CHAPTER 21

RADIO WAVES

ELECTROMAGNETIC WAVE PROPAGATION

2100. Source of Radio Waves

Consider electric current as a flow of electrons through a conductor between points of differing potential/voltage/electric field. A **direct current** flows continuously in one direction. This occurs when the polarity of the electromotive force (EMF, or voltage) causing the electron flow is constant, such as is the case with a battery. If, however, the current is induced by the relative motion between a conductor and a magnetic field, as is the case in a rotating machine (**motor/generator**) or an **electrical oscillator**, signal generator, or **radio transmitter**, then the resulting current changes direction in the conductor as the polarity of the electromotive force changes with a period (or frequency) that ranges from fractions of a hertz (cycles per second) to hundreds of billions of hertz (i.e., gigahertz). This is known as **alternating current**.

The energy contained in the current flowing through the conductor due to a gradient of voltage is either dissipated as heat (an energy loss proportional to both the current flowing through the conductor and the conductor's resistance) or stored in an electromagnetic field oriented symmetrically about the conductor. The orientation of this field is a function of the polarity of the source producing the current. When the current is removed from the conductor, this electromagnetic field will, after a finite time associated with the speed of light, collapse back into the conductor.

What would occur should the polarity of the current source supplying the wire be reversed at a rate which exceeds the finite amount of time required for the electromagnetic field to collapse back upon the wire? In this case, another magnetic field, proportional in strength but exactly opposite in magnetic orientation to the initial field, sometimes referred to as a "back EMF," particularly when associated with an inductor or coil, will be formed upon the wire. This time-varying behavior of the combined electric and magnetic fields creates an electromagnetic wave that propagates, according to Maxwell's famous equations, into space. This is the basic principle of a radio antenna, which transmits a wave at a frequency proportional to the rate of current reversal and at a speed equal, in vacuum, to the speed of light. In materials, such as the dielectrics associated with coaxial cables, the waves travel at a velocity that can be considerably slower than the speed of light in vacuum.

2101. Radio Wave Terminology

The magnetic field strength in the vicinity of a conductor is directly proportional to the magnitude of the current flowing through the conductor. Recall the discussion of alternating current above. A rotating generator produces current in the form of a sine wave. That is, the magnitude of the current varies as a function of the relative position of the rotating conductor and the stationary magnetic field used to induce the current. The current starts at zero, increases to a maximum as the rotor completes one quarter of its revolution, and falls to zero when the rotor completes one half of its revolution. The current then approaches a negative maximum; then it once again returns to zero. This cycle can be represented by a sinusoidal function of time. Note that the electromagnetic waves described above can be represented as sinusoidal functions of both time and position.

The relationship between the current and the magnetic field strength induced in the conductor through which the current is flowing is shown in Figure 2101. Recall from the discussion above that this field strength is proportional to the magnitude of the current; that is, if the current is represented by a sine wave function, then so too will be the magnetic field strength resulting from that current. This characteristic shape of the field strength curve has led to the use of the term "wave" when referring to electromagnetic propagation. The maximum displacement of a peak from zero is called the **amplitude**. The forward side of any wave
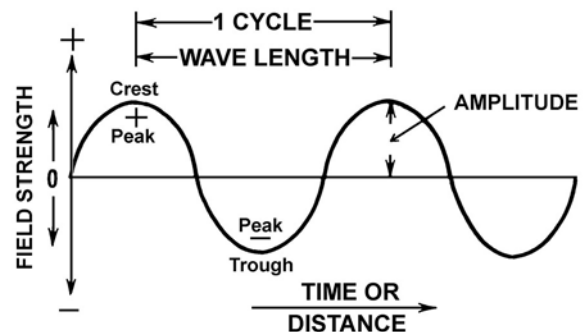

Figure 2101. Radio wave terminology.

is called the **wave front**. For a non-directional antenna, each wave proceeds outward as an expanding sphere (or hemisphere).

One **cycle** is a complete sequence of values, as from crest to crest. The distance traveled by the energy during one cycle is the **wavelength**, usually expressed in metric units (meters, centimeters, etc.). The number of cycles repeated during unit time (usually 1 second) is the **frequency**. This is given in **hertz** (cycles per second). A kilohertz (kHz) is 1,000 cycles per second. A megahertz (MHz) is 1,000,000 cycles per second. Wavelength and frequency are inversely proportional.

The **phase** of a wave is the amount by which the cycle has progressed from a specified origin of time or position. For most purposes it is stated in radians or degrees, a complete cycle being considered 360°. Generally, the origin is not important, with the principal interest being the phase relative to that of some other wave. Thus, two waves having crests 1/4 cycle apart are said to be 90° "out of phase," often being referred to as being in "phase quadrature". If the crest of one wave occurs at the trough of another, the two are 180° out of phase, and are of "opposite polarity."

2102. The Electromagnetic Spectrum

The entire range of electromagnetic radiation frequencies is called the **electromagnetic spectrum**. The frequency range suitable for radio transmission, the **radio spectrum**, extends from 10 kHz to 300,000 MHz, or 300 GHz. It is divided into a number of bands, as shown in Table 2102.

Below the radio spectrum, but overlapping it, is the audio frequency band, extending from 20 to 20,000 Hz. Above the radio spectrum are infrared (often associated with heat), the visible spectrum (light in its various colors), ultraviolet, X-rays, gamma rays, and cosmic rays. These are included in Table 2102. Waves shorter than 30 centimeters are usually called **microwaves**.

Within the frequencies from 1 to 40 GHz (1,000 to 40,000 MHz), additional bands are defined as follows:

Band	Abbreviation	Range of frequency	Range of wavelength
Audio frequency	AF	20 to 20,000 Hz	15,000,000 to 15,000 m
Radio frequency	RF	10 kHz to 300,000 MHz	30,000 m to 0.1 cm
Very low frequency	VLF	10 to 30 kHz	30,000 to 10,000 m
Low frequency	LF	30 to 300 kHz	10,000 to 1,000 m
Medium frequency	MF	300 to 3,000 kHz	1,000 to 100 m
High frequency	HF	3 to 30 MHz	100 to 10 m
Very high frequency	VHF	30 to 300 MHz	10 to 1 m
Ultra high frequency	UHF	300 to 3,000 MHz	100 to 10 cm
Super high frequency	SHF	3,000 to 30,000 MHz	10 to 1 cm
Extremely high frequency	EHF	30,000 to 300,000 MHz	1 to 0.1 cm
Heat and infrared		10^6 to 3.9×10^8 MHz*	0.03 to 7.6×10^{-5} cm*
Visible spectrum		3.9×10^8 to 7.9×10^8 MHz*	7.6×10^{-5} to 3.8×10^{-5} cm*
Ultraviolet		7.9×10^8 to 2.3×10^{10} MHz*	3.8×10^{-5} to 1.3×10^{-6} cm*
X-rays		2.0×10^9 to 3.0×10^{13} MHz*	1.5×10^{-5} to 1.0×10^{-9} cm*
Gamma rays		2.3×10^{12} to 3.0×10^{14} MHz*	1.3×10^{-8} to 1.0×10^{-10} cm*
Cosmic rays		$>4.8 \times 10^{15}$ MHz	$<6.2 \times 10^{-12}$ cm

* values approximate.

Table 2102. Electromagnetic spectrum.

L-band: 1 to 2 GHz (1,000 to 2,000 MHz)

S-band: 2 to 4 GHz (2,000 to 4,000 MHz

C-band: 4 to 8 GHz (4,000 to 8,000 MHz)

X-band: 8 to 12.5 GHz (8,000 to 12,500 MHz)

Lower K-band: 12.5 to 18 GHz (12,500 to 18,000 MHz)

Upper K-band: 26.5 to 40 GHz (26,500 to 40,000 MHz)

Marine radar systems commonly operate in the L, S and X bands, while satellite navigation system signals (e.g., GPS) are found in the L-band.

The break of the K-band into lower and upper ranges is necessary because the resonant frequency of water vapor occurs in the middle region of this band, and severe absorption of radio waves occurs in this part of the spectrum.

2103. Polarization

Radio waves produce both electric and magnetic fields. The direction of the electric component of the field is called the **polarization** of the electromagnetic field. Thus, if the electric component is vertical, the wave is said to be "vertically polarized," and if horizontal, "horizontally polarized." If the horizontal and vertical components of an electric field are equal, and in phase, the polarization is sometimes called "slant right" or "slant left". If the two polarizations vary in phase by 90°, in either time or position, the polarization becomes left hand circularized or right hand circularized.

A radio wave traveling through space may be polarized in any direction. One traveling along the surface of the Earth is always vertically polarized because the Earth, a moderate conductor, "short-circuits" any horizontal component. The magnetic field and the electric field of an electromagnetic wave are always mutually perpendicular.

2104. Reflection

When radio waves strike a surface, the surface reflects them in the same manner as light waves. Radio waves of all frequencies are reflected by the surface of the Earth. The strength of the reflected wave depends upon angle of incidence (the angle between the incident ray and the horizontal), type of polarization, frequency, reflecting properties of the surface, and divergence of the reflected ray. Lower frequencies penetrate the Earth's surface more than higher ones. At very low frequencies, usable radio signals can be received some distance below the surface of the sea.

A phase change occurs when a wave is reflected from the surface of the Earth. The amount of the change varies with the conductivity of the Earth and the polarization of the wave, reaching a maximum of 180° for a horizontally polarized wave reflected from sea water (considered to have infinite conductivity).

When direct waves (those traveling from transmitter to receiver in a relatively straight line, without reflection) and reflected waves arrive at a receiver, the total signal is the vector sum of the two. If the signals are in phase, they reinforce each other, producing a stronger signal. If there is a phase difference, the signals tend to cancel each other, the cancellation being complete if the phase difference is 180° and the two signals have the same amplitude. This interaction of waves is called **wave interference,** and the reflected wave is called a **multipath wave**.

A phase difference may occur because of the change of phase of a reflected wave, or because of the longer path it follows. The second effect decreases with greater distance between transmitter and receiver, for under these conditions the difference in path lengths is smaller.

At lower frequencies there is no practical solution to interference caused in this way. For VHF and higher frequencies, the condition can be improved by elevating the antenna, if the wave is vertically polarized. Additionally, interference at higher frequencies can be more nearly eliminated because of the greater ease of beaming the signal to avoid reflection.

Reflections may also occur from mountains, trees, and other obstacles. Such reflection is negligible for lower frequencies, but becomes more prevalent as frequency increases. In radio communication, it can be reduced by using directional antennas, but this solution is not always available for navigational systems.

Various reflecting surfaces occur in the atmosphere. At high frequencies, reflections take place from rain. At still higher frequencies, reflections are possible from clouds, particularly rain clouds. Reflections may even occur at a sharply defined boundary surface between air masses, as when warm, moist air flows over cold, dry air. When such a surface is roughly parallel to the surface of the Earth, radio waves may travel for greater distances than normal. The principal source of reflection in the atmosphere is the ionosphere.

Figure 2104a. Specular reflection.

Figure 2104b. Specular reflection from bird bath.

Figure 2104a and Figure 2104b show, at optical frequencies (i.e., light), the reflection of electromagnetic waves from water. The first figure shows direct and the reflected views of the sun, in this case known as specular reflection. The reflection of the sun is evident in even a small body of water, in this case a bird-bath. Radio waves exhibit this same behavior.

2105. Refraction

Refraction of radio waves is similar to that of light waves. Thus, as a signal passes from air of one density to that of a different density, the direction of travel is altered. The principal cause of refraction in the atmosphere is the difference in temperature and pressure occurring at various heights and in different air masses of different densities.

Refraction occurs at all frequencies, but below 30 MHz the effect is small as compared with ionospheric effects, diffraction, and absorption. At higher frequencies, refraction in the lower layer of the atmosphere extends the radio horizon to a distance about 15 percent greater than the visible horizon. The effect is the same as if the radius of the Earth were about one-third greater than it is and there were no refraction.

Sometimes the lower portion of the atmosphere becomes stratified. This stratification results in nonstandard temperature and moisture changes with height. If there is a marked temperature inversion or a sharp decrease in water vapor content with increased height, a horizontal radio duct may be formed. High frequency radio waves traveling horizontally within the duct are refracted to such an extent that they remain within the duct, following the curvature of the Earth for phenomenal distances. This is called **super-refraction**. Maximum results are obtained when both transmitting and receiving antennas are within the duct. There is a lower limit to the frequency affected by ducts. It varies from about 200 MHz to more than 1,000 MHz.

At night, surface ducts may occur over land due to cooling of the surface. At sea, surface ducts about 50 feet thick may occur at any time in the trade wind belt. Surface ducts 100 feet or more in thickness may extend from land out to sea when warm air from the land flows over the cooler ocean surface. Elevated ducts from a few feet to more than 1,000 feet in thickness may occur at elevations of 1,000 to 5,000 feet, due to the settling of a large air mass. This is a frequent occurrence in Southern California and certain areas of the Pacific Ocean.

A bending in the horizontal plane occurs when a groundwave crosses a coast at an oblique angle. This is due to a marked difference in the conducting and reflecting properties of the land and water over which the wave travels. The effect is known as **coastal refraction** or **land effect**.

2106. The Ionosphere

Since an atom normally has an equal number of negatively charged electrons and positively charged protons, it is electrically neutral. An **ion** is an atom or group of atoms which has become electrically charged, either positively or negatively, by the loss or gain of one or more electrons.

Loss of electrons may occur in a variety of ways. In the atmosphere, ions are usually formed by collision of atoms with rapidly moving particles, or by the action of cosmic rays or ultraviolet light. In the lower portion of the atmosphere, recombination soon occurs, leaving a small percentage of ions. In "thin" atmosphere far above the surface of the Earth, where the air pressure is considerably lower than that at sea level, atoms are widely separated and a large number of ions may be present. The region of numerous positive and negative ions and unattached electrons is called the **ionosphere**. The extent of ionization depends upon the kinds of atoms present in the atmosphere, the density of the atmosphere, and the position relative to the sun (time of day and season). After sunset, ions and electrons recombine faster than they are separated, decreasing the ionization of the atmosphere.

An electron can be separated from its atom only by the application of greater energy than that holding the electron. Since the energy of the electron depends primarily upon the kind of an atom of which it is a part, and its position relative to the nucleus of that atom, different kinds of radiation may cause ionization of different substances.

In the outermost regions of the atmosphere, the density is so low that oxygen exists largely as separate atoms, rather than combining as molecules as it does nearer the surface of the Earth. At great heights the energy level is low and ionization from solar radiation is intense. This is known as the **F layer**. Above this level the ionization decreases because of the lack of atoms to be ionized. Below this level it decreases because the ionizing agent of appropriate energy has already been absorbed. During daylight, two levels of maximum F ionization can be detected, the F_2
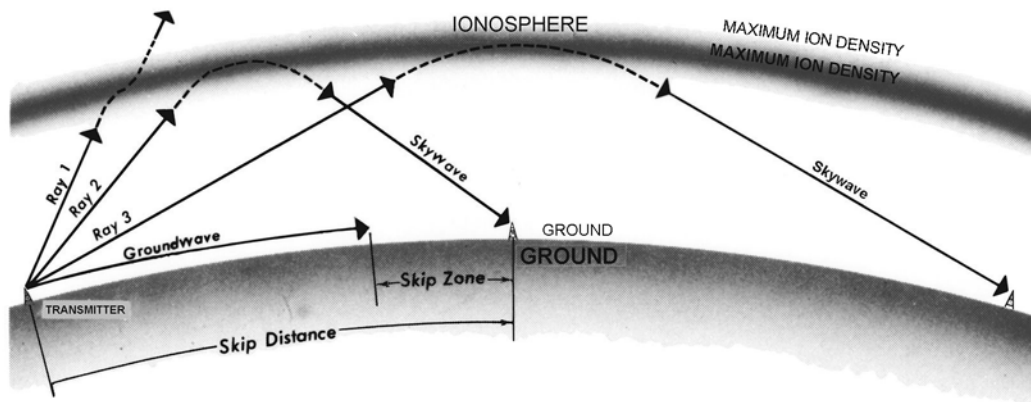
Figure 2106. The effect of the ionosphere on radio waves.

layer at about 125 statute miles above the surface of the Earth, and the F_1 layer at about 90 statute miles. At night, these combine to form a single F layer.

At a height of about 60 statute miles, the solar radiation not absorbed by the F layer encounters, for the first time, large numbers of oxygen molecules. A new maximum ionization occurs, known as the **E layer**. The height of this layer is quite constant, in contrast with the fluctuating F layer. At night the E layer becomes weaker by two orders of magnitude.

Below the E layer, a weak D layer forms at a height of about 45 statute miles, where the incoming radiation encounters ozone for the first time. The D layer is the principal source of absorption of HF waves, and of reflection of LF and VLF waves during daylight.

2107. The Ionosphere and Radio Waves

When a radio wave encounters an atom or molecule having an electric charge, it causes that atom/molecule to vibrate. The vibrating particle absorbs electromagnetic energy from the radio wave and can re-radiate it. The net effect is a change of polarization and an alteration of the path of the wave. That portion of the wave in a more highly ionized region travels faster, causing the wave front to tilt and the wave to be directed toward a region of less intense ionization.

Refer to Figure 2106, in which a single layer of the ionosphere is considered. Ray 1 enters the ionosphere at such an angle that its path is altered, but it passes through and proceeds outward into space. As the angle with the horizontal decreases, a critical value is reached where ray 2 is bent or reflected back toward the Earth. As the angle is still further decreased, such as at ray 3, the return to Earth occurs at a greater distance from the transmitter.

A wave reaching a receiver by way of the ionosphere is called a **skywave**. This expression is also appropriately applied to a wave reflected from an air/mass boundary. In common usage, however, it is generally associated with the ionosphere. The wave which travels along the surface of the Earth is called a **groundwave**. At angles greater than the critical angle, no skywave signal is received. Therefore, there is a minimum distance from the transmitter at which sky waves can be received. This is called the **skip distance**, shown in Figure 2106. If the groundwave extends for less distance than the skip distance, a skip zone occurs, in which no usable signal is received.

The critical radiation angle depends upon the intensity of ionization, and the frequency of the radio wave. As the frequency increases, the angle becomes smaller. At frequencies greater than about 30 MHz, virtually all of the energy penetrates through or is absorbed by the ionosphere. Therefore, at any given receiver there is a maximum usable frequency if sky waves are to be utilized. The strongest signals are received at or slightly below this frequency. There is also a lower practical frequency beyond which signals are too weak to be of value. Within this band the optimum frequency can be selected to give best results. It cannot be too near the maximum usable frequency because this frequency fluctuates with changes of intensity within the ionosphere. During magnetic storms the ionosphere density decreases. The maximum usable frequency decreases, and the lower usable frequency increases. The band of usable frequencies is thus narrowed. Under extreme conditions it may be completely eliminated, isolating the radio receiver and causing a radio blackout.

Skywave signals reaching a given receiver may arrive by any of several paths, as shown in Figure 2107. A signal which undergoes a single reflection is called a "one-hop" signal, one which undergoes two reflections with a ground reflection between is called a "two-hop" signal, etc. A "multi-hop" signal undergoes several reflections. The layer at which the reflection occurs is usually indicated, also, as "one-hop E," "two-hop F," etc.
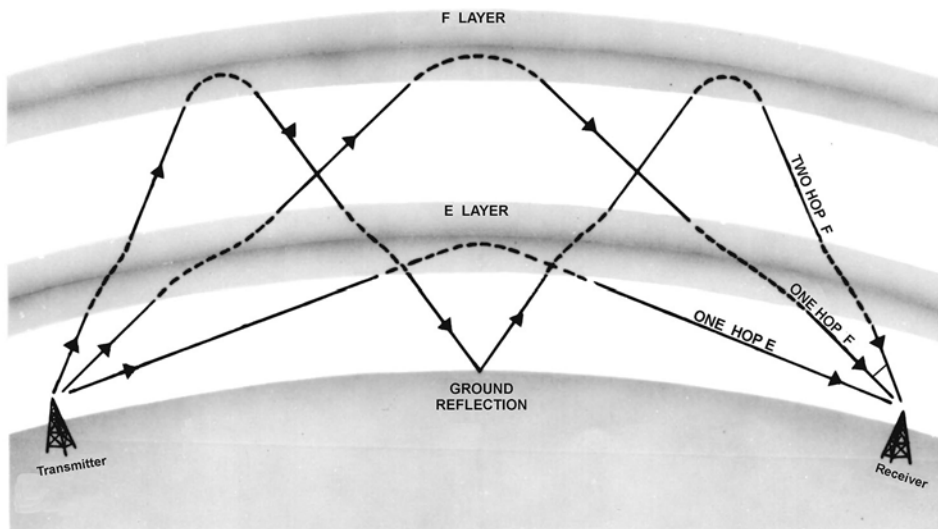
Figure 2107. Various paths by which a skywave signal might be received.

Because of the different paths and phase changes occurring at each reflection, the various signals arriving at a receiver have different phase relationships. Since the density of the ionosphere is continually fluctuating, the strength and phase relationships of the various signals may undergo an almost continuous change. Thus, the various signals may reinforce each other at one moment and cancel each other at the next, resulting in fluctuations of the strength of the total signal received. This is called **fading**. This phenomenon may also be caused by interaction of components within a single reflected wave, or changes in its strength due to changes in the reflecting surface. Ionospheric changes are associated with fluctuations in the radiation received from the sun, since this is the principal cause of ionization. Signals from the F layer are particularly erratic because of the rapidly fluctuating conditions within the layer itself.

The maximum distance at which a one-hop E signal can be received is about 1,400 miles. At this distance the signal leaves the transmitter in approximately a horizontal direction. A one-hop F signal can be received out to about 2,500 miles. At low frequencies, ground waves extend out for great distances.

A skywave may undergo a change of polarization during reflection from the ionosphere, accompanied by an alteration in the direction of travel of the wave. This is called **polarization error**. Near sunrise and sunset, when rapid changes are occurring in the ionosphere, reception may become erratic and polarization error a maximum. This is called **night effect**.

2108. Diffraction

When a radio wave encounters an obstacle, its energy is reflected or absorbed, causing a shadow beyond the obstacle. However, some energy does enter the shadow area because of diffraction. This is explained by Huygens' principle, which states that every point on the surface of a wave front is a source of radiation, transmitting energy in all directions ahead of the wave. No noticeable effect of this principle is observed until the wave front encounters an obstacle, which intercepts a portion of the wave. From the edge of the obstacle, energy is radiated into the shadow area, and also outside of the area. The latter interacts with energy from other parts of the wave front, producing alternate bands in which the secondary radiation reinforces or tends to cancel the energy of the primary radiation. Thus, the practical effect of an obstacle is a greatly reduced signal strength in the shadow area, and a disturbed pattern for a short distance outside the shadow area. This is illustrated in Figure 2108.

The amount of diffraction is inversely proportional to the frequency, being greatest at very low frequencies.

2109. Absorption and Scattering

The amplitude of a radio wave's electric or magnetic field expanding outward through space varies inversely with distance, weakening with increased distance. The combined electric and magnetic fields yield the **power flux density**, or pfd, which varies in inverse proportion to the square of distance. The decrease of strength with distance, due to spherical spreading of the electromagnetic wave, is called **attenuation**. Under certain conditions, typically due to absorption of energy in air, the attenuation is greater than in free space.

A wave traveling along the surface of the Earth loses a certain amount of energy as the wave is diffracted downward and absorbed by the Earth. As a result of this absorption, the remainder of the wave front tilts downward, resulting in further absorption by the Earth. Attenuation is greater over a surface which is a poor conductor. Relatively little absorption occurs over sea water, which is an excellent conductor at low frequencies, and low frequency ground waves travel great distances over water.
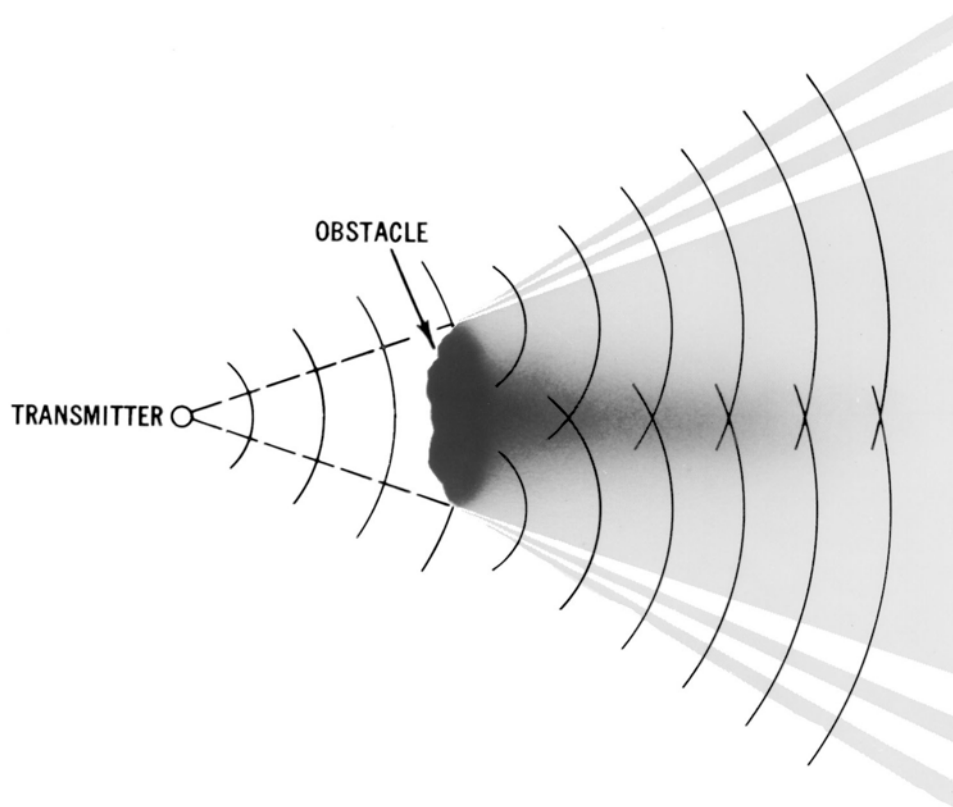
Figure 2108. Diffraction.

A skywave suffers an attenuation loss in its encounter with the ionosphere. The amount depends upon the height and composition of the ionosphere as well as the frequency of the radio wave. Maximum ionospheric absorption occurs at about 1,400 kHz.

In general, atmospheric absorption increases with frequency. It is a problem only in the SHF and EHF frequency range. At these frequencies, attenuation is further increased by scattering due to reflection by oxygen, water vapor, water droplets, and rain in the atmosphere.

2110. Noise

Unwanted signals in a receiver are called **interference**. The intentional production of such interference to obstruct communication is called **jamming**. Unintentional interference is called **noise**.

Noise may originate within the receiver. Hum is usually the result of induction from neighboring circuits carrying alternating current. Irregular crackling or sizzling sounds may be caused by poor contacts or faulty components within the receiver. Stray currents in normal components cause some noise. This source sets the ultimate limit of sensitivity that can be achieved in a receiver. It is the same at any frequency.

Noise originating outside the receiver may be either man-made or natural. Man-made noises originate in electrical appliances, motor and generator brushes, ignition systems, and other sources of sparks which transmit electromagnetic signals that are picked up by the receiving antenna.

Natural noise is caused principally by discharge of static electricity in the atmosphere. This is called **atmospheric noise**, atmospherics, or static. An extreme example is a thunderstorm. An exposed surface may acquire a considerable charge of static electricity. This may be caused by friction of water or solid particles blown against or along such a surface. It may also be caused by splitting of a water droplet which strikes the surface, one part of the droplet requiring a positive charge and the other a negative charge. These charges may be transferred to the surface. The charge tends to gather at points and ridges of the conducting surface, and when it accumulates to a sufficient extent to overcome the insulating properties of the atmosphere, it discharges into the atmosphere. Under suitable conditions this becomes visible and is known as St. Elmo's fire, which is sometimes seen at mastheads, the ends of yardarms, etc.

Atmospheric noise occurs to some extent at all frequencies but decreases with higher frequencies. Above about 30 MHz it is not generally a problem.

2111. Antenna Characteristics

Antenna design and orientation have a marked effect upon radio wave propagation. For a single-wire antenna, strongest signals are transmitted along the perpendicular to the wire, and virtually no signal in the direction of the wire. For a vertical antenna, the signal strength is the same in all horizontal directions. Unless the polarization undergoes a change during transit, the strongest signal received from a vertical transmitting antenna occurs when the receiving antenna is also vertical.

For lower frequencies the radiation of a radio signal takes place by interaction between the antenna and the ground. For a vertical antenna, efficiency increases with greater length of the antenna. For a horizontal antenna, efficiency increases with greater distance between antenna and ground. Near-maximum efficiency is attained when this distance is one-half the wavelength. This is the reason for elevating low frequency antennas to great heights. However, at the lowest frequencies, the required height becomes prohibitively great. At 10 kHz it would be about 8 nautical miles for a half-wavelength antenna. Therefore, lower frequency antennas of practical length are inherently inefficient. This is partly offset by the greater range of a low frequency signal of the same transmitted power as one of higher frequency.

At higher frequencies, the ground is not used, both conducting portions being included in a dipole antenna. Not only can such an antenna be made efficient, but it can also be made sharply directive, thus greatly increasing the strength of the signal transmitted in a desired direction.

