RADAR HANDBOOK

Editor in Chief

MERRILL I. SKOLNIK

Second Edition

McGraw-Hill

A Division of The McGraw-Hill Companies

      12 13 14 15 16 17 18 19   BKM BKM   0 9 8 7 6 5 4 3 2

CONTRIBUTORS

Lamont V. Blake, *Electronics Consultant* (CHAPTER 2)

Michael T. Borkowski, *Raytheon Company* (CHAPTER 5)

Leopold J. Cantafio, *Space and Technology Group TRW* (CHAPTER 22)

Theodore C. Cheston, *Naval Research Laboratory* (CHAPTER 7)

L. J. Cutrona, *Sarcutron, Inc.* (CHAPTER 21)

Daniel Davis, *Electronic Systems Group, Westinghouse Electric Corporation* (CHAPTER 6)

Gary E. Evans, *Electronic Systems Group, Westinghouse Electric Corporation* (CHAPTER 6)

A. Farina, *Radar Department, Selenia S.p.A., Italy* (CHAPTER 9)

Edward C. Farnett, *RCA Electronics Systems Department, GE Aerospace* (CHAPTER 10)

Joe Frank, *Technology Service Corporation* (CHAPTER 7)

V. Gregers-Hansen, *Equipment Division, Raytheon Company* (CHAPTER 15)

J. M. Headrick, *Naval Research Laboratory* (CHAPTER 24)

Dean D. Howard, *Locus, Inc., a subsidiary of Kaman Corp.* (CHAPTER 18)

Alex Ivanov, *Missile Systems Division, Raytheon Company* (CHAPTER 19)

Eugene F. Knott, *The Boeing Company* (CHAPTER 11)

William H. Long, *Westinghouse Electric Corporation* (CHAPTER 17)

David H. Mooney, *Westinghouse Electric Corporation* (CHAPTER 17)

Richard K. Moore, *The University of Kansas* (CHAPTER 12)

David J. Murrow, *General Electric Company* (CHAPTER 20)

William K. Saunders, *formerly of Harry Diamond Laboratories* (CHAPTER 14)

Helmut E. Schrank, *Electronics Systems Group, Westinghouse Electric Corporation* (CHAPTER 6)

Robert J. Serafin, *National Center for Atmospheric Research* (CHAPTER 23)

William W. Shrader, *Equipment Division, Raytheon Company* (CHAPTER 15)

William A. Skillman, *Westinghouse Electric Corporation* (CHAPTER 17)

Merrill I. Skolnik, *Naval Research Laboratory* (CHAPTER 1)

Fred M. Staudaher, *Naval Research Laboratory* (CHAPTER 16)

George H. Stevens, *RCA Electronics Systems Department, GE Aerospace* (CHAPTER 10)

John W. Taylor, Jr., *Westinghouse Electric Corporation* (CHAPTER 3)

G. V. Trunk, *Naval Research Laboratory* (CHAPTER 8)

T. A. Weil, *Equipment Division, Raytheon Company* (CHAPTER 4)

Lewis B. Wetzel, *Naval Research Laboratory* (CHAPTER 13)

Nicholas J. Willis, *Technology Service Corporation* (CHAPTER 25)

PREFACE

This edition has been thoroughly revised to reflect the advances made in radar over the past two decades. There are many new topics not found in the original, and over half of the 25 chapters were written by authors who did not participate in the first edition. The continued growth in radar capability and applications is reflected in much of the new material included in this second edition. The following are some of the many new radar advances that have occurred since the original edition (listed in no particular order):

- The use of digital techniques that allow sophisticated signal processing in MTI and pulse doppler radars, as well as digital data processing to perform automatic detection and tracking.

- The use of the doppler filter bank and the clutter map in MTI radar.

- The reduced dependency on operators for extracting information from a radar and the incorporation of CFAR in automatic detection and tracking systems.

- The emergence of the analog SAW dispersive delay line as the preferred technique for wideband (high-resolution) pulse compression; the use of digital processing for the pulse compression filter when the bandwidth permits; and the introduction of Stretch pulse compression, which allows high resolution, over a limited range interval with considerably reduced processing bandwidth.

