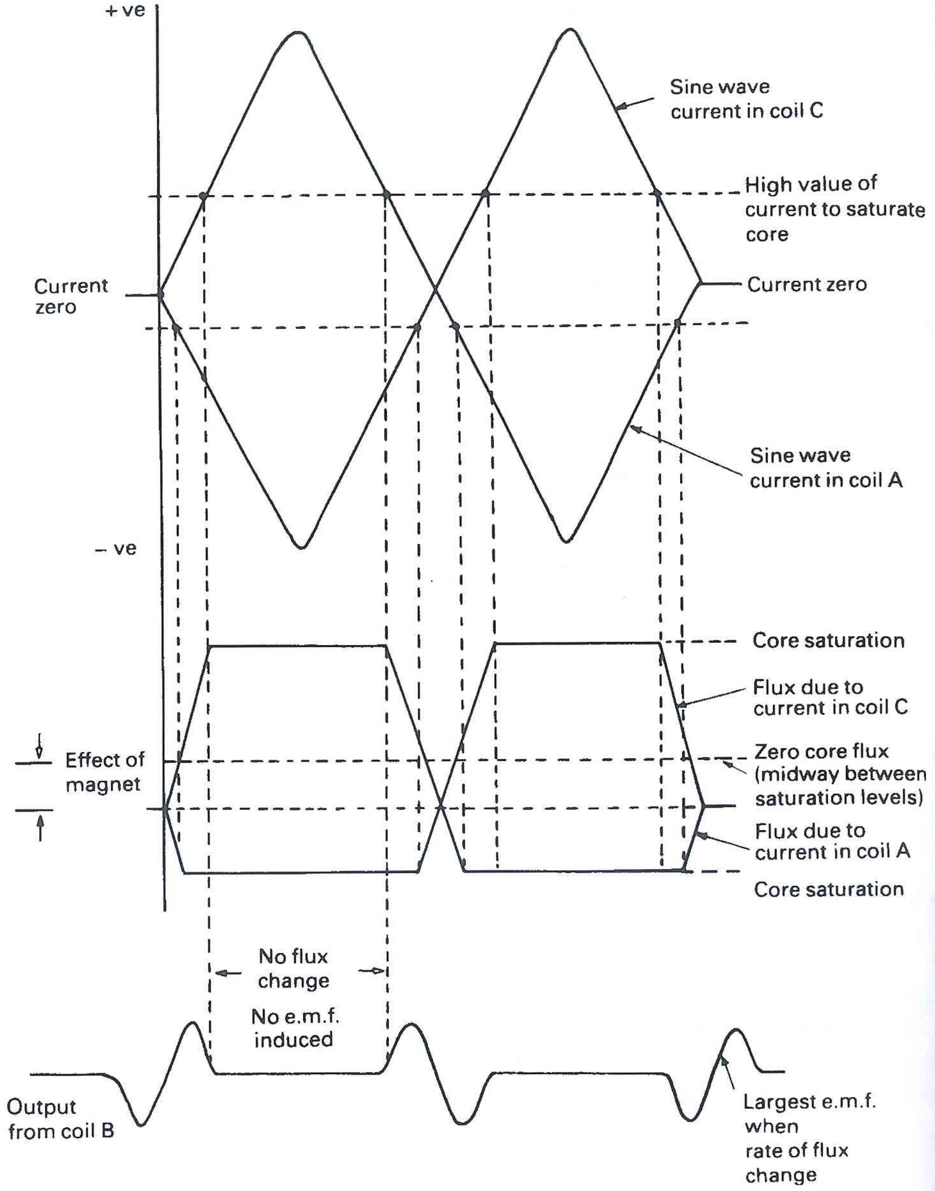


Figure 8.49 An illustration of the fluxes and output e.m.f. produced by an unbalanced flux gate assembly.

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 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.

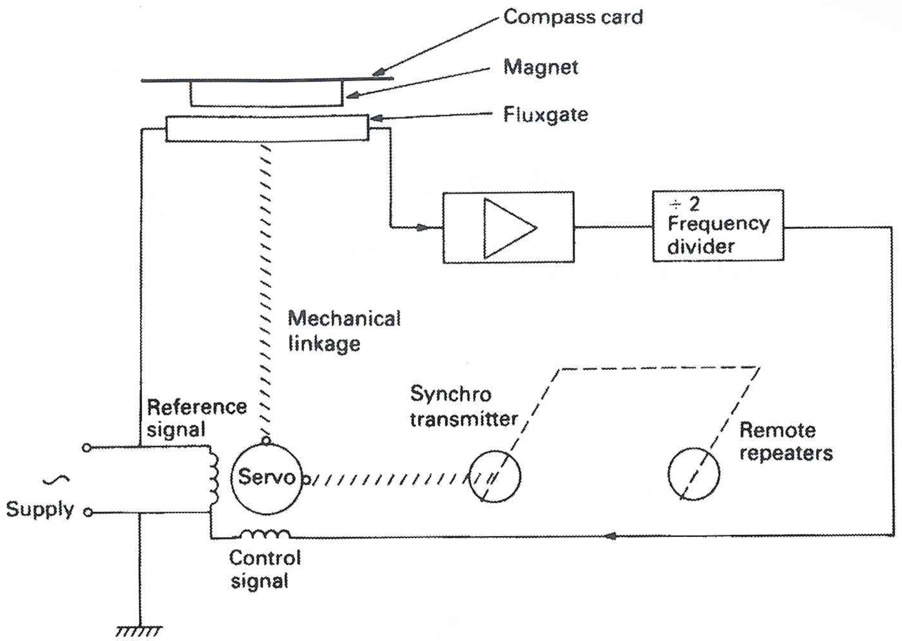


Figure 8.50 A flux gate system used in conjunction with a magnetic compass.

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

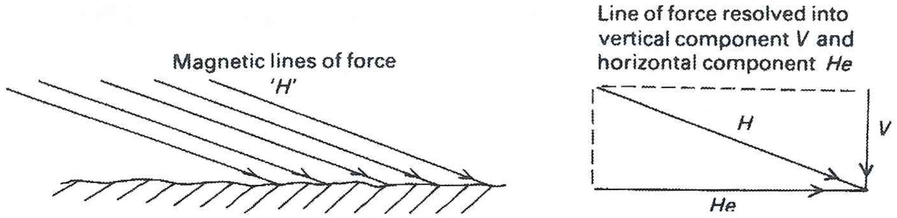


Figure 8.51 An illustration showing how the lines of force of the earth's magnetic field may be resolved into vertical and horizontal components.

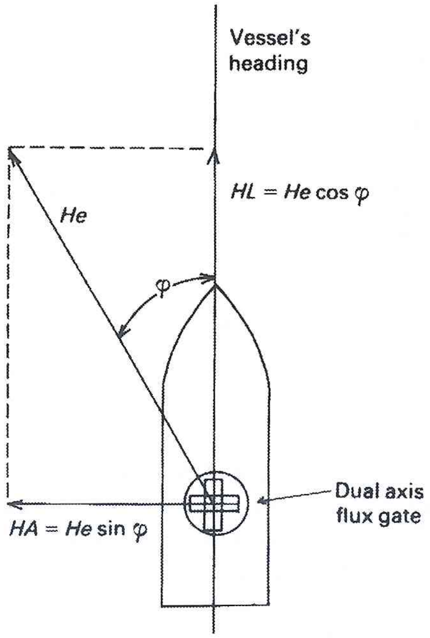


Figure 8.52 The vessel's course shown as a cosine function of *He* – the direction of the earth's magnetic field.

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, 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.

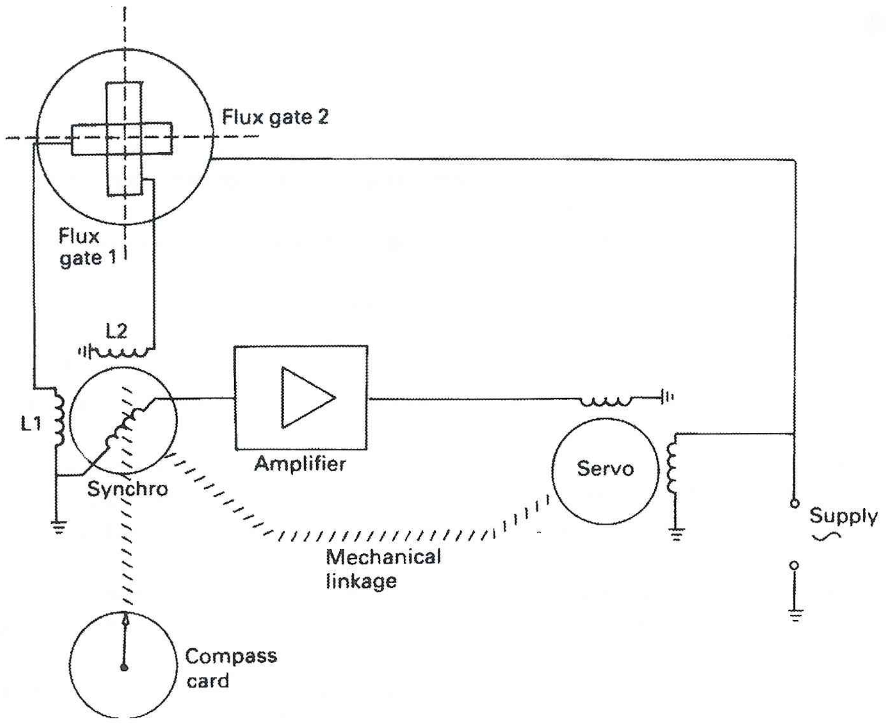


Figure 8.53 A simplified diagram of a dual axis magnetometer type of magnetic compass.

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.

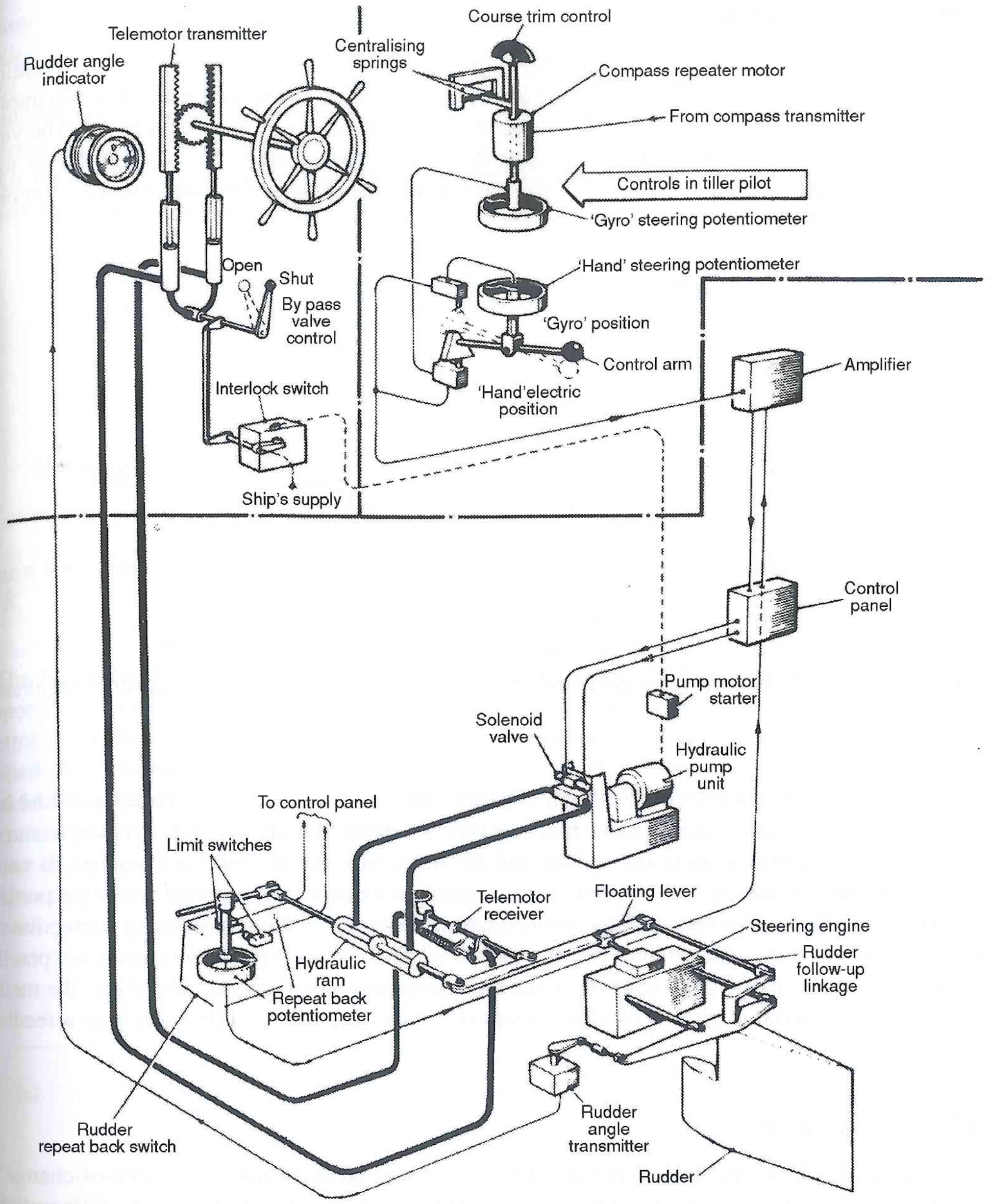


Figure 9.1 An early electro-mechanical autopilot system using telemotors. (Reproduced courtesy of Sperry Ltd.)

- 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.

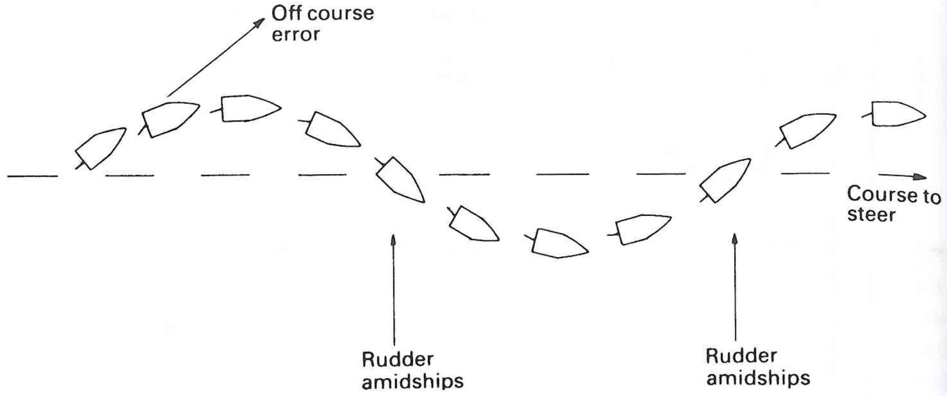


Figure 9.2 The effect of applying proportional control only. The vessel oscillates about the course to steer.

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.

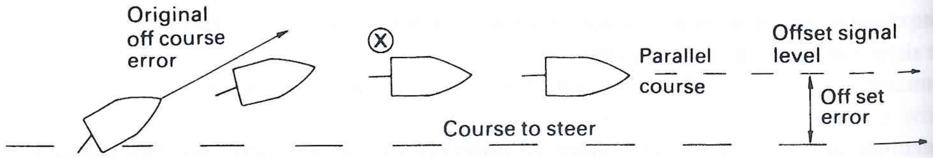


Figure 9.3 The effect of applying derivative control only.

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.

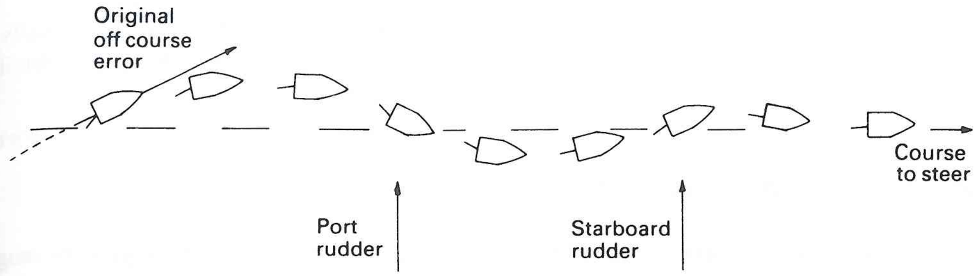


Figure 9.4 Applying a combination of proportional and derivative control brings the vessel back on track.

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.

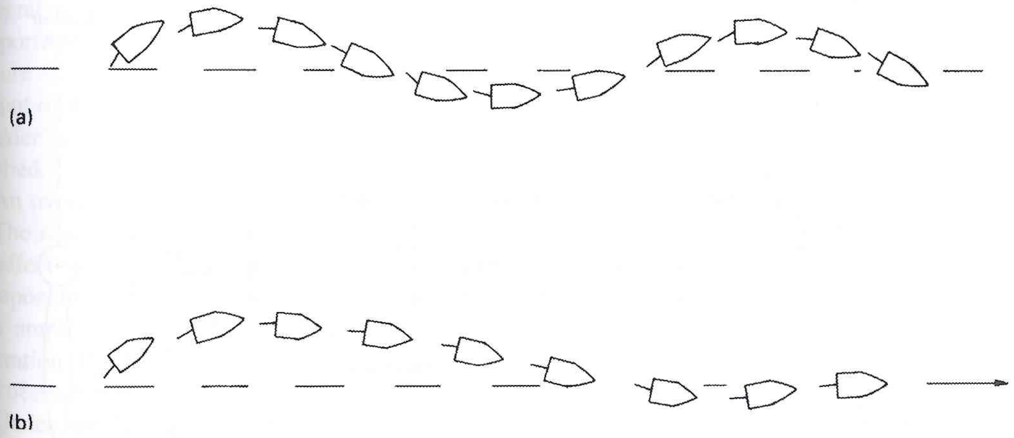


Figure 9.5 (a) If 'counter rudder' and 'rudder' controls are set too high, severe oscillations are produced before the equipment settles.(b) If 'counter rudder' and 'rudder' controls are set too low, there will be little overshoot and a sluggish return to the course.