The power received is inversely proportional to the square of the distance from the transmitter, as described previously, assuming there is no attenuation due to absorption or scattering.

2112. Range

The range at which a usable signal is received depends upon the power transmitted, the sensitivity of the receiver, frequency, route of travel, noise level, and perhaps other factors. For the same transmitted power, both the groundwave and skywave ranges are greatest at the lowest frequencies, but this is somewhat offset by the lesser efficiency of antennas for these frequencies. At higher frequencies, only direct waves are useful, and the effective range is greatly reduced. Attenuation, skip distance, ground reflection, wave interference, condition of the ionosphere, atmospheric noise level, and antenna design all affect the range at which useful signals can be received.

2113. Radio Wave Spectra

Frequency is an important consideration in radio wave propagation. The following summary indicates the principal effects associated with the various frequency bands, starting with the lowest and progressing to the highest usable radio frequency.

Very Low Frequency (VLF, 10 to 30 kHz): The VLF signals propagate between the bounds of the ionosphere and the Earth and are thus guided around the curvature of the Earth to great distances with low attenuation and excellent stability. Diffraction is maximum. Because of the long wavelength, large antennas are needed, and even these are inefficient, permitting radiation of relatively small amounts of power. Magnetic storms have little effect upon transmission because of the efficiency of the "Earth-ionosphere waveguide." During such storms, VLF signals may constitute the only source of radio communication over great distances. However, interference from atmospheric noise may be troublesome. Signals may be received from below the surface of the sea.

Low Frequency (LF, 30 to 300 kHz): As frequency is increased to the LF band and diffraction decreases, there is greater attenuation with distance, and range for a given power output falls off rapidly. However, this is partly offset by more efficient transmitting antennas. LF signals are most stable within groundwave distance of the transmitter. A wider bandwidth permits pulsed signals at 100 kHz. This allows separation of the stable groundwave pulse from the variable skywave pulse up to 1,500 km, and up to 2,000 km for over-water paths. The frequency for Loran C, which is being replaced by enhanced Loran, or eLoran, is in the LF band. This band is also useful for radio direction finding and time dissemination.

Medium Frequency (MF, 300 to 3,000 kHz): Ground waves provide dependable service, but the range for a given power is reduced greatly. This range varies from about 400 miles at the lower portion of the band to about 15 miles at the upper end for a transmitted signal of 1 kilowatt. These values are influenced, however, by the power of the transmitter, the directivity and efficiency of the antenna, and the nature of the terrain over which signals travel. Elevating the antenna to obtain direct waves may improve the transmission. At the lower frequencies of the band, skywaves are available both day and night. As the frequency is increased, ionospheric absorption increases to a maximum at about 1,400 kHz. At higher frequencies the absorption decreases, permitting increased use of sky waves. Since the ionosphere changes with the hour, season, and sunspot cycle, the reliability of skywave signals is variable. By careful selection of frequency, ranges of as much as 8,000 miles with 1 kilowatt of transmitted power are possible, using multihop signals. However, the frequency selection is critical. If it is too high, the signals penetrate the ionosphere and are lost in space. If it is too low, signals are too weak. In general, skywave reception is equally good by day or night, but lower frequencies are needed at night. The standard broadcast band for commercial stations (535 to 1,605 kHz) is in the MF band.

High Frequency (HF, 3 to 30 MHz): As in the higher band, the groundwave range of HF signals is limited to a

few miles, but the elevation of the antenna may increase the direct-wave distance of transmission. Also, the height of the antenna does have an important effect upon skywave transmission because the antenna has an "image" within the conducting Earth. The distance between antenna and image is related to the height of the antenna, and this distance is as critical as the distance between elements of an antenna system. Maximum usable frequencies fall generally within the HF band. By day this may be 10 to 30 MHz, but during the night it may drop to 8 to 10 MHz. The HF band is widely used for ship-to-ship and ship-to-shore communication.

Very High Frequency (VHF, 30 to 300 MHz): Communication is limited primarily to the direct wave, or the direct wave plus a ground-reflected wave. Elevating the antenna to increase the distance at which direct waves can be used results in increased distance of reception, even though some wave interference between direct and ground-reflected waves is present. Diffraction is much less than with lower frequencies, but it is most evident when signals cross sharp mountain peaks or ridges. Under suitable conditions, reflections from the ionosphere are sufficiently strong to be useful, but generally they are unavailable. There is relatively little interference from atmospheric noise in this band. Reasonably efficient directional antennas are possible with VHF. The VHF band used for most communication.

Ultra High Frequency (UHF, 300 to 3,000 MHz): Skywaves are not used in the UHF band because the ionosphere is not sufficiently dense to reflect the waves, which pass through it into space. Ground waves and ground-reflected waves are used, although there is some wave interference. Diffraction is negligible, but the radio horizon extends about 15 percent beyond the visible horizon, due principally to refraction. Reception of UHF signals is virtually free from fading and interference by atmospheric noise. Sharply directive antennas can be produced for transmission in this band, which is widely used for ship-to-ship and ship-to-shore communication.

Super High Frequency (SHF, 3,000 to 30,000 MHz): In the SHF band, also known as the microwave or as the centimeter wave band, there are no sky waves, transmission being entirely by direct and ground-reflected waves. Diffraction and interference by atmospheric noise are virtually nonexistent. Highly efficient, sharply directive antennas can be produced. Thus, transmission in this band is similar to that of UHF, but with the effects of shorter waves being greater. Reflection by clouds, water droplets, dust particles, etc., increases, causing greater scattering, increased wave interference, and fading. The SHF band is used for marine navigational radar.

Extremely High Frequency (EHF, 30,000 to 300,000 MHz): The effects of shorter waves are more pronounced in the EHF band, transmission being free from wave interference, diffraction, fading, and interference by atmospheric noise. Only direct and ground-reflected waves are available. Scattering and absorption in the atmosphere are pronounced and may produce an upper limit to the frequency useful in radio communication.

2114. Regulation of Frequency Use

While the characteristics of various frequencies are important to the selection of the most suitable one for any given purpose, these are not the only considerations. Confusion and extensive interference would result if every user had complete freedom of selection. Some form of regulation is needed. The allocation of various frequency bands to particular uses is a matter of international agreement. Within the United States, the Federal Communications Commission (FCC) has responsibility for authorizing civil use of particular frequencies. However, military and government use is reconciled with the FCC by the National Telecommunication and Information Administration (NTIA) of the Department of Commerce. In some cases a given frequency is allocated to several widely separated transmitters, but only under conditions which minimize interference, such as during daylight hours. Interference between stations is further reduced by the use of channels, each of a narrow band of frequencies. Assigned frequencies are separated by an arbitrary band of frequencies that are not authorized for use. In the case of radio aids to navigation and ship communications bands of several channels are allocated, permitting selection of band and channel by the user. The international allocation and sharing of radio frequencies is regulated by the International Telecommunication Union (ITU), in accordance with decisions made at World Radio Conferences, which are held every three to four years.

2115. Types of Radio Transmission

A series of waves transmitted at constant frequency and amplitude is called a **continuous wave**. This cannot be heard except at the very lowest radio frequencies, when it may produce, in a receiver, an audible hum of high pitch.

Although a continuous wave may be used directly, as in radiodirection finding, it is more commonly modified in some manner. This is called **modulation**. When this occurs, the continuous wave serves as a carrier wave for information. Any of several types of modulation may be used.

In **amplitude modulation (AM)** the amplitude of the carrier wave is altered in accordance with the amplitude of a modulating wave, usually of audio frequency, as shown in Figure 2115a. In the receiver the signal is demodulated by removing the modulating wave and converting it back to its original form. This form of modulation is widely used in voice radio, as in the standard broadcast band of commercial broadcasting.

If the frequency instead of the amplitude is altered in accordance with the amplitude of the impressed signal, as shown in Figure 2115a, **frequency modulation (FM)**
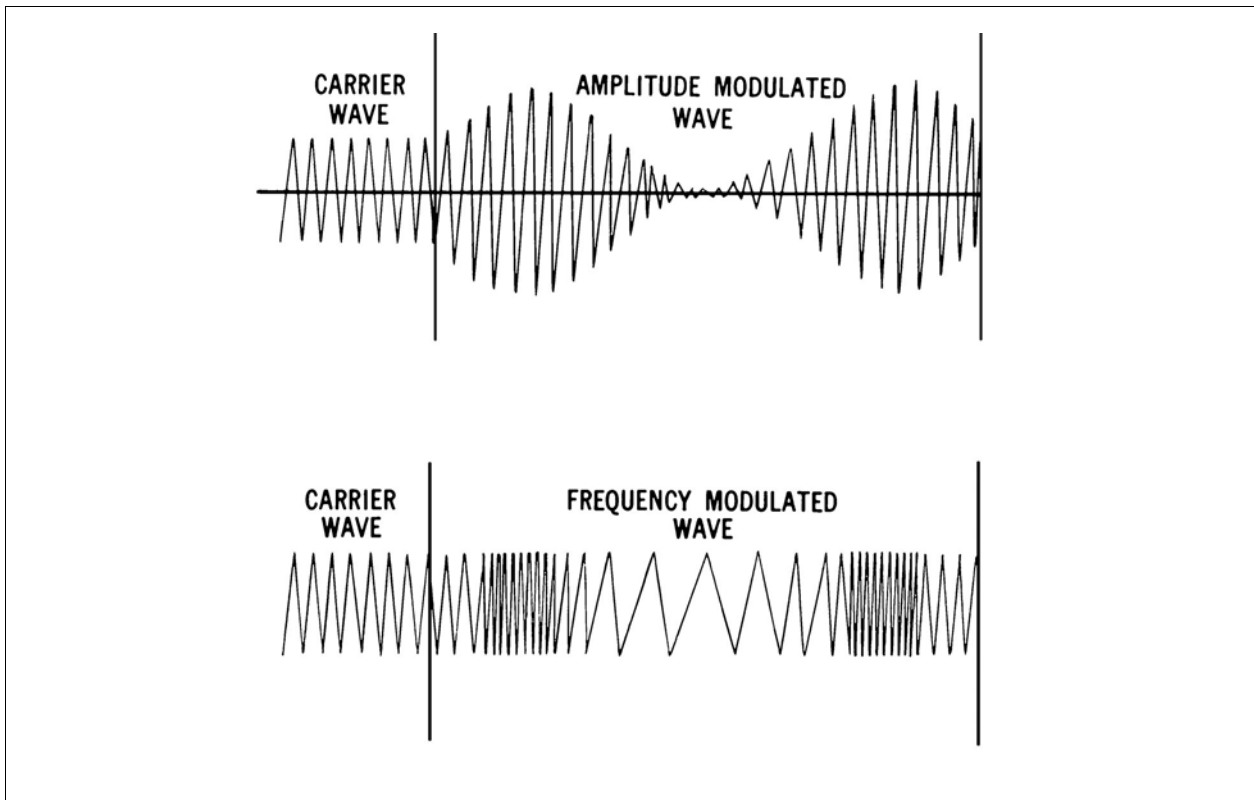
Figure 2115a. Amplitude modulation (upper figure) and frequency modulation (lower figure) by the same modulating wave.

Figure 2115b. Pulse modulation.

occurs. This is used for commercial FM radio broadcasts and the sound portion of television broadcasts.

Pulse modulation (PM) is somewhat different, there being no impressed modulating wave. In this form of transmission, very short bursts of carrier wave are transmitted, separated by relatively long periods of "silence," during which there is no transmission. This type of transmission, illustrated in Figure 2115b, is used in some common radio navigational aids, including radar, and eLORAN.

2116. Transmitters

A radio **transmitter** consists essentially of (1) a power supply to furnish direct current, (2) an oscillator to convert direct current into radio-frequency oscillations (the carrier wave), (3) a device to control the generated signal, and (4) an amplifier to increase the output of the oscillator. For some transmitters a microphone is used with a modulator and final amplifier to modulate the carrier wave. In addition, an antenna and ground (for lower frequencies) are needed to produce electromagnetic radiation. These components are illustrated in Figure 2116.

2117. Receivers

When a radio wave passes a conductor, a current is induced in that conductor. A radio receiver is a device which senses the power thus generated in an antenna, and transforms it into usable form. It is able to select signals of a single frequency (actually a narrow band of frequencies) from among the many which may reach the receiving antenna. The receiver is able to demodulate the signal and provide adequate amplification. The output of a receiver may be presented audibly by earphones or loudspeaker; or visually video display, or as digital data processed by a computer and either displayed or used to command and control other systems. In any case, the useful reception of radio signals requires three components: (1) an antenna, (2) a receiver, and (3) a display or data processing unit.

Radio receivers differ mainly in (1) frequency range, the range of frequencies to which they can be tuned; (2) selectivity, the ability to confine reception to signals of the desired frequency and avoid others of nearly the same frequency; (3) sensitivity, the ability to amplify a weak signal to usable strength against a background of noise; (4)
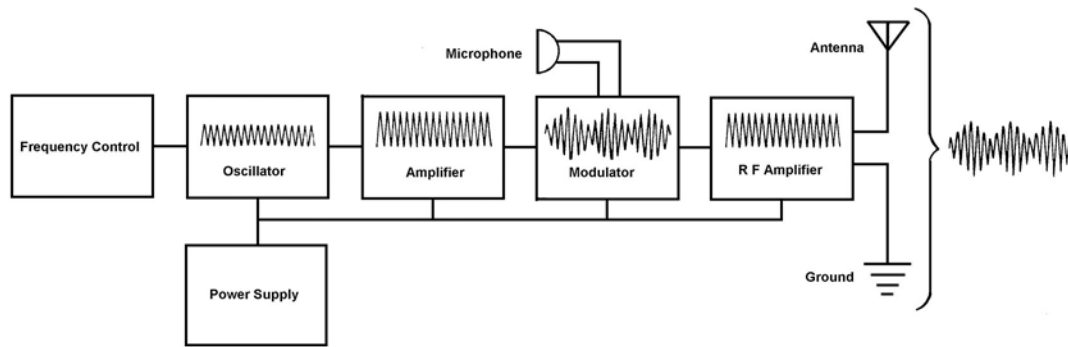
Figure 2116. Components of a radio transmitter.

stability, the ability to resist drift from conditions or values to which set; and (5) fidelity, the completeness with which the essential characteristics of the original signal are reproduced. Receivers may have additional features such as an automatic frequency control, automatic noise limiter, etc.

Some of these characteristics are interrelated. For instance, if a receiver lacks selectivity, signals of a frequency differing slightly from those to which the receiver is tuned may be received. This condition is called spillover, and the resulting interference is called crosstalk. If the selectivity is increased sufficiently to prevent spillover, it may not permit receipt of a great enough band of frequencies to obtain the full range of those of the desired signal. Thus, the fidelity may be reduced.

A **transponder** is a transmitter-receiver capable of accepting the challenge of being interrogated and automatically transmitting an appropriate reply.

U.S. RADIO NAVIGATION POLICY

2118. The Federal Radionavigation Plan

The ideal navigation system should provide three things to the user. First, it should be as accurate as necessary for the job it is expected to do. Second, it should be available 100 percent of the time, in all weather, at any time of day or night. Third, it should have 100 percent integrity, warning the user and shutting itself down when not operating properly. The mix of navigation systems in the U.S. is carefully chosen to provide maximum accuracy, availability, and integrity to all users, marine, aeronautical, and terrestrial, within the constraints of budget and practicality.

Figure 2118. Link to Federal Radionavigation Plan (2014): http://ntl.bts.gov/lib/55000/55100/55108/20150526_Final_ Signed_2014_FRP.pdf

The **Federal Radionavigation Plan** (FRP) is produced by the U.S. Departments of Defense and Transportation. It establishes government policy on the mix of electronic navigation systems, ensuring consideration of national interests and efficient use of resources. It presents an integrated federal plan for all common-use civilian and military radionavigation systems, outlines approaches for consolidation of systems, provides information and schedules, defines and clarifies new or unresolved issues, and provides a focal point for user input. The FRP is a review of existing and planned radionavigation systems used in air, space, land, and marine navigation. It is available from the National Technical Information Service, Springfield, Virginia. The complete 2014 FRP is available via the link provided in Figure 2118.

The first edition of the FRP was released in 1980 as part of a presidential report to Congress. It marked the first time that a joint Department of Transportation (DOT)/Department of Defense (DOD) plan for position, navigation and timing (PNT) had been developed for used by both departments. With the transfer of the United States Coast Guard (USCG) from the DOT to the Department of Homeland Security (DHS), DHS was added as a signatory to the FRP. The FRP was designed to address coordinated planning of federally provided radionavigation systems. Since that time, the Federal planning process has evolved to include other elements of navigation and timing, now referred to as PNT. The FRP has had international impact on navigation systems; PNT services and systems are provided in a manner consistent with the standards and guidelines of in-

ternational groups including the North Atlantic Treaty Organization (NATO) and other allies, the International Maritime Organization (IMO), the International Civil Aviation Organization (ICAO), the International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA), and other international organizations.

During a national emergency, any or all of the systems may be temporarily discontinued by the federal government. The government's policy is to continue to operate radionavigation systems as long as the U.S. and its allies derive greater benefit than adversaries. Operating agencies may shut down systems or change signal formats and characteristics during such an emergency.

The plan is reviewed continually and updated biennially. Industry, advisory groups, and other interested parties provide input. The plan considers governmental responsibilities for national security, public safety, and transportation system economy. It is the official source of radionavigation systems policy and planning for the United States. Systems covered by the FRP include the Global Positioning System (GPS), differential GPS DGPS, Wide area augmentation system (WAAS), Local area augmentation system (LAAS), Loran C or eLORAN, TACAN, Microwave Landing System (MLS), VHF Omnidirectional Range/ Distance Measurement Equipment (VOR/VOR-DME/VORTAC), and Instrument Landing System (ILS).

2119. System Plans

In order to meet both civilian and military needs, the federal government has established a number of different navigation systems. Each system utilizes the latest technology available at the time of implementation and is upgraded as technology and resources permit. The FRP addresses the length of time each system should be part of the system mix. The 2014 FRP sets forth the following system policy guidelines:

RADIOBEACONS: All U.S. marine radiobeacons have been discontinued and most of the stations converted into DGPS sites.

LORAN C: Provides navigation, location, and timing services for both civil and military air, land, and maritime users. LORAN C was replaced in the United States, by GPS, but due to its importance as an alternative to GPS, it is expected to be reactivated as *enhanced* LORAN, or eLORAN.

GPS: The Global Positioning System is the nation's, and the world's, primary radionavigation system. It is operated by the U.S. Air Force.

2120. Enhancements to GPS

Differential GPS (DGPS): DGPS is a ground based system in which differences between observed and calculated GPS signals are broadcast by radiobeacons to users using medium frequencies. The USCG, in cooperation with the U.S. Army Corp of Engineers, operates the maritime DGPS system in U.S. coastal waters including Hawaii, parts of Alaska, portions of the Western Rivers, and the Great Lakes. It provides 10 meter continuous accuracy and integrity alarms, with a typical observed position error of 1-3 meters in a coverage area. In 2016, the inland portion of the National DGPS system, previously funded by the Department of Transportation, was decommissioned and ceased broadcasting.

Wide Area Augmentation System (WAAS): WAAS is a service of the Federal Aviation Administration (FAA), similar to DGPS, and is intended for cross-country and local air navigation, using a series of reference stations and broadcasting correction data through geostationary satellites. WAAS is not optimized for marine use, and while not certified for maritime navigation, may provide additional position accuracy if the signal is unobstructed. Accuracies of a few meters are possible, about the same as with DGPS.

Local Area Augmentation System (LAAS): LAAS is a precision positioning system provided by the FAA for local navigation in the immediate vicinity of airports so equipped. The correctional signals are broadcast on HF radio with a range of about 30 miles. LAAS is not intended or configured for marine use, but can provide extremely accurate position data in a local area.

2121. Factors Affecting Navigation System Mix

The navigator relies on simple, traditional gear, and on some of the most complex and expensive space-based electronic systems man has ever developed. The success of GPS as a robust, accurate, available, and flexible system has made older systems obsolete. Some of the systems which have already met their demise are Transit, Omega, and marine radiobeacons in the U.S. Some might say that the days are numbered for others systems too, as GPS currently retains its primacy in navigation technology.

In the U.S., the DOD, DHS, and DOT continually evaluate the components which make up the federally provided and maintained radionavigation system. Several factors influence the decision on the proper mix of systems; cost, military utility, accuracy requirements, and user requirements all drive the problem of allocating scarce resources to develop and maintain navigation systems.

Many factors influence the choice of PNT systems, which must satisfy an extremely diverse group of users. International agreements must be honored. The current investment in existing systems by both government and users must be considered. The full life-cycle cost of each system must be considered. No system will be phased out without consideration of these factors. The FRP recognizes that GPS may not meet the needs of all users; therefore, some systems are currently being evaluated independently of GPS. The goal is to meet all military and civilian requirements in the most efficient way possible.

RADIO DIRECTION FINDING

2122. Introduction

The simplest use of radio waves in navigation is radio direction finding, in which a MF radio signal is broadcast from a station at a known location. This signal is omnidirectional, but a directional antenna on a vessel is used to determine the bearing of the station. This constitutes a **Line of Position** (LOP), which can be crossed with another LOP to determine a fix.

Once used extensively throughout the world, radiobeacons have been discontinued in the U.S. and many other areas. They are now chiefly used as homing devices by local fishermen, and very little of the ocean's surface is covered by any radiobeacon signal. Because of its limited range, limited availability, and inherent errors, radio direction finding is of limited usefulness to the professional navigator.

In the past, when radiobeacon stations were powerful and common enough for routine ocean navigation, correction of radio bearings was necessary to obtain the most accurate LOP's. The correction process accounted for the fact that, while radio bearings travel along great circles, they are most often plotted on Mercator charts (Aviation charts use the Lambert proportional projection, for which straight lines represent great circles). The relatively short range of those stations remaining has made this process obsolete. Once comprising a major part of NGA *Pub. 117, Radio Navigational Aid*s, radiobeacons are now listed in the back of each volume of the geographically appropriate *List of Lights*.

A radio direction finding station is one which the mariner can contact via radio and request a bearing. Most of these stations are for emergency use only, and a fee may be involved. These stations and procedures for use are listed in NGA *Pub. 117, Radio Navigational Aids*.

2123. Using Radio Direction Finders

Depending upon the design of the **radio direction finder** (**RDF**), the bearings of the radio transmissions are measured as relative bearings, or as both relative and true bearings. The most common type of marine radiobeacon transmits radio waves of approximately uniform strength in all directions. Except during calibration, radiobeacons operate continuously, regardless of weather conditions. Simple combinations of dots and dashes comprising Morse code letters are used for station identification. All radiobeacons superimpose the characteristic on a carrier wave, which is broadcast continuously during the period of transmission. A 10-second dash is incorporated in the characteristic signal to enable users of the aural null type of radio direction finder to refine the bearing.

Bearing measurement is accomplished with a directional antenna. Nearly all types of receiving antennas have some directional properties, but the RDF antenna is designed to be as directional as possible. Simple small craft RDF units usually have a ferrite rod antenna mounted directly on a receiver, with a 360 degree graduated scale. To get a bearing, align the unit to the vessel's course or to true north, and rotate the antenna back and forth to find the exact null point. The bearing to the station, relative or true according to the alignment, will be indicated on the dial. Some small craft RDF's have a portable hand-held combination ferrite rod and compass, with earphones to hear the null.

Two types of loop antenna are used in larger radio direction finders. In one of these, the crossed loop type, two loops are rigidly mounted in such manner that one is placed at 90 degrees to the other. The relative output of the two antennas is related to the orientation of each with respect to the direction of travel of the radio wave, and is measured by a device called a **goniometer**.

2124. Errors of Radio Direction Finders

RDF bearings are subject to certain errors. Quadrantal error occurs when radio waves arrive at a receiver and are influenced by the immediate shipboard environment.

A radio wave crossing a coastline at an oblique angle experiences a change of direction due to differences in conducting and reflecting properties of land and water known as coastal refraction, sometimes called land effect. It is avoided by not using, or regarding as of doubtful accuracy, bearings which cross a shoreline at an oblique angle.

In general, good radio bearings should not be in error by more than two or three degrees for distances under 150 nautical miles. However, conditions vary considerably, and skill is an important factor. By observing the technical instructions for the equipment and practicing frequently when results can be checked, one can develop skill and learn to what extent radio bearings can be relied upon under various conditions. Other factors affecting accuracy include range, the condition of the equipment, and the accuracy of calibration.

The strength of the signal determines the usable range of a radiobeacon. The actual useful range may vary considerably from the published range with different types of RDFs and during varying atmospheric conditions. The sensitivity of a RDF determines the degree to which the full range of a radiobeacon can be utilized. Selectivity varies with the type of receiver and its condition.

CHAPTER 22

SATELLITE NAVIGATION

INTRODUCTION

2200. Development

The idea that led to development of the satellite navigation systems dates back to 1957 and the first launch of an artificial satellite into orbit, Russia's Sputnik I. Dr. William H. Guier and Dr. George C. Wieffenbach at the Applied Physics Laboratory of the Johns Hopkins University were monitoring the famous "beeps" transmitted by the passing satellite. They plotted the received signals at precise intervals, and noticed that a characteristic Doppler curve emerged. Since satellites generally follow fixed orbits, they reasoned that this curve could be used to describe the satellite's orbit. They then demonstrated that they could determine all of the orbital parameters for a passing satellite by Doppler observation of a single pass from a single fixed station. The Doppler shift apparent while receiving a transmission from a passing satellite proved to be an effective measuring device for establishing the satellite orbit.

Dr. Frank T. McClure, also of the Applied Physics Laboratory, reasoned in reverse: If the satellite orbit was known, Doppler shift measurements could be used to determine one's position on Earth. His studies in support of this hypothesis earned him the first National Aeronautics and Space Administration award for important contributions to space development.

In 1958, the Applied Physics Laboratory proposed exploring the possibility of an operational satellite Doppler navigation system. The Chief of Naval Operations then set forth requirements for such a system. The first successful launching of a prototype system satellite in April 1960 demonstrated the Doppler system's operational feasibility.

The **Navy Navigation Satellite System** (**NAVSAT**, also known as **TRANSIT**) was the first operational satellite navigation system. The system's accuracy was better than 0.1 nautical mile anywhere in the world, though its availability was somewhat limited. It was used primarily for the navigation of surface ships and submarines, but it also had some applications in air navigation. It was also used in hydrographic surveying and geodetic position determination.

The transit launch program ended in 1988 and the system was disestablished when the Global Positioning System became operational in 1996.

THE GLOBAL POSITIONING SYSTEM

2201. System Description

The Federal Radionavigation Plan has designated the **Navigation System using Timing And Ranging (NAVSTAR) Global Positioning System (GPS)** as the primary navigation system of the U.S. government. GPS is a spaced-based radio positioning system which provides suitably equipped users with highly accurate position, velocity, and time data. It consists of three major segments: a **space segment**, a **control segment**, and a **user segment**.

Code/Frequency	L1 (1575.42 MHz)	L2 (1227.60 MHz)	L5 (1176.45 MHz)
C/A	X		
L1C	X		
P(y)	X	X	
M-Code	X	X	
L2 CM		X	
L2 CL		X	
L5 I			X
L5Q			X

Figure 2201. GPS Satellite Code by Broadcast Frequency.

The space segment consists of 31 GPS satellites with at least 24 operational 95% of the time. Spacing of the satellites in their orbits is arranged so that at least four satellites are in view to a user at any time, anywhere on the Earth, including the North and South Poles. Each satellite transmits signals on three radio frequencies, superimposed on which are navigation and system data. Included in this data are predicted satellite ephemeris, atmospheric propagation correction data, satellite clock error information and satellite health data. The satellites orbit at an altitude of 20,200 km, in six separate orbital planes, each plane inclined 55° relative to the equator. The satellites complete an orbit approximately once every 12 hours.

GPS satellites transmit **pseudorandom noise (PRN)** sequence-modulated radio frequencies, designated L1 (1575.42 MHz), L2 (1227.60 MHz) and L5 (1176.45 MHz). Various transmissions are sent on these channels as shown in Table 2201.

Superimposed on both the legacy C/A and P(y) codes is the navigation message. This message contains the satellite ephemeris data, atmospheric propagation correction data, and satellite clock bias. In addition, four additional new messages have been introduced by the so called GPS modernization: L2-CNAV, CNAV-2, L5-CNAV and MNAV. The "legacy" message and the first three of the modernized GPS are civil messages, while the MNAV is a military message. In modernized GPS, the same type of contents as the legacy navigation message (NAV) is transmitted but at a higher rate and with improved robustness.

The messages L2-CNAV, L5-CNAV and MNAV have a similar structure and (modernized) data format. The new format allows more flexibility, better control and improved content. Furthermore, the MNAV includes new improvements for the security and robustness of the military message. The CNAV-2 is modulated onto L1CD, sharing the same band as the "legacy" navigation message.

GPS assigns a unique C/A code and a unique P code to each satellite. This practice, known as **code division multiple access (CDMA)**, allows all satellites the use of a common carrier frequency while still allowing the receiver to determine which satellite is transmitting. CDMA also allows for easy user identification of each GPS satellite. Since each satellite broadcasts using its own unique C/A and P code combination, it can be assigned a unique **PRN sequence number**. This number is how a satellite is identified when the GPS control system communicates with users about a particular GPS satellite.