- The increased use of 3D radar for military applications.

- The introduction of the ultralow-sidelobe antenna for airborne pulse doppler radar and, later, for ECCM.

- The replacement of the parabolic reflector antenna with the planar-aperture array antenna for 3D radar, ultralow-sidelobe antennas, and airborne radar.

- The high-power solid-state transmitter that consists of many transistor modules distributed on the rows of a 3D radar (such as the AN/TPS-59), or employed at the elements of a phased array (as in PAVE PAWS), or configured as a transmitter for a conventional radar (as in the AN/SPS-40 or the Canadian ATC radar known as RAMP).

- The serial production of phased arrays for the Patriot, Aegis, PAVE PAWS, and B1-B radar systems.

- The interest in the radar cross section of targets brought about by the attempts to reduce the cross section of military vehicles; and advances in computer methods for predicting the cross section of complex targets.

- The increased capability of military airborne radar (airborne intercept, AWACS, and AEW) due to advances in components and technology that permitted the application of AMTI and pulse doppler to the detection of aircraft in the midst of large clutter.

- The use of radar in space for rendezvous and landing, remote sensing of the earth's environment, planetary exploration, and the detection of targets on the oceans of the world.

- The use of semiactive radar for the guidance of military missile systems.

- The extraction of the doppler frequency shift in meteorological radars that permits the recognition of hazardous weather phenomena not possible with previous weather radars.

- The use of radar operating in the HF portion of the spectrum for long-range over-the-horizon detection of aircraft, ships, and missiles, as well as to provide the direction of the surface winds and the sea state over wide areas of the ocean.

- The development of electronic counter-countermeasures (ECCM) in military radars to thwart attempts to negate radar capability by hostile electronic radiations.

- The increased range resolution and doppler resolution in synthetic aperture radars (SAR) for the imaging of a scene, the use of inverse SAR (ISAR) for the imaging of targets, and the replacement of optical processing with digital processing for SAR imaging.

- The adaptive antenna for application in sidelobe cancelers (as an ECCM) and AMTI radar.

- The use of computers to reliably and quickly predict the capability and coverage of radar systems in the real environment.

The purpose of the above listing is to indicate that radar is dynamic. Not all the new advances made since the first edition of this handbook are listed, nor does the list include all the new material discussed in this edition. There continue to be significant advances in the application of new technology and in the appearance of new applications. Radar grows and is viable since it satisfies important societal, economic, and military needs. It has no serious competitor for most of its many applications.

The size of this edition of the handbook is smaller than the original edition. This is more an indication of the problems involved in technical book publishing rather than problems with the health of radar or the

availability of material to include. It was not an easy task to constrain the chapter authors to a limited page budget, and I am appreciative of their efforts to keep the size of their chapters within the allocated number of pages. It would have been easy to double the size of this edition even without increasing the number of chapters. The limitation on size was one of the reasons a number of chapters found in the first edition do not appear here. Some of the omitted chapters were concerned with subjects for which there is not as much interest as there had been or whose technology has not advanced as much as other areas of radar. It is with some regret that 16 of the chapters in the original could not be included in this second edition.

As in the first edition, no attempt was made to utilize a standard notation throughout the book. Each particular subspecialty of radar seems to have developed its own nomenclature, and it is not appropriate in a book such as this to force authors to use notation that is foreign to their field even though it might be commonplace in some other aspect of radar.

Each chapter author was instructed to assume that the average reader has a general knowledge of radar but is not necessarily an expert in the particular subject covered by the chapter.

After a general introduction to radar in Chap. 1, there is a review of the methodology for predicting the range of a radar that has evolved over the years. This is followed by several chapters on the major subsystems of a radar: the receiver, transmitter, solid-state transmitters, reflector antennas, phased array antennas, data processing, ECCM, and pulse compression. Next are discussions of the target cross section and the nature of the radar echoes from the ground and the sea. The various types of radar systems are then discussed: CW and FM-CW, MTI, AMTI, pulse doppler, tracking, missile guidance, height finding and 3D, and synthetic aperture radar. This is followed by three specialized examples of radar that have their own unique character: radar in space, meteorological (weather) radar, and HF over-the horizon radar. The book closes with a treatment of the bistatic radar, which was the first type of radar explored during the onset of radar development in the 1930s.