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 (ψ) produced as the difference between the selected heading and the 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.

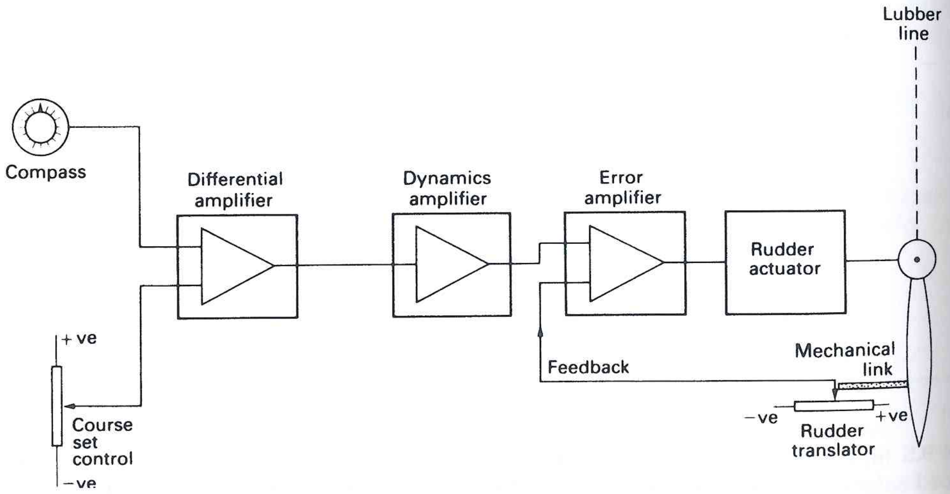


Figure 9.6 A simple autopilot system.

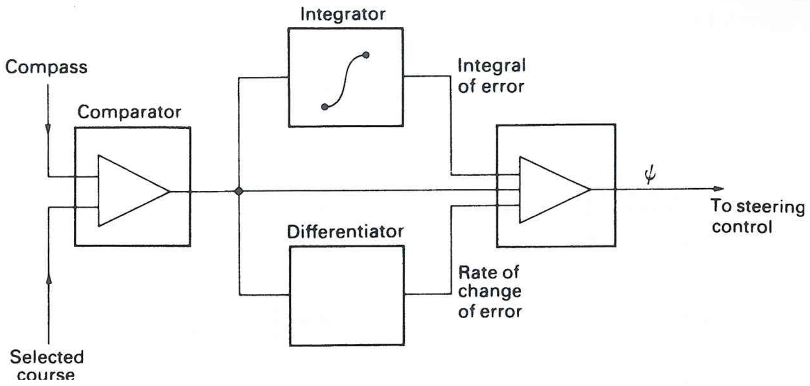


Figure 9.7 Error signal summing circuit.

The demanded rudder error signal (ψ) 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 (ψ) 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

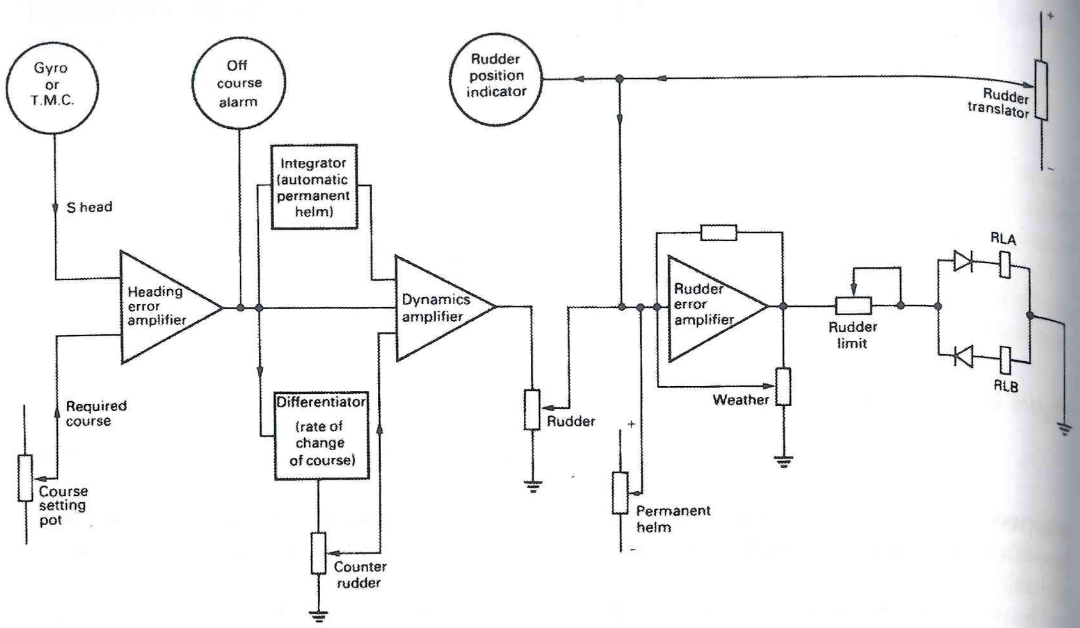


Figure 9.8 A simplified diagram of an autopilot system.

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.

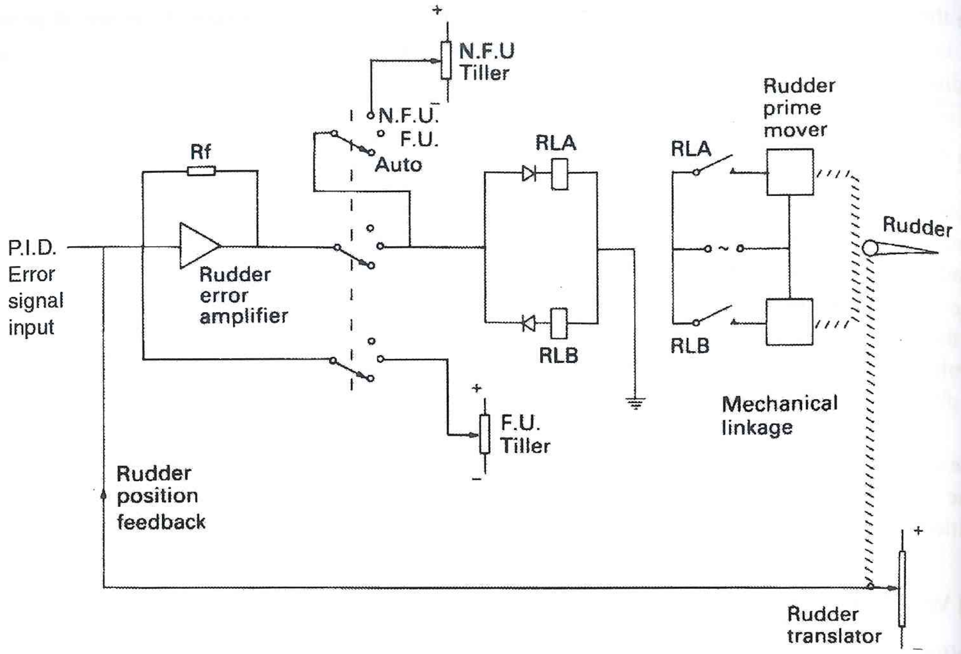


Figure 9.9 FU and NFU control of tiller operation. (Reproduced courtesy of Racal Marine Controls.)

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.

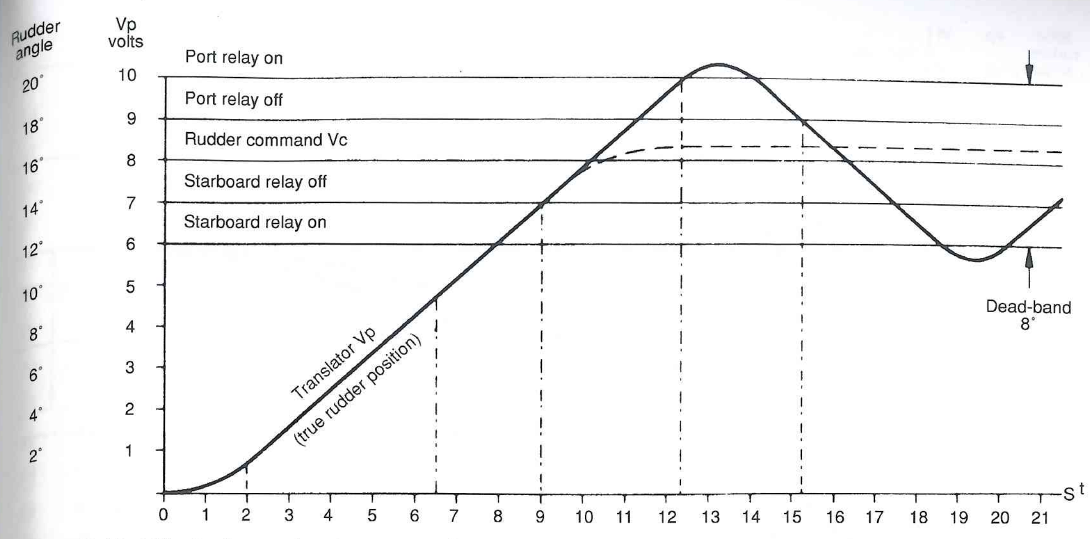


Figure 9.10 Effect of overshoot on control system stability. (Reproduced courtesy of Racal Marine Controls.)

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

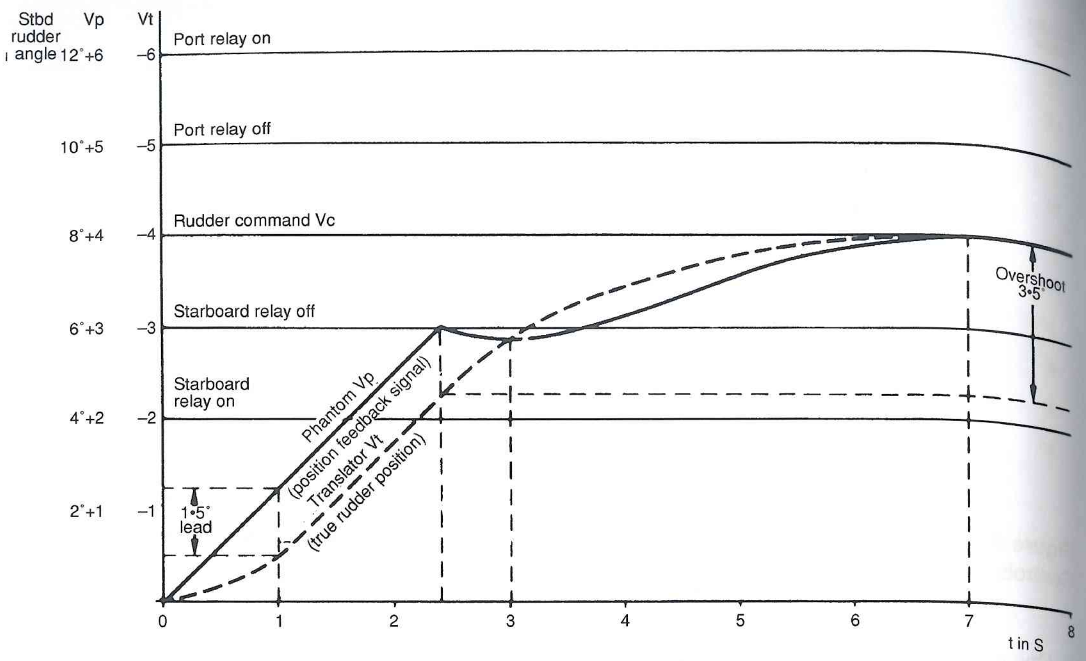


Figure 9.11 Operational principle of a phantom rudder. (Reproduced courtesy of Racal Marine Controls.)

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 the potentiometers during constantly varying weather conditions; thus the system cannot be absolutely efficient.

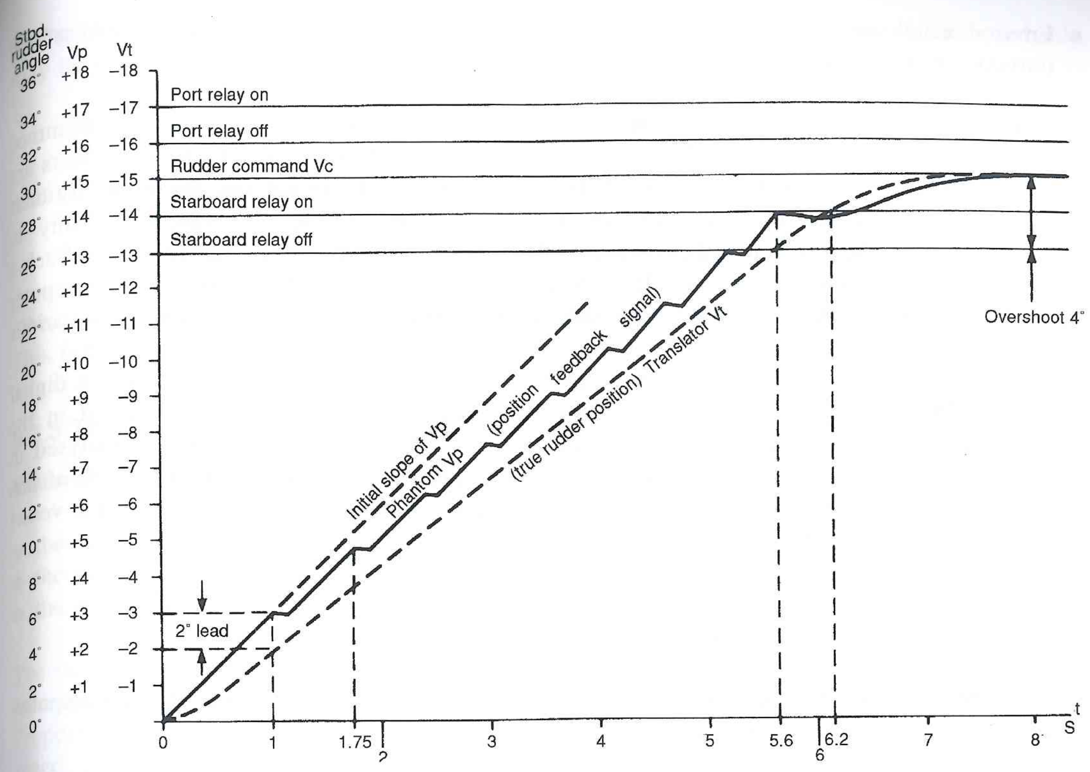


Figure 9.12 Characteristics of a practical application of phantom rudder. (Reproduced courtesy of Racal Marine Controls.)

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 ψ = ship's heading error,
δ = rudder angle, and
λ = 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

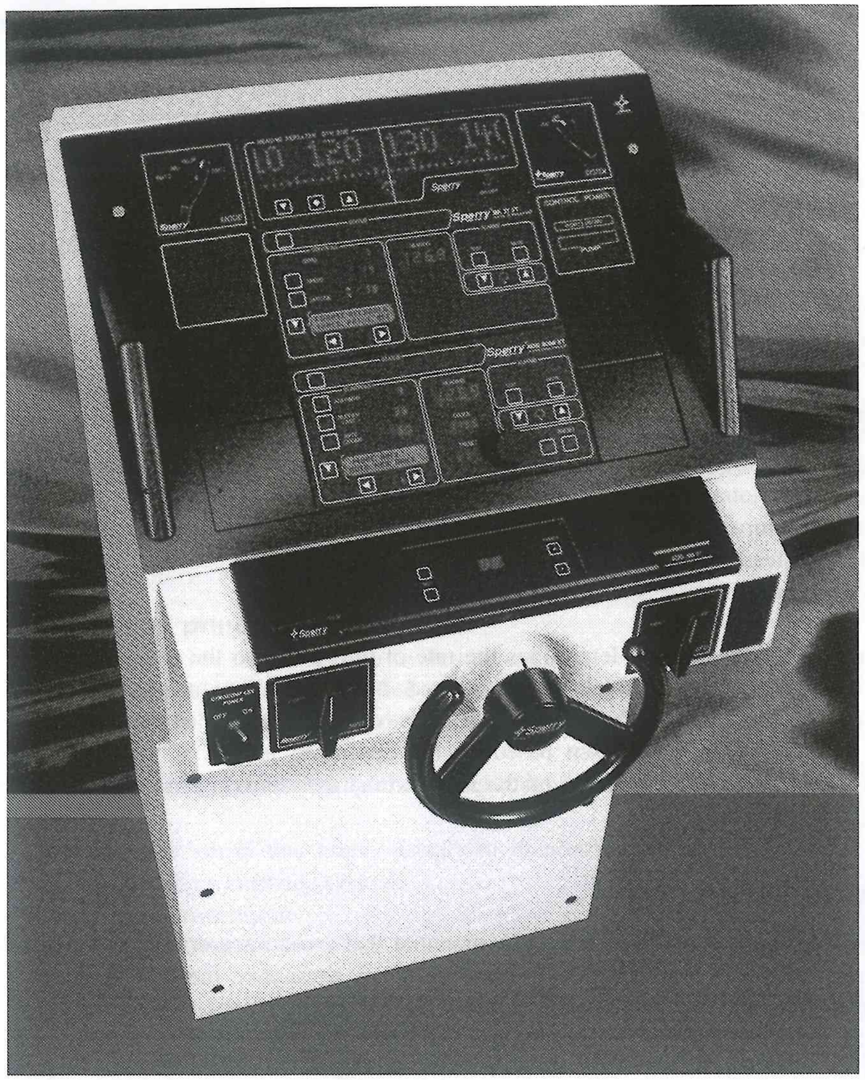


Figure 9.13 The Sperry Adaptive Digital Gyropilot" Steering Control Console. (Reproduced courtesy Litton Marine Systems.)

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.