The control segment includes a **master control station** (MCS), a number of monitor stations, and ground antennas located throughout the world. The master control station, located in Colorado Springs, Colorado, consists of equipment and facilities required for satellite monitoring, telemetry, tracking, commanding, control, uploading, and navigation message generation. The monitor stations, located in Hawaii, Colorado Springs, Kwajalein, Diego Garcia, and Ascension Island, passively track the satellites, accumulating ranging data from the satellites' signals and relaying them to the MCS. The MCS processes this information to determine satellite position and signal data accuracy, updates the navigation message of each satellite and relays this information to the ground antennas. The ground antennas then transmit this information to the satellites. The ground antennas, located at Ascension Island, Diego Garcia, and Kwajalein, are also used for transmitting and receiving satellite control information.

The user equipment is designed to receive and process signals from four or more orbiting satellites either simultaneously or sequentially. The processor in the receiver then converts these signals to navigation information. Since GPS is used in a wide variety of applications, from marine navigation to land surveying, these receivers can vary greatly in function and design.

2202. System Capabilities

GPS provides multiple users with accurate, continuous, worldwide, all-weather, common-grid, three-dimensional positioning and navigation information.

To obtain a navigation solution of position (latitude, longitude, and altitude) and time (four unknowns), four satellites must be used. The GPS user measures pseudorange and pseudorange rate by synchronizing and tracking the navigation signal from each of the four selected satellites. Pseudorange is the true distance between the satellite and the user plus an offset due to the user's clock bias. Pseudorange rate is the true slant range rate plus an offset due to the frequency error of the user's clock. By decoding the ephemeris data and system timing information on each satellite's signal, the user's receiver/processor can convert the pseudorange and pseudorange rate to three-dimensional position and velocity. Four measurements are necessary to solve for the three unknown components of position (or velocity) and the unknown user time (or frequency) bias.

The navigation accuracy that can be achieved by any user depends primarily on the variability of the errors in making pseudorange measurements, the instantaneous geometry of the satellites as seen from the user's location on Earth, and the presence of **Selective Availability (SA)**. Selective Availability is discussed further below.

2203. Global Positioning System Concepts

GPS receivers (or user equipment) measure distances between satellites in orbit and a receiver on Earth, and computes spheres of position from those distances. The intersections of those spheres of position then determine the receiver's position.

The distance measurements described above are done by comparing timing signals generated simultaneously by the satellites' and receiver's internal clocks. These signals, charac-

terized by a special wave form known as the pseudo-random code, are generated in phase with each other. The signal from the satellite arrives at the receiver following a time delay proportional to its distance traveled. This time delay is detected by the phase shift between the received pseudo-random code and the code generated by the receiver. Knowing the time required for the signal to reach the receiver from the satellite allows the receiver to calculate the distance from the satellite. The receiver, therefore, must be located on a sphere centered at the satellite with a radius equal to this distance measurement. The intersection of three spheres of position yields two possible points of receiver position. One of these points can be disregarded since it is hundreds of miles from the surface of the Earth. Theoretically, then, only three time measurements are required to obtain a fix from GPS.

In practice, however, a fourth measurement is required to obtain an accurate position from GPS. This is due to receiver clock error. Timing signals travel from the satellite to the receiver at the speed of light; even extremely slight timing errors between the clocks on the satellite and in the receiver will lead to tremendous range errors. The satellite's atomic clock is accurate to 10^{-9} seconds; installing a clock that accurate on a receiver would make the receiver prohibitively expensive. Therefore, receiver clock accuracy is sacrificed, and an additional satellite timing measurement is made. The fix error caused by the inaccuracies in the receiver clock is reduced by simultaneously subtracting a constant timing error from four satellite timing measurements until a pinpoint fix is reached.

Assuming that the satellite clocks are perfectly synchronized and the receiver clock's error is constant, the subtraction of that constant error from the resulting distance determinations will reduce the fix error until a "pinpoint" position is obtained. It is important to note here that the number of lines of position required to employ this technique is a function of the number of lines of position required to obtain a fix. GPS determines position in three dimensions; the presence of receiver clock error adds an additional unknown. Therefore, four timing measurements are required to solve for the resulting four unknowns.

2204. GPS Signal Coding

The GPS L1 band (1575.42 MHz) has turned out to be the most important band for navigation purposes. Indeed most of the applications in the world today are based on the signals transmitted at this frequency. Three signals are transmitted at the moment by GPS in L1: C/A Code, P(Y) Code and M-Code. In the future, an additional new civil signal, known as L1C, will also be transmitted.

GPS is transmitting in the L2 band (1227.60 MHz). It is modernized civil signal known as L2C together with the P(Y) Code and the M-Code. The P(Y) Code and M-Code were already described shortly in the previous chapter and the properties and parameters are thus similar to those in the L1 band. In addition, for Block IIR-M, IIF, and subsequent blocks of SVs, two additional PRN ranging codes will be transmitted. They are the L2 Civil Moderate (L2 CM) code and the L2 Civil Long (L2 CL) code. These two signals are time multiplexed so that the resulting chipping rate is double as high as that of each individual signal.

The GPS L5 (1176.45 MHz) signal will be transmitted for the first time on board IIF satellites. The GPS carriers of the L5 band are modulated by two bit trains in phase quadrature: the L5 data channel and the L5 pilot channel. Moreover, two PRN ranging codes are transmitted on L5.

For a more detailed analysis of GPS signal coding see Appendix C in Volume I.

2205. The Correlation Process

The correlation process compares the signal received from the satellites with the signal generated by the receiver by comparing the square wave function of the received signal with the square wave function generated by the receiver. The computer logic of the receiver recognizes the square wave signals as either a +1 or a 0 depending on whether the signal is "on" or "off." The signals are processed and matched by using an **autocorrelation function**.

This process defines the necessity for a "pseudo-random code." The code must be repeatable (i.e., non-random) because it is in comparing the two signals that the receiver makes its distance calculations. At the same time, the code must be random for the correlation process to work; the randomness of the signals must be such that the matching process excludes all possible combinations except the combination that occurs when the generated signal is shifted a distance proportional to the received signal's time delay. These simultaneous requirements to be both repeatable (non-random) and random give rise to the description of "pseudo-random"; the signal has enough repeatability to enable the receiver to make the required measurement while simultaneously retaining enough randomness to ensure incorrect calculations are excluded.

2206. Precise Positioning Service and Standard Positioning Service

Two levels of navigational accuracy are provided by the GPS: the **Precise Positioning Service (PPS)** and the **Standard Positioning Service (SPS)**. GPS was designed, first and foremost, by the U.S. Department of Defense as a United States military asset; its extremely accurate positioning capability is an asset access to which the U.S. military may need to limit during time of war to prevent use by enemies. Therefore, the PPS is available only to authorized users, mainly the U.S. military and authorized allies. SPS, on the other hand, is available worldwide to anyone possessing a GPS receiver. The accuracy of the GPS signal in space is actually the same for both the civilian GPS service and the military GPS service.

However, SPS broadcasts on one frequency, while PPS uses two. This means military users can perform ionospheric correction, a technique that reduces radio degradation caused by the Earth's atmosphere. With less degradation, PPS provides better accuracy than the basic SPS.

The ongoing GPS modernization program is adding new civilian signals and frequencies to the GPS satellites, enabling ionospheric correction for all users. Eventually, the accuracy difference between military and civilian GPS will disappear. But military GPS will continue to provide important advantages in terms of enhanced security and jam resistance.

Anti-spoofing (A-S) is designed to negate any hostile imitation of GPS signals. The technique alters the P code into another code, designated the Y code. The C/A code remains unaffected. The U.S. employs this technique to the satellite signals at random times and without warning; therefore, civilian users are unaware when this P code transformation takes place. Since anti-spoofing is applied only to the P code, the C/A code is not protected and can be spoofed.

GPS PPS receivers can use either the P code or the C/A code, or both, in determining position. Maximum accuracy is obtained by using the P code on both L1 and L2. The difference in propagation delay is then used to calculate ionospheric corrections. The C/A code is normally used to acquire the satellite signal and determine the approximate P code phase. Some PPS receivers possess a clock accurate enough to track and lock on the P code signal without initially tracking the C/A code. Some PPS receivers can track only the C/A code and disregard the P code entirely. Since the C/A code is transmitted on only one frequency, the dual frequency ionosphere correction methodology is unavailable and an ionospheric modeling procedure is required to calculate the required corrections.

SPS receivers, as mentioned above, provide positions with a degraded accuracy. The A-S feature denies SPS users access to the P code when transformed to the Y code. Therefore, the SPS user cannot rely on access to the P code to measure propagation delays between L1 and L2 and compute ionospheric delay corrections. Consequently, the typical SPS receiver uses only the C/A code because it is unaffected by A-S. Like PPS, the C/A is transmitted only on L1, the dual frequency method of calculating ionospheric corrections is unavailable; an ionospheric modeling technique must be used. This is less accurate than the dual frequency method; this degradation in accuracy is accounted for in the 100-meter accuracy calculation.

2207. Selective Availability Discontinued

In May 2000, President Bill Clinton directed the Department of Defense to turn off the GPS **Selective Availability (SA)** feature. In 2007, the U.S. government announced plans to permanently eliminate SA by building the GPS III satellites without it. SA was a method to degrade GPS accuracy to civilian users.

2208. GPS Receiver Operations

In order for the GPS receiver to navigate, it has to track satellite signals, make pseudorange measurements, and collect navigation data.

A typical satellite tracking sequence begins with the receiver determining which satellites are available for it to track. Satellite visibility is determined by user-entered predictions of position, velocity, and time, and by almanac information stored internal to the receiver. If no stored almanac information exists, then the receiver must attempt to locate and lock onto the signal from any satellite in view. When the receiver is locked onto a satellite, it can demodulate the navigation message and read the almanac information about all the other satellites in the constellation. A carrier tracking loop tracks the carrier frequency while a code tracking loop tracks the C/A and P code signals. The two tracking loops operate together in an iterative process to acquire and track satellite signals.

The receiver's carrier tracking loop will locally generate an L1 carrier frequency which differs from the satellite produced L1 frequency due to a Doppler shift in the received frequency. This Doppler offset is proportional to the relative velocity along the line of sight between the satellite and the receiver, subject to a receiver frequency bias. The carrier tracking loop adjusts the frequency of the receiver-generated frequency until it matches the incoming frequency. This determines the relative velocity between the satellite and the receiver. The GPS receiver uses this relative velocity to calculate the velocity of the receiver. This velocity is then used to aid the code tracking loop.

The code tracking loop is used to make pseudorange measurements between the GPS receiver and the satellites. The receiver's tracking loop will generate a replica of the targeted satellite's C/A code with estimated ranging delay. In order to match the received signal with the internally generated replica, two things must be done: 1) the center frequency of the replica must be adjusted to be the same as the center frequency of the received signal; and 2) the phase of the replica code must be lined up with the phase of the received code. The center frequency of the replica is set by using the Doppler-estimated output of the carrier tracking loop. The receiver will then slew the code loop generated C/A code though a millisecond search window to correlate with the received C/A code and obtain C/A tracking.

Once the carrier tracking loop and the code tracking loop have locked onto the received signal and the C/A code has been stripped from the carrier, the navigation message is demodulated and read. This gives the receiver other information crucial to a pseudorange measurement. The navigation message also gives the receiver the handover word, the code that allows a GPS receiver to shift from C/A code tracking to P code tracking.

The handover word is required due to the long phase (seven days) of the P code signal. The C/A code repeats every millisecond, allowing for a relatively small search window. The seven day repeat period of the P code requires that the receiver be given the approximate P code phase to narrow its search window to a manageable time. The handover word provides this P code phase information. The handover word is repeated every subframe in a 30 bit long block of data in the navigation message. It is repeated in the second 30 second data block of each subframe. For some receivers, this handover word is unnecessary; they can acquire the P code directly. This normally requires the receiver to have a clock whose accuracy approaches that of an atomic clock. Since this greatly increases the cost of the receiver, most receivers for non-military marine use do not have this capability.

Once the receiver has acquired the satellite signals from four GPS satellites, achieved carrier and code tracking, and has read the navigation message, the receiver is ready to begin making pseudorange measurements. Recall that these measurements are termed *pseudorange* because a receiver clock offset makes them inaccurate; that is, they do not represent the true range from the satellite, only a range biased by a receiver clock error. This clock bias introduces a fourth unknown into the system of equations for which the GPS receiver must solve (the other three being the x coordinate, y coordinate, and z coordinate of the receiver position). The receiver solves this clock bias problem by making a fourth pseudorange measurement, resulting in a fourth equation to allow solving for the fourth unknown. Once the four equations are solved, the receiver has an estimate of the receiver's position in three dimensions and of GPS time. The receiver then converts this position into coordinates referenced to an Earth model based on the World Geodetic System (1984).

2209. User Range Errors and Geometric Dilution of Precision

There are two formal position accuracy requirements for GPS:

1) The PPS spherical position accuracy shall be 16 meters SEP (spherical error probable) or better.

2) The SPS user two dimensional position accuracy shall be 100 meters 2 DRMS (distance root mean squared) or better.

Assume that a universal set of GPS pseudorange measurements results in a set of GPS position measurements. The accuracy of these measurements will conform to a normal (i.e. values symmetrically distributed around a mean of zero) probability function because the two most important factors affecting accuracy, the **geometric dilution of precision (GDOP)** and the **user equivalent range error (UERE)**, are continuously variable.

The UERE is the error in the measurement of the pseudoranges from each satellite to the user. The UERE is the product of several factors, including the clock stability, the predictability of the satellite's orbit, errors in the 50 Hz navigation message, the precision of the receiver's correlation process, errors due to atmospheric distortion and the calculations to compensate for it, and the quality of the satellite's signal. The UERE, therefore, is a random error which is the function of errors in both the satellites and the user's receiver.

The GDOP depends on the geometry of the satellites in relation to the user's receiver. It is independent of the quality of the broadcast signals and the user's receiver. Generally speaking, the GDOP measures the "spread" of the satellites around the receiver. The optimum case would be to have one satellite directly overhead and the other three spaced 120° around the receiver on the horizon. The worst GDOP would occur if the satellites were spaced closely together or in a line overhead.

There are special types of DOP's (dilution of precision) for each of the position and time solution dimensions; these particular DOP's combine to determine the GDOP. For the vertical dimension, the **vertical dilution of precision (VDOP)** describes the effect of satellite geometry on altitude calculations. The **horizontal dilution of precision (HDOP)** describes satellite geometry's effect on position (latitude and longitude) errors. These two DOP's combine to determine the **position dilution of precision (PDOP)**. The PDOP combined with the **time dilution of precision (TDOP)** results in the GDOP. See Figure 2209.

2210. Ionospheric Delay Errors

Section 2209 covered errors in GPS positions due to errors inherent in the satellite signal (UERE) and the geometry of the satellite constellation (GDOP). Another major cause of accuracy degradation is the effect of the ionosphere on the radio frequency signals that comprise the GPS signal.

A discussion of a model of the Earth's atmosphere will be useful in understanding this concept. Consider the Earth as surrounded by three layers of atmosphere. The first layer, extending from the surface of the Earth to an altitude of approximately 10 km, is known as the **troposphere**. Above the troposphere and extending to an altitude of approximately 50 km is the **stratosphere**. Finally, above the stratosphere and extending to an altitude that varies as a function of the time of day is the **ionosphere**. Though radio signals are subjected to effects which degrade its accuracy in all three layers of this atmospheric model, the effects of the ionosphere are the most significant to GPS operation.

The ionosphere, as the name implies, is that region of the atmosphere which contains a large number of ionized molecules and a correspondingly high number of free electrons. These charged molecules have lost one or more
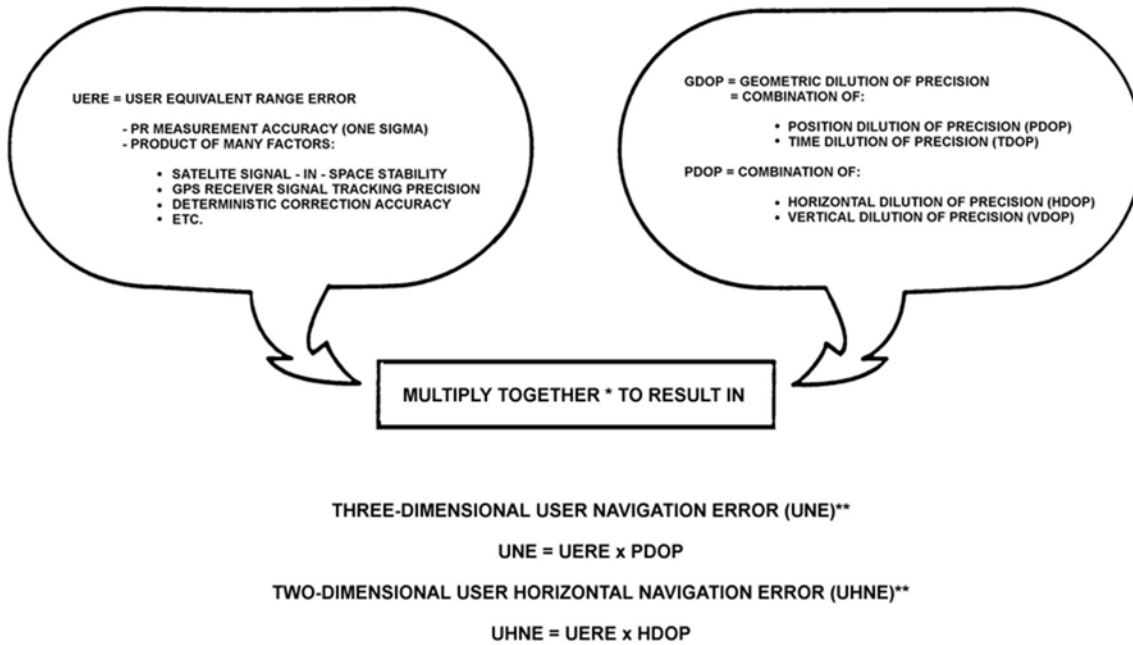
Figure 2209. Position and time error computations.

electrons. No atom will lose an electron without an input of energy; the energy input that causes the ions to be formed in the ionosphere comes from the **ultraviolet (U-V)** radiation of the Sun. Therefore, the more intense the Sun's rays, the larger the number of free electrons which will exist in this region of the atmosphere.

The largest effect that this ionospheric effect has on GPS accuracy is a phenomenon known as **group time delay**. As the name implies, group time delay results in a delay in the time a signal takes to travel through a given distance. Obviously, since GPS relies on extremely accurate timing measurement of these signals between satellites and ground receivers, this group time delay can have a noticeable effect on the magnitude of GPS position error.

The group time delay is a function of several elements. It is inversely proportional to the square of the frequency at which the satellite transmits, and it is directly proportional to the atmosphere's **total electron content (TEC)**, a measure of the degree of the atmosphere's ionization. The general form of the equation describing the delay effect is:

$$\Delta t = \frac{(K \quad TEC)}{f^2}$$

where

Δt = group time delay
f = operating frequency
K = constant

Since the Sun's U-V radiation ionizes the molecules in the upper atmosphere, it stands to reason that the time delay value will be highest when the Sun is shining and lowest at night. Experimental evidence has borne this out, showing that the value for TEC is highest around 1500 local time and

lowest around 0500 local time. Therefore, the magnitude of the accuracy degradation caused by this effect will be highest during daylight operations. In addition to these daily variations, the magnitude of this time delay error also varies with the seasons; it is highest at the vernal equinox. Finally, this effect shows a solar cycle dependence. The greater the number of sunspots, the higher the TEC value and the greater the group time delay effect. The solar cycle typically follows an eleven year pattern. The current solar cycle began on January 4, 2008 with minimal activity until early 2010. The cycle is on track to have the lowest recorded sunspot activity since cycle 14 which reached maximum in 1906. See Figure 2210 Solar cycle 24 prediction.

Given that this ionospheric delay introduces a serious accuracy degradation into the system, how does GPS account for it? There are two methods used: (1) the dual frequency technique, and (2) the ionospheric delay method.

2211. Dual Frequency Correction Technique

As the term implies, the dual frequency technique requires the ability to acquire and track both the L1 and L2 frequency signals. Recall from the discussion in Section 2204 that the C/A and P codes are transmitted on carrier frequency L1, but only the P code is transmitted on L2. Recall also that only authorized operators with access to DOD cryptographic material are able to copy the P code. It follows, then, that only those authorized users are able to copy the L2 carrier frequency. Therefore, only those authorized users are able to use the dual frequency correction method. The dual frequency method measures the distance between the satellite and the user based on both the L1 and L2 carrier signal. These ranges will be
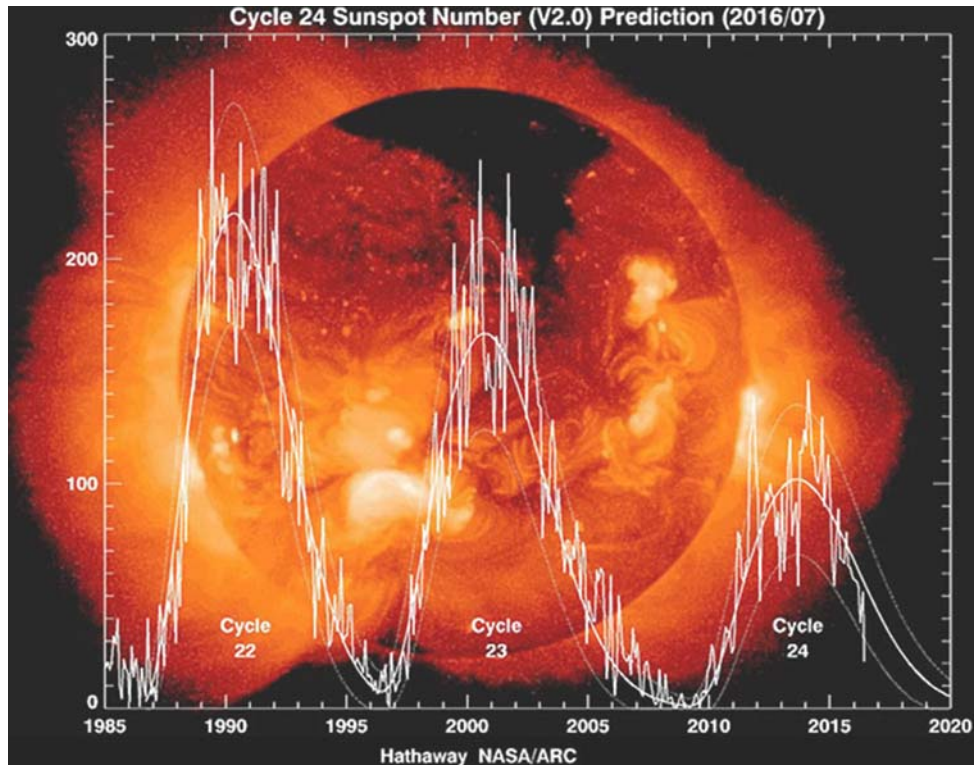
Figure 2210. Solar cycle prediction. Courtesy of NASA.

different because the group time delay for each signal will be different. This is because of the frequency dependence of the time delay error. The range from the satellite to the user will be the true range combined with the range error caused by the time delay, as shown by the following equation:

$$R(f) \ = \ R_{actual} + \text{error term}$$

where R(f) is the range which differs from the actual range as a function of the carrier frequency. The dual frequency correction method takes two such range measurements, R(L1) and R(L2). Recall that the error term is a function of a constant divided by the square of the frequency. By combining the two range equations derived from the two frequency measurements, the constant term can be eliminated and one is left with an equation in which the true range is simply a function of the two carrier frequencies and the measured ranges R(L1) and R(L2). This method has two major advantages over the ionospheric model method: (1) it calculates corrections from real-time measured data, therefore, it is more accurate; (2) it alleviates the need to include ionospheric data on the navigation message. A significant portion of the data message is devoted to ionospheric correction data. If the receiver is dual frequency capable, then it does not need any of this data.

The vast majority of maritime users cannot copy dual frequency signals. For them, the ionospheric delay model provides the correction for the group time delay.

2212. The Ionospheric Delay Model

The ionospheric delay model mathematically models the diurnal ionospheric variation. The value for this time delay is determined from a cosinusoidal function into which coefficients representing the maximum value of the time delay (i.e., the amplitude of the cosine wave representing the delay function), the time of day, the period of the variation and a minimum value of delay are introduced. This model is designed to be most accurate at the diurnal maximum. This is obviously a reasonable design consideration because it is at the time of day when the maximum diurnal time delay occurs that the largest magnitude of error appears. The coefficients for use in this delay model are transmitted to the receiver in the navigation data message. As stated in Section 2211, this method of correction is not as accurate as the dual frequency method; however, for the non-military user, it is the only method of correction available.

2213. Multipath Reflection Errors

Multipath reflection errors occur when the receiver detects parts of the same signal at two different times. The first reception is the direct path reception, the signal that is received directly from the satellite. The second reception is from a reflection of that same signal from the ground or any other reflective surface. The direct path signal arrives first, the reflected signal, having had to travel a longer distance to the receiver, arrives later. The GPS signal is designed to

minimize this multipath error. The L1 and L2 frequencies used demonstrate a diffuse reflection pattern, lowering the signal strength of any reflection that arrives at the receiver. In addition, the receiver's antenna can be designed to reject a signal that it recognizes as a reflection. In addition to the properties of the carrier frequencies, the high data frequency of both the P and C/A codes and their resulting good correlation properties minimize the effect of multipath propagation.

The design features mentioned above combine to reduce the maximum error expected from multipath propagation to less than 20 feet.

DIFFERENTIAL GPS

2214. Differential GPS Concept

The discussions above make it clear that the Global Positioning System provides the most accurate positions available to navigators today. They should also make clear that the most accurate positioning information is available to only a small fraction of the using population: U.S. and allied military. For most open ocean navigation applications, the degraded accuracy inherent in selective availability and the inability to copy the precision code presents no serious hazard to navigation. A mariner seldom if ever needs greater than 100 meter accuracy in the middle of the ocean.

It is a different situation as the mariner approaches shore. Typically for harbor approaches and piloting, the mariner will shift to visual piloting. The increase in accuracy provided by this navigational method is required to ensure ship's safety. The 100 meter accuracy of GPS in this situation is not sufficient. Any mariner who has groped his way through a restricted channel in a thick fog will certainly appreciate the fact that even a degraded GPS position is available for them to plot. However, 100 meter accuracy is not sufficient to ensure ship's safety in most piloting situations. In this situation, the mariner needs P code accuracy. The problem then becomes how to obtain the accuracy of the Precise Positioning Service with due regard to the legitimate security concerns of the U.S. military. The answer to this seeming dilemma lies in the concept of **Differential GPS (DGPS)**.

Differential GPS is a system in which a receiver at an accurately surveyed position utilizes GPS signals to calculate timing errors and then broadcasts a correction signal to account for these errors. This is an extremely powerful concept. The errors which contribute to GPS accuracy degradation, ionospheric time delay and selective availability, are experienced simultaneously by both the DGPS receiver and a relatively close user's receiver. The extremely high altitude of the GPS satellites means that, as long as the DGPS receiver is within 100-200 km of the user's receiver, the user's receiver is close enough to take advantage of any DGPS correction signal.

The theory behind a DGPS system is straightforward. Located on an accurately surveyed site, the DGPS receiver already knows its location. It receives data which tell it where the satellite is. Knowing the two locations, it then calculates the theoretical time it should take for a satellite's signal to reach it. It then compares the time that it actually takes for the signal to arrive. This difference in time between the theoretical and the actual is the basis for the DGPS receiver's computation of a timing error signal; this difference in time is caused by all the errors to which the GPS signal is subjected; errors, except for receiver error and multipath error, to which both the DGPS and the user's receivers are simultaneously subject. The DGPS system then broadcasts a timing correction signal, the effect of which is to correct for selective availability, ionospheric delay, and all the other error sources the two receivers share in common.

For suitably equipped users, DGPS results in positions at least as accurate as those obtainable by the Precise Positioning Service. This capability is not limited to simply displaying the correct position for the navigator to plot. The DGPS position can be used as the primary input to an electronic chart system, providing an electronic readout of position accurate enough to pilot safely in the most restricted channel.

SATELLITE BASED AUGMENTATION SYSTEMS (SBAS)

2215. WAAS/LAAS for Aeronautical Use

The **Wide Area Augmentation System (WAAS)** program, which corrects for GPS signal errors caused by ionospheric disturbances, timing, and satellite orbit errors, provides vital integrity information regarding the health of each GPS satellite. The concept is similar to the DGPS concept, except that correctional signals are sent from geostationary satellites via HF signals directly to the user's GPS receiver. This eliminates the need for a separate receiver and antenna, as is the case with DGPS. WAAS is intended for en route air navigation, with 25 reference stations widely spaced across the United States that monitor GPS satellite data and two master stations on either coast, and creates a GPS correction message. WAAS provides coverage of the entire U.S. and parts of Mexico and Canada.

The **Local Area Augmentation System (LAAS)** is intended for precision airport approaches, with reference stations located at airports and broadcasting their correction message on VHF radio frequencies.