I would hope that readers who refer in their own writings to material from this book do so by chapter author and title and not by citing just the "Radar Handbook." This will give proper credit to the individual authors who created the work.

It is with much pleasure that I acknowledge the contributions of the individual chapter authors. I enjoyed working with these talented radar engineers and having the opportunity to learn so much from my association with them. A handbook such as this exists only because of the dedicated efforts of the many experts who took the time and energy to

prepare the individual chapters. I appreciate their hard work in committing to writing their knowledge and experience, and I am grateful that they have shared this with us.

Merrill Skolnik

ABOUT THE EDITOR IN CHIEF

Merrill I. Skolnik, known worldwide for his leadership in radar research and development, has been affiliated with the Johns Hopkins Radiation Laboratory, Sylvania, MIT Lincoln Laboratory, the Research Division of Electronic Communications Inc., the Institute for Defense Analyses, and the U.S. Naval Research Laboratory. He received his doctorate in electrical engineering from Johns Hopkins University, where he also earned B.E. and M.S.E. degrees. He is the author of the leading college textbook on radar, *Introduction to Radar Systems* (McGraw-Hill), now in its second edition, and the editor of *Radar Applications*.

He is a member of the National Academy of Engineering, a Fellow of the IEEE, and has served as editor of the *Proceedings of the IEEE*.

Contents

CHAPTER 1
AN INTRODUCTION TO RADAR

Merrill I. Skolnik

1.1 DESCRIPTION OF RADAR

The basic concept of radar is relatively simple even though in many instances its practical implementation is not. A radar operates by radiating electromagnetic energy and detecting the echo returned from reflecting objects (targets). The nature of the echo signal provides information about the target. The range, or distance, to the target is found from the time it takes for the radiated energy to travel to the target and back. The angular location of the target is found with a directive antenna (one with a narrow beamwidth) to sense the angle of arrival of the echo signal. If the target is moving, a radar can derive its track, or trajectory, and predict the future location. The shift in frequency of the received echo signal due to the doppler effect caused by a moving target allows a radar to separate desired moving targets (such as aircraft) from undesired stationary targets (such as land and sea clutter) even though the stationary echo signal may be many orders of magnitude greater than the moving target. With sufficiently high resolution, a radar can discern something about the nature of a target's size and shape. Radar resolution may be obtained in range or angle, or both. Range resolution requires large bandwidth. Angle resolution requires (electrically) large antennas. Resolution in the cross-range dimension is usually not as good as the resolution that can be obtained in range. However, when there is relative motion between the individual parts of a target and the radar, it is possible to use the inherent resolution in doppler frequency to resolve in the cross-range dimension. The cross-range resolution of a synthetic aperture radar (SAR) for imaging a scene such as terrain can be explained as being due to resolution in doppler, although a SAR is usually thought of as generating a large "synthetic" antenna by storing received signals in a memory. The two views—doppler resolution and synthetic antenna—are equivalent. Resolution in the doppler domain is a natural way to envision the cross-range resolution achieved by the inverse synthetic aperture radar (ISAR) used for the imaging of a target.

Radar is an active device in that it carries its own transmitter and does not depend on ambient radiation, as do most optical and infrared sensors. Radar can detect relatively small targets at near or far distances and can measure their range with precision in all weather, which is its chief advantage when compared with other sensors.

The principle of radar has been applied from frequencies of a few megahertz

(HF, or high-frequency region of the electromagnetic spectrum) to well beyond the optical region (laser radar). This is a frequency extent of about 1 billion to 1. The particular techniques for implementing a radar differ greatly over this range of frequencies, but the basic principles remain the same.

Radar was originally developed to satisfy the needs of the military for surveillance and weapon control. Military applications have funded much of the development of its technology. However, radar has seen significant civil applications for the safe travel of aircraft, ships, and spacecraft; the remote sensing of the environment, especially the weather; and law enforcement and many other applications.