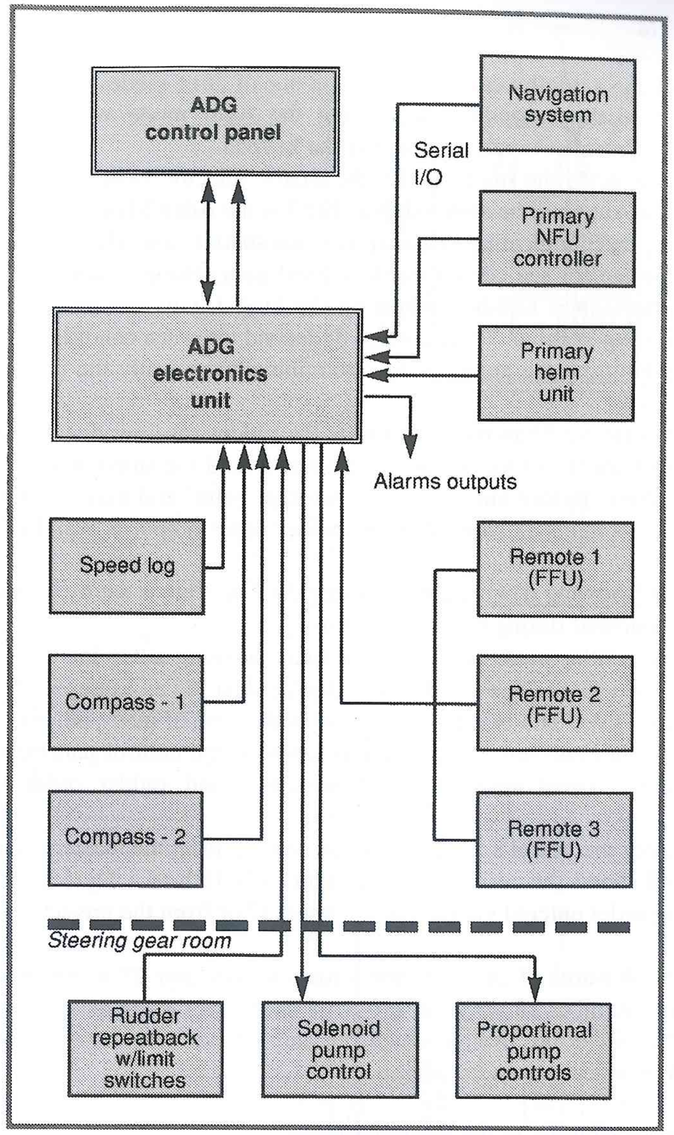


Figure 9.14 Sperry ADG 3000VT autopilot bridge configuration. (Reproduced courtesy of Litton Marine Systems).

- 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

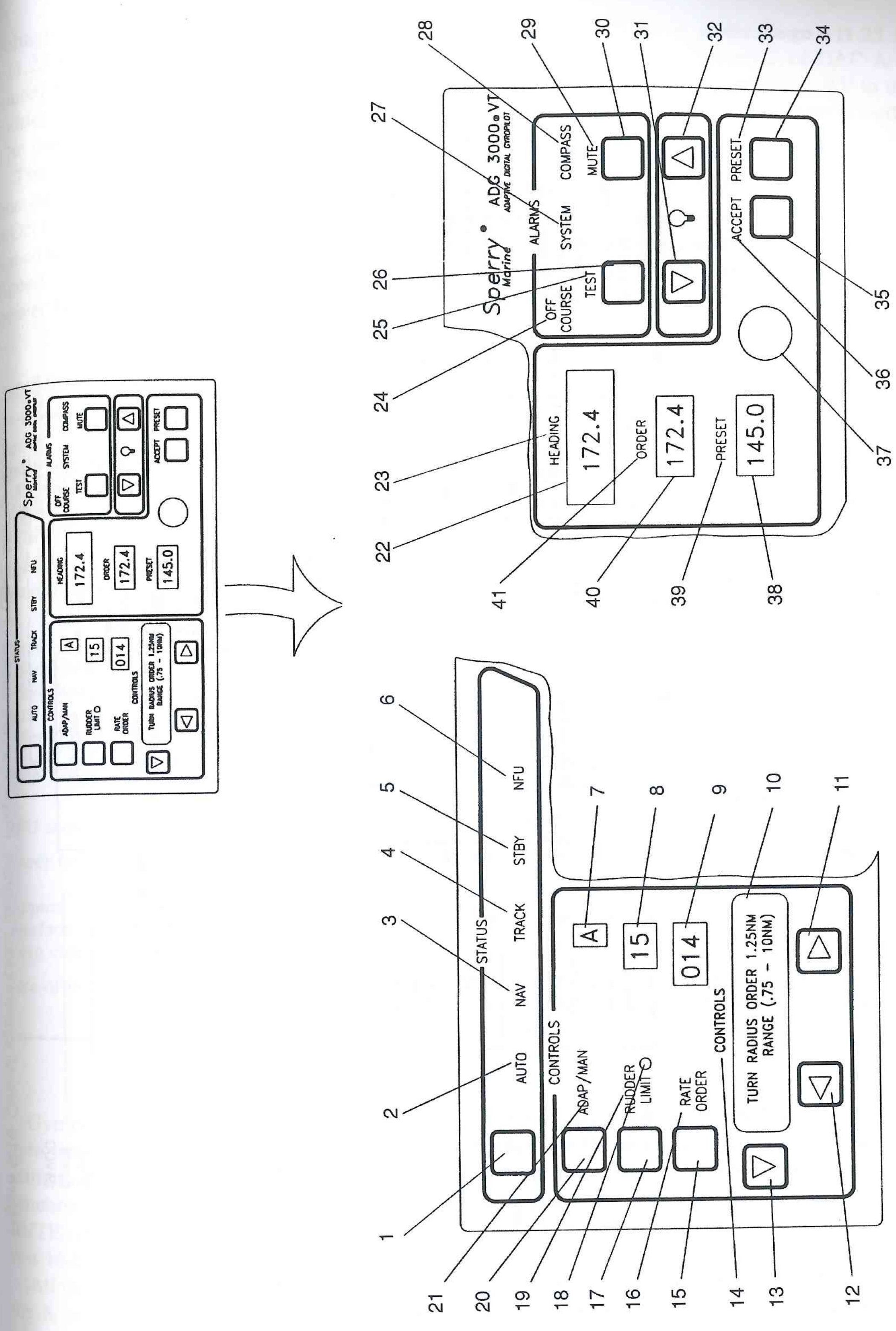


Figure 9.15 Display unit. (Reproduced courtesy Litton Marine Systems.)

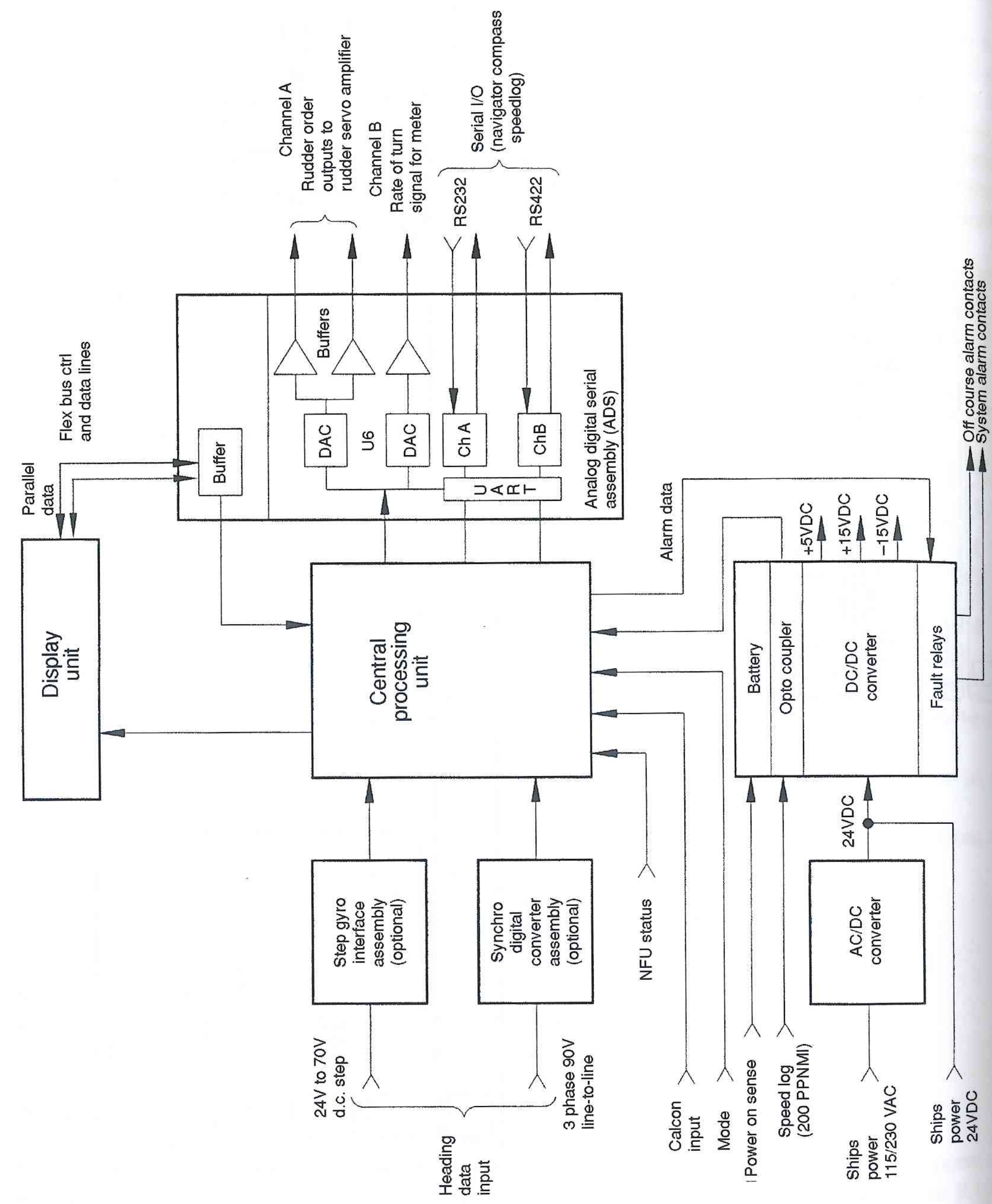


Figure 9.16 Overall system block diagram. (*Reproduced courtesy of Litton Marine Systems.*)

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. ± 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	½	15
HTC	Heading of course to next waypoint	½	15
HTR	True rate order	½	15
XTE	Cross track error	½	15
APB	Alternate order command and cross track error	½	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 (See Above)	14	Loss of receiver interrupts for 15 s (when NMEA 0183 device installed on RS-422 port (Channel B))	a. Check source. b. Check wire connection. c. Replace ADS Assembly. d. Replace CPU Assembly.
(DEVICE B) ERROR (See Above)	15	Framing error; invalid message bit format (on RS-422)	a. Check serial channel wire protocol of source. b. Replace ADS Assembly.
(DEVICE B) ERROR (See Above)	16	Overrun error (on RS-422)	a. Check serial channel wire protocol of source. b. Replace ADS Assembly.
(DEVICE B) ERROR (See Above)	17	Loss of transmitter interrupts for 1 s after character sent (on RS-422)	Replace ADS Assembly.
(DEVICE C) ERROR (See Above)	18	Overrun error; input too fast (on RS-232)	a. Check serial channel wire protocol of source. b. Replace ADS Assembly.
CALCON I/O ERROR Frame/Overrun/Noise	18	Framing, overrun, or noise error (if CALCON connected)	a. Check CALCOM cable connection. b. Try running CALCON from a different PC. c. Replace CPU Assembly.
(DEVICE C) ERROR (See Above)	19	Loss of transmitter interrupts for 1 s after character sent (on RS232)	Replace ADS Assembly.
CALCON I/O ERROR Loss of interrupts	19	Loss of transmitter interrupts for 1 second after character sent	a. Check CALCON cable connection. b. Try running CALCON from a different PC. c. Replace CPU Assembly.
SPEED LOG ERROR Enter Manual Speed	40	No VBW/VHW message received for 4 s	a. Check source. b. Check wire connection. c. Replace ADS Assembly. d. Replace CPU Assembly.
SPEED LOG ERROR Enter Manual Speed	41	Invalid format in VBW/VHW sentence	a. Check message string output by the source. b. Check connection. c. Replace ADS Assembly. d. Replace CPU Assembly.
SPEED LOG ERROR 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). b. Check wire connection speed log. c. Replace DC/DC Assembly. d. Replace CPU Assembly for pulse log.
SPEED LOG ERROR Enter Manual Speed	43	Speed data null or marked invalid	Check speed log source (log data strings from source).
RADIUS DISABLED Log Speed Required	44	Speed setting changed to a manual entry while in RADIUS control	Operator misuse.

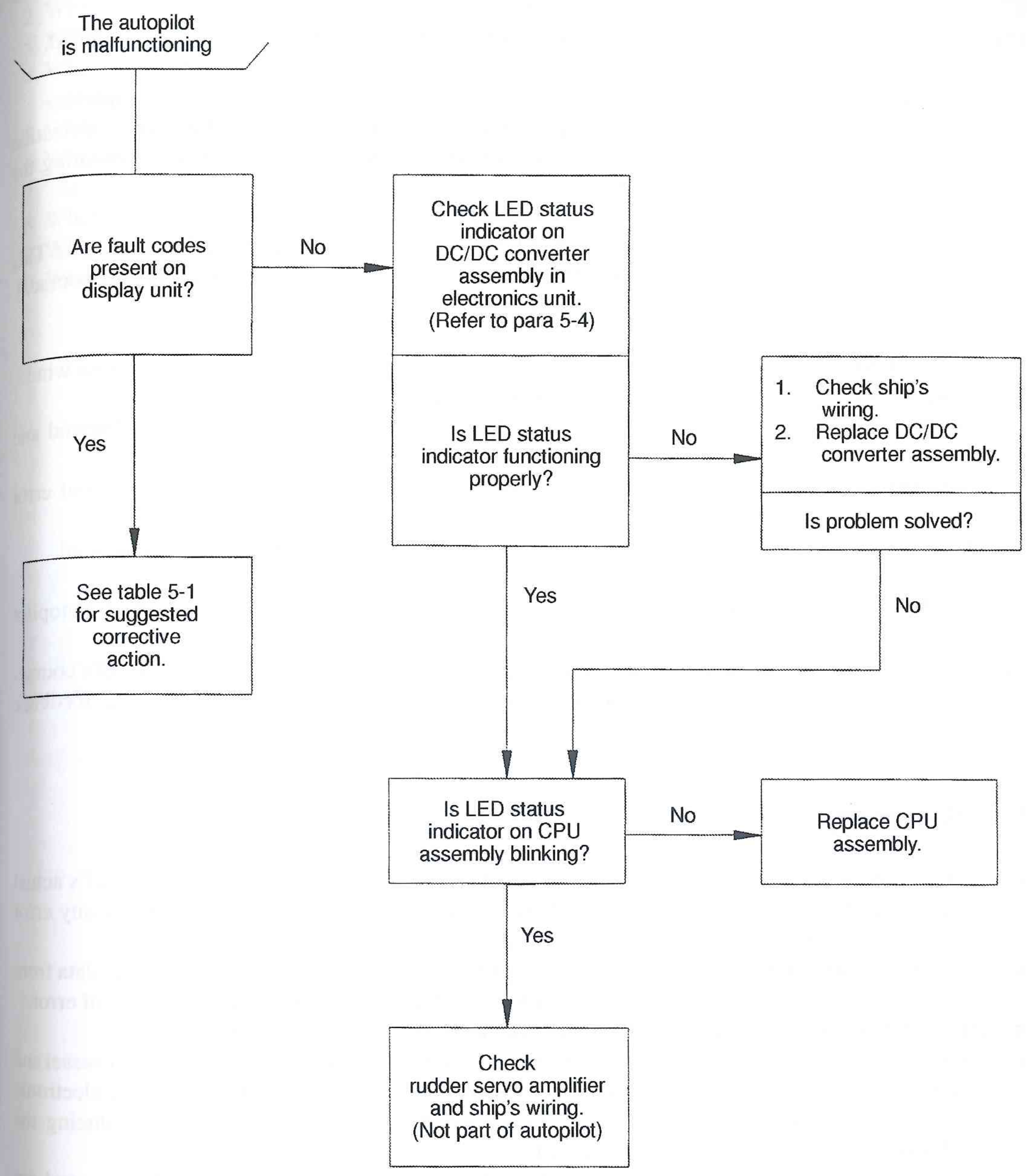


Figure 9.17 Sperry ADG 3000 VT fault logic diagram. (Reproduced courtesy of Litton Marine Systems.)

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.

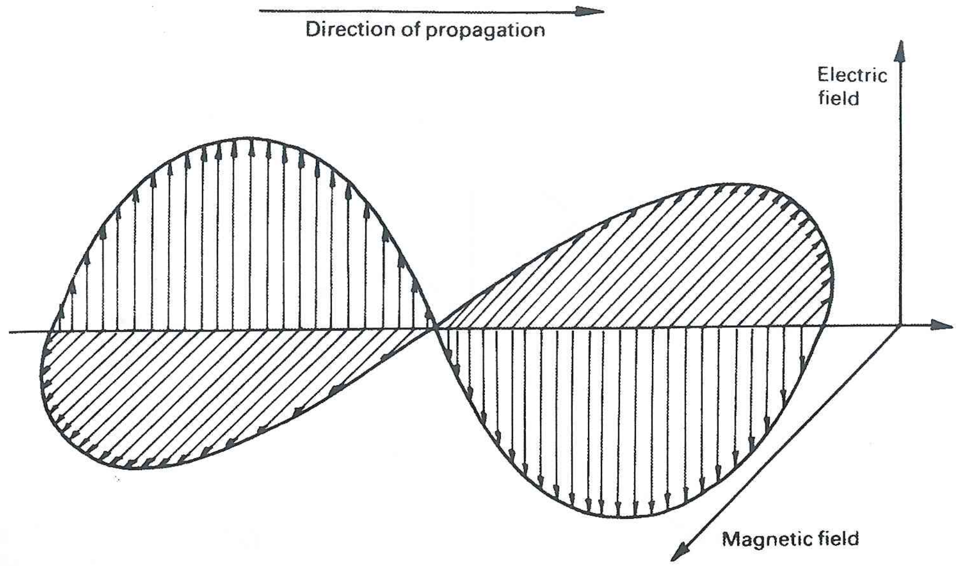


Figure 10.1 A propagated radio wave illustrating the relationship between both fields of energy and the direction of propagation.