While many marine GPS receivers incorporate WAAS circuitry (but not the more accurate, shorter-range LAAS), WAAS is not optimized for surface navigation because the HF radio signals are line-of-sight and are transmitted from geostationary satellites. At low angles to the horizon, the WAAS signal may be blocked and the resulting GPS position accuracy significantly degraded with no warning. The DGPS signal, on the other hand, is a terrain-following signal that is unaffected by objects in its path. It simply flows around them and continues on unblocked.

The accuracy of WAAS and DGPS is comparable, on the order of a few meters. Any GPS receiver equipped to receive WAAS has its accuracy improved to less than 3 meters. Both systems have been found in actual use to provide accuracies somewhat better than designed. DGPS was designed to provide 10 meter accuracy 95% of the time, but in actual use one can expect about 1-3 meter accuracy when the user is within 100 miles of the DGPS transmitter. Over 100 miles, DGPS accuracy will commonly degrade by an additional 1 meter per 100 miles from the transmitter site.

The WAAS signal, while not certified for use in the marine environment as is DGPS, can be a very useful navigational tool if its limitations are understood. In open waters of the continental U.S., the WAAS signal can be expected to be available and useful, provided the receiver has WAAS circuitry and is programmed to use the WAAS data. Outside the U.S., or in any area where tall buildings, trees, or other obstructions rise above the horizon, the WAAS signal may be blocked, and the resulting GPS fix could be in error by many meters. Since the highest accuracy is necessary in the most confined waters, WAAS should be used with extreme caution in these areas.

WAAS can enhance the navigator's situational awareness when available, but availability is not assured. Further, a marine receiver will provide no indication when WAAS data is not a part of the fix. [Aircraft GPS receivers may contain Receiver Autonomous Integrity Monitoring (RAIM) software, which provides warning of WAAS satellite signal failure, and removes the affected signal from the fix solution.]

LAAS data, broadcast on VHF, is less subject to block-ing, but is only available in selected areas near airports. Its range is about 30 miles. It is therefore not suitable for general marine navigational use.

2216. More Information

For more information on the Global Positioning System (GPS) and related topics see the link provided in Figure 2216.

Figure 2216. GPS.gov at http://www.gps.gov

2217. Foreign SBAS

SBAS systems are spreading out all over the world. More and more, it is believed that upon dual-frequency SBAS service provision, a seamless navigation will be possible from and to any two locations in the world.

Presently, three foreign SBAS systems are operational. These are Japan's Multi-functional Transport Satellite based Augmentation System (MSAS), the European Geostationary Navigation Overlay Service (EGNOS) and India's GPS and Geo-Augmented Navigation System (GAGAN),

Other foreign SBAS are under implementation such as SDCM (System of Differential Correction and Monitoring) in Russia and SNAS (Satellite Navigation Augmentation System) in China. Still others are under development or feasibility studies; SACCSA (Solucion de Aumentacion para Caribe, Centro y Sudamerica) would cover Central & South America including the Caribbean. Member States to SACCSA include Argentina Bolivia, Colombia, Costa Rica, Guatemala, Panama, Spain and Venezuela. Malaysia, much of Africa and South Korean SBAS are also studying SBAS particularly for aeronautical navigation.

NON-U.S. SATELLITE NAVIGATION SYSTEMS

2218. The Galileo System

Galileo is the **global navigation satellite system** (GNSS) that is currently being created by the European Union (EU) through the European Space Agency (ESA) and the European GNSS Agency (GSA), with two ground operations centers in Germany and Italy.

One of the aims of Galileo is to provide an indigenous alternative high-precision positioning system upon which European nations can rely, independently from other country systems, in case they were disabled by their operators.

The use of basic (low-precision) Galileo services will be free and open to everyone. The high-precision capabilities will be available for paying commercial users. Galileo is intended to provide horizontal and vertical position measurements within one meter precision, and better positioning services at high latitudes than other positioning systems.

Galileo is to provide a new global search and rescue (SAR) function as part of the Medium-altitude Earth Orbit Search and Rescue (MEOSAR) system. Satellites will be equipped with a transponder which will relay distress sig-

nals from emergency beacons to a rescue coordination center, which will then initiate a rescue operation. At the same time, the system is projected to provide a signal, the Return Link Message (RLM), to the emergency beacon, informing victims that their situation has been detected and help is on the way. This latter feature is new and is considered a major upgrade compared to the existing international search and rescue system (Cospas-Sarsat), which does not provide feedback to the user.

Galileo will also provide an important feature for civilian use that GPS does not: integrity monitoring. Currently, a civilian GPS user receives no indication that his unit is not receiving proper satellite signals, there being no provision for such notification in the code. However, Galileo will provide such a signal, alerting the user that the system is operating improperly.

The first Galileo test satellite, the GIOVE-A, was launched 28 December 2005, while the first satellite to be part of the operational system was launched on 21 October 2011. As of May 2016 the system has 14 of 30 satellites in orbit. Galileo will start offering Early Operational Capability (EOC) from 2016, go to Initial Operational Capability (IOC) in 2017-18 and reach Full Operational Capability (FOC) in 2019. The complete 30-satellite Galileo system (24 operational and 6 active spares) is expected by 2020.

For detailed information on the Galileo signal structure see Appendix C in Volume I.

2219. GLONASS

The **Global Navigation Satellite System (GLON-ASS)**, under the control of the Russian military, has been in use since 1993, and is based on the same principles as GPS. The space segment consists of 24 satellites in three orbital planes, the planes separated by 120 degrees and the individual satellites by 45 degrees. The orbits are inclined to the equator at an angle of 64.8 degrees, and the orbital period is about 11hours, 15minutes at an altitude of 19,100 km (10,313 nm). The designed system fix accuracy for civilian use is 100 meters horizontal (95%), 150 meters vertical, and 15 cm/sec. in velocity. Military codes provide accuracies of some 10-20 meters horizontal.

The ground segment of GLONASS lies entirely within the former Soviet Union. Reliability has been an ongoing problem for the GLONASS system, but new satellite designs with longer life spans are addressing these concerns. The user segment consists of various types of receivers that provide position, time, and velocity information.

GLONASS signals are in the L-band, operating in 25 channels with 0.5625 MHz separation in 2 bands: from 1602.5625 MHz to 1615.5 MHz, and from 1240 to 1260 MHz.

For detailed information on the GLONASS signal structure see Appendix C in Volume I.

2220. BeiDou

The **BeiDou Navigation Satellite System (BDS)**, also known as BeiDou-2, is China's second-generation satellite navigation system that will be capable of providing positioning, navigation, and timing services to users on a continuous worldwide basis.

Although the evolution of its regional navigation system towards a global solution started in 1997, the formal approval by the Government of the development and deployment of BDS System was done in 2006 and it is expected to provide global navigation services by 2020, similarly to the GPS, GLONASS or Galileo systems.

As of December 2011, the BeiDou system was officially announced to provide Initial Operational Service providing initial passive positioning navigation and timing services for the whole Asia-Pacific region with a constellation of 10 satellites (5 GEO satellites and 5 Inclined Geosynchronous Satellite Orbit (IGSO) satellites). During 2012, 5 additional satellites (1 GEO satellite and 4 Medium-Earth Orbit (MEO) satellites) were launched increasing to 14 the number of satellites of the constellation. In 2020, the system is going to launch the remaining satellites and evolve towards global navigation capability.

The BeiDou Space Segment consists of a constellation of 35 satellites, which include 5 geostationary earth orbit (GEO) satellites and 30 non-GSO satellites. The system is currently under development evolving from a regional system called BeiDou-1, and in the first phase will provide global navigation services by 2020, similarly to the GPS, GLONASS or Galileo systems.

For detailed information on the BeiDou signal plan see Appendix C in Volume I.

2221. IRNSS

The **Indian Regional Navigational Satellite System (IRNSS)** is a regional satellite navigation system owned by the Indian government. The system is being developed by Indian Space Research Organization (ISRO).

In April 2016, with the last launch of the constellation's satellite, IRNSS was renamed Navigation Indian Constellation (NAVIC) by India's Prime Minister Narendra Modi.

IRNSS will be an independent and autonomous regional navigation system aiming a service area of about 1500 kilometers around India. The system will be under complete Indian control, with the space segment, ground segment and user receivers all being built in India. It will have a range of applications including personal navigation.

For detailed information on the IRNSS signal plan see Appendix C in Volume I.

2222. QZSS

The **Quasi-Zenith Satellite System (QZSS)** is a regional navigation satellite system commissioned by the Japanese Government as a National Space Development Program.

QZSS was authorized by the Japanese government in 2002. At the beginning the system was developed by the Advanced Space Business Corporation (ASBC) team, including Mitsubishi Electric Corp., Hitachi Ltd., and GNSS Technologies Inc. When in 2007 ASBC collapsed, the work was taken over by JAXA together with Satellite Positioning Research and Application Center (SPAC), established in February 2007 and approved by the Ministers associated with QZSS research and development.

The QZSS service area covers East Asia and Oceania region and its platform is multi-constellation GNSS. The QZSS system is not required to work in a stand-alone mode, but together with data from other GNSS satellites.

For detailed information on the QZSS signal plan see Appendix C in Volume I.

2223. References

Ávila Rodríguez, José Ángel. (2008). *On Generalized Signal Waveforms for Satellite Navigation*. University FAF, Munich. Retrieved from: https://athene-forschung.unibw.de/doc/86167/86167.pdf

CHAPTER 23

RADAR NAVIGATION

PRINCIPLES OF RADAR OPERATION

2300. Introduction

Radar determines distance to an object by measuring the time required for a radio signal (moving at the speed of light) to travel from a transmitting antenna to the object, reflect off that object, and return as a received echo.

Distance, or range, can be found by the simple formula:

range = 1/2 (C x t)

where range is in nautical miles,
C = the speed of light in nautical miles per second, and
t = the time in seconds from the time of pulse transmission to echo reception.

Because the value of C is very large (162,000 NM/sec), t is very small, 0.0001 sec for a target at a range of 10 miles for example.

Such measurements can be converted into lines of position (LOP's) comprised of circles with radius equal to the distance to the object. Since marine radars use directional antennae, they can also determine an object's bearing. However, due to its design, radar's bearing measurements are much less accurate than its distance measurements. Understanding this concept is crucial to ensuring the optimal employment of the radar for safe navigation.

2301. Signal Characteristics

In most marine navigation applications, the radar signal is pulse modulated. Signals are generated by a timing circuit so that energy leaves the antenna in very short pulses, usually less than one millionth of a second (or 1 μsec) in duration. When transmitting, the antenna is connected to the transmitter but not the receiver. As soon as the pulse leaves, an electronic switch disconnects the antenna from the transmitter and connects it to the receiver. Another pulse is not transmitted until after the preceding one has had time to travel to the most distant target within range and return. Since the interval between pulses is long compared with the length of a pulse, strong signals can be provided with low average power. The duration or length of a single pulse is called **pulse length**, **pulse duration**, or **pulse width**. This pulse emission sequence repeats a great many times, perhaps 1,000 per second. This rate defines the **pulse repetition rate (PRR)**. The returned pulses are displayed on an indicator screen or *display*.

2302. The Transmitter

In traditional marine radar sets, those produced since the 1940s, the transmitter is a special electronic oscillator diode tube known as a magnetron. The magnetron produces very high power microwaves (25 KW and greater) for very short periods of time.

Recently, another type of radar has been introduced into the commercial marine industry know as solid state or *coherent* radar. In modern solid state radars, the pulses generated by special circuitry in the transmitter are of much less power, much longer in length, and of varying frequency. This type of radar does not use a magnetron and generates an entirely different waveform. Presently, solid state radar is only available in the S-Band and will be further discussed in the following sections.

2303. The Receiver

The function of the receiver is to amplify the strength of the very weak return echoes. The enhanced signals can then be used to produce video signals which are presented as targets on the display. The amplifiers in a traditional magnetron radar have to deal with only one frequency, either 3000 MHz or 10000 MHz depending on the radar set.

A solid state radar receiver however, must process a much more complex signal with changing frequency. This variable frequency, or chirp, necessitates signal processing within the receiver known as pulse compression, which shortens the comparatively long, 5 - 18 microsecond transmitted pulse into a pulse of similar length to traditional radars (0.05 - 1.0 μsec), while at the same time increasing signal amplitude, thus yielding the same detection and range measuring capabilities. A very great advantage of solid state radars over magnetron radars is their superior ability to filter out rain and sea clutter effects and therefore assist the radar observer in identification of land targets used in radar navigation.

2304. The Antenna

Nearly all modern commercial marine radars use a type of antenna known as a slotted waveguide. See Figure 2304 for a depiction of a slotted waveguide antenna. Both solid state and magnetron radar sets utilize this antenna configuration because it is simple, efficient, and produces a beam that minimizes unwanted side lobes (side lobes will be discussed later in this chapter).
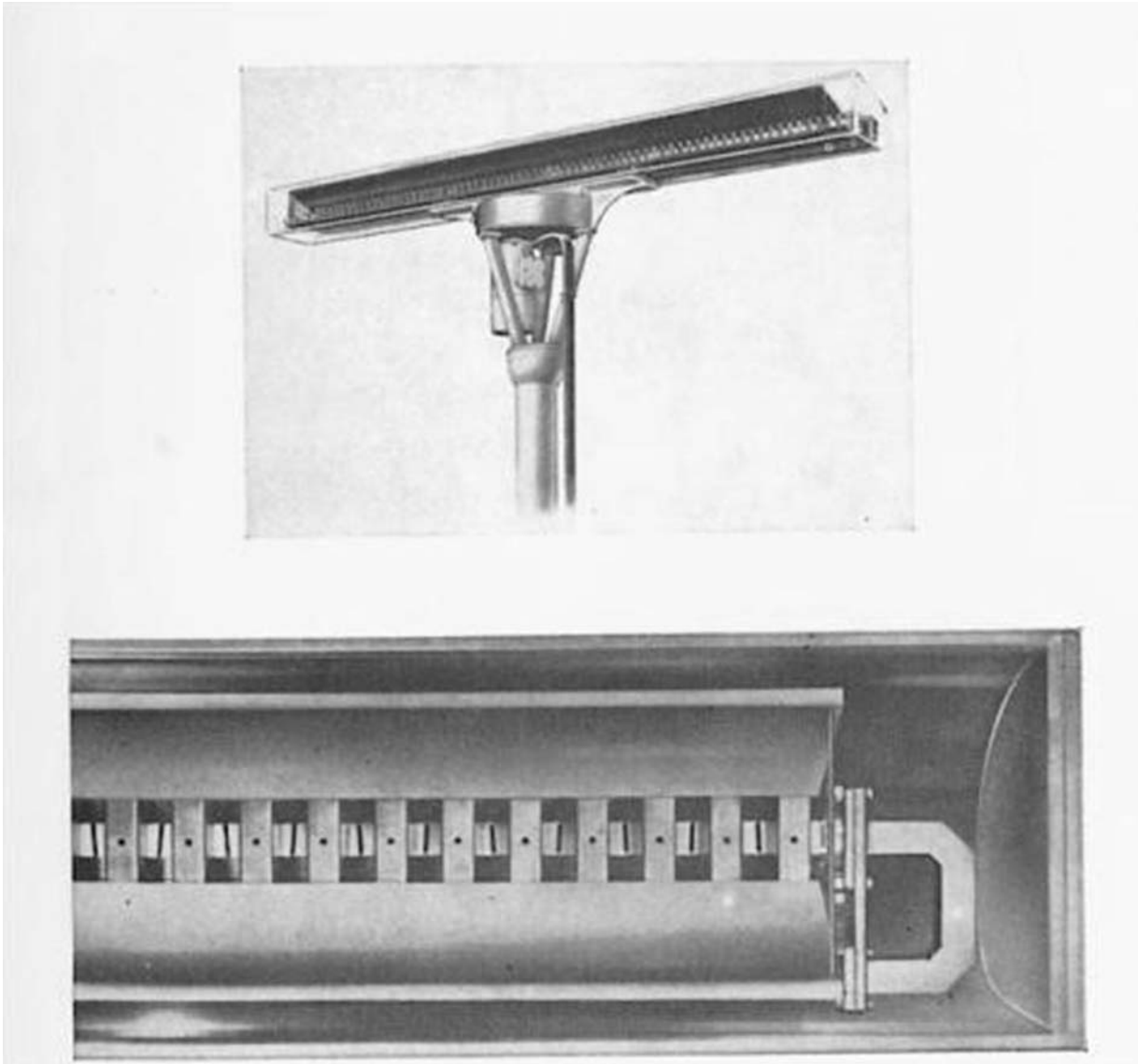
Figure 2304. Slotted waveguide antenna.

2305. The Display

The radar display is often referred to as the **plan position indicator (PPI)**. On a PPI, the sweep appears as a radial line, centered at the center of the scope and rotating in synchronization with the antenna. Any returned echo causes a brightening of the display screen at the bearing and range of the object. The glow continues after the sweep rotates past the target.

On a PPI, a target's actual range is proportional to its distance from the center of the scope. A movable cursor helps to measure ranges and bearings. In the "heading-upward" presentation, which indicates relative bearings, the top of the scope represents the direction of the ship's head. In this destabilized presentation, the orientation changes as the ship changes heading. In the stabilized "north-upward" presentation, gyro north is always at the top of the scope.

2306. The Radar Beam

The pulses of energy comprising the radar beam would form a single lobe-shaped pattern of radiation if emitted in free space. Figure 2306a shows this free space radiation pattern, including the undesirable minor lobes or side lobes associated with practical antenna design. This radiation pattern, as well as the effects of diffraction, reflection and attenuation described below, are common to both magnetron and solid state generated radar signals. Although the radiated energy is concentrated into a relatively narrow

main beam by the antenna, there is no clearly defined enve-lope of the energy radiated, although most of the energy is concentrated along the axis of the beam.

The radiation diagram shown in Figure 2306b depicts relative values of power in the same plane existing at the same distances from the antenna or the origin of the radar beam. Maximum power is in the direction of the axis of the beam. Power values diminish rapidly in directions away from the axis. The beam width is taken as the angle between the half-power points.

Figure 2306a. Freespace radiation pattern.
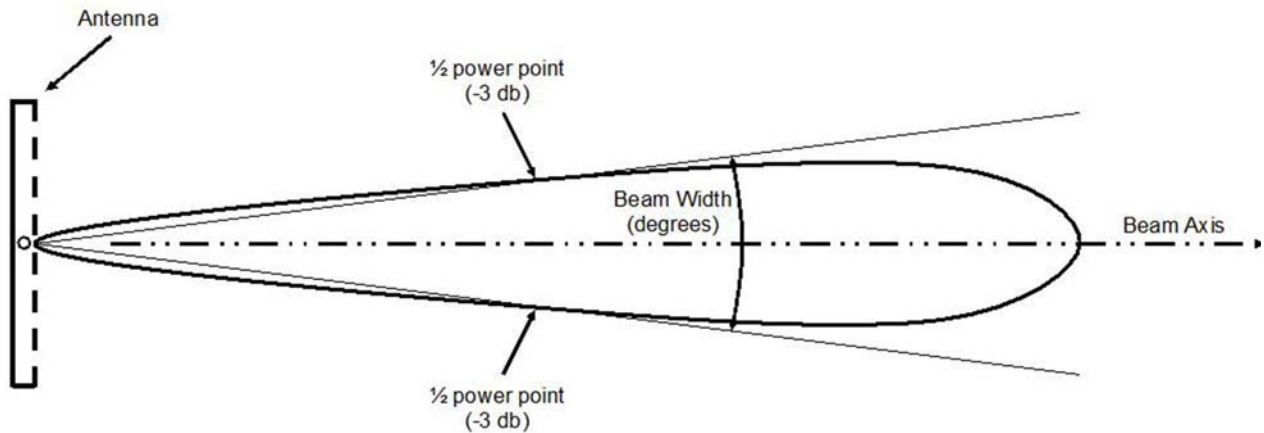
Figure 2306b. Radiation diagram.

The beam width depends upon the frequency or wave-length of the transmitted energy, antenna design, and the dimensions of the antenna. For a given antenna size (an-tenna aperture), narrower beam widths result from using shorter wavelengths. For a given wavelength, narrower beam widths result from using larger antennas, or i.e., beam width is inversely proportional to antenna aperture. Because marine radar antennas are long in the horizontal dimension and narrow in the vertical dimension, they pro-duce a beam that is narrow in the horizontal direction and somewhat wider in the vertical direction. The narrow hor-izontal beam is desirable for bearing accuracy while the wide vertical beam is needed to account for the pitching and rolling of a vessel in a seaway. If the vertical beam was as narrow as the horizontal beam, a vessel in rough weather would experience intermittent target response as

the beam would not intersect the horizon at all times.

The main lobe of the radar beam is composed of a number of separate lobes in the vertical dimension, as op-posed to the single lobe-shaped pattern of radiation as emitted in free space. This phenomenon is the result of in-terference between radar waves taking a direct line-of-sight path to a target, and those waves that are reflected from the surface of the sea before striking the target. There is a slight difference in distance between which the direct and indirect waves must travel. See Figure 2306c. These reflected (indirect) waves interfere either constructively or destructively with the direct waves depending upon the waves' phase relationship. This sets up the possibility of poor target response for objects at certain ranges from own ship.
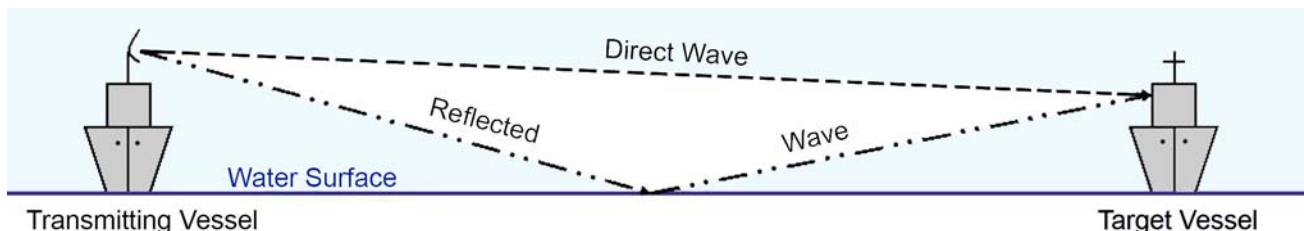

Figure 2306c. Direct and indirect waves.

2307. Effects of Distance, Target Response, Attenuation and Diffraction

Just as a light source reflected in a mirror appears much dimmer than the direct image, radar echoes are much weaker than the transmitted pulses due to the general spreading out of the radar signal energy with distance. The strengths of these echoes are also dependent upon the amount of transmitted energy striking the targets and the size and reflecting properties of the targets known as *radar cross section*.

Attenuation is the scattering and absorption of the energy in the radar beam as it passes through the atmosphere. It causes a decrease in echo strength. Attenuation is greater in 3-cm rather than 10-cm radar. Atmospheric water particles (heavy fog, rain and snow) can significantly degrade the performance of a 3-cm radar system. During periods of heavy precipitation, the radar observer should switch to the 10-cm set if one is available.

Diffraction is the bending of a wave as it passes an obstruction. Because of diffraction there is some illumination of the region behind an obstruction or target by the radar beam. Diffraction effects are greater at the lower frequencies with longer wavelengths (S-Band). Thus, the radar beam of 10-cm radar tends to illuminate more of the shadow region behind an obstruction than the beam of X-Band radar of 3-cm wavelength.

2308. Refraction

If the radar waves traveled in straight lines, the distance to the radar horizon would be dependent only on the power output of the transmitter and the height of the antenna. In other words, the distance to the radar horizon would be the same as that of the geometrical horizon for the antenna height. However, atmospheric density gradients bend radar rays as they travel to and from a target. This bending is called **refraction**.

The distance to the radar horizon does not always limit the distance from which echoes may be received from targets. Assuming that adequate power is transmitted, echoes may be received from targets beyond the radar horizon if their reflecting surfaces extend above it. The distance to the radar horizon is the distance at which the radar rays pass tangent to the surface of the Earth.

The following formula, where h is the height of the antenna in feet, gives the theoretical distance to the radar horizon in nautical miles:

$$D = 1.22 \sqrt{h}.$$

D = the range in nautical miles
h = height of the antenna.

2309. Factors Affecting Radar Interpretation

Radar's value as a navigational aid depends on the navigator's understanding its characteristics and limitations. Whether measuring the range to a single reflective object or trying to discern a shoreline lost amid severe clutter, knowledge of the characteristics of the individual radar used are crucial. Some of the factors to be considered in interpretation are discussed below:

- **Resolution in Range.** In Part A of Figure 2309a, a transmitted pulse has arrived at the second of two targets of insufficient size or density to absorb or reflect all of the energy of the pulse. While the pulse has traveled from the first to the second target, the echo from the first has traveled an equal distance in the opposite direction. At B, the transmitted pulse has continued on beyond the second target, and the two echoes are returning toward the transmitter. The distance between leading edges of the two echoes is twice the distance between targets and so the display will indicate two distinct targets. The correct distance between targets will be shown on the display, which is calibrated to show half the distance traveled out and back. At C the targets are closer together and the pulse length has been increased. The two echoes merge, and on the scope they will appear as a single, large target. At D the pulse length has been decreased, and the two echoes appear separated. The ability of a radar to separate targets close together on the same bearing is called *resolution in range*. It is related primarily to pulse length. The minimum distance between targets that can be distinguished as separate is one half the pulse length. This (half the pulse length) is the apparent depth or thickness of a target but in no way represents that actual size of a small isolated target like a buoy or boat. Thus, several ships close together on nearly the same bearing may appear as an island. Echoes from a number of small boats, piles, breakers, or even a single large ship close to the shore may blend with echoes from the shore, resulting in an incorrect indication of the position and shape of the shoreline.

- **Resolution in Bearing**. Echoes from two or more targets close together at the same range may merge to form a single, wider echo. The ability to separate targets close together at the same range is called *resolution in bearing*. Bearing resolution is a function of two variables: horizontal beam width and range to the targets. A narrower horizontal beam and/or a shorter distance to the objects will allow for better bearing resolution.

- **Height of Antenna and Target**. If the radar horizon is between the transmitting vessel and the target, the lower part of the target will not be visible. A large vessel may appear as a small craft, or a shoreline may appear at some distance inland.
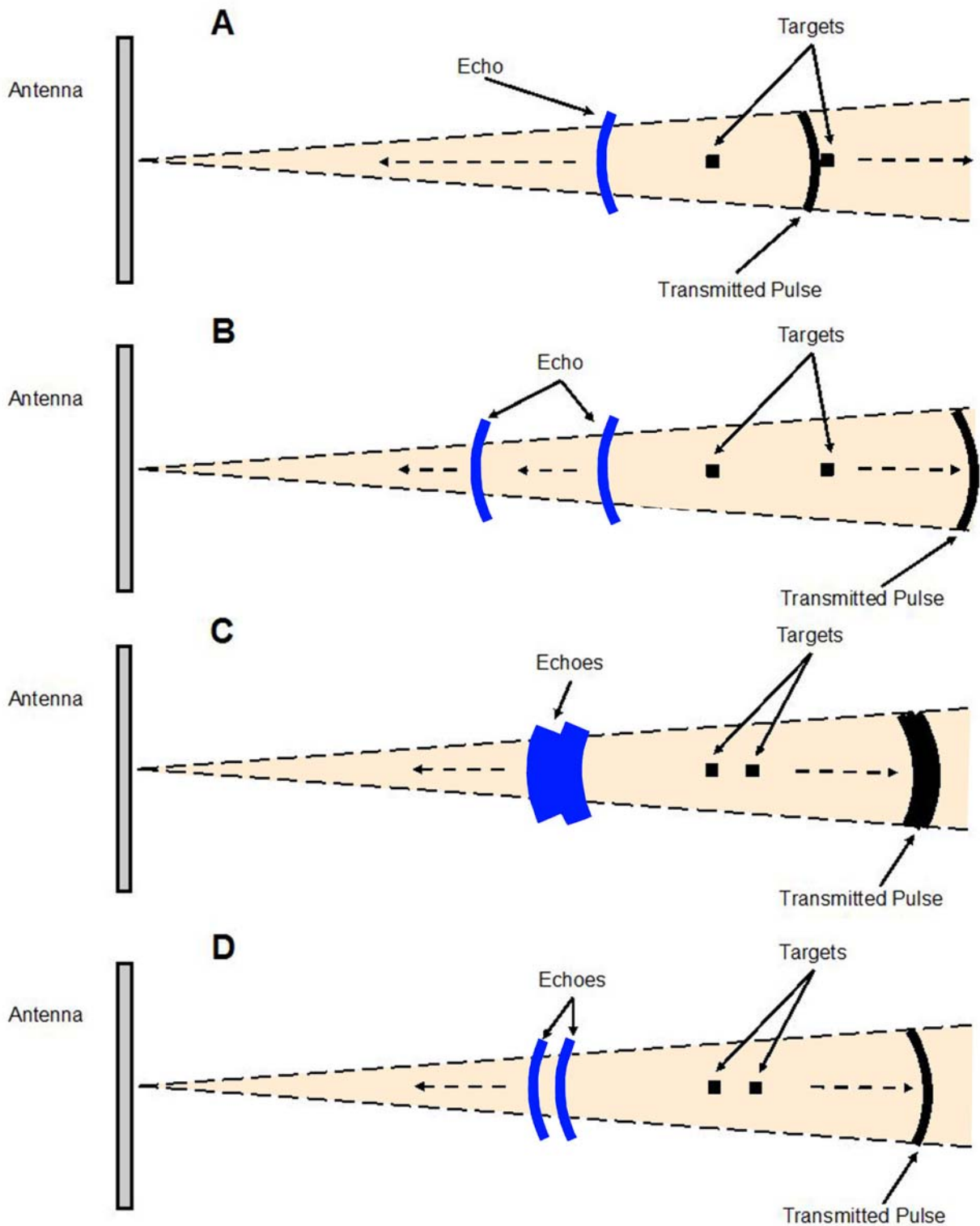
Figure 2309a. Resolution in range.