Radar Block Diagram. The basic parts of a radar system are illustrated in the simple block diagram of Fig. 1.1. (Other examples of radar block diagrams can be found throughout the handbook.) The radar signal, usually a repetitive train of short pulses, is generated by the transmitter and radiated into space by the antenna. The duplexer permits a single antenna to be time-shared for both transmission and reception. Reflecting objects (targets) intercept and reradiate a portion of the radar signal, a small amount of which is returned in the direction of the radar. The returned echo signal is collected by the radar antenna and amplified by the receiver. If the output of the radar receiver is sufficiently large, detection of a target is said to occur. A radar generally determines the location of a target in range and angle, but the echo signal also can provide information about the nature of the target. The output of the receiver may be presented on a display to an operator who makes the decision as to whether or not a target is present, or the receiver output can be processed by electronic means to automatically recognize the presence of a target and to establish a track of the target from detections made over a period of time. With automatic detection and track (ADT) the operator usually is presented with the processed target track rather than the raw radar detections. In some applications, the processed radar output might be used to directly control a system (such as a guided missile) without any operator intervention.
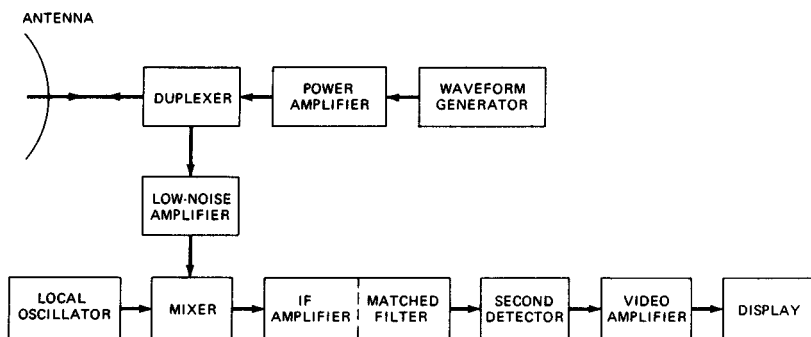
FIG. 1.1 Simple block diagram of a radar employing a power amplifier transmitter and a superheterodyne receiver.

The operation of the radar is described in more detail, starting with the transmitter.

Transmitter. The transmitter (Chap. 4) in Fig. 1.1 is shown as a power amplifier, such as a klystron, traveling-wave tube, crossed-field amplifier, or solid-state device (Chap. 5). A power oscillator such as a magnetron also can be used as the transmitter; but the magnetron usually is of limited average power compared with power amplifiers, especially the klystron, which can produce much larger average power than can a magnetron and is more stable. (It is the *average* power, rather than the *peak* power, which is the measure of the capability of a radar.) Since the basic waveform is generated at low power before being delivered to the power amplifier, it is far easier to achieve the special waveforms needed for pulse compression and for coherent systems such as moving-target indication (MTI) radar and pulse doppler radar. Although the magnetron oscillator can be used for pulse compression and for MTI, better performance can be obtained with a power amplifier configuration. The magnetron oscillator might be found in systems where simplicity and mobility are important and where high average power, good MTI performance, or pulse compression is not required.

The transmitter of a typical ground-based air surveillance radar might have an average power of several kilowatts. Short-range radars might have powers measured in milliwatts. Radars for the detection of space objects (Chap. 22) and HF over-the-horizon radars (Chap. 24) might have average powers of the order of a megawatt.

The radar equation (Sec. 1.2 and Chap. 2) shows that the range of a radar is proportional to the fourth root of the transmitter power. Thus, to double the range requires that the power be increased by 16. This means that there often is a practical, economical limit to the amount of power that should be employed to increase the range of a radar.

Transmitters not only must be able to generate high power with stable waveforms, but they must often operate over a wide bandwidth, with high efficiency and with long, trouble-free life.

Duplexer. The duplexer acts as a rapid switch to protect the receiver from damage when the high-power transmitter is on. On reception, with the transmitter off, the duplexer directs the weak received signal to the receiver rather than to the transmitter. Duplexers generally are some form of gas-discharge device and may be used with solid-state or gas-discharge receiver protectors. A solid-state circulator is sometimes used to provide further isolation between the transmitter and the receiver.