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

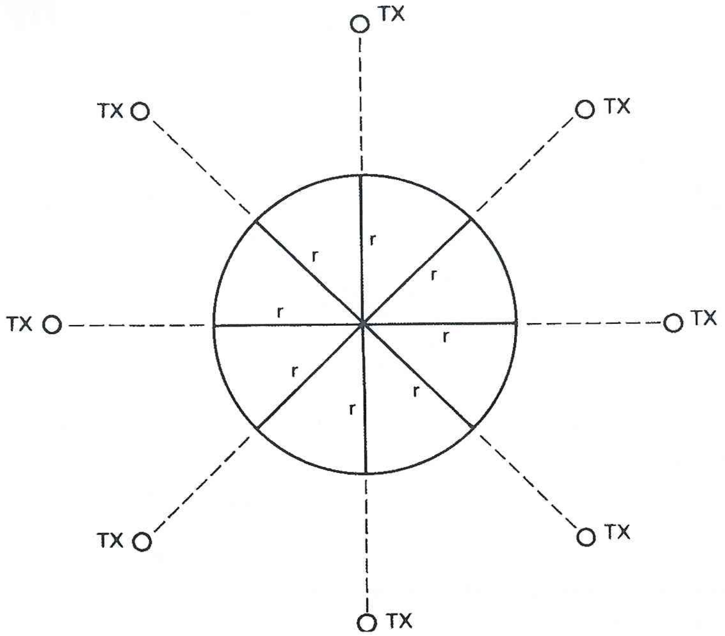


Figure 10.2 The AGP reception plot for a vertical antenna. The antenna is at the centre of the circle.

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

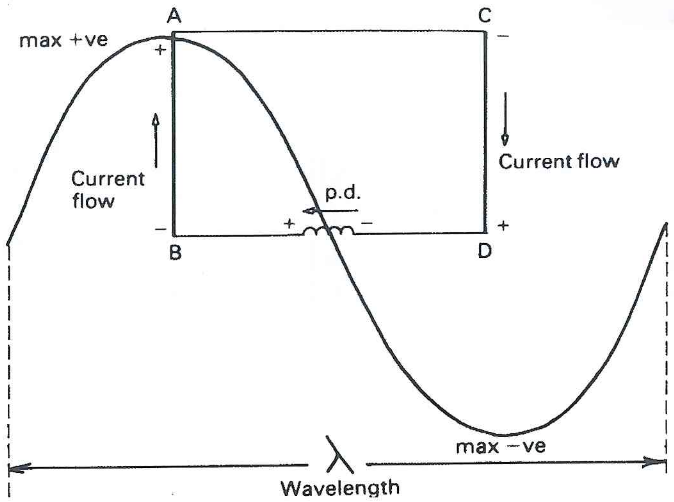


Figure 10.3 Signal currents induced in the vertical arms of a loop antenna produce a resultant potential difference across the input coil to a receiver.

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

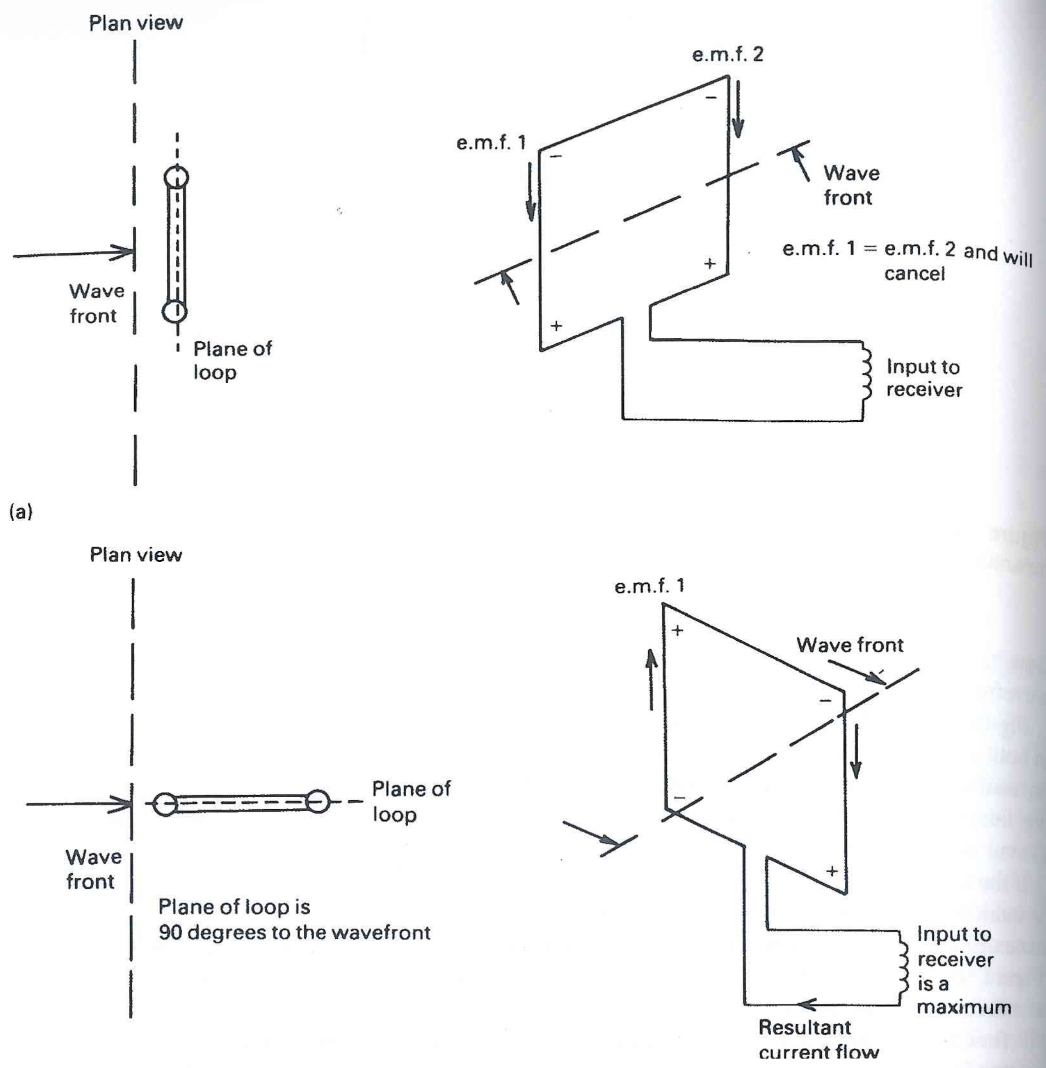


Figure 10.4 (a)The resultant input to a receiver is zero if the plane of the loop is parallel with the travelling wavefront.(b)The input is a maximum if the loop plane is at 90° to the received signal.

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

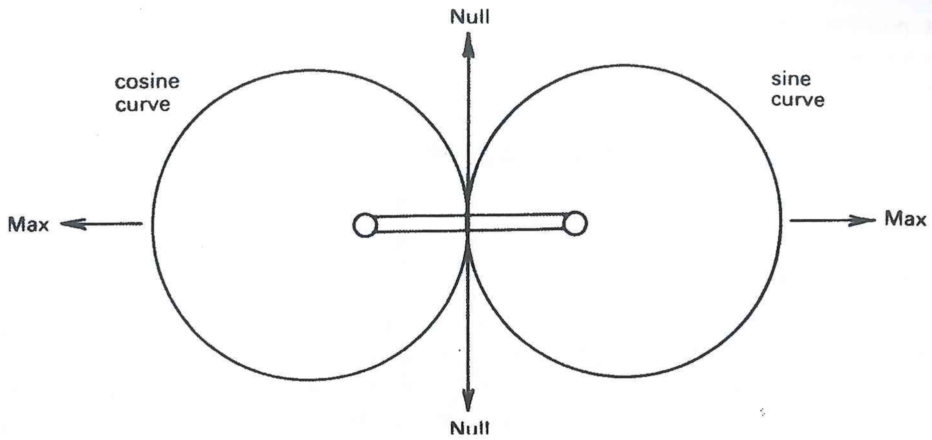


Figure 10.5 The figure-of-eight azimuth gain plot for a loop antenna.

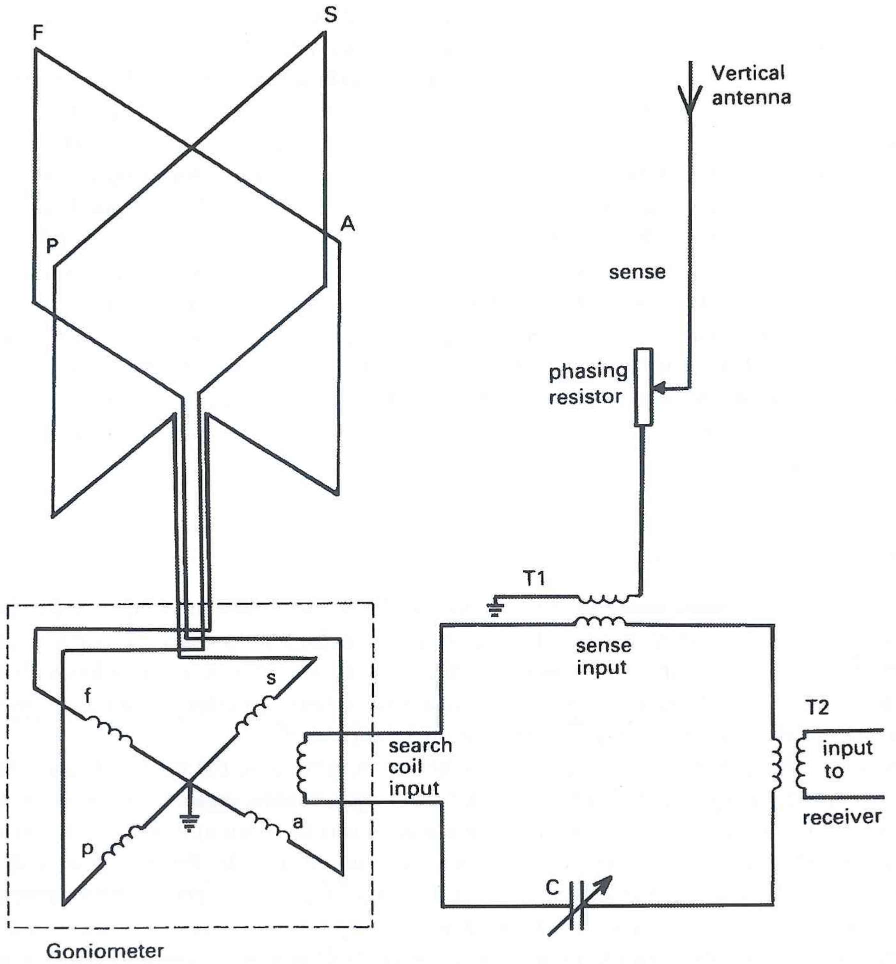


Figure 10.6 A simple receiver input circuitry for a fixed loops system.

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.

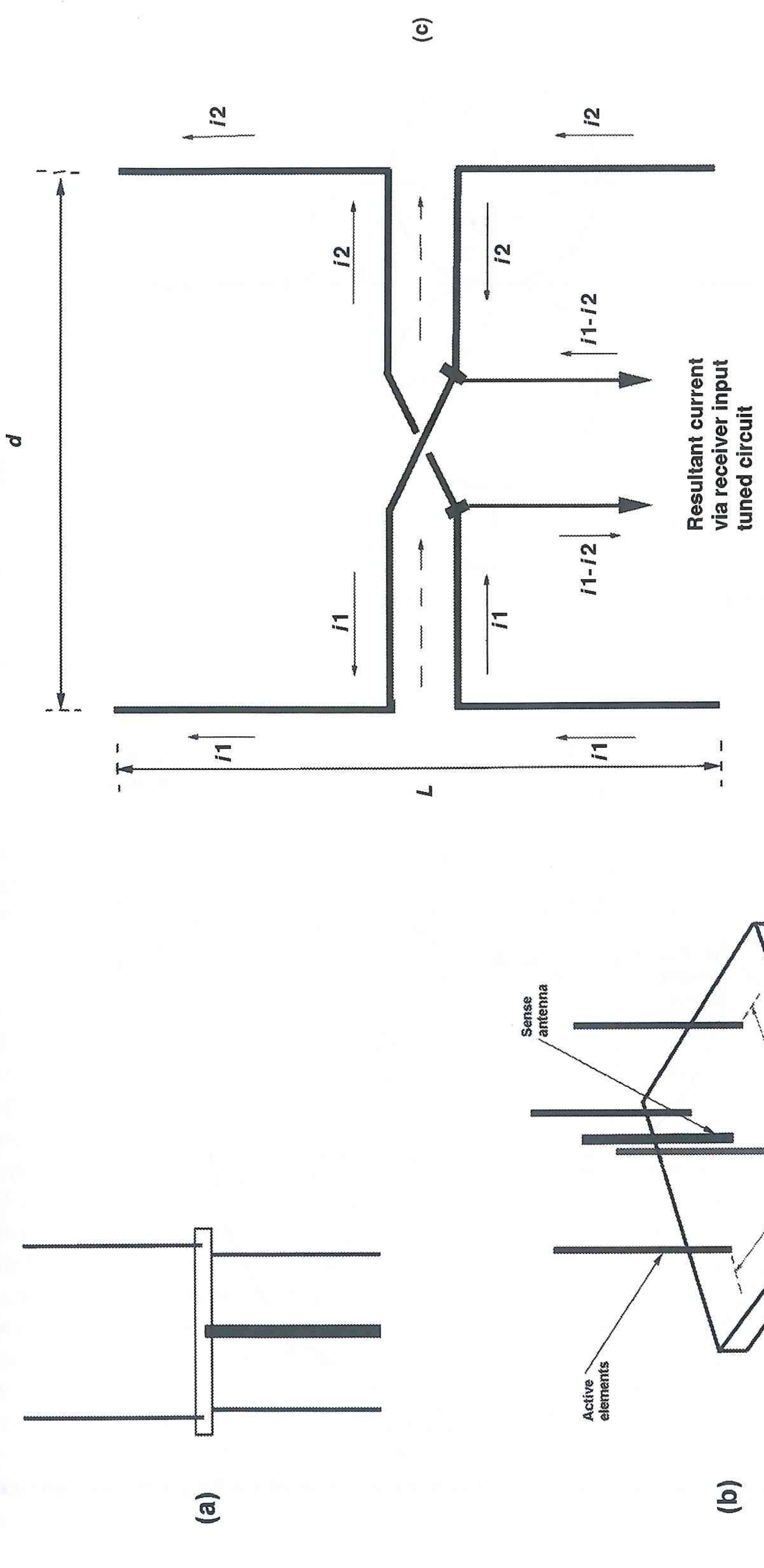


Figure 10.7 (a) A pole-mounted Adcock antenna and (c) its electrical equivalent. (b) A base plate-mounted Adcock array.

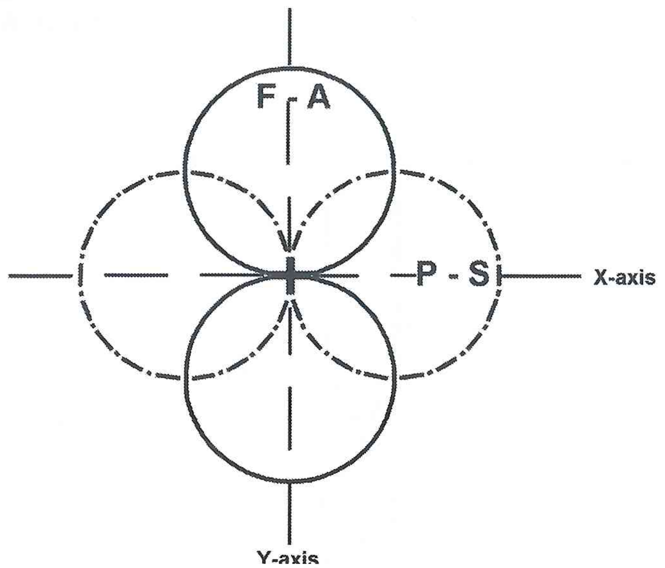


Figure 10.8 AGP diagram for an Adcock (or a crossed loop) pair.

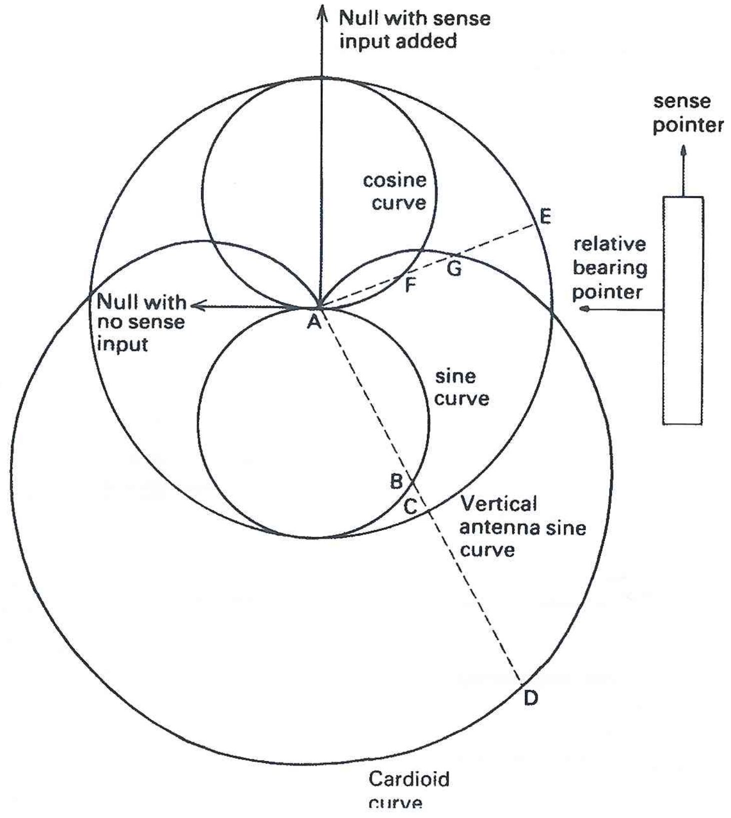


Figure 10.9 The resultant cardioid AGP produced by the addition of the figure-of-eight and circular plots.