- **Reflecting Quality and Aspect of Target**. Echoes from several targets of the same size may be quite different in appearance. A metal surface reflects radio waves more strongly than a wooden surface. A surface perpendicular to the beam returns a stronger echo than a non-perpendicular one. A vessel seen broadside returns a stronger echo than one heading directly toward or away. Some surfaces absorb most radar energy rather that reflecting it.

- **Frequency**. A 3-cm radar has the ability to discern smaller targets than a 10-cm set. For example, a very small boat or a submarine periscope might be invisible in S-Band but detectable in X-Band. In a calm sea, a 3-cm radar, properly tuned, can detect a single bird or even a soda can.

Atmospheric noise, sea return, and precipitation complicate radar interpretation by producing **clutter**. Clutter is usually strongest near the vessel. Strong echoes from targets of interest can sometimes be discerned by reducing receiver gain to eliminate weaker signals. By watching the display during several rotations of the antenna, the operator can often discriminate between clutter and a target even when the signal strengths from clutter and the target are equal. The echoes from real targets will remain relatively stationary on the display while those caused by clutter will appear to move around randomly with each sweep.

Another major problem lies in determining which features in the vicinity of the shoreline are actually represented by echoes shown on the display. Particularly in cases where a low lying shore remains below the radar horizon, there may be considerable uncertainty.

A related problem is that certain features on the shore will not return echoes because they are blocked or shadowed from the radar beam by other physical features or obstructions. This shadowing effect in turn causes the image painted on the display to differ from the charted image of the area.

If the navigator is to be able to interpret the presentation on the radar display, he or she must understand the characteristics of radar propagation, the capabilities of his radar set, the reflecting properties of different types of radar targets, and the ability to analyze his chart to determine which charted features are most likely to reflect the transmitted pulses or to be shadowed. Experience gained during clear weather comparison between radar and visual images is invaluable.

Land masses are generally recognizable because of the steady brilliance of the relatively large areas painted on the PPI. Also, land should be at positions expected from the ship's navigational position. Although land masses are readily recognizable, the primary problem is the identification of specific land features. Identification of specific features can be quite difficult because of various factors in addition to shadowing, including distortion resulting from beam width and pulse length, and uncertainty as to just which charted features are reflecting the echoes

Sand spits and smooth, clear beaches normally do not appear on the PPI at ranges beyond 1 or 2 miles because these targets have almost no area that can reflect energy back to the radar. Such a smooth horizontal surface will reflect all radar signals away from the antenna and so are essentially invisible. If waves are breaking over a sandbar, echoes may be returned from the surf. Waves may, however, break well out from the actual shoreline, so that ranging on the surf may be misleading.

Mud flats and marshes normally reflect radar pulses only a little better than a sand spit. The weak echoes received at low tide disappear at high tide. Mangroves and other thick growth may produce a strong echo. Areas that are indicated as swamps on a chart, therefore, may return either strong or weak echoes, depending on the density type, and size of the vegetation growing in the area.

Sand dunes covered with vegetation are usually well back from a low, smooth beach, and the apparent shoreline determined by radar appears at the line of the dunes rather than the true shoreline. This can lead navigators to believe they are farther away from the beach than they really are, a potentially hazardous situation.

Lagoons and inland lakes usually appear as blank areas on a PPI because the smooth water surface returns no energy to the radar antenna. In some instances, even the sandbar or reef surrounding the lagoon may not appear on the PPI because it lies too close to the water.

Coral atolls and long chains of islands may produce long lines of echoes when the radar beam is directed perpendicular to the line of the islands. This indication is especially true when the islands are closely spaced. The reason is that the spreading resulting from the width of the radar beam exceeds the radar's resolution in bearing and causes the echoes to blend into continuous lines. When the same chain of islands is viewed lengthwise, or obliquely, however, each island may produce a separate return if the distance between the islands does not exceed the radar's resolution in range.

Surf breaking on a reef around an atoll produces a ragged, variable line of echoes. Even the smallest of rocks projecting above the surface of the water may be discerned depending on their shape and distance from own ship.

If the land rises in a gradual, regular manner from the shoreline, no part of the terrain produces an echo that is stronger than the echo from any other part. As a result, a general haze of echoes appears on the PPI, and it is difficult to ascertain the range to any particular part of the land.

Blotchy echoes are returned from hilly ground, because the crest of each hill returns a good echo though the area beyond is in a radar shadow. If high receiver gain is used, the pattern may become solid except for very deep depressions.

Low islands ordinarily produce small echoes. When thick palm trees or other foliage grow on the island, strong echoes often are produced because the horizontal surface of the water around the island forms a sort of corner reflector with the vertical surfaces of the trees. As a result, wooded islands give good echoes and can be detected at a much greater range than barren islands.

Sizable land masses may be missing from the radar display because of shadowing. A shoreline which is continuous on the PPI display when the ship is at one position, may not appear continuous when the ship is at another position and scanning the same shoreline. The radar beam may be blocked from a segment of this shoreline by an obstruction such as a

promontory. An indentation in the shoreline, such as a cove or bay, appearing on the PPI when the ship is at one position, may not appear when the ship is at another position nearby. Radar shadowing alone can cause considerable differences between the PPI display and the chart presentation. This effect in conjunction with beam width and pulse length distortion of the PPI display can cause even greater differences, possibly leading to confusion and navigational error.

The returns of objects close to shore may merge with the shoreline image on the PPI, because of distortion effects of horizontal beam width and pulse length. Target images on the PPI are distorted angularly by an amount equal to the effective horizontal beam width. Also, the target images always are distorted radially by an amount at least equal to one-half the pulse length (150 meters per microsecond of pulse length).

See Figure 2309b. It illustrates the effects of own ship position, horizontal beam width, and pulse length on the radar image of a coastline. Because of beam width distortion, a straight, or nearly straight shoreline often appears crescent-shaped on the PPI. This effect is greater with the wider beam widths. Note that this distortion increases as the angle between the beam axis and the shoreline decreases. Figure 2309c, Figure 2309d and Figure
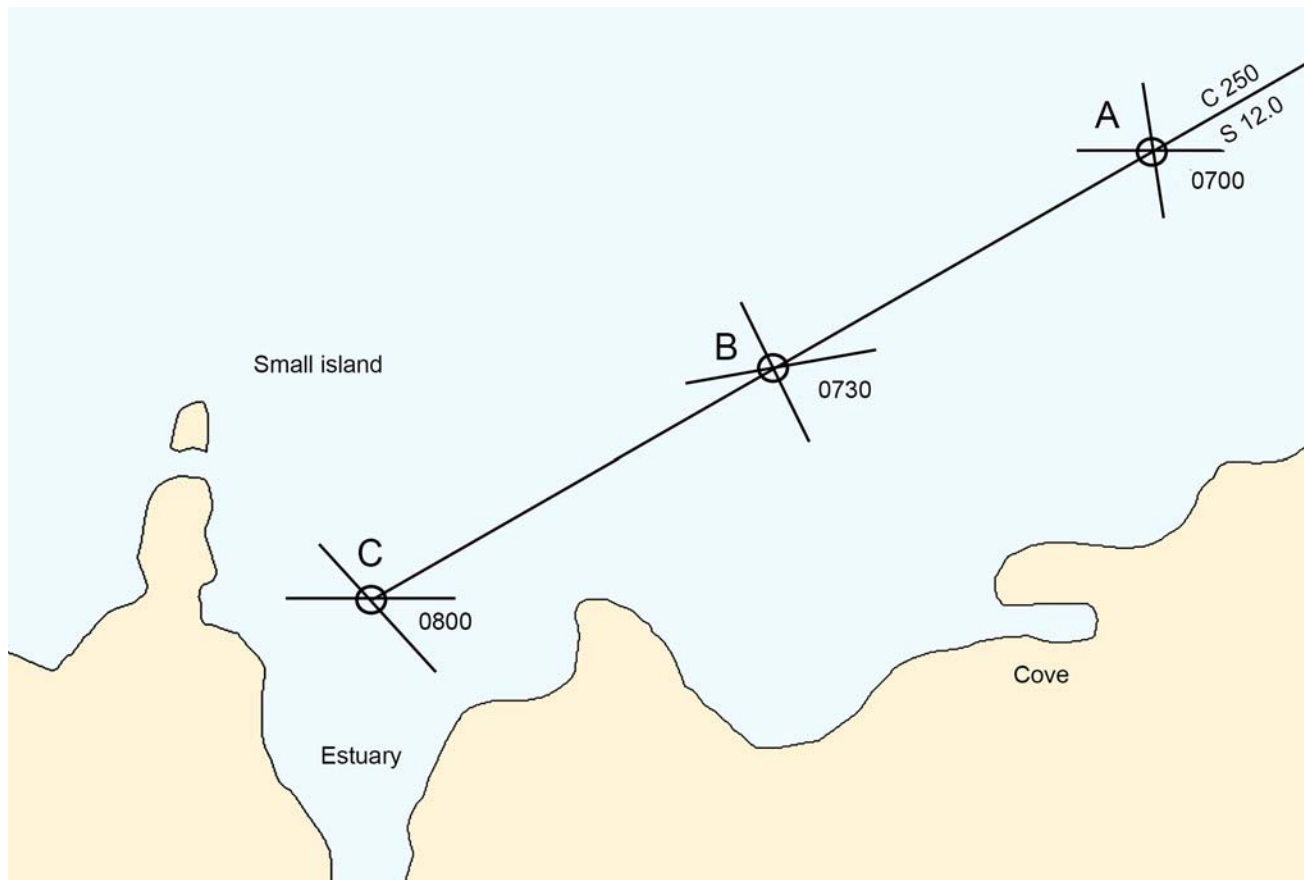
Figure 2309b. Effects of ship's position, beam width, and pulse length on radar shoreline. Figure 2309c, Figure 2309d and Figure 2309e correspond to position A, B and C in the image above.

2309e correspond to positions A, B and C in Figure 2309b.

See Figure 2309f. View A shows the actual shape of the shoreline and the land behind it. Note the steel tower on the low sand beach and the two ships at anchor close to shore. The heavy line in View B represents the shoreline on the PPI. The dotted lines represent the actual position and shape of all targets. Note in particular:

1. The low sand beach is not detected by the radar.
2. The tower on the low beach is detected, but it looks like a ship in a cove. At closer range the land would be detected and the cove-shaped area would begin to fill in; then the tower could not be seen without reducing the receiver gain.

3. The radar shadow behind both mountains. Distortion owing to radar shadows is responsible for more confusion than any other cause. The small island does not appear because it is in the radar shadow.

4. The spreading of the land in bearing caused by beam width distortion. Look at the upper shore of the peninsula. The shoreline distortion is greater to the west because the angle between the radar beam and the shore is smaller as the beam seeks out the more westerly shore.

5. Ship No. 1 appears as a small peninsula. Its return has merged with the land because of the beam width distortion.
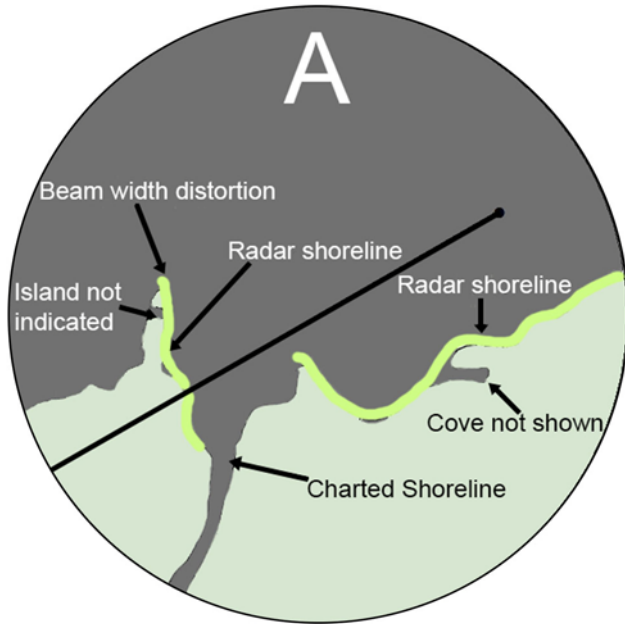
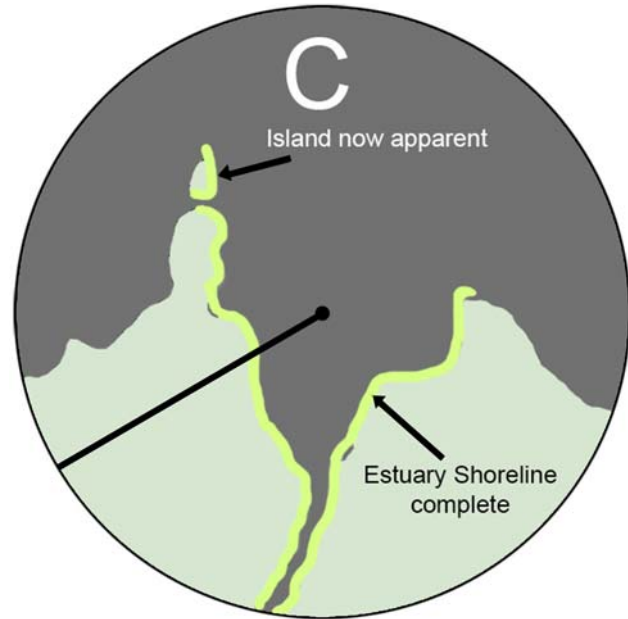*Figure 2309c. 12 mile scale (off-center display) at 0700 position. See position A in* Figure 2309b.

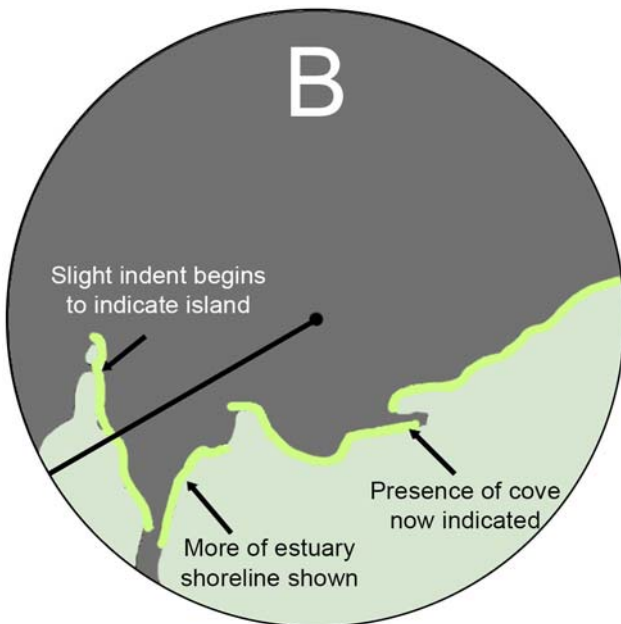*Figure 2309e. 6 mile scale (display center) at 0800 position. See position C in* Figure 2309b.

2310. Recognition of Unwanted Echoes

Indirect or **false echoes** are caused by reflection of the main lobe of the radar beam off own ship's structures such as masts, stacks, kingposts or deck cargo, especially containers. When such reflection from obstructions does occur, the echo will return from a legitimate radar contact to the antenna by the same indirect path. Consequently, the echo will appear on the PPI at the bearing of the reflecting surface. As shown in Figure 2310a, the indirect echo will appear on the PPI at the same range as the direct echo received, assuming that the additional distance by the indirect path is negligible.

Characteristics by which indirect echoes may be recognized are summarized as follows:

1. Indirect echoes will often occur in shadow sectors.
2. They are received on substantially constant relative bearings (the direction of the obstruction), although the true bearing of the radar contact may change appreciably.
3. They appear at the same ranges as the corresponding direct echoes.
4. When plotted, their movements are usually abnormal.
5. Their distorted or fuzzy shapes may indicate that they are not direct echoes.

Side-lobe effects are readily recognized in that they produce a series of echoes (See Figure 2310b) on each side of the main lobe echo at the same range as the latter. Semi-circles, or even complete circles, may be produced.
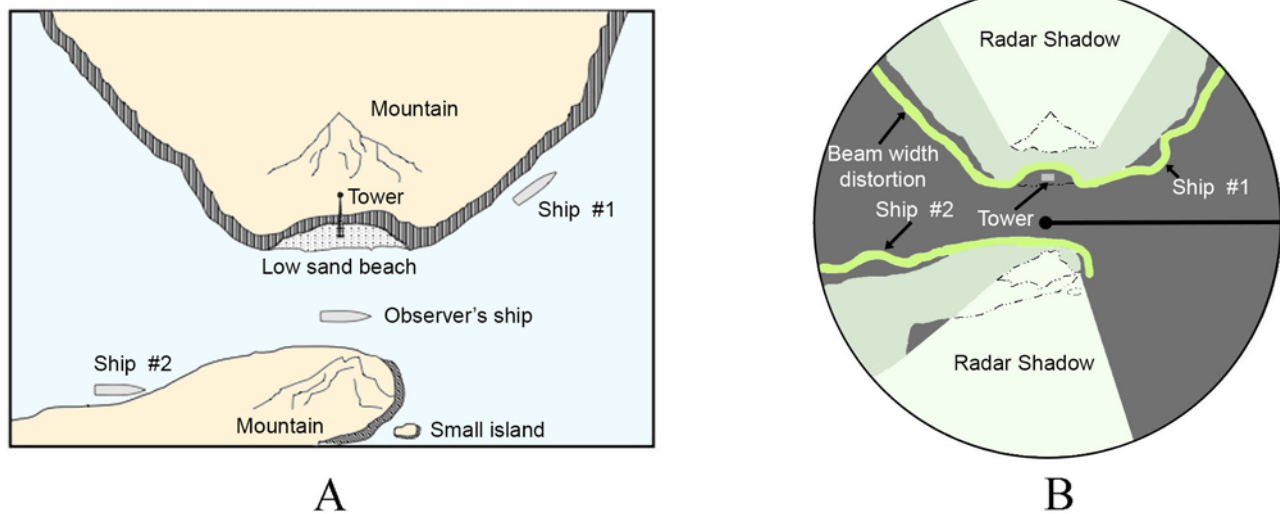

Figure 2309d. 12 mile scale (display centered) at 0730 position. See position B in Figure 2309b.

6. Ship No. 2 also merges with the shoreline and forms a bump. This bump is caused by pulse length and beam width distortion. Reducing receiver gain might cause the ship to separate from land, provided the ship is not too close to the shore. The rain clutter control could also be used to attempt to separate the ship from land by effectively reducing the pulse lengths within the receiver.

Figure 2309f. Distortion effects of radar shadow, beam width, and pulse length.

Because of the low energy of the side-lobes, these effects will normally occur only at the shorter ranges. The effects may be minimized or eliminated, through use of the gain and anti-clutter controls, but always at the risk of failing to detect weaker targets like buoys or small boats. The introduction of slotted wave guide antennas has drastically reduced the side-lobe problem. Nevertheless, when strong reflecting targets are present at close range, side lobe effects will still be encountered and may be difficult to eliminate entirely without severely reducing gain.

Multiple echoes may occur when a strong echo is received from another ship at close range. A second or third or more echoes may be observed on the radarscope at double, triple, or other multiples of the actual range of the radar contact (Figure 2310c).

Second-trace echoes (multiple-trace echoes) are echoes received from a contact at an actual range greater than the radar range setting. If an echo from a distant target is received after the next pulse has been transmitted, the echo will appear on the display at the correct bearing but not at the true range. Second-trace echoes are unusual, except under abnormal atmospheric conditions, or conditions under which super-refraction or ducting is present. Second-trace echoes may be recognized through changes in their positions on the display when changing range scales with different pulse repetition rates (PRR), their hazy, streaky, or distorted shapes (especially noticeable with large land targets), and their erratic movements on plotting.

As illustrated in Figure 2310d, a target echo is detected on a true bearing of 090° at a distance of 7.5 miles. On changing the PRR from 2,000 to 1,800 pulses per second in Figure 2310e, the same target is detected on a bearing of 090° at a distance of 3 miles. The change in the position of the target indicates that the echo is a second-trace echo. The actual distance of the target is the distance as indicated on the PPI plus half the distance the radar waves travel be-

tween pulses. In this case, (162,000 NM/sec ÷ 2000 PPS ÷ 2) + 7.5 = 48 nautical miles.

Naturally, since we are on the 12-mile scale, the target should not be visible and so must be a second-trace echo.

Electronic interference effects, which may occur when near another radar operating in the same frequency band as that of own ship, are usually seen on the radar as a large number of small bright dots either scattered at random or in the form of curving dotted lines extending from the center to the edge of the PPI.

Interference effects are greater at the longer radar range scale settings. Interference effects can be distinguished easily from normal echoes because they do not appear in the same places on successive rotations of the antenna. Most radar systems have interference rejection controls (IR) that eliminate most of the unwanted interference effects.

Stacks, masts, containers, and other structures, may cause a reduction in the intensity of the radar beam beyond these obstructions, especially if they are close to the radar antenna. If the angle at the antenna subtended by the obstruction is more than a few degrees, the reduction of the intensity of the radar beam beyond the obstruction may produce a blind sector. Less reduction in the intensity of the beam beyond the obstructions may produce shadow sectors. Within a shadow sector, small targets at close range may not be detected, while larger targets at much greater ranges will appear.

The echo from an overhead power cable can be wrongly identified as the echo from a ship on a steady bearing and decreasing range. Course changes to avoid the contact are ineffective; the contact remains on a steady bearing, decreasing range. This phenomenon is particularly apparent for the power cable spanning the Straits of Messina.
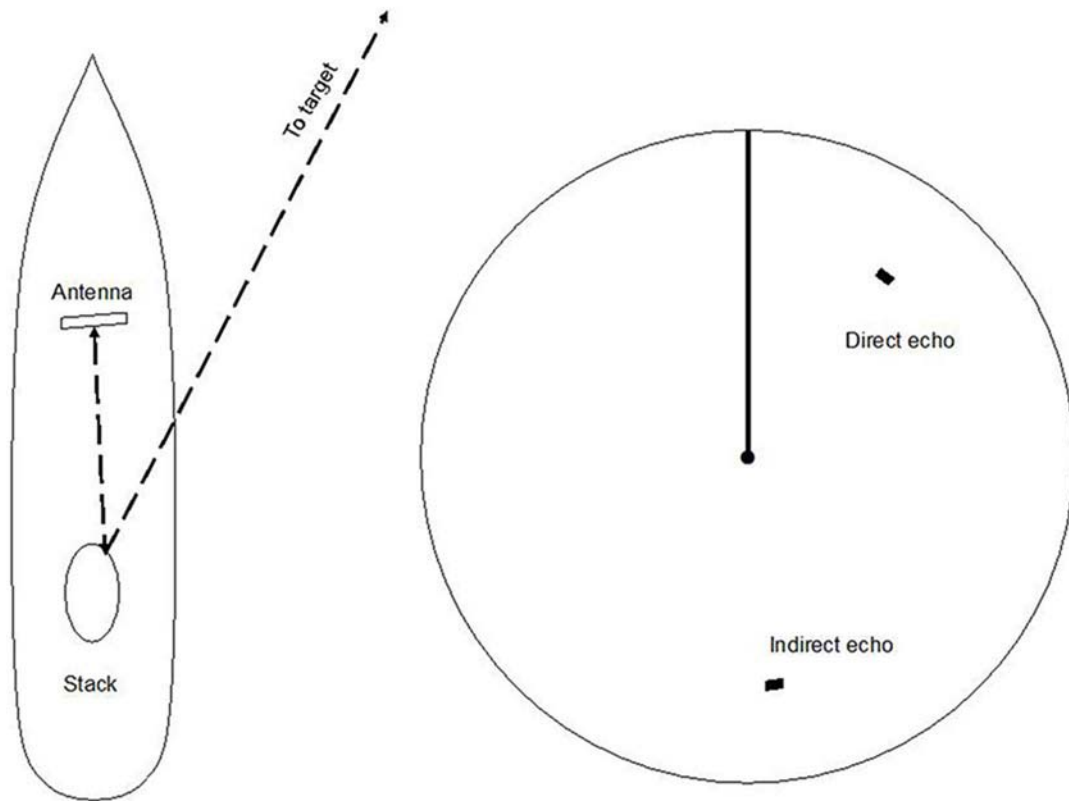
Figure 2310a. Indirect echo.

Figure 2310b. Side lobe effects.

2311. Aids to Radar Navigation

Radar navigation aids help identify radar targets and increase echo signal strength from otherwise poor radar targets.

Buoys are particularly poor radar targets. Weak, fluctuating echoes received from these targets are easily lost in the sea clutter. To aid in the detection of these targets, **radar reflectors**, designated corner reflectors, may be used. These reflectors may be mounted on the tops of buoys or designed into the structure.

Each corner reflector, as shown in Figure 2311a, consists of three mutually perpendicular flat metal surfaces. A radar wave striking any of the metal surfaces or plates will be reflected back in the direction of its source. Maximum energy will be reflected back to the antenna if the axis of the radar beam makes equal angles with all the metal surfaces. Frequently, corner reflectors are assembled in clusters to maximize the reflected signal.

Although radar reflectors are used to obtain stronger echoes from radar targets, other means are required for more positive identification of radar targets. **Radar beacons** are transmitters operating in the marine radar frequency band, which produce distinctive indications on the radar displays of ships within range of these beacons. There are two general
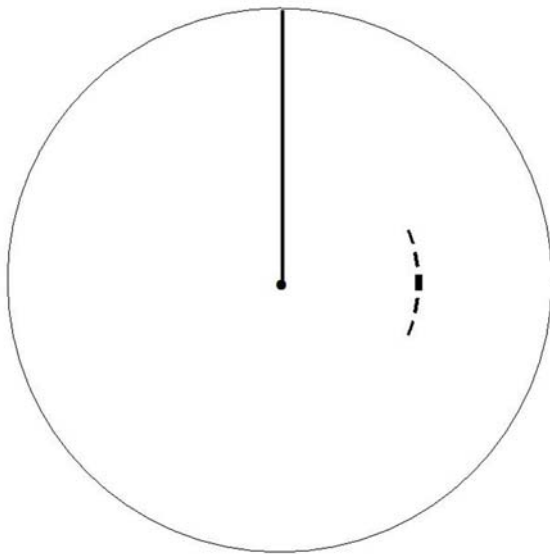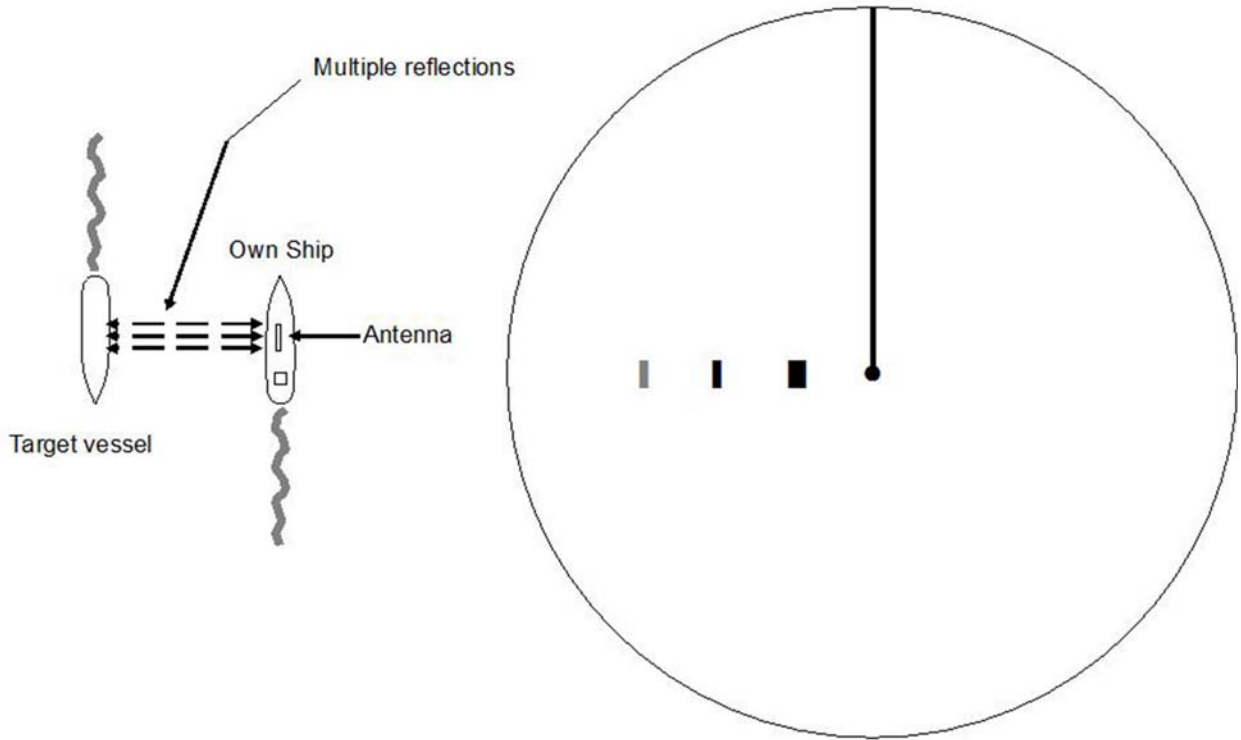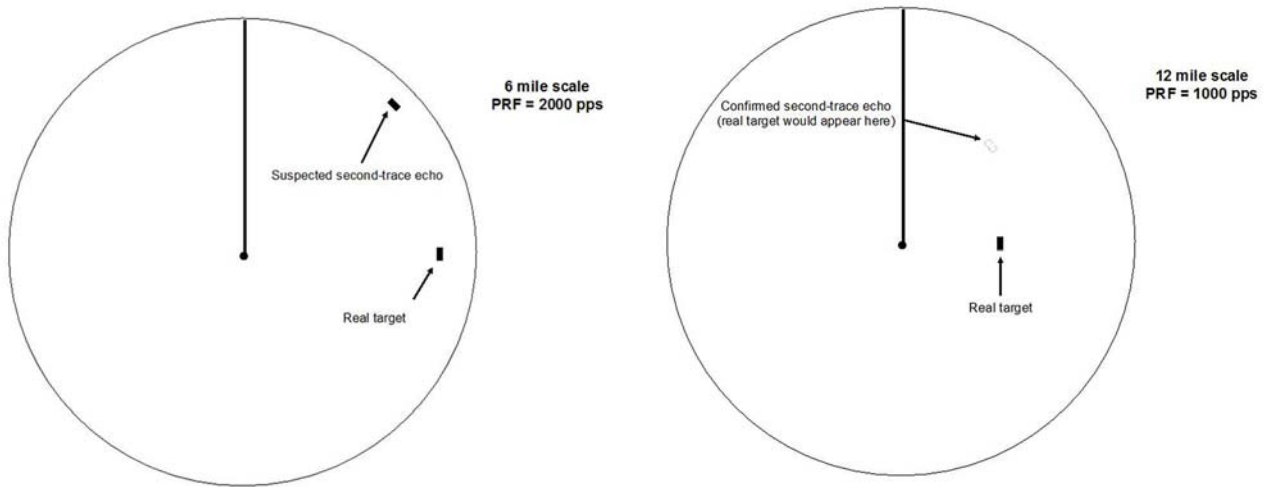
Figure 2310c. Multiple echoes.