Antenna. The transmitter power is radiated into space by a directive antenna which concentrates the energy into a narrow beam. Mechanically steered parabolic reflector antennas (Chap. 6) and planar phased arrays (Chap. 7) both find wide application in radar. Electronically steered phased array antennas (Chap. 7) are also used. The narrow, directive beam that is characteristic of most radar antennas not only concentrates the energy on target but also permits a measurement of the direction to the target. A typical antenna beamwidth for the detection or tracking of aircraft might be about 1 or 2°. A dedicated tracking radar (Chap. 18) generally has a symmetrical antenna which radiates a pencil-beam pattern. The usual ground-based air surveillance radar that provides the range and azimuth of a target generally uses a mechanically rotated reflector antenna with a fan-shaped beam, narrow in azimuth and broad in elevation. Airborne radars and surface-based 3D air surveillance radars (those that rotate mechanically in azimuth to measure the azimuth angle but use some form of electronic steering or beamforming to obtain the elevation angle, as discussed in Chap. 20) often employ planar array apertures. Mechanical scanning of the radar antenna is usually quite acceptable for the vast majority of radar applications. When it is

necessary to scan the beam more quickly than can be achieved with mechanical scanning and when high cost can be tolerated, the electronically steered phased array antenna can be employed. (Beam steering with electronically steered phased arrays can be accomplished in microseconds or less if necessary.)

The size of a radar antenna depends in part on the frequency, whether the radar is located on the ground or on a moving vehicle, and the environment in which it must operate. The lower the frequency, the easier it is to produce a physically large antenna since the mechanical (and electrical) tolerances are proportional to the wavelength. In the ultrahigh-frequency (UHF) band, a *large* antenna (either reflector or phased array) might have a dimension of 100 ft or more. At the upper microwave frequencies (such as X band), radar antennas greater than 10 or 20 ft in dimension can be considered large. (Larger antennas than the above examples have been built, but they are not the norm.) Although there have been microwave antennas with beamwidths as small as 0.05°, radar antennas rarely have beamwidths less than about 0.2°. This corresponds to an aperture of approximately 300 wavelengths (about 31 ft at X band and about 700 ft at UHF).

Receiver. The signal collected by the antenna is sent to the receiver, which is almost always of the superheterodyne type (Chap. 3). The receiver serves to (1) separate the desired signal from the ever-present noise and other interfering signals and (2) amplify the signal sufficiently to actuate a display, such as a cathode-ray tube, or to allow automatic processing by some form of digital device. At microwave frequencies, the noise at the receiver output is usually that generated by the receiver itself rather than external noise which enters via the antenna. The input stage of the receiver must not introduce excessive noise which would interfere with the signal to be detected. A transistor amplifier as the first stage offers acceptably low noise for many radar applications. A first-stage receiver noise figure (defined in Sec. 1.2) might be, typically, 1 or 2 dB. A low-noise receiver front end (the first stage) is desirable for many civil applications, but in military radars the lowest noise figure attainable might not always be appropriate. In a high-noise environment, whether due to unintentional interference or to hostile jamming, a radar with a low-noise receiver is more susceptible than one with higher noise figure. Also, a low-noise amplifier as the front end generally will result in the receiver having less dynamic range—something not desirable when faced with hostile electronic countermeasures (ECM) or when the doppler effect is used to detect small targets in the presence of large clutter. When the disadvantages of a low-noise-figure receiver are to be avoided, the RF amplifier stage is omitted and the mixer stage is employed as the receiver front end. The higher noise figure of the mixer can then be compensated by an equivalent increase in the transmitter power.

The mixer of the superheterodyne receiver translates the receiver RF signal to an intermediate frequency. The gain of the intermediate-frequency (IF) amplifier results in an increase of the receiver signal level. The IF amplifier also includes the function of the matched filter: one which maximizes the output signal-to-noise ratio. Maximizing the signal-to-noise ratio at the output of the IF maximizes the detectability of the signal. Almost all radars have a receiver which closely approximates the matched filter.

The second detector in the receiver is an envelope detector which eliminates the IF carrier and passes the modulation envelope. When doppler processing is employed, as it is in CW (continuous-wave), MTI, and pulse doppler radars, the envelope detector is replaced by a phase detector which extracts the doppler frequency by comparison with a reference signal at the transmitted frequency.