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.

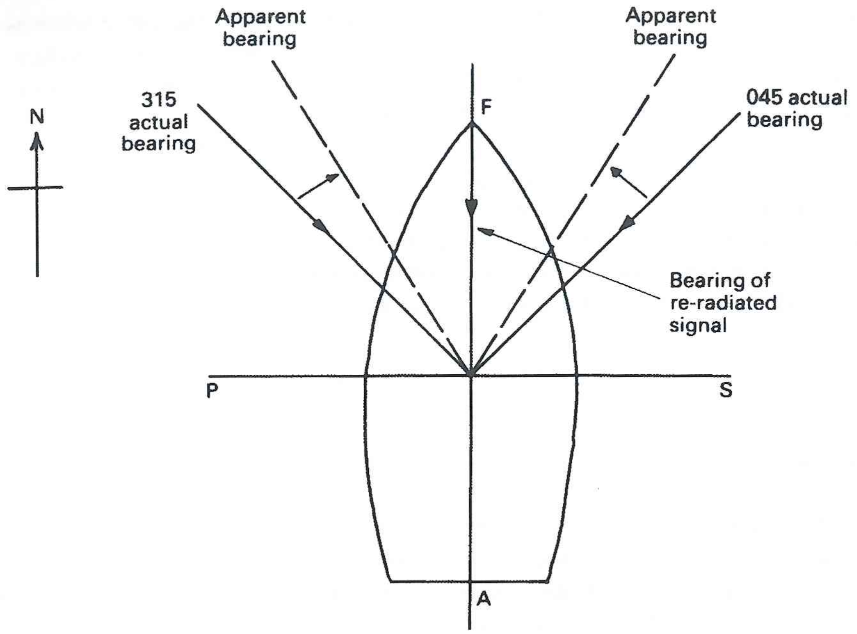


Figure 10.10 The effects of quadrantal error are to pull the bearing indication towards the vessel's lubber line.

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.

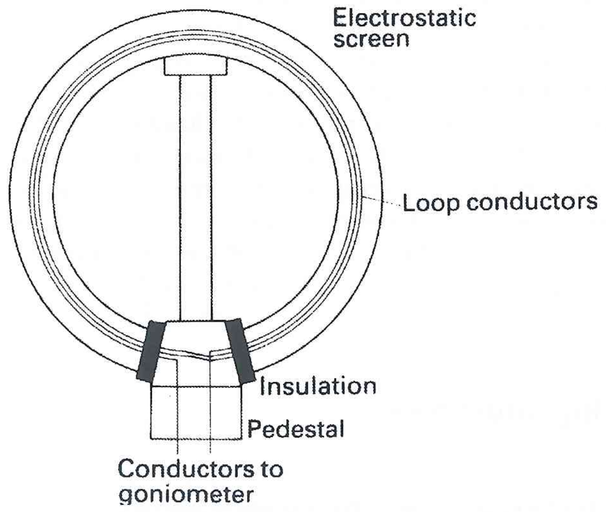


Figure 10.11 Electrostatic screening of a single loop to minimize vertical error.

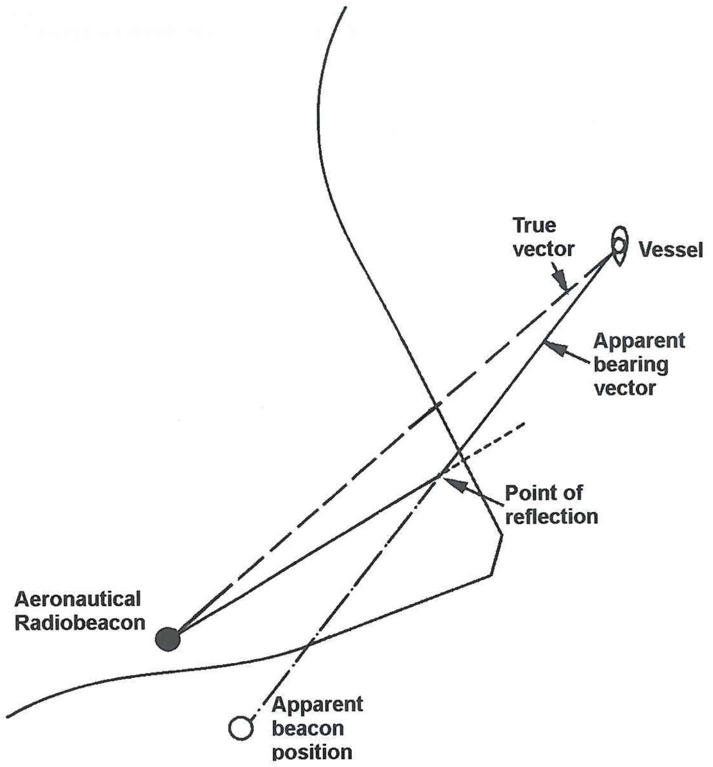


Figure 10.12 Error introduced by a reflected VHF radio wave.

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

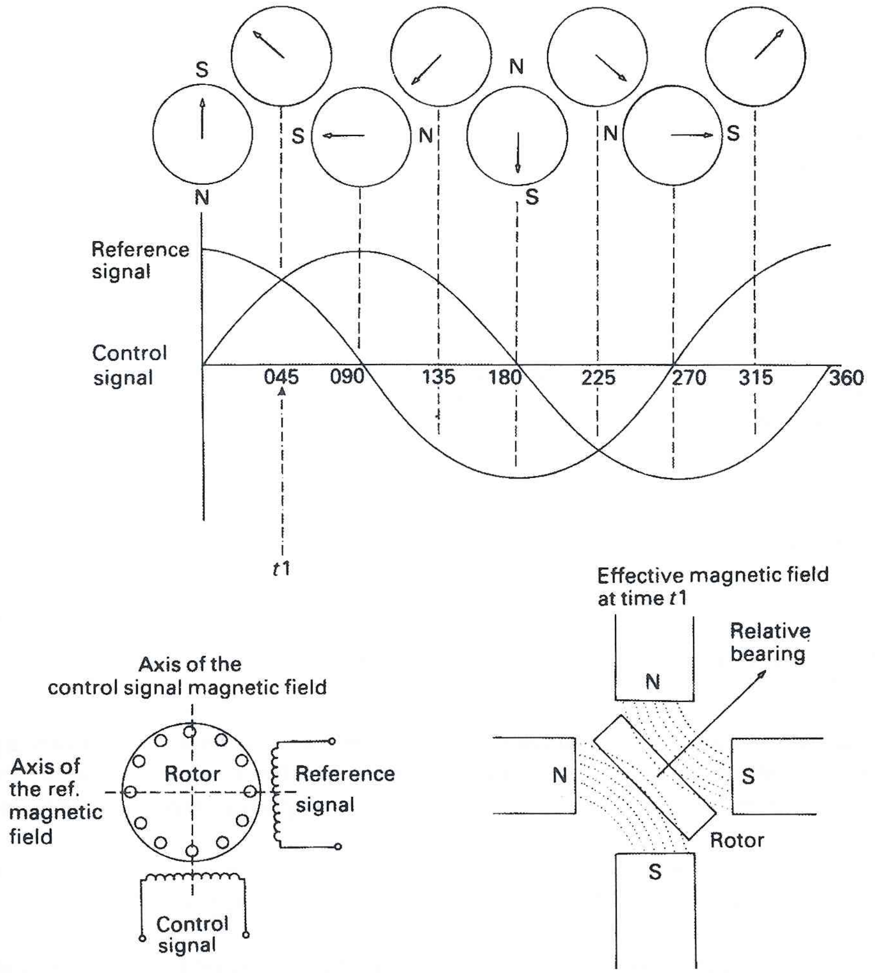


Figure 10.13 The rotating magnetic field produced in the stator windings of a two-phase induction system.

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.

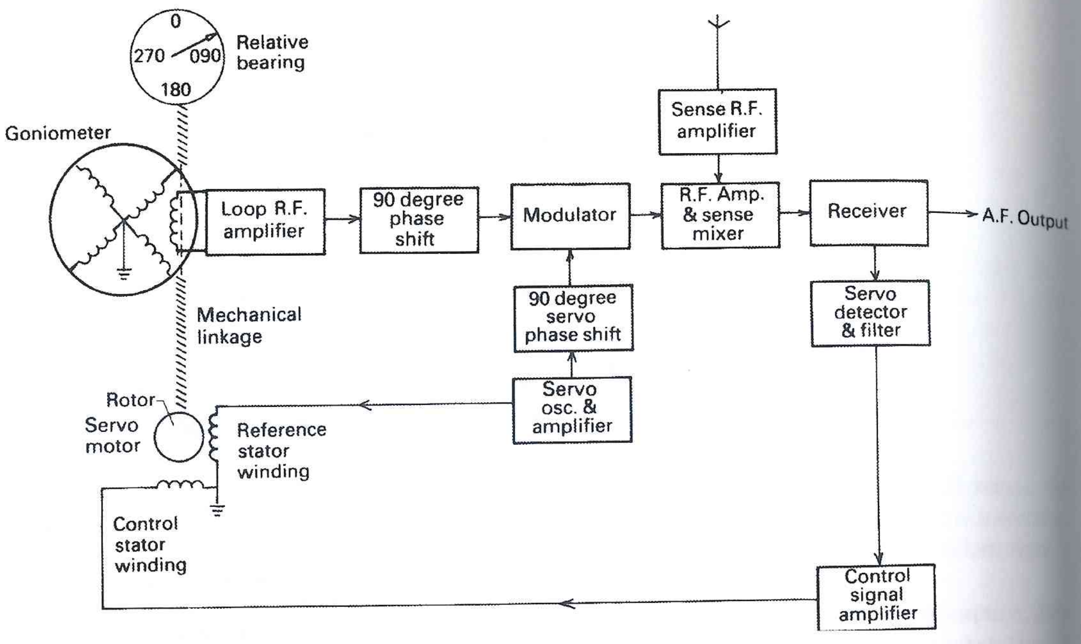


Figure 10.14 System diagram of a servo-controlled automatic RDF system.

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 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.

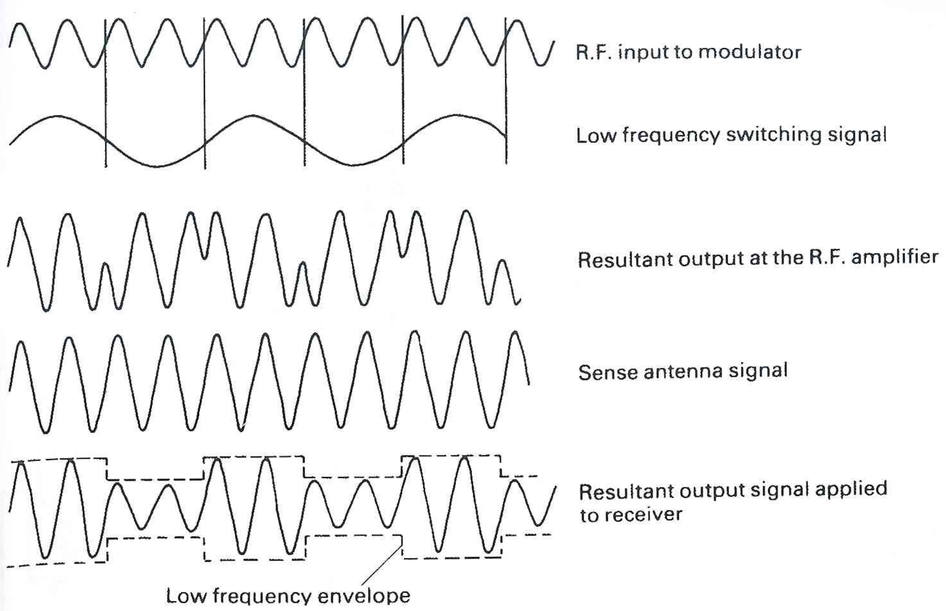


Figure 10.15 Illustration of the waveform mixing process to produce the servo control signal envelope.

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.

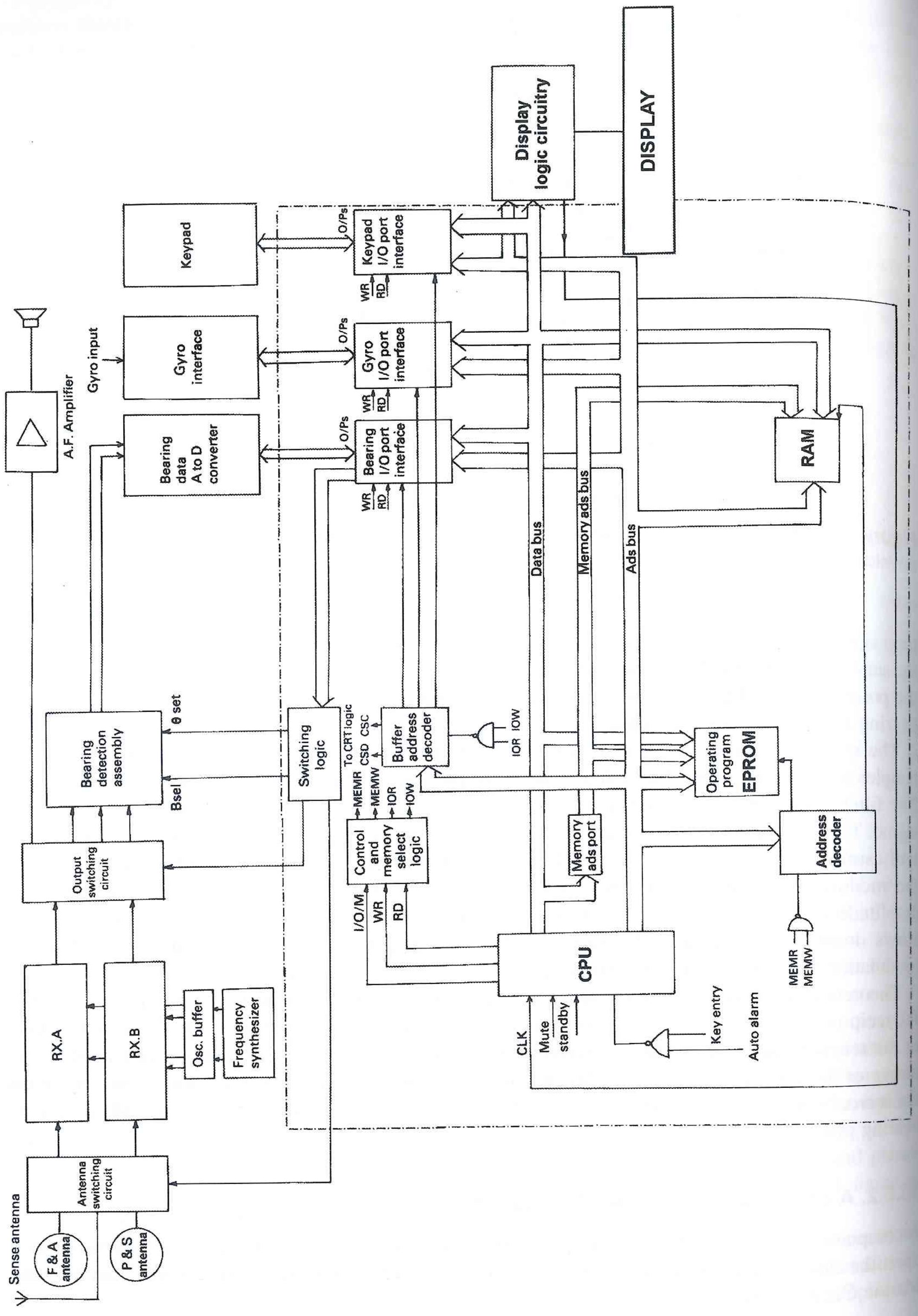


Figure 10.16 A system diagram for a computer-controlled RDF.

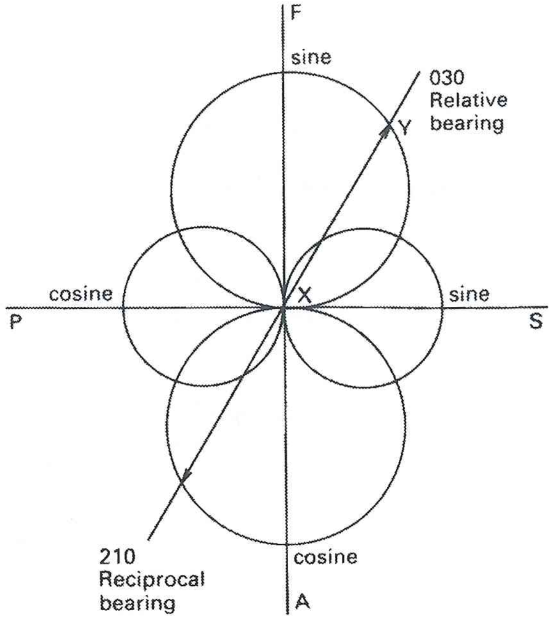


Figure 10.17 AGP plots of the input to the receiver.

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 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.