Figure 2310d. Second-trace echo.

Figure 2310e. Second-trace echo after altering PRR.

classes of these beacons: **racons**, which provide both bearing and range information to the target, and **ramarks** which provide bearing information only. However, if the ramark installation is detected as an echo on the display, the range will be available also.

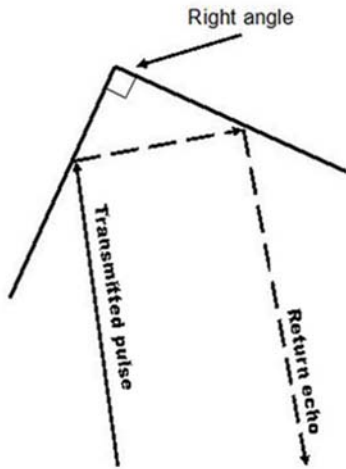A racon is a radar transponder which emits a characteristic signal when triggered by a ship's radar. The signal is emitted on the same frequency as that of the triggering radar, in which case it is superimposed on the ship's radar display automatically. However, the only racons in service are "in band" beacons which transmit in one of the marine radar bands, usually only the 3-centimeter band.

The racon signal appears on the PPI as a radial line originating at a point just beyond the position of the radar
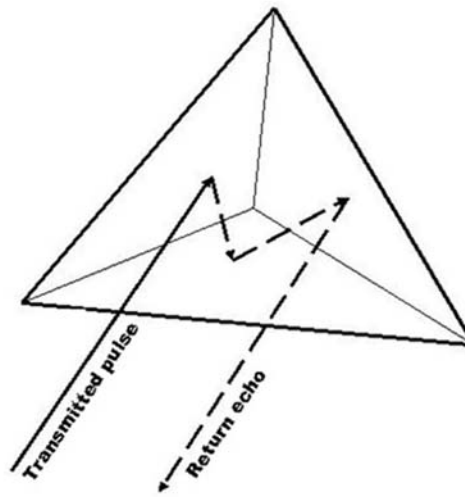
beacon, or as a Morse Code signal as shown in Figure 2311b, emanating from the beacon in a direction radially outward from the center of the display. The Morse Code symbol of the racon signal helps to identify important navigational aids on the navigator's chart.

A ramark is a radar beacon which transmits either continuously or at intervals. The latter method of transmission is used so that the PPI can be inspected without any clutter introduced by the ramark signal on the scope. The ramark signal as it appears on the PPI is a radial line from the center. The radial line may be a continuous narrow line, a broken line a series of dots, or a series of dots and dashes (See Figure 2311c). Ramarks are not as common as racons and are not as useful for navigational purposes as they do not indicate the range to the transmitting beacon.
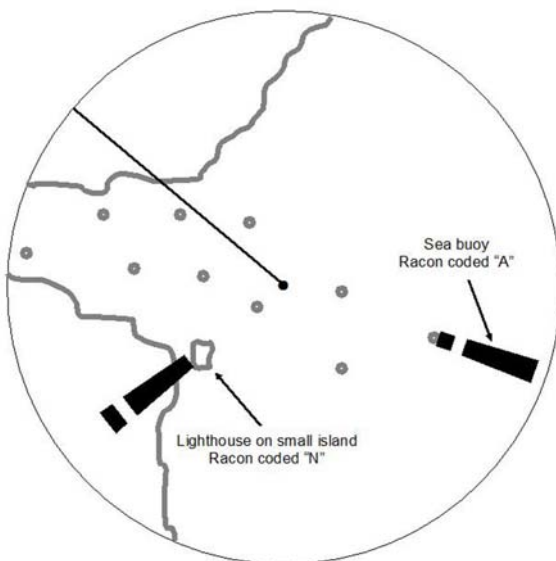

Figure 2311a. Corner reflectors.
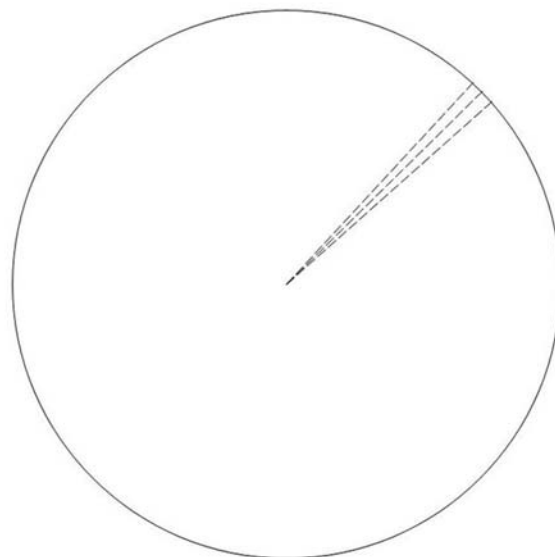

Figure 2311b. Coded racon signal.

Figure 2311c. Ramark appears a broken radial line.

RADAR PILOTING

2312. Introduction

When navigating in restricted waters, a mariner most often relies on visual piloting to provide the accuracy required to ensure ship safety. Visual piloting, however, requires clear weather; often, mariners must navigate through fog or other conditions of restricted visibility. When weather conditions render visual piloting impossible on a vessel not equipped with ECDIS, radar navigation provides a method of fixing a vessel's position with sufficient accuracy to allow safe passage. See Chapter 10 Piloting for a detailed discussion of integrating radar into a piloting procedure on a vessel using paper charts. However, even on ECDIS equipped vessels, radar provides a vital positional cross-checking capability that is paramount to the practice of safe and prudent navigation.

2313. Fix by Radar Ranges

Since radar can more accurately determine ranges than bearings, the most accurate radar fixes result from measuring and plotting a series of ranges to two or more objects. If one measures the range to objects directly ahead or astern first and objects closest to the beam last, the time of the fix will be the time the ranges were measured to objects ahead or astern. In other words, the fix time is the time that distances were measured to objects with the greatest rate of change of range (range rate) due to own ship's motion. This minimizes measurement time delay errors without resorting to the use of running fixes. Record the ranges to the navigation aids used and lay the resulting range arcs down on the chart. Theoretically, these lines of position should intersect at a point coincident with the ship's position at the time of the fix. Where possible, use objects widely separated in bearing (60°-90°) for the greatest accuracy. See Figure 2313.
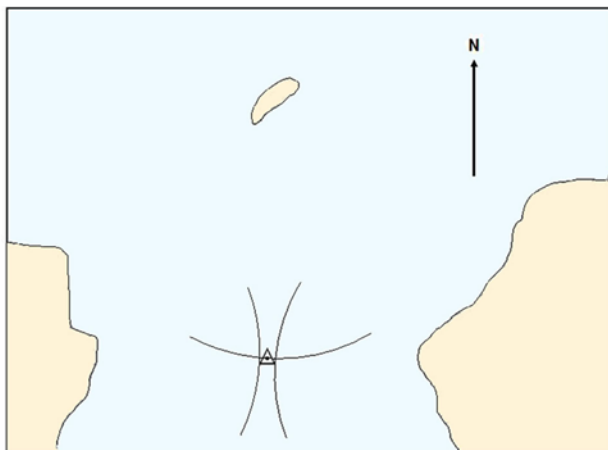

Figure 2313. Fix by radar ranges.

Though verifying soundings is always a good practice in all navigation scenarios, its importance increases when piloting using only radar. One of the most common and serious errors in radar navigation involves object misidentification. These errors can be discovered through correlation of fathometer readings with expected charted depths. Assuming proper operation of the fathometer, soundings give the navigator invaluable conformation on the reliability of radar fixes.

2314. Fix by Radar Bearings

When determining a fix by radar bearings (or visual bearings) take bearings of objects on the beam first and those ahead or astern last. The time of the fix will be the time that the objects abeam were measured. This is because the rate of change of bearing is highest for objects on the beam and lowest for those ahead and astern. Again, this procedure minimizes the fix error due to the time delay in taking a round of bearings.

But the inherent inaccuracy of fixes composed solely of radar bearings as discussed above makes this method less accurate than fixing position by radar ranges. Use this method to plot a position quickly on the chart when approaching restricted waters to obtain an approximate ship's position for evaluating radar targets to use for range measurements. This method is not suitable while piloting in restricted waters and should only be used if no more accurate method (combining visual bearings with radar ranges for example) is available.

2315. Fix by Range and Bearing to One Object

Visual piloting requires bearings from at least two objects; radar, with its ability to determine both bearing and range from one object, allows the navigator to obtain a fix where only a single navigation aid is available. An example of using radar in this fashion occurs in approaching a harbor whose entrance is marked with a single, prominent object such as Chesapeake Light at the entrance of the Chesapeake Bay. Well beyond the range of any land-based visual navigation aid, and beyond the visual range of the light itself, a shipboard radar can detect the light and provide bearings and ranges for the ship's piloting party. But care should be taken. Navigators must ensure they take fixes on the navigation aid and not some nearby stationary vessel.

This methodology is limited by the inherent inaccuracy associated with radar bearings; typically, a radar bearing is accurate to within about 5° of the true bearing due to factors such as beam width distortion. Therefore, the navigator must carefully evaluate the resulting position, possibly checking it with a sounding. If a visual bearing is available from the object, use that bearing instead of the radar bearing

when laying down the fix. This illustrates the basic concept discussed above: radar ranges are inherently more accurate than radar bearings. One must also be aware that even though the radar is gyro stabilized, there may be a gyro error of more than a degree or so. Radar and visual bearings will be in error by that amount.

Prior to using this method, navigators must ensure they have correctly identified the object from which the bearing and range are to be taken. Using only one navigation aid for both lines of position can lead to disaster if the navigation aid is not properly identified.

2316. Fix Using Tangent Bearings and Range

This method combines bearings tangent to an object with a range measurement from some point on that object. The object must be large enough to provide sufficient bearing spread between the tangent bearings; often an island or peninsula works well. Identify some prominent feature of the object that is displayed on both the chart and the radar display. Take a range measurement from that feature and plot it on the chart. Then determine the tangent bearings to the feature and plot them on the chart. The range LOP should not intersect where the tangent bearing LOPs intersect but somewhat farther out. The fix position will be the point midway between the tangent bearing lines along the range LOP (see Figure 2316).

Steep-sided features work the best. Tangents to low, sloping shorelines will seriously reduce accuracy, as will tangent bearings in areas of excessively high tides, which can change the location of the apparent shoreline by many meters.
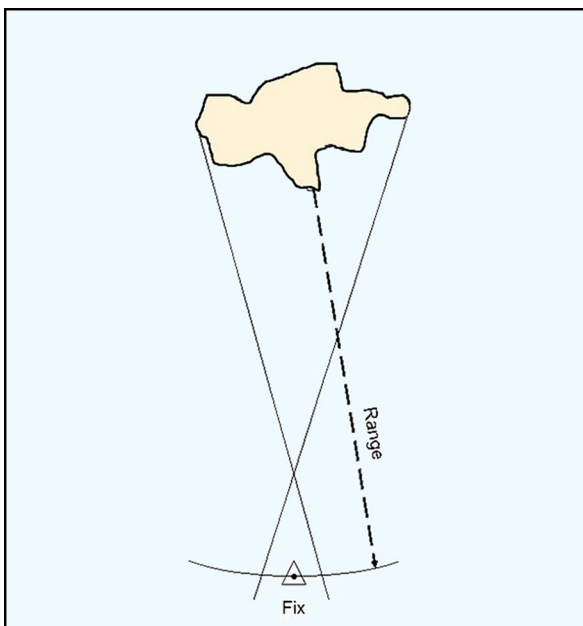

Figure 2316. Fix using tangent bearings and range.

2317. Parallel Indexing

Whenever a vessel is being navigated in confined waters, traditional position fixing methods become inadequate. The time lag inherent in taking a visual bearing, radar bearing or radar range, plotting positions on a nautical chart, obtaining a fix, and then acting on the information with a possible course change may be as much as five minutes or more, even for experienced navigators. If sea room is severely restricted and there are hazards to navigation in the area, such delays could lead to disaster. What we must do in this unforgiving situation is to monitor the vessel's position constantly through continuous position fixes. ECDIS is of course greatly preferable to paper chart navigation in these circumstances but suffers from complete reliance on GPS position fixes. Radar can provide similar real-time navigation capability not reliant on GPS utilizing a technique known as parallel indexing.

A properly prepared parallel indexing plot will quickly show the navigator when the vessel begins to deviate from the desired track. This will enable corrective measures to be taken immediately without resorting to time-consuming standard fixing methods. Parallel indexing can be indispensable when a vessel must be navigated through confined waters during restricted visibility or when executing a critical turn. Also, in areas with few or unreliable navigational aids, parallel indexing can prove decisive to safe navigation.

The first step in setting up a parallel indexing plot is to examine the nautical chart where the piloting will take place. Imagine that we wish to follow a track line that leaves a small island or rock to starboard at a distance of 2 miles off when abeam. The track line course is 045° (see Figure 2317). If we are able to place an electronic line on the radar screen bearing 045°-225° at a range of 2 miles to starboard, all we will have to do when the island comes onto the radar display is to maneuver the ship to keep the island on that line which in turn locates (indexes) the vessel on the track line.

One way to conduct parallel indexing on a modern radar display is to utilize the Electronic Bearing Line (EBL) feature. Most radars have the ability to offset the EBL from the center of the display. This allows it to be used as a single parallel index line. Once the EBL bearing is set to that of the vessel's track line, the origin can be floated out to the desired distance tangent to a Variable Range Marker (VRM) set to that distance.

Modern radar sets are usually fitted with a dedicated parallel indexing (PI) feature that may take many forms depending on the radar manufacturer, and are easier to use than the floating EBL. While the details of these PI features may be quite different, they all have the following in common:

1. The display of an electronic PI line, wholly or partially across the radar screen

2. The PI line is adjustable in direction (bearing) and distance (range) from own ship
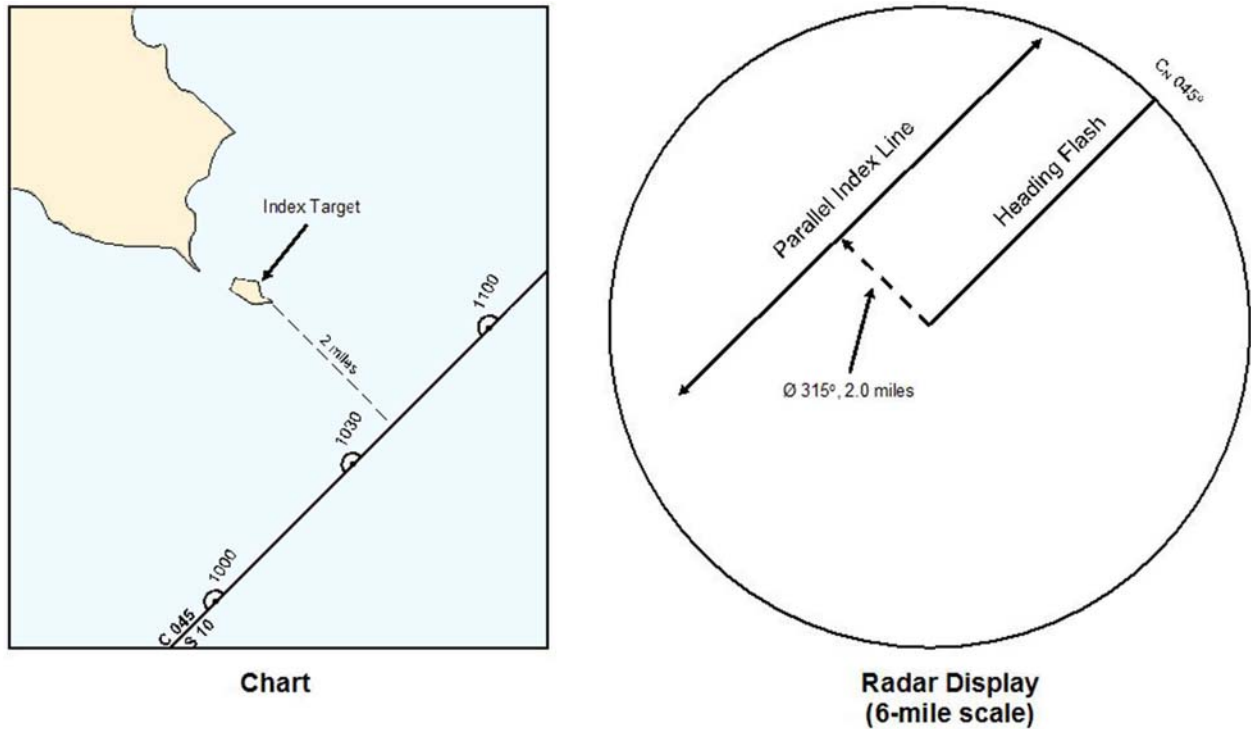
Figure 2317. Parallel indexing setup.

3. Once set at desired bearing and range, the PI line is fixed relative to own ship

It is vital that when placing a single PI line on the radar display, the bearing of the line is set first, then the range. If done in reverse order, the distance of the PI line from own ship to target will be less than desired.

The method described above is very basic and utilizes only a single index line and a single index target. But the level of sophistication of indexing required varies with the situation. A passage may call for many lines on different scales, multiple index targets, margin lines, danger zones and wheel over points. The more complicated the setup, of course, the more time and effort on the part of the navigator is demanded. More complex indexing schemes, however

elegant, also carry a greater risk of error in construction. A point will be reached where there is little to be gained by an excessively elaborate setup because it may also lead to a more cluttered and confusing radar display. A vessel that routinely makes passages through navigationally challenging waters would be better advised to rely more on the ECDIS and use a simpler parallel indexing setup on the radar as a backup and for cross checking.

2318. References

Pecota, S., (2006). *Radar Observer Manual*, 6th. Marine Education Textbooks. Section 2317 **reprinted with permission**.

CHAPTER 24

LORAN NAVIGATION

INTRODUCTION TO LORAN

2400. History and Role of Loran

The theory behind the operation of hyperbolic navigation systems was known in the late 1930s, but it took the urgency of World War II to speed development of the system into practical use. By early 1942, the British had an operating hyperbolic system in use designed to aid in long-range bomber navigation. This system, named Gee, operated on frequencies between 30 MHz and 80 MHz and employed "master" and "slave" transmitters spaced approximately 100 miles apart. The Americans were not far behind the British in development of their own system. By 1943, the U. S. Coast Guard was operating a chain of hyperbolic navigation transmitters that became **Loran-A** (The term Loran was originally an acronym for LOng RAnge Navigation). By the end of the war, the network consisted of over 70 transmitters providing coverage over approximately 30% of the earth's surface.

In the late 1940s and early 1950s, experiments in low frequency Loran produced a longer range, more accurate system. Using the 90-110 kHz band, Loran developed into a 24-hour-a-day, all-weather radionavigation system named **Loran-C**. From the late 1950s, Loran-A and Loran-C systems were operated in parallel until the mid-1970s when the U.S. Government began phasing out Loran-A. The United States continued to operate Loran-C in a number of areas around the world, including Europe, Asia, the Mediterranean Sea, and parts of the Pacific Ocean until the mid-1990s when it began closing its overseas Loran-C stations or transferring them to the governments of the host countries. This was a result of the U.S. Department of Defense adopting the Global Positioning System (GPS) as its primary radionavigation service.

From the 1990s until 2010, Loran served the 48 contiguous states within the United States, their coastal areas, Alaska, and nine of 13 provinces in Canada. North American Loran-C signals, however, were terminated in 2010 in accordance with the 2010 Department of Homeland Security Appropriations Act. The United States Coast Guard ceased transmitting Loran-C signals on 08 FEB 2010 across most of the United States. On 03 AUG 2010, US stations that operated in concert with Canadian stations, and the Canadian stations themselves, ceased transmitting. The Unites States government began dismantling former Loran-C stations until 2014 when the "Howard Coble Coast Guard and Maritime Transportation Act of 2014" was signed into law.

The "Coast Guard Authorization Act of 2015" extended this provision until the Secretary of the agency overseeing the Coast Guard could justify that the Loran-C infrastructure was not needed as a backup to GPS.

As of early 2014, various countries still had operational Loran-C transmitters (or Loran-C equivalents such as the Russian Chayka system) including China, India, Japan, Northwest Europe (e.g., United Kingdom, France, Norway, Germany, and Denmark), Russia, Saudi Arabia, and South Korea. In 2014, Norway and France announced that they would shut down their transmitters on 31 December 2015. Sites in Denmark, Germany, and the U.K. subsequently decided to shut down transmitters as well though the Anthorn transmitter in Cumbria (U.K.) remains active.

In 2001, the "Volpe" report (United States Department of Transportation 2001) outlined key vulnerabilities in the reliance of GPS for critical infrastructure needs. This report (United States Department of Transportation 2001) was the first to mention the use of **Enhanced Loran** or **eLoran** as it is now called. eLoran was conceived and designed as a modern, 21st century replacement to Loran-C. eLoran was outlined as a backup navigational and timing method to a Global Navigation Satellite System (GNSS) such as GPS in instances where a GNSS system may be unavailable or untrustworthy. It was conceived as a result of the "Loran Modernization Program" and has greater accuracy than Loran-C and new features (International Loran Association 2007). The eLoran definition document, stating the design of the eLoran system, was released on 16 October 2007 (International Loran Association 2007) outlining the requirements that this new method must have and how it differs from Loran-C. As of 2016 eLoran is currently being tested at stations across the United States (UrsaNav 2015). South Korea is set to build eLoran stations in response to North Korean GPS jamming (GPS World 2016) and other countries are seeking to build eLoran infrastructure. With the cessation of signals in Northwest Europe on 31 December 2015, eLoran is no longer available for navigational use anywhere in the world. The UK continues to operate their Anthorn eLoran station for the provision of data communications and timing. eLoran signals are also transmitted from the former USCG Loran Support Unit in Wildwood, New Jersey as part of a Cooperative Research and Development Agreement (CRADA) between the DHS, USCG, UrsaNav, and Harris Corporation.

Additional information on eLoran may be found at the end of this chapter. See Section 2418.

LORAN-C DESCRIPTION

2401. Summary of Operation

The Loran-C signal is still transmitted on a continuous basis from stations in China, South Korea, and the Kingdom of Saudi Arabia. Additionally, the Chayka signal is still transmitted from stations in Russia. Modern Enhanced Loran (eLoran) is intermittently tested in the UK and US. Legacy Loran-C receivers can be used with eLoran. However, legacy receivers cannot take advantage of the Loran Data Channel, a key component of eLoran that is necessary to achieve the enhanced capabilities. All of the information presented about Loran-C is given because it is the basis of Loran-C navigation and all Loran-C navigation methods would also apply using eLoran. Some information and aids, such as Loran-C charts are not directly available or maintained by the United States government.

The Loran-C (hereafter referred to simply as Loran) system consists of **transmitting stations**, which are placed several hundred miles apart and organized into **chains**. Within a Loran chain, one station is designated as the master station and the others as **secondary stations**. Every Loran chain contains at least one master station and two secondary stations in order to provide at least two lines of position (LOP).

The master and secondary stations transmit radio pulses at precise time intervals. A Loran receiver measures the **time difference** (or time delay) **(TD)** between the vessel's receipt of the master and secondary station signal transmissions. The elapsed time is converted to distance, the locus of points having the same TD between the master and secondary forms the hyperbolic LOP. The navigator records the delayed TD values and applies them to the chart by interpolating between the printed lattice lines, manually plotting the LOPs parallel to lattice lines. The intersection of two or more of these LOPs produces a fix of the vessel's position.

There are two methods by which the navigator can convert this information into a geographic position. The first involves the use of a chart overprinted with a Loran **time delay lattice** consisting of hyperbolic TD lines spaced at convenient intervals. The navigator plots the displayed TDs by interpolating between the lattice lines printed on the chart, manually plotting the fix where lines intersect to determine latitude and longitude. In the second method, computer algorithms in the receiver's software convert the TDs to latitude and longitude for display.

As with other computerized navigation receivers, a typical Loran receiver can accept and store **waypoints**. Waypoints are sets of coordinates that describe either locations of navigational interest or points along a planned route. Waypoints may be entered by visiting the spot of interest and pressing the appropriate receiver control key, or by keying in the waypoint coordinates manually, either as a TD or latitude-longitude pair. If using waypoints to mark a planned route, the navigator can use the receiver to monitor the vessel's progress in relation to the track between each waypoint. By continuously providing parameters such as cross-track error, course over ground, speed over ground, and bearing and distance to next waypoint, the receiver continually serves as a check on the primary navigation plot.

2402. Components of the Loran System

For the marine navigator, the components of the Loran system consist of the land-based transmitting stations, the Loran **receiver** and **antenna**, and the **Loran charts**. In addition to the master and secondary transmitting stations, land-based Loran facilities also include the primary and secondary **system area monitor sites**, the **control station** and a precise time reference. The transmitters emit Loran signals at precisely timed intervals. The monitor sites and control stations continually measure and analyze the characteristics of the Loran signals received to detect any anomalies or out-of-specification conditions. Some transmitters serve only one function within a chain (i.e., either master or secondary). However, in many instances, one transmitter transmits signals for each of two adjacent chains. This practice is termed **dual rating**.

Loran receivers exhibit varying degrees of sophistication, but their signal processing is similar. The first processing stage consists of **search and acquisition**, during which the receiver searches for the signal from a particular Loran chain and establishes the approximate time reference of the master and secondaries with sufficient accuracy to permit subsequent settling and tracking.

After search and acquisition, the receiver enters the **settle** phase. In this phase, the receiver searches for and detects the front edge of the Loran pulse. After detecting the front edge of the pulse, it selects the correct cycle of the pulse to track.

Having selected the correct tracking cycle, the receiver begins the **tracking and lock** phase, in which the receiver maintains synchronization with the selected received signals. Once this phase is reached, the receiver displays either the time difference of the signals or the computed latitude and longitude.