There must also be included filters for rejecting the stationary clutter and passing the doppler-frequency-shifted signals from moving targets.

The video amplifier raises the signal power to a level where it is convenient to display the information it contains. As long as the video bandwidth is not less than half of the IF bandwidth, there is no adverse effect on signal detectability.

A threshold is established at the output of the video amplifier to allow the detection decision to be made. If the receiver output crosses the threshold, a target is said to be present. The decision may be made by an operator, or it might be done with an automatic detector without operator intervention.

Signal Processing.   There has not always been general agreement as to what constitutes the signal-processing portion of the radar, but it is usually considered to be the processing whose purpose is to reject undesired signals (such as clutter) and pass desired signals due to targets. It is performed prior to the threshold detector where the detection decision is made. Signal processing includes the matched filter and the doppler filters in MTI and pulse doppler radar. Pulse compression, which is performed before the detection decision is made, is sometimes considered to be signal processing, although it does not fit the definition precisely.

Data Processing.   This is the processing done after the detection decision has been made. Automatic tracking (Chap. 8) is the chief example of data processing. Target recognition is another example. It is best to use automatic tracking with a good radar that eliminates most of the unwanted signals so that the automatic tracker only has to deal with desired target detections and not undesired clutter. When a radar cannot eliminate all nuisance echoes, a means to maintain a constant false-alarm rate (CFAR) at the input to the tracker is necessary.

The CFAR portion of the receiver is usually found just before the detection decision is made. It is required to maintain the false-alarm rate constant as the clutter and/or noise background varies. Its purpose is to prevent the automatic tracker from being overloaded with extraneous echoes. It senses the magnitude of the radar echoes from noise or clutter in the near vicinity of the target and uses this information to establish a threshold so that the noise or clutter echoes are rejected at the threshold and not confused as targets by the automatic tracker.

Unfortunately, CFAR reduces the probability of detection. It also results in a loss in signal-to-noise ratio, and it degrades the range resolution. CFAR or its equivalent is necessary when automatic tracking computers cannot handle large numbers of echo signals, but it should be avoided if possible. When an operator is used to make the threshold decision, CFAR is not a necessity as in limited-capacity automatic systems because the operator can usually recognize echoes due to clutter or to increased noise (such as jamming) and not confuse them with desired targets.

Displays.   The display for a surveillance radar is usually a cathode-ray tube with a PPI (plan position indicator) format. A PPI is an intensity-modulated, maplike presentation that provides the target's location in polar coordinates (range and angle). Older radars presented the video output of the receiver (called *raw video*) directly to the display, but more modern radars generally display *processed* video, that is, after processing by the automatic detector or the automatic detector and tracker (ADT). These are sometimes called *cleaned-up displays* since the noise and background clutter are removed.

Radar Control.   A modern radar can operate at different frequencies within its band, with different waveforms and different signal processing, and with different polarizations so as to maximize its performance under different environ-

mental conditions. These radar parameters might need to be changed according to the local weather, the clutter environment (which is seldom uniform in azimuth and range), interference to or from other electronic equipment, and (if a military radar) the nature of the hostile ECM environment. The different parameters, optimized for each particular situation, can be programmed into the radar ahead of time in anticipation of the environment, or they can be chosen by an operator in real time according to the observed environmental conditions. On the other hand, a *radar control* can be made to automatically recognize when environmental conditions have changed and automatically select, without the aid of an operator, the proper radar operating parameters to maximize performance.

Waveform. The most common radar waveform is a repetitive train of short pulses. Other waveforms are used in radar when particular objectives need to be achieved that cannot be accomplished with a pulse train. CW (a continuous sine wave) is employed on some specialized radars for the measurement of radial velocity from the doppler frequency shift. FM/CW (frequency-modulated CW) is used when range is to be measured with a CW waveform (Chap. 14). Pulse compression waveforms (Chap. 10) are used when the resolution of a short pulse but the energy of a long pulse is desired. MTI radars (Chaps. 15 and 16) with low pulse repetition frequencies (PRFs) and pulse doppler radars (Chap. 17) with high PRFs often use waveforms with multiple pulse repetition intervals in order to avoid range and/or doppler ambiguities.