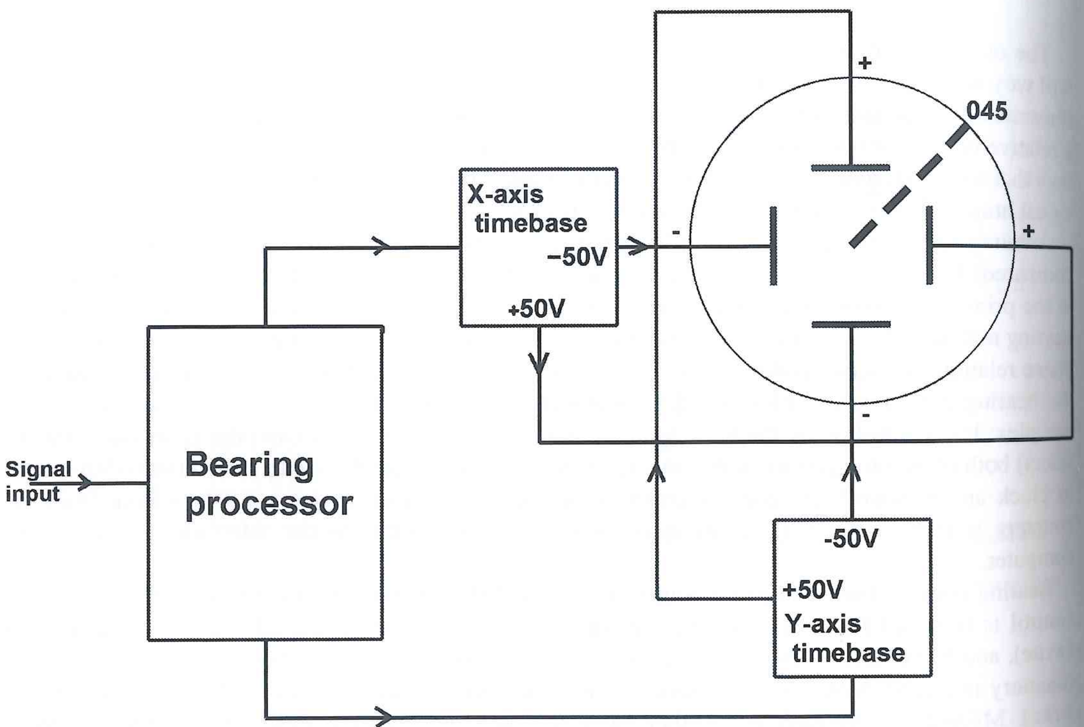


Figure 10.18 The magnitude of each x and y voltage determines both the azimuth indication and the strength of the signal as shown by the length of the vector.

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).

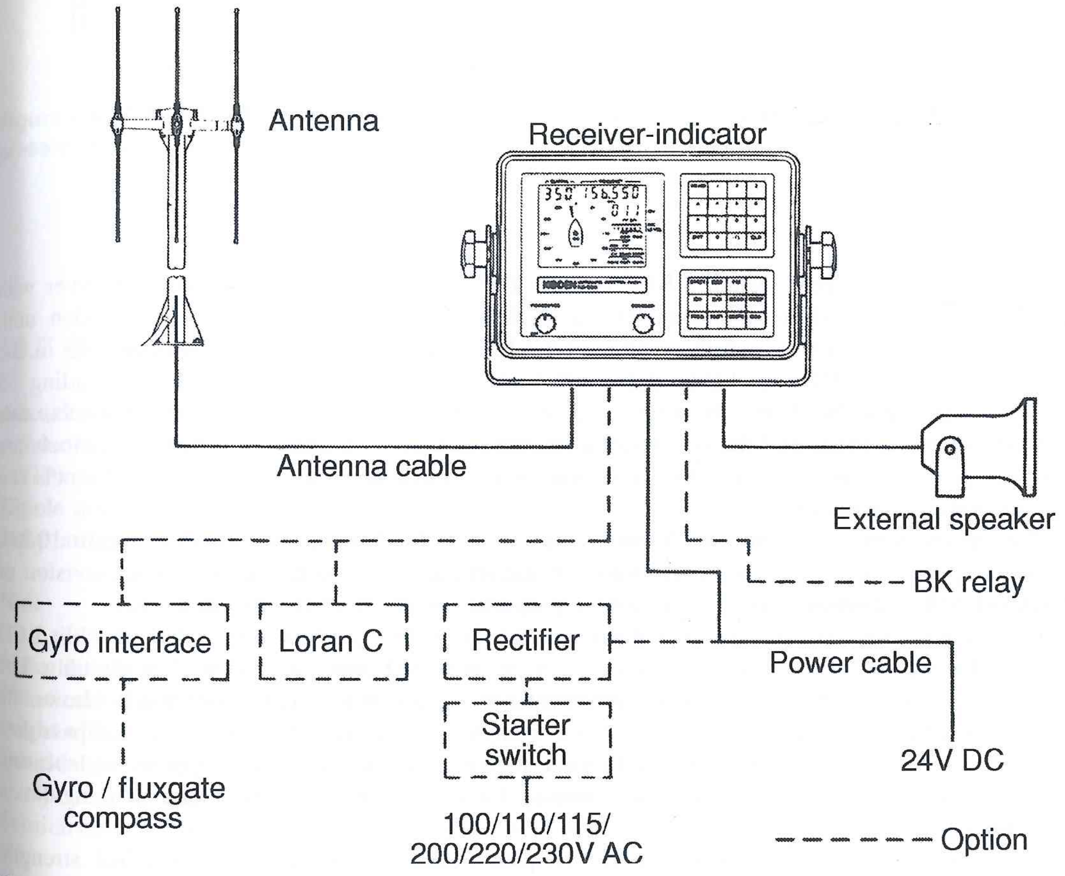


Figure 10.19 A modern RDF installation showing interface details. (Reproduced courtesy of Koden Electronics Co. Ltd.)

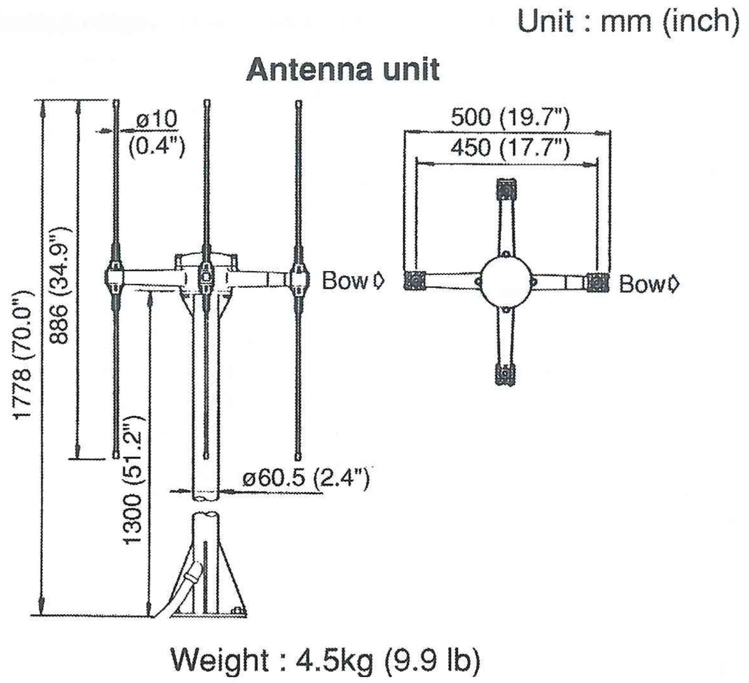


Figure 10.20 Construction detail of the Adcock antenna unit. (Reproduced courtesy of Koden Electronics Co. Ltd.)

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.

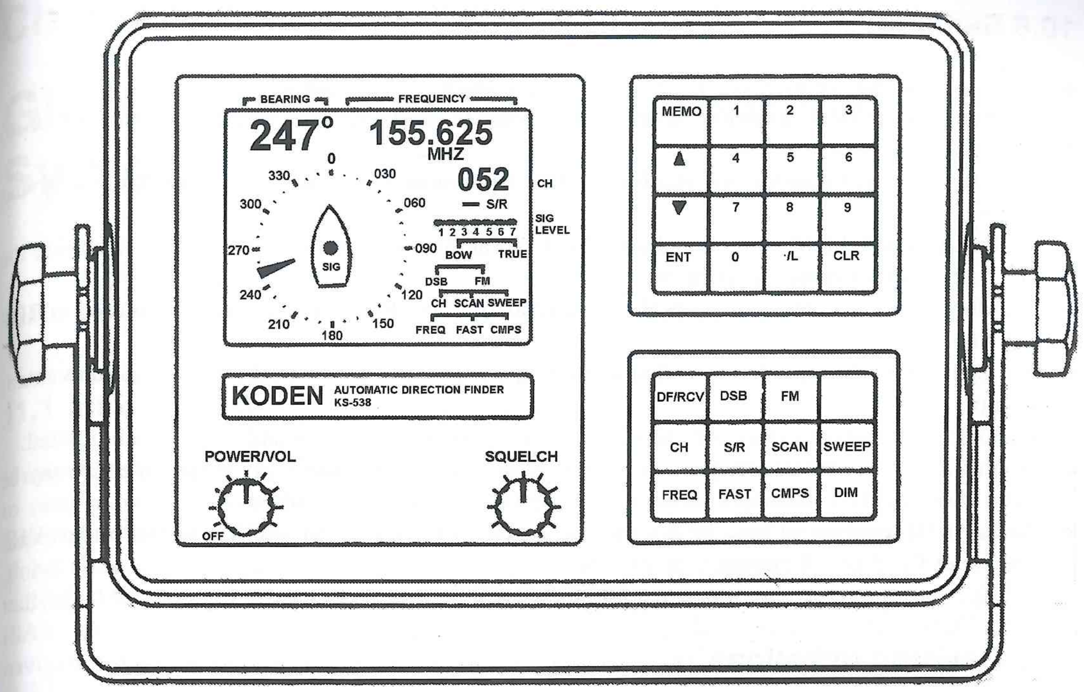


Figure 10.21 The Koden KS-538 RDF system display is a good indication of the information presented to the user of a modern equipment. (Reproduced courtesy of Koden Electronics Co. Ltd.)

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

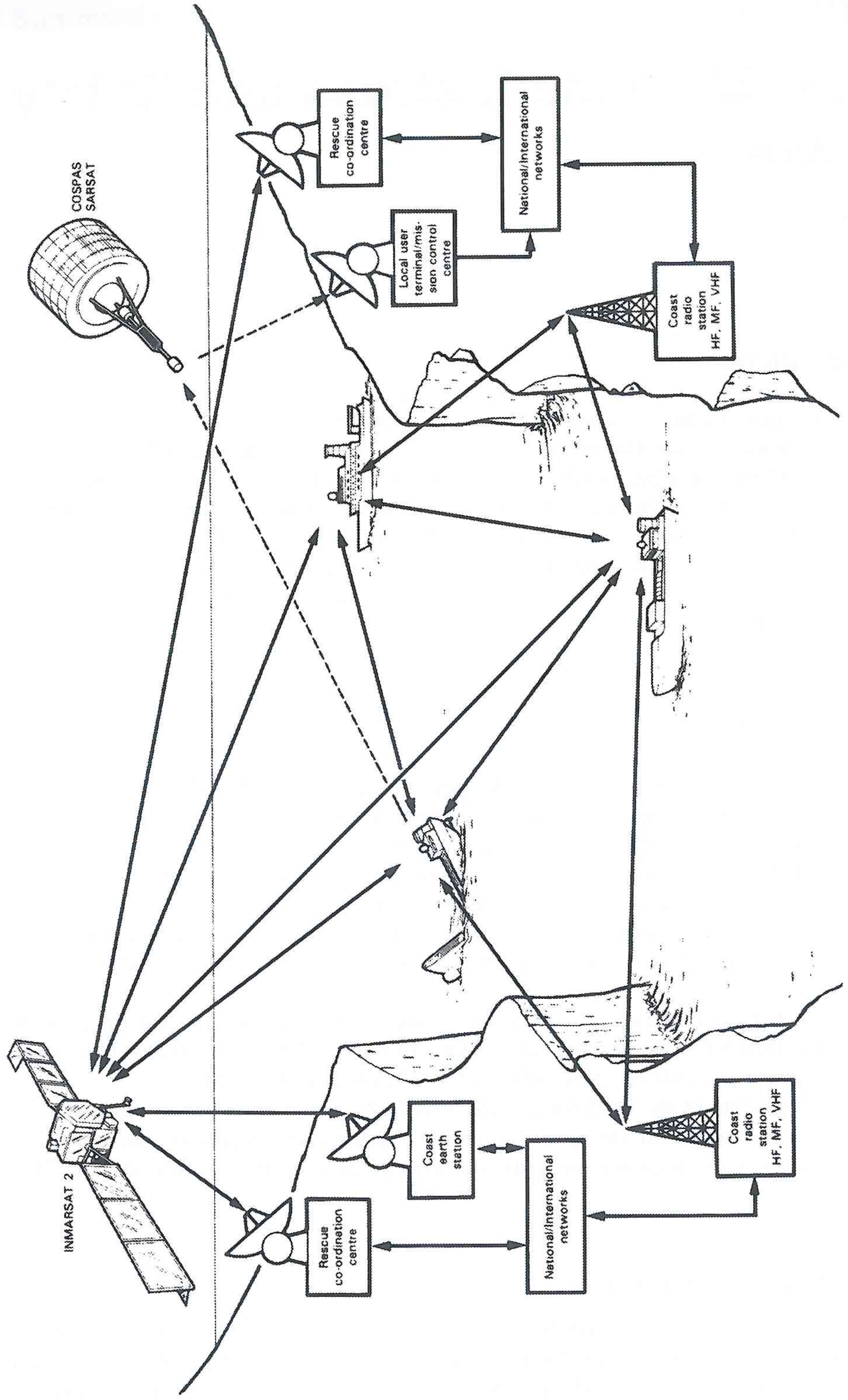


Figure 11.1 General concept of the GMDSS. (Reproduced courtesy of the IMO.)

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.

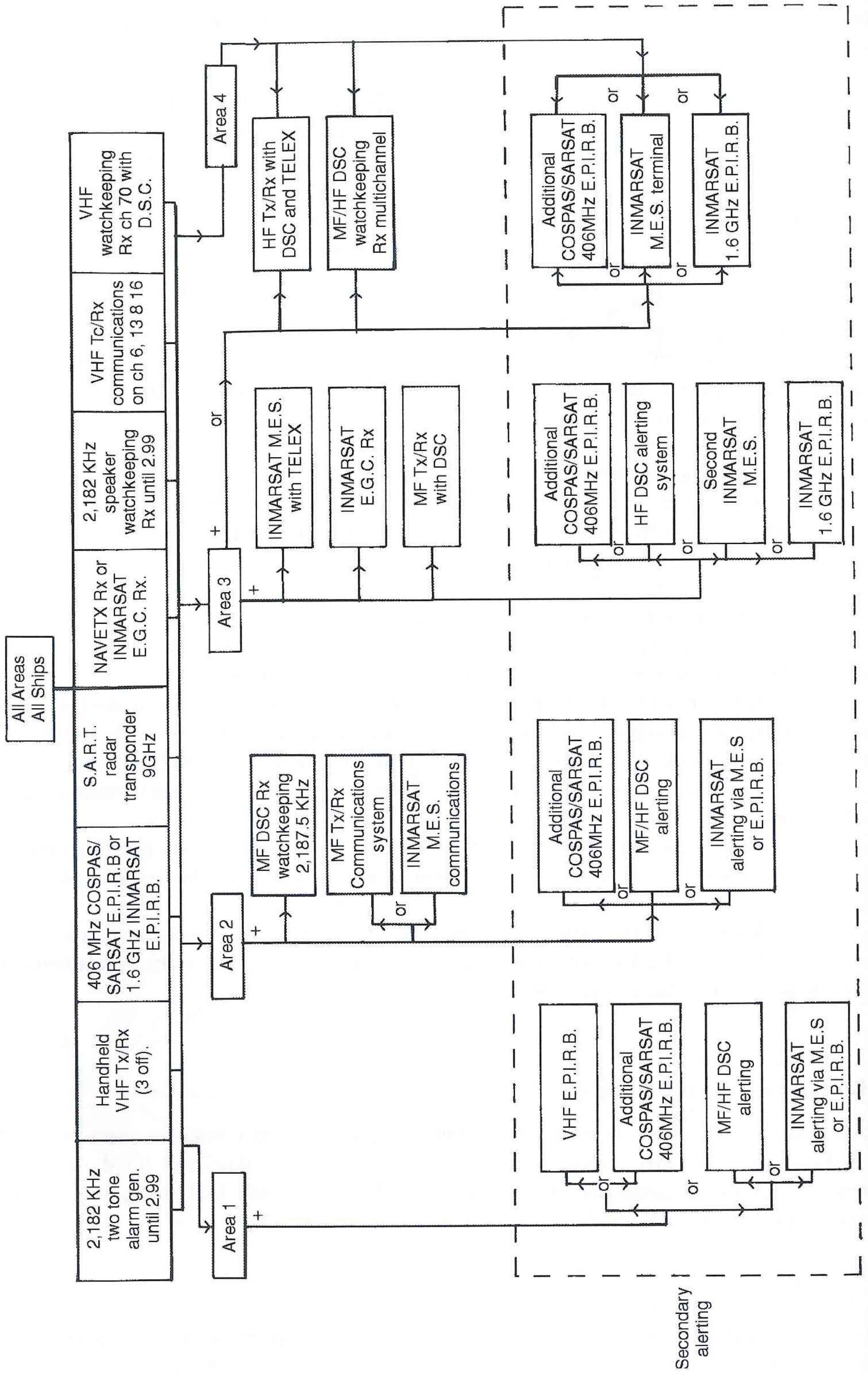


Figure 11.2 GMDSS equipment carriage requirement table.

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.