2403. The Loran Signal

The Loran signal consists of a series of 100 kHz pulses sent first by the master station and then, in turn, by the secondary stations. Both the shape of the individual pulse and the pattern of the entire pulse sequence are shown in Figure 2403a. As compared to a carrier signal of constant amplitude, pulsed transmission allows the same signal range to be achieved with a lower average output power. Pulsed transmission also yields better signal identification properties and more precise timing of the signals.
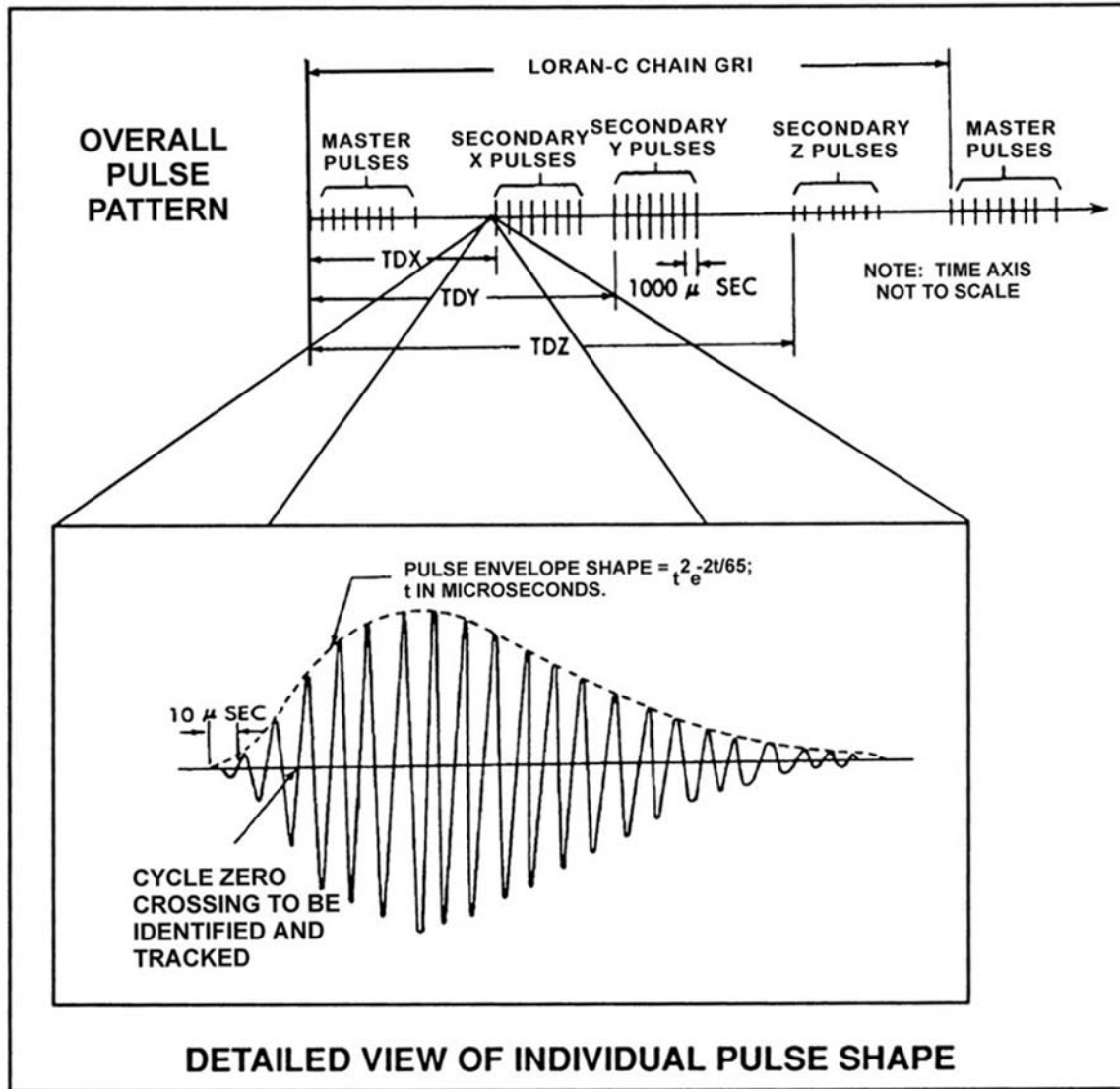
Figure 2403a. Pulse pattern and shape for Loran C transmission.

The individual sinusoidal Loran pulse exhibits a steep rise to its maximum amplitude within 65 μsec of emission and an exponential decay to zero within 200 to 300 μsec. The signal frequency is nominally defined as 100 kHz; in actuality, the signal is designed such that 99% of the radiated power is contained in a 20 kHz band centered on 100 kHz.

The Loran receiver is programmed to track the signal on the cycle corresponding to the carrier frequency's third positive crossing of the x-axis. This occurrence, termed the **standard zero crossing**, is chosen for two reasons. First, it is late enough for the pulse to have built up sufficient signal strength for the receiver to detect it. Second, it is early enough in the pulse to ensure that the receiver is detecting the transmitting station's ground wave pulse and not its sky wave pulse. Sky wave pulses are affected by atmospheric refraction and, if used unknowingly, would introduce large

errors into positions determined by a Loran receiver. The pulse architecture described here reduces this major source of error.

Another important parameter of the pulse is the **envelope-to-cycle difference (ECD)**. This parameter indicates how propagation of the signal causes the pulse shape envelope (i.e., the imaginary line connecting the peak of each sinusoidal cycle) to shift in time relative to the zero crossings. The ECD is important because Loran-C receivers use the precisely shaped pulse envelope to identify the correct zero crossing. Transmitting stations are required to keep the ECD within defined limits. Many receivers display the received ECD as well.

Next, individual pulses are combined into sequences. For the master signal, a series of nine pulses is transmitted, the first eight spaced 1000 μsec apart followed by a ninth transmitted 2000 μsec after the eighth. Secondary stations
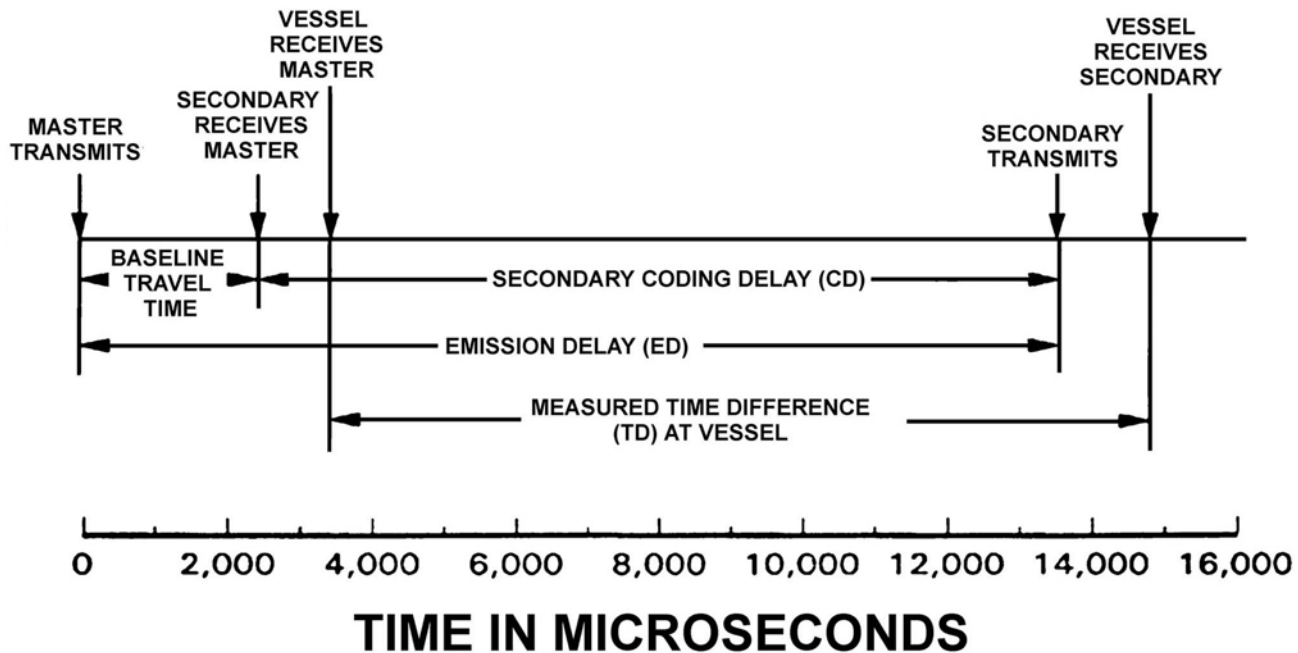
Figure 2403b. The time axis for Loran TD for point "A."

transmit a series of eight pulses, each spaced 1000 μsec apart. Secondary stations are given letter designations of V, W, X, Y, and Z; this letter designation indicates the order in which they transmit following the master. If a chain has two secondaries, they will be designated Y and Z. If a chain has three secondaries, they are X, Y and Z, and so on. Some exceptions to this general naming pattern exist (e.g., W, X and Y for some 3-secondary chains).

The spacing between the master signal and each of the secondary signals is governed by several parameters as illustrated in Figure 2403b. The general idea is that each of the signals must clear the entire chain coverage area before the next one is transmitted, so that no signal can be received out of order. The time required for the master signal to travel to the secondary station is defined as the average **baseline travel time (BTT)**, or **baseline length (BLL)**. To this time interval is added an additional delay defined as the **secondary coding delay (SCD)**, or simply **coding delay (CD)**. The total of these two delays is termed the **emission delay (ED)**, which is the exact time interval between the transmission of the master signal and the transmission of the secondary signal. Each secondary station has its own ED value. To ensure the proper sequence, the ED of secondary Y is longer than that of X, and the ED of Z is longer than that of Y.

Once the last secondary has transmitted, the master transmits again, and the cycle is repeated. The time to complete this cycle of transmission defines an important characteristic for the chain: the **group repetition interval (GRI)**. The group repetition interval divided by ten yields the chain's numeric designator. For example, the interval

between successive transmissions of the master pulse group for the Northeast U.S. Chain (commonly referred to as "NEUS") is 99,600 μsec, just less than one tenth of a second. From the definition above, the GRI designator for this chain is defined as 9960. As mentioned previously, the GRI must be sufficiently large to allow the signals from the master and secondary stations in the chain to propagate fully throughout the region covered by the chain before the next cycle of pulses begins.

Two additional characteristics of the pulse group are **phase coding** and **blink coding**. In phase coding, the phase of the 100 kHz carrier signal is reversed from pulse to pulse in a preset pattern that repeats every two GRIs. Phase coding allows a receiver to remove skywave contamination from the groundwave signal. Loran-C signals travel away from a transmitting station in all possible directions. Groundwave is the Loran energy that travels along the surface of the earth. Skywave is Loran energy that travels up into the sky. The ionosphere reflects some of these skywaves back to the earth's surface. The skywave always arrives later than the groundwave because it travels a greater distance. The skywave of one pulse can thus contaminate the ground wave of the next pulse in the pulse group. Phase coding ensures that this skywave contamination will always "cancel out" when all the pulses of two consecutive GRIs are averaged together.

Blink coding provides integrity to the received Loran signal. When a signal from a secondary station is out of tolerance and therefore temporarily unsuitable for navigation, or **out-of-tolerance (OOT)**, the affected secondary station will blink; that is, the first two pulses of the affected second-

ary station are turned off and on in a repeating cycle, 3.6 seconds off and 0.4 seconds on. The receiver detects this condition and displays it to the operator. When the blink indication is received, the operator should not use the affected secondary station. If a station's signal will be temporarily shut down for maintenance, interruption notifications will be promulgated by responsible local authorities. When a secondary station is blinking, the master station will also blink its ninth pulse in a predetermined pattern that identifies the out-of-tolerance secondary or secondaries. If a master station is out of tolerance, all secondaries in the affected chain will blink. If the entire chain is OOT, then the master and all secondaries will blink.

Two other concepts important to the understanding of Loran operation are the **baseline** and **baseline extension**. The geographic line connecting a master to a particular secondary station is defined as the station pair baseline. The baseline is, in other words, that part of a great circle on which lie all the points connecting the two stations. The extension of this line beyond the stations to encompass the points along this great circle not lying between the two stations defines the baseline extension. The optimal region for hyperbolic navigation occurs in the vicinity of the baseline, while the most care must be exercised in the regions near the baseline extension. These concepts are further developed in the next few articles.

2404. Loran Theory of Operation

In Loran navigation, the locus of points having a constant difference in distance between an observer and each of two transmitter stations defines a hyperbola, which is a line of position.

Assuming a constant speed of propagation of electromagnetic radiation in the atmosphere, the time difference in the arrival of electromagnetic radiation from the two transmitter sites is proportional to the distance between each of the transmitting sites, thus creating the hyperbola on the earth's surface. The following equations demonstrate this proportionality between distance and time:

Distance = Velocity x Time

or, using algebraic symbols

d = v x t

Therefore, if the velocity (v) is constant, the distance between a vessel and each of two transmitting stations will be directly proportional to the time delay detected at the vessel between pulses of electromagnetic radiation transmitted from the two stations.

An example illustrates the concept. As shown in Figure 2404, let us assume that two Loran transmitting stations, a master and a secondary, are located along with an observer in a Cartesian coordinate system whose units are in nautical miles. We assume further that the master station, designated "M", is located at coordinates $(x,y) = (-200,0)$ and the secondary, designated "X," is located at $(x,y) = (+200,0)$. An observer with a receiver capable of detecting electromagnetic radiation is positioned at any point "A" whose coordinates are defined as (x_a, y_a).

Note that for mathematical convenience, these hyperbola labels have been normalized so that the hyperbola perpendicular to the baseline is labeled zero, with both negative and positive difference values. In actual practice, all Loran TDs are positive.

The Pythagorean theorem can be used to determine the distance between the observer and the master station; similarly, one can obtain the distance between the observer and the secondary station:

$$distance_{am} = \left[(x_a + 200)^2 + y_a^2 \right]^{0.5}$$

$$distance_{ax} = \left[(x_a - 200)^2 + y_a^2 \right]^{0.5}$$

The difference between these distances (D) is:

$$D = distance_{am} - distance_{ax}$$

Substituting,

$$D = [(x_a + 200)^2 + y_a^2]^{0.5} - \left[(x_a - 200)^2 + y_a^2 \right]^{0.5}$$

With the master and secondary stations in known geographic positions, the only unknowns are the two geographic coordinates of the observer.

Each hyperbolic line of position in Figure 2404 represents the locus of points for which (D) is held constant. For example, if the observer above were located at point A (271.9, 200) then the distance between that observer and the secondary station (the point designated "X" in Figure 2404) would be 212.5 NM. In turn, the observer's distance from the master station would be 512.5 NM. The function D would simply be the difference of the two, or 300 NM. For every other point along the hyperbola passing through A, distance D has a value of 300 NM. Adjacent LOPs indicate where D is 250 NM or 350 NM.

To produce a fix, the observer must obtain a similar hyperbolic line of position generated by another master-secondary pair. Let us say another secondary station "Y" is placed at point (50,500). Mathematically, the observer will then have two equations corresponding to the M-X and M-Y TD pairs:

$$D_1 = [(x_a + 200)^2 + y_a^2]^{0.5} - [(x_a - 200)^2 + y_a^2]^{0.5}$$

$$D_2 = [(x_a + 200)^2 + y_a^2]^{0.5} - [(x_a - 50)^2 + (y_a - 500)^2]^{0.5}$$

Distances D_1 and D_2 are known because the time differences have been measured by the receiver and converted to these distances. The two remaining unknowns, x_a and y_a, may then be solved.

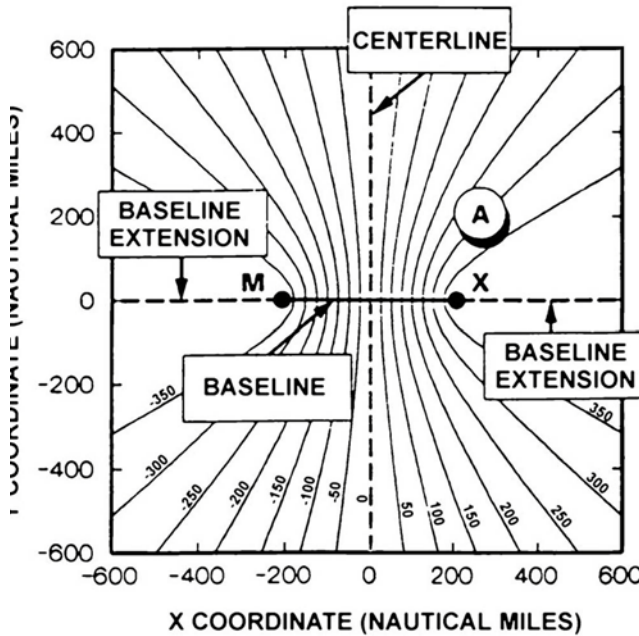The above example is expressed in terms of distance in

Figure 2404. Depiction of Loran LOP's.

nautical miles. Because the navigator uses TDs to perform Loran hyperbolic navigation, let us rework the example for the M-X TD pair in terms of time rather than distance, adding timing details specific to Loran. Let us assume that electromagnetic radiation travels at the speed of light (one nautical mile traveled in 6.18 μsec). The distance from master station M to point A was 512.5 NM. From the relationship just defined between distance and time, it would take a signal (6.18 μsec/NM) 512.5 NM = 3,167 μsec to travel from the master station to the observer at point A. At the arrival of this signal, the observer's Loran receiver would start the TD measurement. Recall from the general discussion above that a secondary station transmits after an emission delay equal to the sum of the baseline travel time and the secondary coding delay. In this example, the master and the secondary are 400 NM apart; therefore, the baseline travel time is (6.18 μsec/NM) 400 NM = 2,472 μsec. Assuming a secondary coding delay of 11,000

μsec, the secondary station in this example would transmit (2,472 + 11,000) μsec or 13,472 μsec after the master station. The secondary signal then propagates over a distance 212.5 NM to reach point A, taking (6.18 μsec/NM) 212.5 NM = 1,313 μsec to do so. Therefore, the total time from *transmission* of the master signal to the *reception* of the secondary signal by the observer at point A is (13,472 + 1,313) μsec = 14,785 μsec.

Recall, however, that the Loran receiver measures the time delay between *reception* of the master signal and *reception* of the secondary signal. Therefore, the time quantity above must be corrected by subtracting the amount of time required for the signal to travel from the master transmitter to the observer at point A. This amount of time was 3,167 μsec. Therefore, the TD observed at point A in this hypothetical example would be (14,785 - 3,167) μsec or 11,618 μsec. Once again, this time delay is a function of the simultaneous differences in distance between the observer and the two transmitting stations, and it gives rise to a hyperbolic line of position which can be crossed with another LOP to fix the observer's position.

2405. Allowances for Non-Uniform Propagation Rates

The initial calculations above assumed the speed of light in free space; however, the actual speed at which electromagnetic radiation propagates on earth is reduced both by the atmosphere through which it travels and by the conductive surfaces—sea and land—over which it passes. The specified accuracy needed from Loran therefore requires three corrections to the propagation speed of the signal.

The reduction in propagation speed caused by the atmosphere is represented by the first correction term: the **Primary Phase Factor (PF)**. Similarly, a **Secondary Phase Factor (SF)** accounts for the reduced propagation speed caused by traveling over seawater. These two corrections are transparent to the operator because they are uniformly incorporated into all calculations represented on charts and in Loran receivers.

Because land surfaces have lower conductivity than seawater, the propagation speed of the Loran signal passing over land is further reduced as compared to the signal passing over seawater. A third and final correction, the **Additional Secondary Phase Factor (ASF)**, accounts for the delay caused by the land conductivity when converting time delays to distances and then to geographic coordinates. Depending on the mariner's location, signals from some Loran transmitters may have traveled hundreds of miles over land and must be corrected to account for this non-seawater portion of the signal path. Of the three corrections mentioned in this section, this is the most complex and the most important one to understand, and is accordingly treated in detail in Section 2410.

LORAN ACCURACY

2406. Defining Accuracy

Specifications of Loran and other radionavigation systems typically refer to three types of accuracy: **absolute, repeatable** and **relative.**

Absolute accuracy, also termed predictable or geodetic accuracy, is the accuracy of a position with respect to the geographic coordinates of the earth. For example, if the navigator plots a position based on the Loran latitude and longitude (or based on Loran TDs) the difference between the Loran position and the actual position is a measure of the system's absolute accuracy.

Repeatable accuracy is the accuracy with which the navigator can return to a position whose coordinates have been measured previously with the same navigational system. For example, suppose a navigator were to travel to a buoy and note the TDs at that position. Later, suppose the navigator, wanting to return to the buoy, returns to the previously measured TDs. The resulting position difference between the vessel and the buoy is a measure of the system's repeatable accuracy.

Relative accuracy is the accuracy with which a user can measure position relative to that of another user of the same navigation system at the same time. If one vessel were to travel to the TDs determined by another vessel, the difference in position between the two vessels would be a measure of the system's relative accuracy.

The distinction between absolute and repeatable accuracy is the most important one to understand. With the correct application of ASFs and within the **coverage area** defined for each chain, the absolute accuracy of the Loran system varies from between 0.1 and 0.25 nautical miles. However, the repeatable accuracy of the system is much better, typically between 18 and 90 meters (approximately 60 to 300 feet) depending on one's location in the coverage area. If the navigator has been to an area previously and noted the TDs corresponding to different navigational aids (e.g., a buoy marking a harbor entrance), the high repeatable accuracy of the system enables location of the buoy in adverse weather. Similarly, selected TD data for various harbor navigational aids and other locations of interest have been collected and recorded and is generally commercially available. This information provides an excellent backup navigational source to conventional harbor approach navigation.

2407. Limitations to Loran Accuracy

There are limits on the accuracy of any navigational system, and Loran is no exception. Several factors that contribute to limiting the accuracy of Loran as a navigational aid are listed in Table 2407 and are briefly discussed in this section. Even though all these factors except operator error are included in the published accuracy of Loran, the mari-

ner's aim should be to have a working knowledge of each one and minimize any that are under their control so as to obtain the best possible accuracy.

The geometry of LOPs used in a Loran fix is of prime importance to the mariner. Because understanding of this factor is so critical to proper Loran operation, the effects of crossing angles and gradients are discussed in detail in the Section 2408. The remaining factors are briefly explained as follows.

The age of the North American (i.e. US and Canadian) Loran transmitting equipment varies from station to station. When some older types of equipment are switched from standby to active and vice versa, a slight timing shift as large as tens of nanoseconds may be seen. This is so small that it is undetectable by most marine receivers, but since all errors accumulate, it should be understood as part of the Loran "error budget."

The effects of actions to control chain timing are similar. The timing of each station in a chain is controlled based on data received at the primary system area monitor site. Signal timing errors are kept as near to zero as possible at the primary site, making the absolute accuracy of Loran generally the best in the vicinity of the primary site. Whenever, due to equipment casualty or to accomplish system maintenance, the control station shifts to the secondary system area monitor site, slight timing shifts may be introduced in parts of the coverage area.

Atmospheric noise, generally caused by lightning, reduces the **signal-to-noise ratio (SNR)** available at the receiver. This in turn degrades accuracy of the LOP. Man-made noise has a similar effect on accuracy. In rare cases, a man-made noise source whose carrier signal frequency or harmonics are near 100 kHz (such as the constant carrier control signals commonly used on high-tension power lines) may also interfere with lock-on and tracking of a Loran receiver. In general, Loran stations that are the closest to the user will have the highest SNR and will produce LOPs with the lowest errors. Geometry, however, remains a key factor in producing a good fix from combined LOPs. Therefore, the best LOPs for a fix may not all be from the very nearest stations.

The user should also be aware that the propagation speed of Loran changes with time as well. Temporal changes may be seasonal, due to snow cover or changing groundwater levels, or diurnal, due to atmospheric and surface changes from day to night. Seasonal changes may be as large as 1 μsec and diurnal changes as large as 0.2 μsec, but these vary with location and chain being used. Passing cold weather fronts may have temporary effects as well.

Disturbances on the sun's surface, most notably solar flares, disturb the earth's atmosphere as well. These Sudden Ionospheric Disturbances (SIDs) increase attenuation of radio waves and thus disturb Loran signals and reduce SNR. Such a disturbance may interfere with Loran reception for periods of hours or even longer.

Factor	Has effect on	
	Absolute Accuracy	**Repeatable Accuracy**
Crossing angles and gradients of the Loran LOPs	Yes	Yes
Stability of the transmitted signal (e.g., transmitter effect)	Yes	Yes
Loran chain control parameters (e.g., how closely actual ED is maintained to published ED, which system area monitor is being used, etc.)	Yes	Yes
Atmospheric and man-made ambient electronic noise	Yes	Yes
Factors with temporal variations in signal propagation speed (e.g., weather, seasonal effects, diurnal variations, etc.)	Yes	Yes
Sudden ionospheric disturbances	Yes	Yes
Receiver quality and sensitivity	Yes	Yes
Shipboard electric noise	Yes	Yes
Accuracy with which LOPs are printed on nautical charts	Yes	No
Accuracy of receiver's computer algorithms for coordinate conversion	Yes	No
Operator error	Yes	Yes

Table 2407. Selected Factors that Limit Loran Accuracy.

The factors above all relate to the propagated signal before it reaches the mariner. The remaining factors discussed below address the accuracy with which the mariner receives and interprets the signal.

Receivers vary in precision, quality and sophistication. Some receivers display TDs to the nearest 0.1 μsec; others to 0.01 μsec. Internal processing also varies, whether in the analog "front end" or the digital computer algorithms that use the processed analog signal. By referencing the user manual, the mariner may gain an appreciation for the advantages and limitations of the particular model available, and may adjust operator settings to maximize performance.

The best receiver available may be hindered by a poor installation. Similarly, electronic noise produced by electric motors, other electronic equipment or even fluorescent lighting may hinder the performance of a Loran receiver if the noise source is close to the receive antenna. The mariner should consult documentation supplied with the receiver for proper installation. Generally, proper installation and placement of the of the receive antenna will mitigate these problems. In some cases, contacting the manufacturer or obtaining professional installation assistance may be appropriate.

The raw TDs obtained by the receiver must be corrected with ASFs and then translated to position. Whether the receiver performs this entire process or the mariner assists by translating TDs to position manually using a Loran overprinted chart, published accuracies take into account the small errors involved in this conversion process.

Finally, as in all endeavors, operator error when using Loran is always possible. This can be minimized with alertness, knowledge and practice.

2408. The Effects of Crossing Angles and Gradients

The hyperbolic nature of Loran requires the operator to pay special attention to the geometry of the fix, specifically to crossing angles and gradients, and to the possibility of fix ambiguity. We begin with crossing angles.

As discussed above, the TDs from any given master-secondary pair form a family of hyperbolas. Each hyperbola in this family can be considered a line of position; the vessel must be somewhere along that locus of points which forms the hyperbola. A typical family of hyperbolas is shown in Figure 2408a.
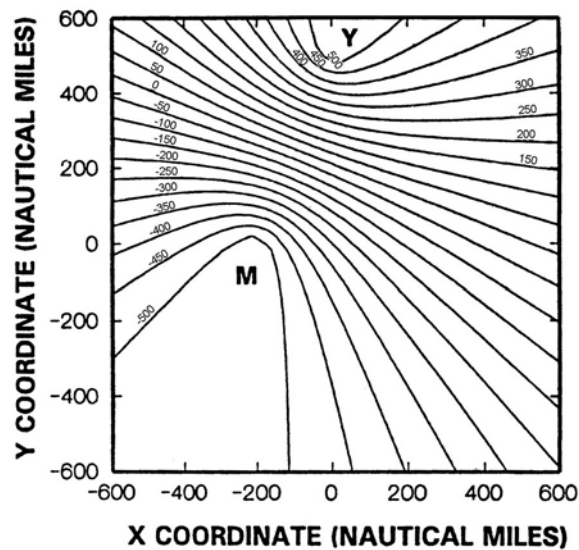

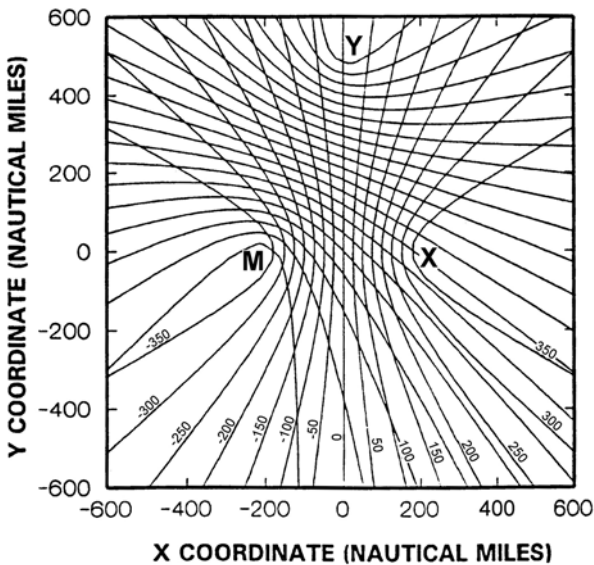*Figure 2408a. A family of hyperbolic lines generated by Loran signals.*

Figure 2408b. A hyperbolic lattice formed by station pairs M-X and M-Y.

Now, suppose the hyperbolic family from the Master-Xray station pair shown in Figure 2404 were superimposed upon the family shown in Figure 2408a. The results would be the hyperbolic lattice shown in Figure 2408b.

As has been noted, Loran LOPs for various chains and secondaries are printed on nautical charts. Each of the sets of LOPs is given a separate color and is denoted by a characteristic set of symbols. For example, an LOP might be designated 9960-X-25750. The designation is read as follows: the chain GRI designator is 9960, the TD is for the Master-Xray pair (M-X), and the time difference along this LOP is 25750 μsec. The chart shows only a limited number of LOPs to reduce clutter on the chart. Therefore, if the observed time delay falls between two charted LOPs, interpolation between them is required to obtain the precise LOP. After having interpolated (if necessary) between two TD measurements and plotted the resulting LOPs on the chart, the navigator marks the intersection of the LOPs and labels that intersection as the Loran fix. Note also in Figure 2408b the various angles at which the hyperbolas cross each other.