1.2 RADAR EQUATION

Perhaps the single most useful description of the factors influencing radar performance is the radar equation which gives the range of a radar in terms of the radar characteristics. One form of this equation gives the received signal power P_r as

$$P_r = \frac{P_t G_t}{4\pi R^2} \times \frac{\sigma}{4\pi R^2} \times A_e \tag{1.1}$$

The right-hand side has been written as the product of three factors to represent the physical processes taking place. The first factor is the power density at a distance R meters from a radar that radiates a power of P_t watts from an antenna of gain G_t. The numerator of the second factor is the target cross section σ in square meters. The denominator accounts for the divergence on the return path of the electromagnetic radiation with range and is the same as the denominator of the first factor, which accounts for the divergence on the outward path. The product of the first two terms represents the power per square meter returned to the radar. The antenna of effective aperture area A_e intercepts a portion of this power in an amount given by the product of the three factors. If the maximum radar range $R_{\max}$ is defined as that which results in the received power P_r being equal to the receiver minimum detectable signal $S_{\min}$, the radar equation may be written

$$R^4{}_{\max} = \frac{P_t G_t A_e \sigma}{(4\pi)^2 S_{\min}} \tag{1.2}$$

When the same antenna is used for both transmitting and receiving, the transmitting gain G_t and the effective receiving aperture A_e are related by $G_t = 4\pi A_e/\lambda^2$, where λ is the wavelength of the radar electromagnetic energy. Substituting into Eq. (1.2) gives two other forms of the radar equation:

$$R^4_{max} = \frac{P_t G_t^2 \lambda^2 \sigma}{(4\pi)^3 S_{min}} \tag{1.3a}$$

$$R^4_{max} = \frac{P_t A_e^2 \sigma}{4\pi\lambda^2 S_{min}} \tag{1.3b}$$

The examples of the radar equation given above are useful for rough computations of range performance but they are overly simplified and do not give realistic results. The predicted ranges are generally overly optimistic. There are at least two major reasons why the simple form of the radar equation does not predict with any accuracy the range of actual radars. First, it does not include the various losses that can occur in a radar. Second, the target cross section and the minimum detectable signal are statistical in nature. Thus the specification of the range must be made in statistical terms. The elaboration of the simple range equation to yield meaningful range predictions is the subject of Chap. 2. Although the range enters as the fourth power in Eq. (1.3), it can appear as the cube, as the square, or as the first power in specific situations, some of which are described later in this section and in other chapters.

In addition to its use for range prediction, the radar equation forms a good basis for preliminary system design by providing a guide to the possible tradeoffs among the various parameters that enter into radar performance.

The minimum detectable signal S_{min}, which appears in the radar equation, is a statistical quantity and must be described in terms of the probability of detection and the probability of a false alarm. This is discussed in more detail in Chap. 2; for present purposes it suffices to state that for a signal to be reliably detected it must be larger than noise (generally by 10 to 20 dB) at the point in the receiver where the detection decision is made. The minimum detectable signal can be expressed as the signal-to-noise ratio (S/N) required for reliable detection times the receiver noise. The receiver noise is expressed relative to the thermal noise that would be produced by an ideal receiver. The thermal noise is equal to kTB, where k is Boltzmann's constant, T is the temperature, and B is the receiver bandwidth. The receiver noise is the thermal noise multiplied by the factor F_n, the receiver noise figure. The receiver noise figure is measured relative to a reference temperature $T_0 = 290$ K (approximately room temperature), and the factor kT_0 becomes 4×10^{-21} W/Hz. The minimum detectable signal in the radar equation can be written

$$S_{min} = kT_0 B F_n \frac{S}{N} \tag{1.4}$$

Sometimes the factor $T_0 F_n$ is replaced with T_s, the system noise temperature.

The above discussion of the radar equation was in terms of the signal power. Although power is a well-understood characteristic of the usual radar waveform consisting of a rectangular pulse, with more complicated waveforms the total sig-

nal energy is often a more convenient measure of waveform detectability. It also is more appropriate for theoretical reasons. The ratio of signal energy to noise energy, denoted E/N_0, is a more fundamental parameter than the signal-to-noise (power) ratio in theoretical analyses based on statistical detection theory. No matter what the shape of the received waveform, if the receiver is designed as a matched filter the peak signal-to-noise (power) ratio at the output of the matched filter is $2E/N_0$.