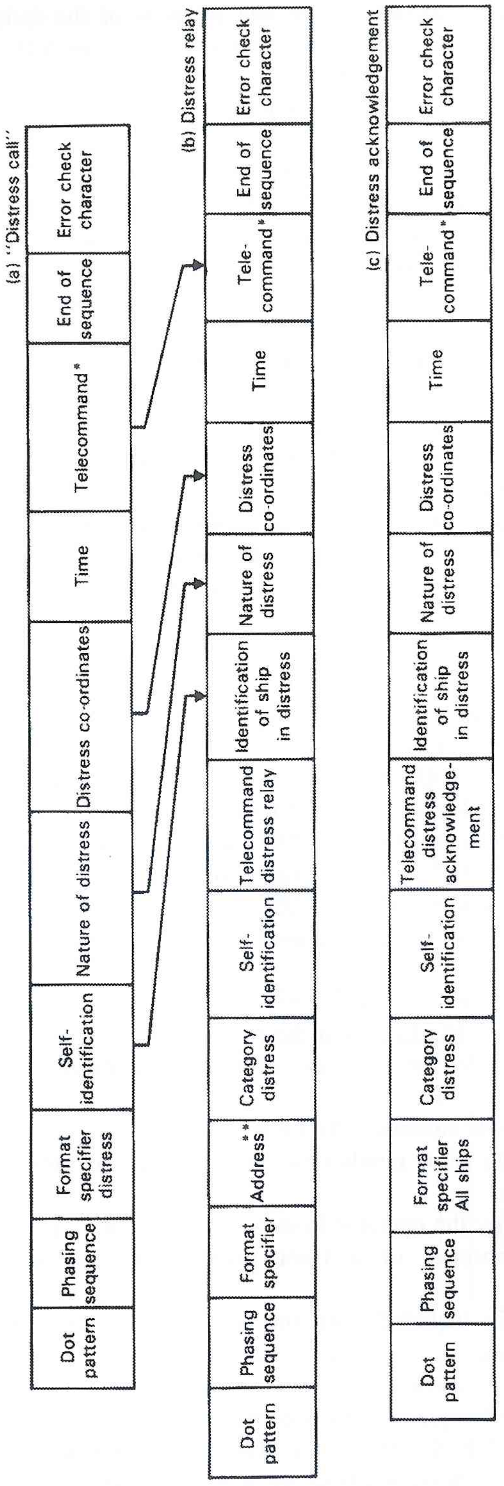
By using the ship's call number, a coastal radio station is able to call a selected vessel. Collective calls may also be made to all ships, or ships belonging to one company or trading in one area of the world. Selective calling is, depending upon the propagational characteristics of the radio wave, 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.

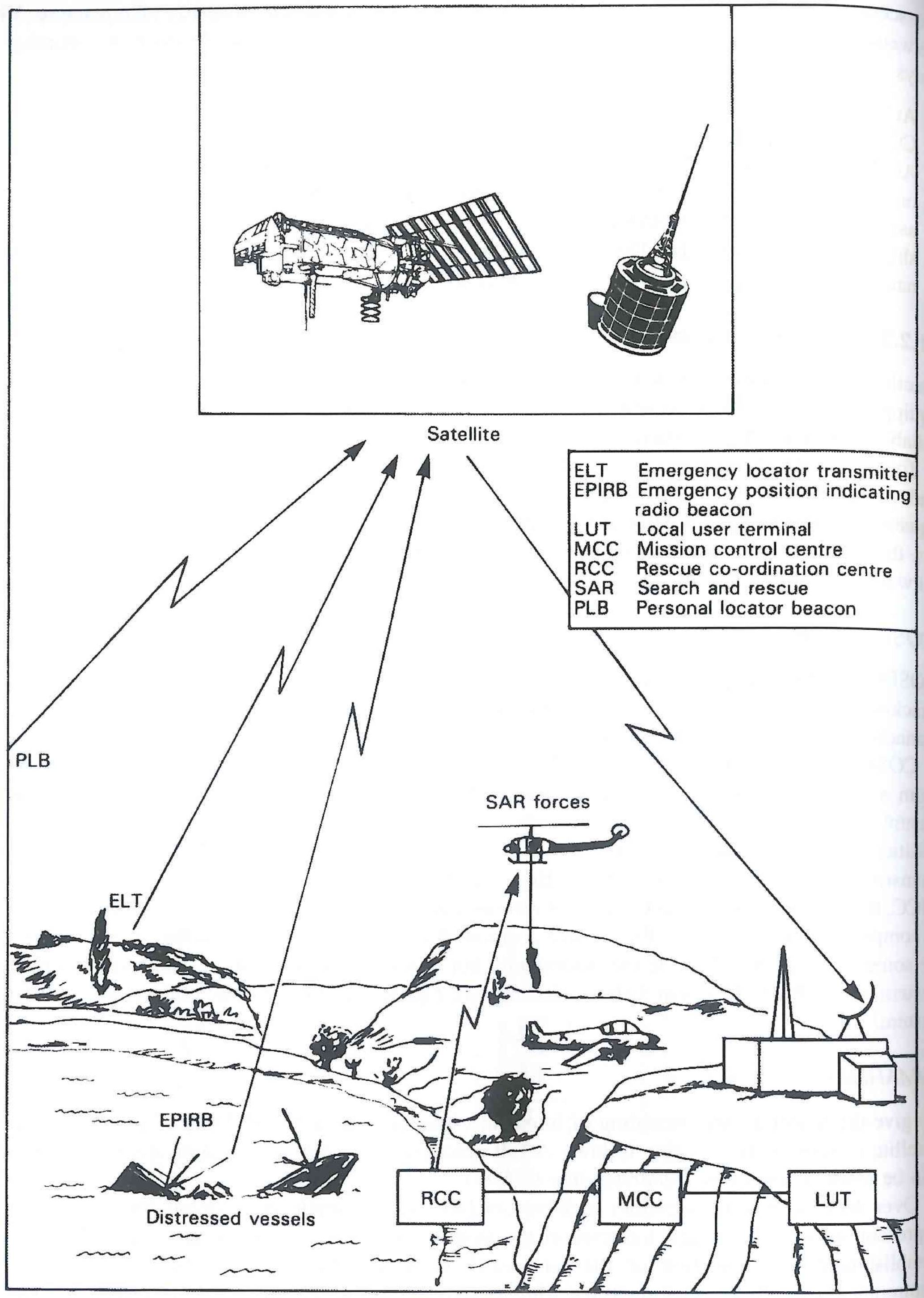


Figure 11.4 Basic concept of the COSPAS/SARSAT alerting system. (Reproduced courtesy of the IMO.)

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.

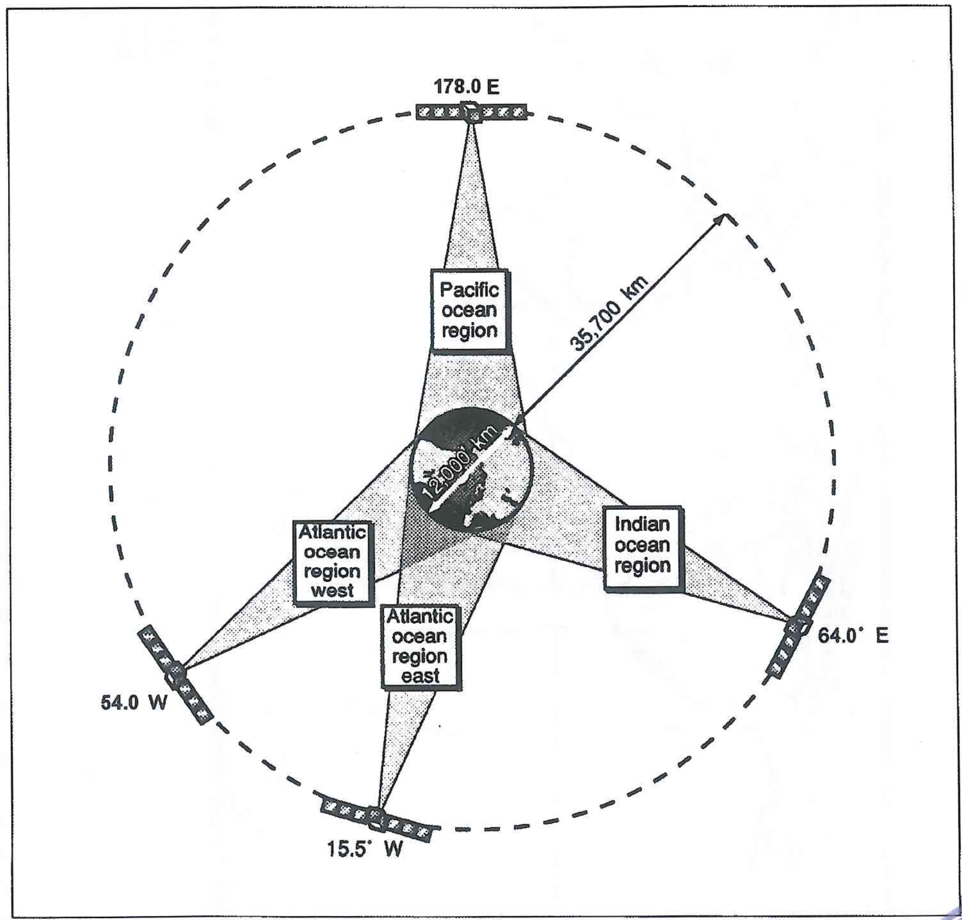


Figure 11.5 Earth total coverage from Inmarsat's four geostationary satellite configuration. (Reproduced courtesy of Inmarsat.)

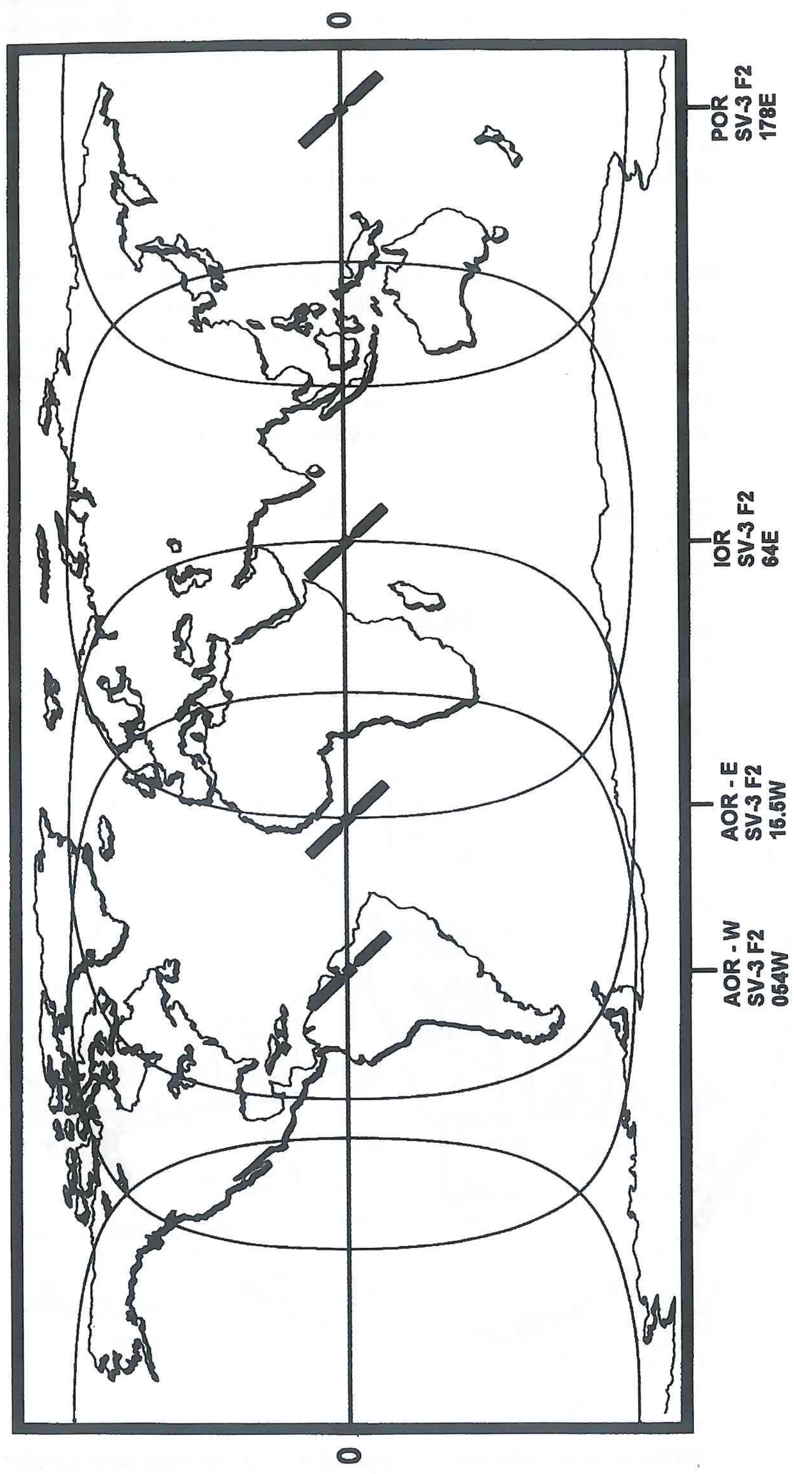


Figure 11.6 Footprint coverage of the earth's surface from Inmarsat-3 four geostationary satellites. (Reproduced courtesy of Inmarsat.)

- 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

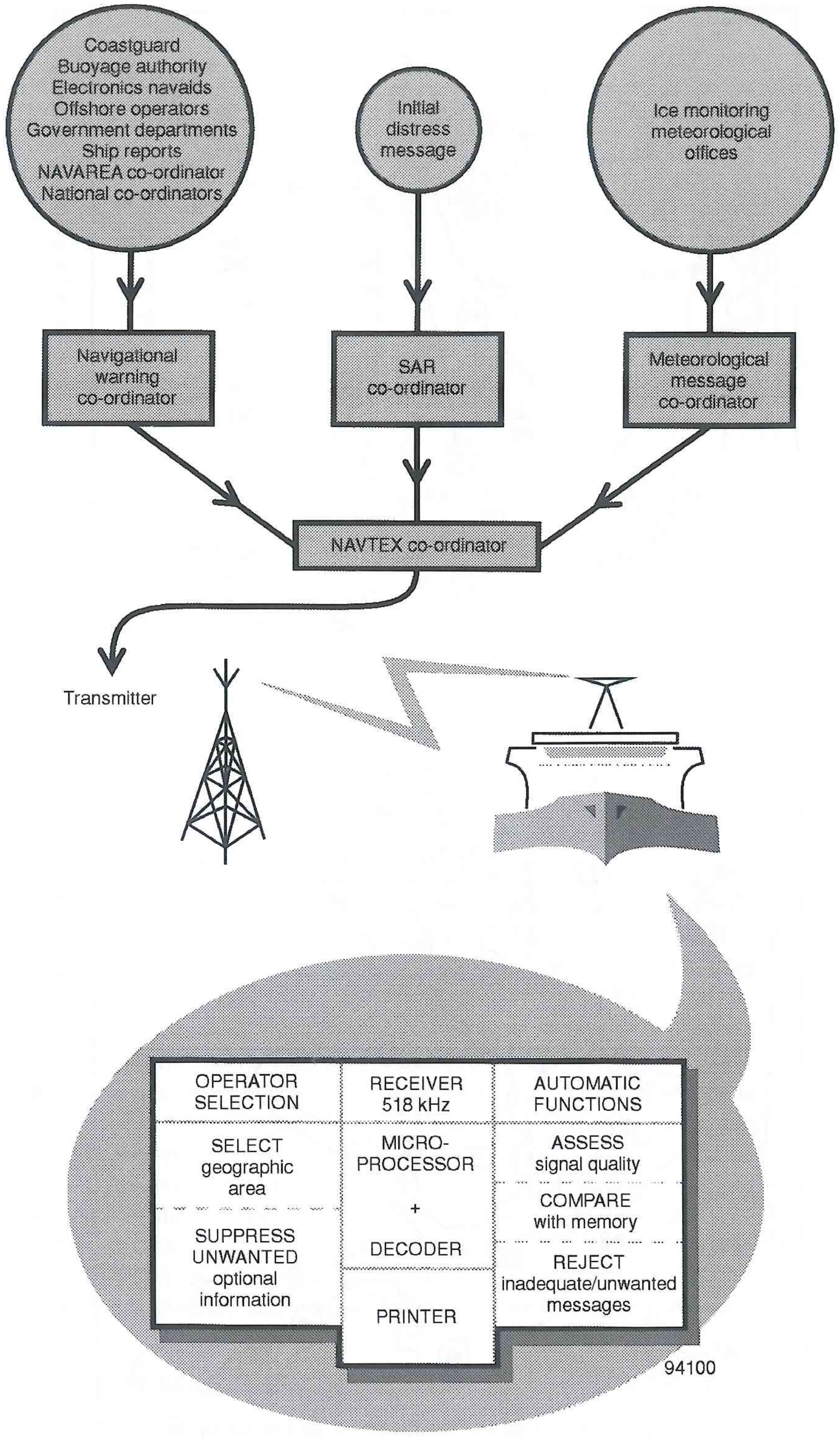


Figure 11.7 Structure of the NAVTEX service. (Reproduced courtesy of the IMO.)