Figure 2408c shows graphically how error magnitude varies as a function of crossing angle. Assume that LOP 1 is known to contain no error, while LOP 2 has an uncertainty as shown. As the crossing angle (i.e., the angle of intersection of the two LOPs) approaches 90°, range of possible positions along LOP 1 (i.e., the position uncertainty or fix error) approaches a minimum; conversely, as the crossing angle decreases, the position uncertainty increases; the line defining the range of uncertainty grows longer. This illustration demonstrates the desirability of choosing LOPs for which the crossing angle is as close to 90° as possible.

The relationship between crossing angle and fix uncertainty can be expressed mathematically:

$$\sin(x) \ = \ \frac{\text{LOP error}}{\text{fix uncertainty}}$$

where x is the crossing angle.

Rearranging algebraically,

$$\text{fix uncertainty} \ = \ \frac{\text{LOP error}}{\sin(x)}$$

Assuming that LOP error is constant, then position uncertainty is inversely proportional to the sine of the crossing angle. As the crossing angle increases from 0° to 90°, the sine of the crossing angle increases from 0 to 1. Therefore, the error is at a minimum when the crossing angle is 90°, and increases thereafter as the crossing angle decreases.

Understanding and proper use of TD gradients are also important to the navigator. The gradient is defined as the rate of change of distance with respect to TD. Put another way, this quantity is the ratio of the spacing between adjacent Loran TDs (usually expressed in feet or meters) and the difference in microseconds between these adjacent LOPs. For example, if at a particular location two printed TD lines differ by 20 μsec and are 6 NM apart, the gradient is.

$$Gradient \ = \ \frac{6NM \quad 6076\,\text{ft/NM}}{20\mu\text{sec}} = \ 1822.8 \ \text{ft/}\mu\text{sec}$$

The smaller the gradient, the smaller the distance error that results from any TD error. Thus, the best accuracy from Loran is obtained by using TDs whose gradient is the smallest possible (i.e. the hyperbolic lines are closest together). This occurs along the baseline. Gradients are much larger (i.e. hyperbolic lines are farther apart) in the vicinity of the baseline extension. Therefore, the user should select TDs having the smallest possible gradients.

Another Loran effect that can lead to navigational error in the vicinity of the baseline extension is fix ambiguity. Fix ambiguity results when one Loran LOP crosses another LOP in two separate places. Near the baseline extension, the "ends" of a hyperbola can wrap around so that they cross another LOP twice, once along the baseline, and again along the baseline extension. A third LOP would resolve the ambiguity.

Most Loran receivers are equipped with an ambiguity alarm to alert the navigator to this occurrence. However, both fix ambiguity and large gradients necessitate that the navigator avoid using a master-secondary pair when operating in the vicinity of that pair's baseline extension.

2409. Coverage Areas

The 0.25 NM absolute accuracy specified for Loran is valid within each chain's coverage area. This area, whose limits define the maximum range of Loran for a
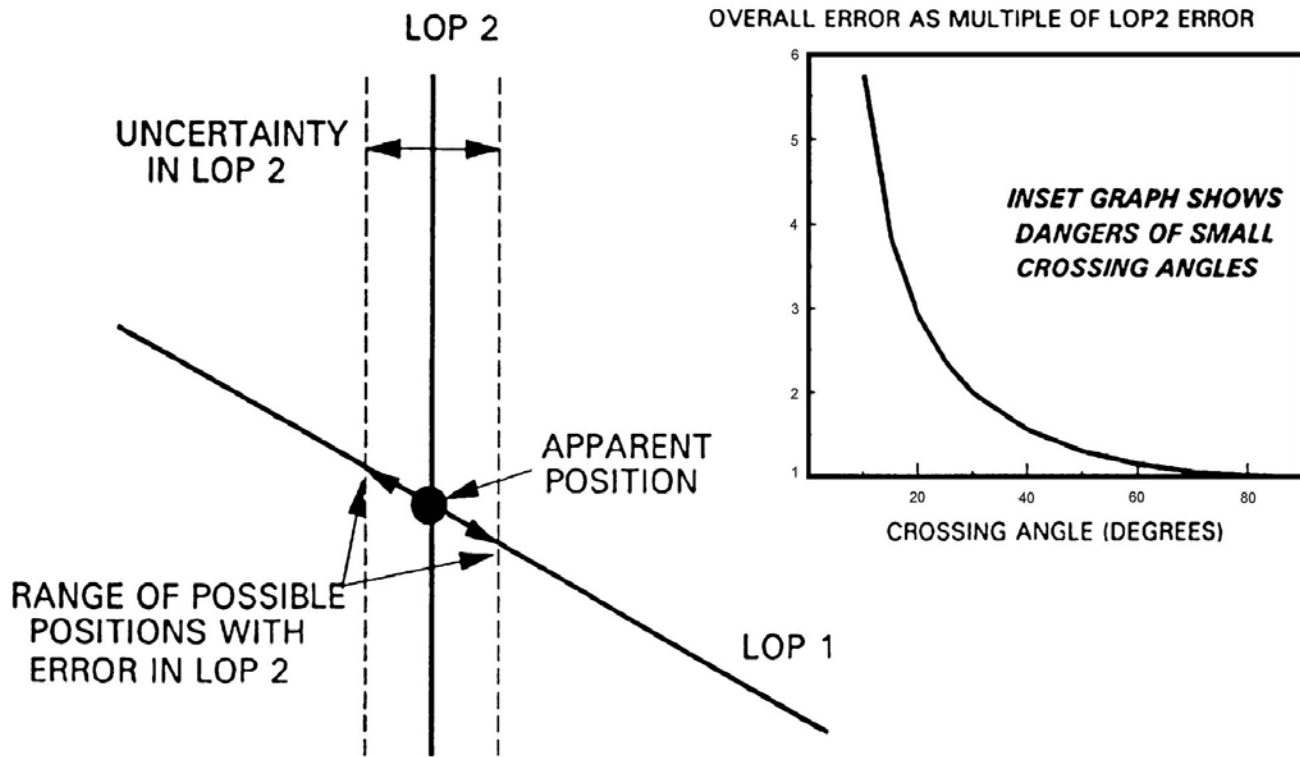
Figure 2408c. Error in Loran LOP's is magnified if the crossing angle is less than 90°.

particular chain, is the region in which both accuracy and SNR criteria are met. The National Oceanic and Atmospheric Administration (NOAA) has generally followed these coverage area limits when selecting where to print particular Loran TD lines on Loran overprinted charts.

One caveat to remember when considering coverage areas is that the 0.25 NM accuracy criteria is modified inside the coverage area in the vicinity of the coastline due to ASF effects. The following section describes this more fully.

2410. Understanding Additional Secondary Factors (ASF's)

Mathematically, calculating the reduction in propagation speed of an electromagnetic signal passing over a land surface of known conductivity is relatively straightforward. In practice, however, determining this Loran ASF correction accurately for use in the real world can be complex.

There are at least four reasons for this complexity. First, the conductivity of ground varies from region to region, so the correction to be applied is different for every signal path. Moreover, ground conductivity data may not take into account all the minor variations within each region. Second, methods used to compute ASFs vary. ASFs can be determined from either a mathematical model based on known approximate ground conductivities, or from empirical time delay measurements in various locations, or a combination of both. Methods incorporating empirical measurements tend to yield more accurate results. One receiver manufacturer may not use exactly the same correction method as another, and neither may use exactly the same method as those incorporated into time differences printed on a particular nautical chart. While such differences are minor, a user unaware of these differences may not obtain the best accuracy possible from Loran. Third, relatively large local variations in ASF variations may not be fully accounted for in the ASF models applied to the coverage area. Over the years, even empirically measured ASFs may change slightly in these areas with the addition of buildings, bridges and other structures to coastal areas. Fourth and finally, ASFs vary seasonally with changes in groundwater levels, snow pack depths and similar factors. However, ASFs are generally consistent year-on-year for a given area.

Designers of the Loran system, including Loran receiver manufacturers, have expended a great deal of effort to include ASFs in error calculations and to minimize these effects. Indeed, inaccuracies in ASF modeling are accounted for in published accuracy specifications for Loran. What then does the marine navigator need to know about ASFs beyond this? To obtain the 0.25 NM absolute accuracy advertised for Loran, the answer is clear. One must know *where* in the coverage area ASFs affect published accura-

cies, and one must know *when* ASFs are being incorporated, both in the receiver and on any chart in use.

With respect to *where* ASFs affect published accuracies, one must remember that local variations in the vicinity of the coastline are the most unpredictable of all ASF related effects because that is where rapid transitions from water to land occur. As a result, even though fixes determined by Loran may satisfy the 0.25 NM accuracy specification in these areas, such accuracy is not "guaranteed" for Loran within 10 NM of the coast. Users should also avoid relying solely on the lattice of Loran TDs in inshore areas.

With respect to *when* ASFs are being applied, one should realize that the default mode in most receivers combines ASFs with raw TD measurements. This is because the inclusion of ASFs is required to meet the 0.25 NM accuracy criteria. The navigator should verify which mode the receiver is in, and ensure the mode is not changed unknowingly.

The key point to remember there is that the "ASF included" and "ASF not included" modes must not be mixed. In other words, the receiver and any chart in use must handle ASFs in the same manner. If the receiver includes them, any chart in use must also include them. If operating on a chart that does not include ASFs-Loran coverage areas in another part of the world, for example-the receiver must be set to the same mode. If the navigator desires to correct ASFs manually, tables for U.S. Loran chains may be used although are not currently directly available from the U.S. Government. These documents also provide a fuller explanation of manual ASF corrections. When viewing ASF tables, remember that although the ASF correction for a single signal is always positive (indicating that the signal is always slowed and never speeded by its passage over land), the ASF correction for a time difference may be negative because two signal delays are included in the computation.

The U.S. Government does not guarantee the accuracy of ASF corrections incorporated into Loran receivers by their respective manufacturers. The prudent navigator will regularly check Loran TDs against charted LOPs when in a known position, and will compare Loran latitude and longitude readouts against other sources of position information. Ensuring the proper configuration and operation of the Loran receiver remains the navigator's responsibility.

Up to this point, our discussion has largely focused on correctly understanding and using Loran in order to obtain published accuracies. In some portions of the coverage areas, accuracy levels actually obtainable may be significantly better than these minimum published values. The following articles discuss practical techniques for maximizing the absolute, repeatable and relative accuracy of Loran.

2411. Maximizing Loran's Absolute Accuracy

Obtaining the best possible absolute accuracy from Loran rests primarily on the navigator's selection of TDs, particularly taking into account geometry, SNR and proximity to the baseline and baseline extension. As a vessel transits the coverage area, these factors gradually change and, except for SNR, are not visible on the display panel of the Loran receiver. Most receivers track an entire chain and some track multiple chains simultaneously, but the majority of installed marine receivers still use only two TDs to produce a latitude and longitude. Some receivers monitor these factors and may automatically select the best pair. The best way for the navigator, however, to monitor these factors is by referring to a Loran overprinted chart, even if not actually plotting fixes on it. The alert navigator will frequently reevaluate the selection of TDs during a transit and make adjustments as necessary.

Beyond this advice, two additional considerations may help the navigator maximize absolute accuracy. The first is the realization that Loran TD error is not evenly distributed over the coverage area. Besides the effects of transmitter station location on geometry and fix error, the locations of the primary and secondary monitor sites also have a discernible effect on TD error in the coverage area. As ASFs change daily and seasonally, the Loran control stations continually adjust the emission delay of each secondary station to keep it statistically at its nominal value as observed at the primary monitor site. What this means is that, on average, the Loran TD is more stable and more accurate in the absolute sense in the vicinity of the primary monitor site. The primary system area monitor for stations 9960-M, 9960-X and 9960-Y was placed at the entrance to New York Harbor at Sandy Hook, New Jersey for just this reason. A switch by the control station to the secondary monitor site will shift the error distribution slightly within the coverage area, reducing it near the secondary site and slightly increasing it elsewhere.

The second consideration in maximizing absolute accuracy is that most Loran receivers may be manually calibrated using a feature variously called "bias," "offset," "homeport" or a similar term. When in homeport or another known location, the known latitude and longitude (or in some cases, the difference between the current Loran display and the known values) is entered into the receiver. This forces the receiver's position error to be zero at that particular point and time.

The limitation of this technique is that this correction becomes less accurate with the passage of time and with increasing distance away from the point used. Most published sources indicate the technique to be of value out to a distance of 10 to 100 miles of the point where the calibration was performed. This correction does not take into account local distortions of the Loran grid due to bridges, power lines or other such man-made structures. The navigator should evaluate experimentally the effectiveness of this technique in good weather conditions before relying on it for navigation at other times. The bias should also be adjusted regularly to account for seasonal Loran variations; using the same value throughout the year is not the most ef-

fective application of this technique. Also, entering an offset into a Loran receiver alters the apparent location of waypoints stored prior to establishing this correction.

Finally, receivers vary in how this feature is implemented. Some receivers save the offset when the receiver is turned off; others zero the correction when the receiver is turned on. Some receivers replace the internal ASF value with the offset, while others add it to the internal ASF values. Refer to the owner's manual for the receiver in use.

2412. Maximizing Loran's Repeatable Accuracy

Many users consider the high repeatable accuracy of Loran its most important characteristic. To obtain the best repeatable accuracy consistently, the navigator should use measured TDs rather than latitude and longitude values supplied by the receiver.

The reason for this lies in the ASF conversion process. Recall that Loran receivers use ASFs to correct TDs. Recall also that the ASFs are a function of the terrain over which the signal must pass to reach the receiver. Therefore, the ASFs for one station pair are different from the ASFs for another station pair because the signals from the different pairs must travel over different terrain to reach the receiver.

This consideration matters because a Loran receiver may not always use the same pairs of TDs to calculate a fix. Suppose a navigator marks the position of a channel buoy by recording its latitude and longitude using the TD pair selected automatically by the Loran receiver. If, on the return trip, the receiver is using a different TD pair, the latitude and longitude readings for the exact same buoy would be slightly different because the new TD pair would be using a different ASF value. By using previously-measured TDs and not previously-measured latitudes and longitudes, this ASF-introduced error is avoided. The navigator should also record the values of all secondary TDs at the waypoint and

not just the ones used by the receiver at the time. When returning to the waypoint, other TDs will be available even if the previously used TD pair is not. Recording the time and date the waypoint is stored will also help evaluate the cyclical seasonal and diurnal variations that may have since occurred.

2413. Maximizing Loran's Relative Accuracy

The classical application of relative accuracy involves two users finding the same point on the earth's surface at the same time using the same navigation system. The maximum relative Loran accuracy would be theoretically be achieved by identical receivers, configured and installed identically on identical vessels, tracking the same TDs. In practice, the two most important factors are tracking the same TDs and ensuring that ASFs are being treated consistently between the two receivers. By attending to these, the navigator should obtain relative accuracy close to the theoretical maximum.

Another application of relative accuracy is the current practice of converting old Loran TDs into latitude and longitude for use with GPS and DGPS receivers. Several commercial firms sell software applications that perform this tedious task. One key question posed by these programs is whether or not the Loran TDs include ASFs. The difficulty in answering this question depends on how the Loran TDs were obtained, and of course an understanding of ASFs. If in doubt, the navigator can perform the conversion once by specifying "with" ASFs and once "without," and then carefully choosing which is the valid one, assisted by direct observation underway if needed.

To round out the discussion of Loran, the following section briefly describes present and possible future uses for this system beyond the well-known hyperbolic navigation mode.

NON-HYPERBOLIC USES OF LORAN-C

2414. Precise Timing with Loran-C

Because Loran is fundamentally a **precise timing system**, a significant segment of the user community uses Loran for the propagation of Coordinated Universal Time (UTC). The accessibility of UTC at any desired location enables such applications as the synchronization of telephone and data networks. Because the timing of each secondary station is relative to the master, its timing accuracy derives from that of the master.

The start of each Loran station's GRI periodically coincides with the start of the UTC second. This is termed the Time of Coincidence (TOC). Because one Loran station is sufficient to provide an absolute timing reference, timing receivers do not typically rely on the hyperbolic mode or use TDs per se.

A noteworthy feature of Loran is that each transmitter station has an independent timing reference consisting of one or more Primary Reference Standards. Timing equipment at the transmitter stations constantly compares these signals and adjusts to minimize oscillator drift. The end result is a nationwide system with a large ensemble of independent timing sources. This strengthens the U.S. technology infrastructure. As another cross-check of Loran time, daily comparisons are made with UTC, as disseminated via GPS.

2415. Loran-C Time of Arrival (TOA) Mode

With the advent of the powerful digital processors and compact precise oscillators now embedded in user receivers, technical limitations that dictated Loran's hyperbolic architecture decades ago have been overcome. A receiver

can now predict in real time the exact point in time a Loran station will transmit its signal, as well as the exact time the signal will be received at any assumed position.

An alternate receiver architecture that takes advantage of these capabilities uses Loran **Time of Arrival (TOA)** measurement, which are measured relative to UTC rather than to an arbitrary master station's transmission. A receiver operating in TOA mode can locate and track all Loran signals in view, prompting the descriptor "all in view" for this type of receiver. This architecture steps beyond the limitations of using only one Loran chain at a time. As a result, system availability can be improved across all the overlapping coverage areas. Coupled with advanced **Receiver Autonomous Integrity Monitor (RAIM)** algorithms, this architecture can also add an additional layer of integrity at the user level, independent of Loran blink.

2416. Loran-C in an Integrated Navigation System

An exponential worldwide increase in reliance on electronic navigation systems, most notably GPS, for positioning and timing has fueled a drive for more robust systems immune from accidental or intentional interference. Even a short outage of GPS, for example, would likely have severe safety and economic consequences for users.

In this environment, integrated navigation systems are attractive options as robust sources of position and time. The ideal **integrated navigation system** (INS) can tolerate the degradation or failure of any component system without degradation as a whole.

Loran offers several advantages to an integrated system based on GPS. Although Loran relies on radio propagation and is thus similarly vulnerable to large-scale atmospheric events such as ionospheric disturbances, at 100 kHz it occupies a very different portion of the spectrum than the 1.2 GHz to 1.6 GHz band used by GPS. Loran is a high-power system whose low frequency often uses a very large antenna for efficient propagation. Therefore, jamming Loran over a broad area is much more difficult than jamming GPS over the same area. Loran signals are present in urban and natural canyons and under foliage, where GPS signals may be partially or completely blocked. Loran's independent timing source also provides an additional degree of robustness to an integrated system. In short, the circumstances that cause failure or degradation of Loran are very different from those that cause failure or degradation of GPS. When the absolute accuracy of Loran is continually calibrated by GPS, the repeatable accuracy of Loran could ensure near-GPS performance of an integrated system in several possible navigation and timing scenarios, for periods of several hours to a few days after a total loss of GPS, depending upon the capability of the INS.

2417. Loran-C as a Data Transfer Channel

Low data rate transmission using Loran signals began in the 1960s with a system known as Clarinet Pilgrim (CP). CP was followed in the 1970s with a similar system termed "Two-Pulse Communications (TPC)". The two primary uses of this capability were Loran chain control and backup military communications. In all cases, the data superimposed on the Loran signal were transparent to the users, who were nearly universally unaware of this dual use.

In the late 1990s, the Northwest European Loran System (NELS) implemented a pulse-position modulation scheme termed Eurofix to provide differential GPS corrections via the Loran signal to certain areas in western and northern Europe. Eurofix successfully incorporated sophisticated data communications techniques to broadcast GPS corrections in real time while allowing traditional Loran users to operate without interruption.

ENHANCED LORAN (E-LORAN)

2418. eLoran Improvements over Loran-C

As of 2016, eLoran is not available for navigational use anywhere in the world.

While eLoran is currently only broadcast in North America from the former USCG Loran Support Unit transmitting site in Wildwood, New Jersey, the system specifications have been developed and tested. The information presented here comes from various sources involved in the development of eLoran and gives an overview to the enhanced capabilities that eLoran will provide.

eLoran was designed such that new capabilities were added to increase system performance while retaining all of the previous hyperbolic navigation characteristics of Loran-C (Helwig, Offermans, Stout, & Schue, 2011). Any Loran-C receiver can be used with an eLoran transmitting station although Loran-C receivers cannot take advantage of the new capabilities built into eLoran.

eLoran was designed to have improved accuracy, availability, continuity, and integrity over Loran-C (FAA, 2004). eLoran will be a stratum-1 source of UTC time within 50ns such that clocks can be calibrated using eLoran (Helwig, 2011). When fully deployed, it would be the most accurate broadcast source of UTC time independent of a Global Navigation Satellite System (GNSS) such as GPS.

eLoran will be more accurate than Loran-C with a designed position accuracy within 8-20 meters provided the receiver is set up properly and any additional secondary factor corrections are applied (International Loran Association 2007) (Helwig, 2011). eLoran will be able to achieve an increased accuracy over Loran-C because the transmit-

ted signal has tighter tolerances between the GRIs, pulses, and zero-crossings which result in less error in the transmitted signal (International Loran Association 2007) (Helwig, 2011). eLoran also contains a data channel which transmits messages indicating error corrections and precise timing information (International Loran Association 2007) (Helwig, 2011). eLoran's increased position and timing accuracy over Loran-C will allow it to meet modern Maritime Harbor Entrance and Approach (HEA) and Aviation En Route and Non-Precision Approach (RNA) requirements (International Loran Association 2007) (Helwig, 2011).

eLoran also includes one or more Loran Data Channels (LDC), which use various means of modulation to transmit messages (Schue et al., 2000). Current LDC modulation schemes include either the 3-state Eurofix approach or the 9th pulse modulation approach, or both (Helwig, 2011). These messages are very short in nature because the LDC has a low data throughput (slow rate of message transmission). The LDC continuously transmits a series of messages when the system broadcasts (Schue et al., 2000) (Helwig, 2011).

Each pulse position modulation technique accomplishes transmission of a full message within 3s, though the internal structure of each message is slightly different. Alternative modulation techniques provide higher data rates (Schue et al., 2000). The Eurofix approach independently modulates each one of the last six pulses of the GRI by ±1μs. Many possible configurations (combinations of -1μs, 0, or +1μs shifted pulse) of the last six pulses can be created using this modulation technique; 128 are used for encoding messages. Each sixth-pulse modulation represents seven bits of information. Every message is 210 bits long, containing 30 seven-bit parts. One complete message takes 30 GRIs to receive and a new message begins broadcasting every 30 GRIs. A full message would take a maximum of 3s (assuming a GRI of 9999) to receive using the Eurofix method (Offermans, Helwig, Van Willigen, 1996) (Offermans, Helwig, Van Willigen, 1997).

The 9th pulse modulation technique adds an extra pulse approximately 1000μs after the 8th pulse of the Master station (which is also 1000μs before the final pulse in the master station) and an extra (9th) pulse in the secondary station approximately 1000μs after the 8th secondary pulse (making the modulated 9th pulse the final pulse in the secondary GRI). 32 possible states (states 0 through state 31) are defined by moving the position of this pulse in each GRI. The zero-state is defined when this pulse occurs exactly 1000μs after the 8th pulse. The 31 remaining symbols are positioned in the GRI using the formula: D_x μ=1.25mod[x,8]+50.625floor(x/8) where "x" is the possible state (0,31) and Dx is the pulse's time-offset from the zero-state position in the GRI. A receiver would obtain the offset distance of the 9th pulse and use the inverse of the above formula to determine message state. Each GRI can carry 5-bits of information and each 9th pulse modulated message is 120-bits long; so an entire message is transmitted over 24 GRIs. A full message would take a maximum of 2.4s (assuming a GRI of 9999) to receive using the 9th pulse modulation method (Peterson, Dykstra, Lown & Shmihluk 2006).

A standard eLoran receiver should have the capability of reading messages from the LDC encoded with any type of standardized LDC technique. The message types will be standardized and repeat at regular intervals. When operational. eLoran will be capable of transmitting the following message types and additional message types may be defined in the future (The Radio Technical Commission for Maritime Services 2008) (International Loran Association 2007) (Dykstra & Peterson, 2006) (Helwig, 2011):

- ASF corrections.
- Almanac information containing station specific information such as: station position, station name and station status (replacing Loran blink codes).
- UTC Time of Day expressed as number of seconds since the Loran epoch of 0h0m0s-01 JAN 1958. The number of seconds from the Loran epoch to the time of transmission of the message can be calculated as: $T = 24(GRI)(MEC)+ED$ where MEC is the Message-Epoch-Count which is the number of 24-GRI intervals since the Loran epoch.
- Various Government-Use only messages.

The source of timing for the transmission of eLoran pulses is independent of monitor sites and control centers; the eLoran signals are synchronized to an identifiable, independent UTC source at each site (Helwig, 2011). All time of transmissions for both the master and secondary stations are determined using the independent clocks at each station synchronized to UTC so that a user can obtain/calculate timing information from the strongest signal available instead of just needing the master station fix (The Radio Technical Commission for Maritime Services 2008). The synchronization of all stations with an independent UTC time source allows for greater position accuracy.

eLoran pulses are synchronized independent of any GNSS system using a clock at each transmitter site (Offermans, et al. 2013). One can obtain current UTC time by reading the time of day message from the LDC. Another method of calculating the UTC time involves knowing the receiver's position and ASF corrections (Offermans, Johannessen, Schue, Hirchauer & Powers, 2013). An eLoran receiver measures the time of arrival (the time when a pulse is received). Knowing the receiver's position along with some ASF corrections, one can obtain a synchronized UTC time of transmission (Offermans, et al. 2013) (Helwig, 2011). Since each eLoran transmission is locked to UTC time and each transmitter is an independent source of UTC time, then UTC time may be obtained accurately from any eLoran transmitting station fix.

2419. eLoran Definition Document

The Enhanced Loran (eLoran) Definition Document was developed in 2006 at the United States Coast Guard Navigation Center by an international team of authors and was published by the ILA in 2007.

The document provides an overview, background and introduction to eLoran along with a detailed description of the eLoran system (eLoran signal, transmitting stations, control centers, monitoring & reference stations and user equipment. The document includes a description of the maritime application for eLoran along with a broader overall service provision for the system. See the following ling for access to the document:

Figure 2419. Enhance Loran Definition Document (2007). https://rntfnd.org/wp-content/uploads/eLoran-Definition-Document-0-1-Released.pdf

2420. References

Dykstra, K., Peterson, B., (2006). Loran data channel: Progress to date and future plans. *35th Annual Convention and Technical Symposium of the International Loran Association*. Groton, CT.

Helwig, A., Offermans, G., Stout C., & Schue, C., (2011). eLoran system definition and signal specification tutorial. *40th International Loran Association Conference*. Busan.

International Loran Association (ILA), 2007. *Enhanced Loran (eLoran) Definition*.

Offermans, G., Helwig, A., & Van Willigen, D., (1996). The Eurofix datalink concept: Reliable data transmission using Loran-C. *Proceedings of the 25th Annual Technical Meeting of the International Loran Association*, San Diego, CA.

Offermans, G., Helwig, A., & Van Willigen, D., (1997). Eurofix DGPS Service through the Sylt Loran-C transmitter: Static and dynamic test results. *Proceedings of the 26th Annual Technical Meeting of the International Loran Association*, Ottawa, Canada.

Offermans, G., Johannessen, E., Schue, C., Hirschauer, J., & Powers., E., (2013). UTC synchronization and stratum-1 frequency recovery using eLoran: The alternate basket for your eggs. *45th Annual Precise Time and Time Interval Systems and Applications Meeting. The Institute of Navigation*. Bellevue, WA.

Peterson, B., Dykstra, K., Lown, D., & Shmilhluk, K., (2006, vers 1.3). Loran Data Channel Communications Using 9th Pulse Modulation.

Schue, C., Peterson, B., et al, (2000). Enhanced Loran-C Data Channel Project, *German Institute of Navigation (DGON) Symposium and Technical Conference*, Bonn, Germany.

Schue, C., Peterson, B, Celano, T., (2004). Low Cost Digitally Enhanced Loran for Tactical Applications (LC DELTA), *Thirty-third Annual International Loran Association Technical Symposium*.

Schue, C. A. III, (2008). The next generation LF transmitter and its impact on Loran, eLoran, and tactical (e)Loran systems, *Royal Institute of Navigation NAV08 conference and Thirty-seventh Annual International Loran Association Technical Symposium*.

The Radio Technical Commission for Maritime Services, (2008). *Specification of the eLoran System Preliminary, Tentative and Unofficial*. Unofficial Specification.

United States Coast Guard, (2012). *Loran-C General Information*. http://www.navcen.uscg.gov/?pageName=loranMain (accessed 09 15, 2016).

United States Department of Transportation, (2001). *Vulnerability assessment of the transportation infrastructure relying on the global positioning system*. Government Report.

UrsaNav. (2015). *eLoran Points of Light*. Q&A Paper.

UrsaNav. *Loran-C*. http://www.ursanav.com/solutions/technology/loran-c/ (accessed September15, 2016).