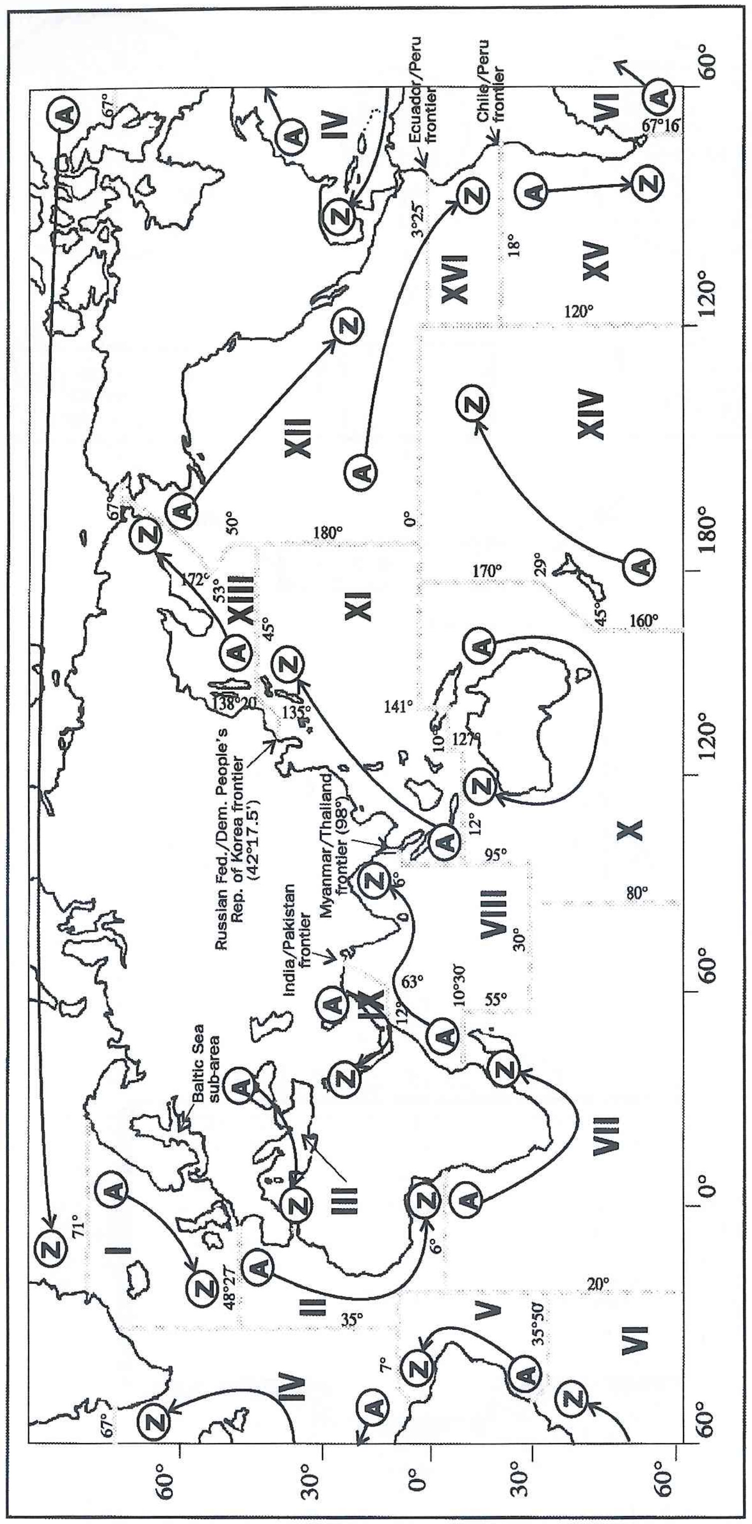


Figure 11.8 NAVAREAS of the World Wide Navigational Warnings Service (WWNWS) showing the basic scheme for allocation of transmitter identification characteristics. (Reproduced courtesy of the IMO.)

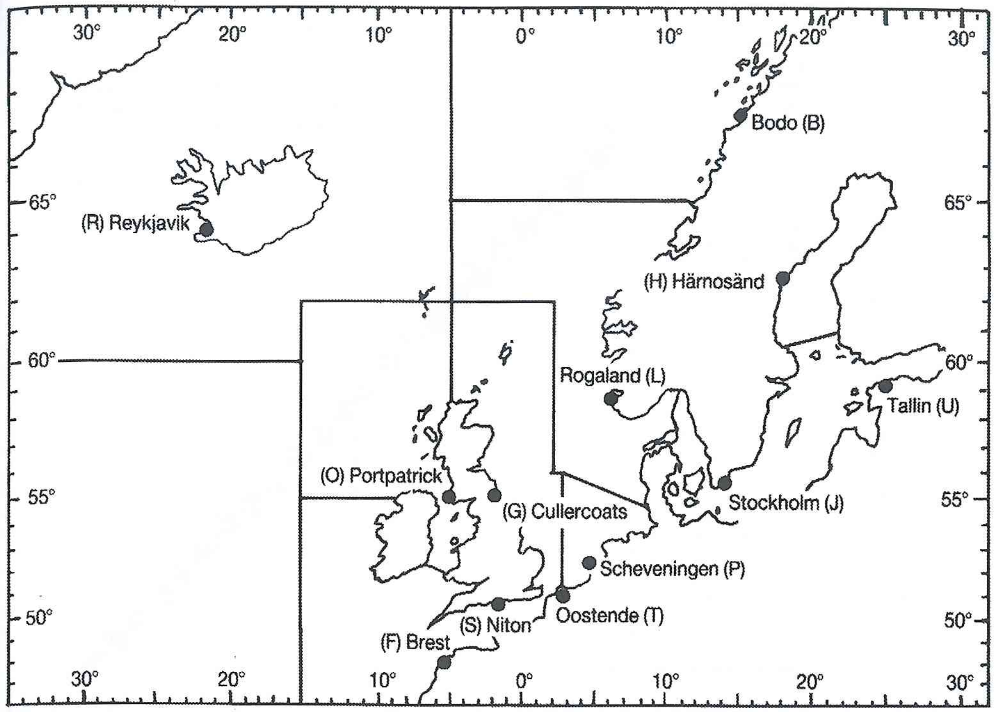


Figure 11.9 NAVTEX coverage areas within NAVAREA1.

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.

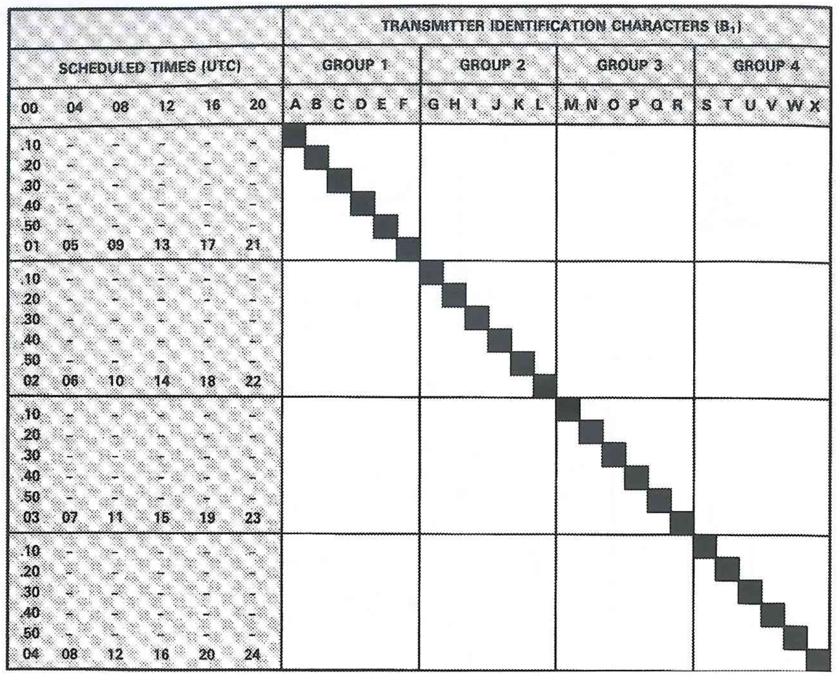


Figure 11.10 Scheme for the allocation of transmission schedules. (Reproduced courtesy of the IMO.)

- 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.

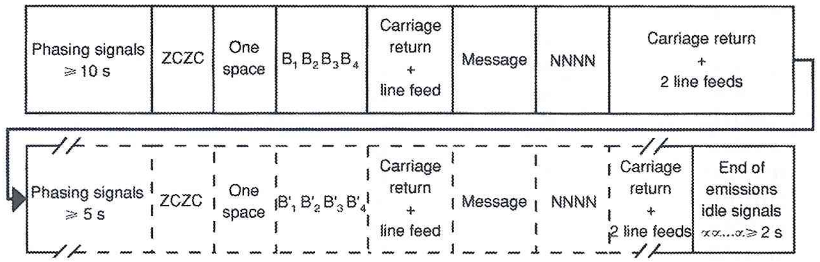


Figure 11.11 Data format of NAVTEX transmissions. (Reproduced courtesy of the IMO.)

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

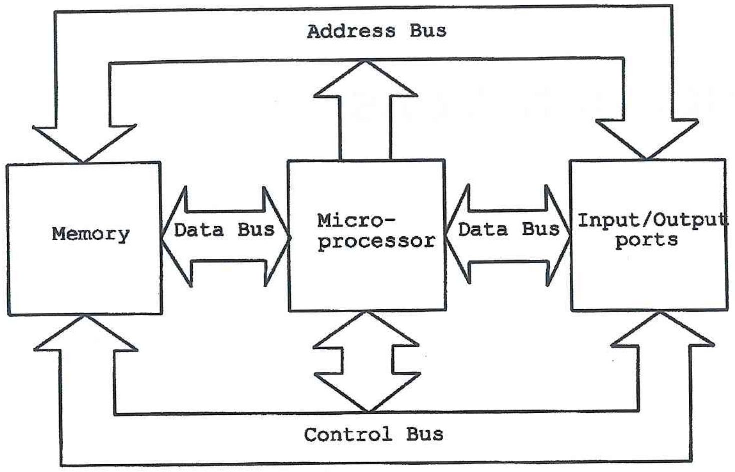


Figure A1.1 A basic microcomputer system.

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/$\overline{W}$) 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., $\overline{W}$, 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.

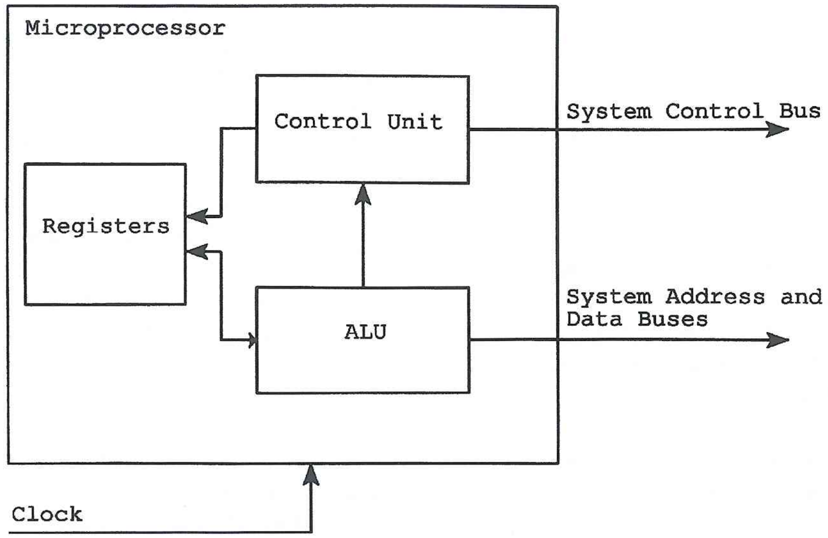


Figure A1.2 Block diagram of a microprocessor.

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.

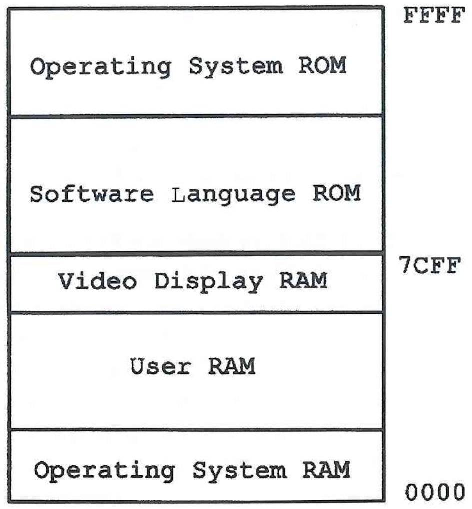


Figure A1.3 Memory map for a system with a 16-bit address bus.

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.

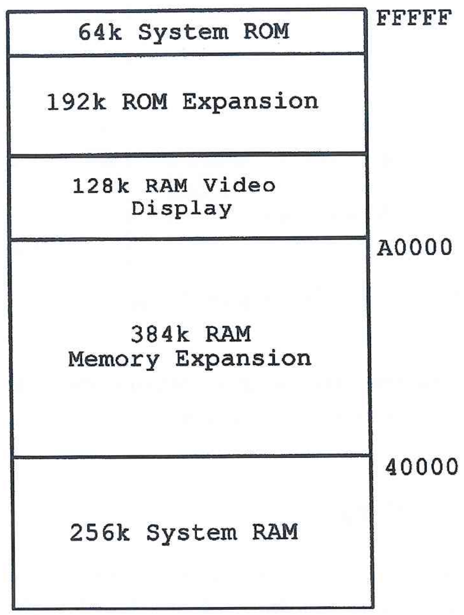


Figure A1.4 Memory map for a system with a 20-bit address bus.

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.

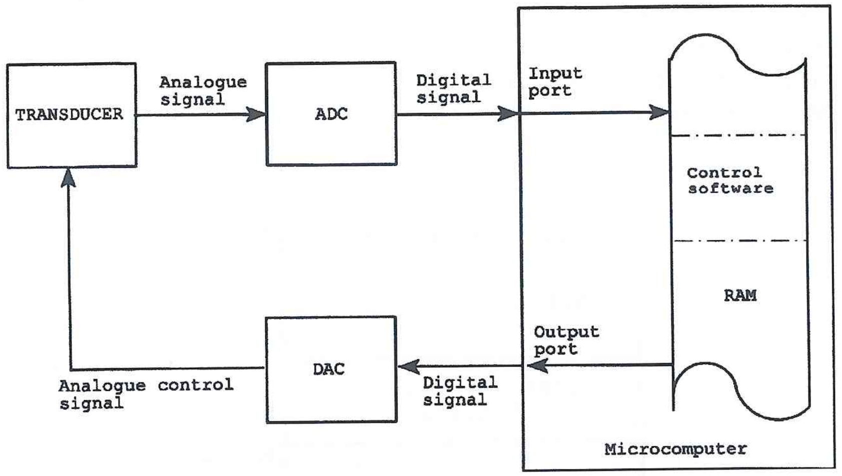


Figure A1.5 Control of an external transducer parameter by a microcomputer.

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.

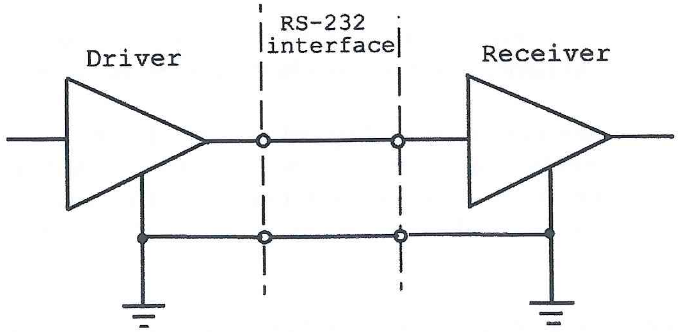


Figure A3.1 Driver and receiver circuit connected via an RS-232 interface.

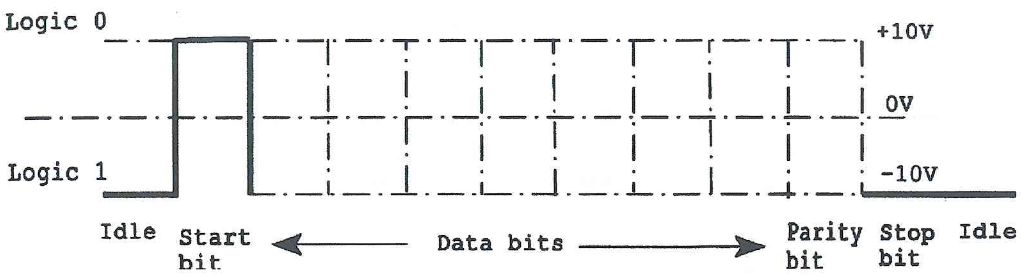


Figure A3.2 Typical arrangement for the transmission of an ASCII character using the RS-232 standard.

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.

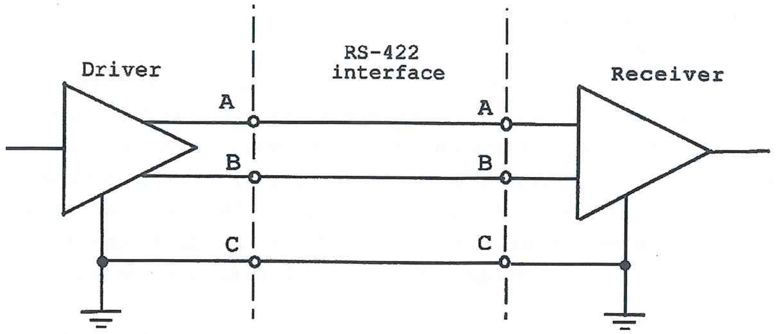


Figure A3.3 Driver and receiver circuit connected via an RS-422 interface.

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.

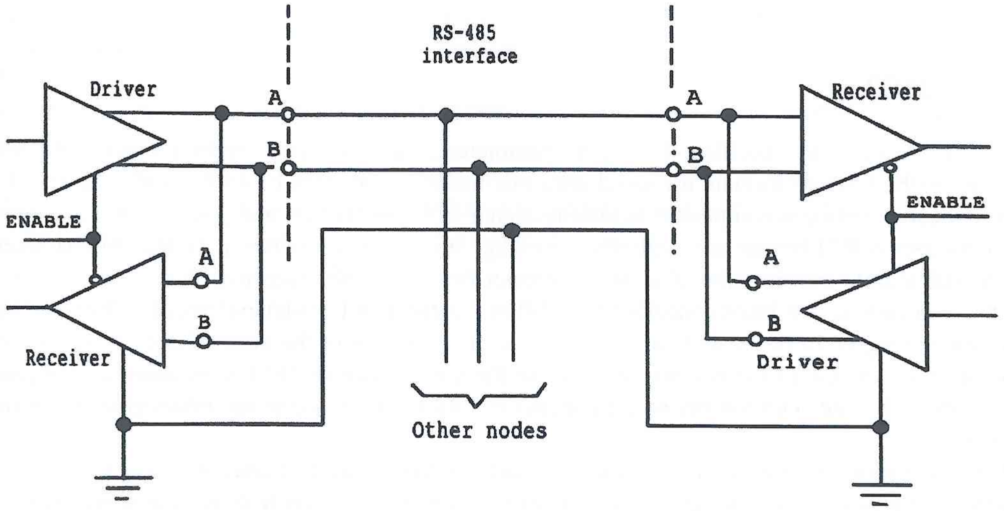


Figure A3.4 Typical arrangement for an RS-485 two-wire multidrop network.

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 opto-isolator 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 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