For a rectangular pulse of width τ the signal power is E/τ and the noise power is $N_0 B$, where E = signal energy, N_0 = noise energy, or noise power per unit bandwidth (provided the noise is uniform with frequency), and B = receiver bandwidth. With these substitutions, S_{min} becomes $kT_0 F_n(E/N_0)/\tau$. Substituting into Eq. (1.2) gives

$$R^4{}_{max} = \frac{E_t G_t \, A_e \sigma}{(4\pi)^2 kT_0 F_n(E/N_0)} \tag{1.5}$$

where $E_t = P_t\tau$ is the energy contained in the transmitted waveform. Although Eq. (1.5) assumes a rectangular pulse, it can be applied to any waveform provided that E_t is interpreted as the energy contained in the transmitted waveform and that the receiver of noise figure F_n is designed as a matched filter. Some of the published results of radar detection theory give the probability of detection and probability of false alarm in terms of S/N rather than E/N_0. When these results assume optimum (matched-filter) processing, the required values of E/N_0 for use in the radar equation can be obtained from the published results for S/N or the visibility factor as described in Chap. 2.

The radar equation can be manipulated into various forms, depending on the particular application. Several examples are given below.

Tracking. In this situation the radar is assumed to track continuously or "searchlight" a target for an interval of time t_0. Equation (1.5) applies, so that the tracking-, or searchlighting-, radar equation is

$$R^4{}_{max} = \frac{P_{av}t_0 G_t \, A_e \sigma}{4\pi kT_0 F_n(E/N_0)} \tag{1.6}$$

where $P_{av}t_0 = E_t$. Thus, in a tracking radar that must "see" to a long range, the average power must be high, the time on target must be long, and the antenna must be of large electrical size (G_t) and large physical size (A_e). The frequency does not enter explicitly. Since it is easier mechanically to move a small antenna than a large one, tracking radars are usually found at the higher frequencies, where small apertures can have high gain and thus an adequate $G_t A_e$ product.

The radar equation is based on detectability. A tracking radar must also be designed for good angular accuracy. Good angle accuracy is achieved with narrow beamwidths (large G_t) and with high E/N_0 (large A_e). Thus a large $G_t A_e$ product is consistent with good tracking accuracy as well as good detectability.

Volume Search. Assume that the radar must search an angular volume of Ω steradians in the time t_s. If the antenna beam subtends an angle of Ω_b steradians, the antenna gain G_t is approximately $4\pi/\Omega_b$. If the antenna beam dwells a time t_0 in each direction subtended by the beam, the total scan time is $t_s = t_0\Omega/\Omega_b$. Substituting these expressions into Eq. (1.5) and noting that $E_t = P_{av}t_0$,

$$R^4_{\max} = \frac{P_{av} A_e \sigma}{4\pi k T_0 F_n (E/N_0)} \frac{t_s}{\Omega}$$ (1.7)

Thus for a volume search radar the two important parameters for maximizing range are the average transmitter power and the antenna aperture. Any decrease in time to scan the volume or any increase in the volume searched must be accompanied by a corresponding increase in the product $P_{av} A_e$. Note that the frequency does not enter explicitly.

Jamming. When the detection of the radar signal is limited by an external noise source, such as a deliberate noise jammer rather than by receiver noise, the parameters of importance in determining range performance are slightly different from those presented above (Chap. 9). The receiver noise power per unit bandwidth is now that determined by the jammer rather than the receiver noise figure. When a radar is performing volume search and jamming power enters from a particular direction via the sidelobes, the maximum range can be written

$$R^4_{\max} = \frac{P_{av}}{g_s} \frac{t_s}{\Omega} \frac{\sigma}{E/N_0} \frac{R_J^2 B_J}{P_J G_J}$$ (1.8)

where g_s = sidelobe level relative to main beam (number less than unity)
   R_J = jammer range
   B_J = jammer bandwidth
   P_J = jammer power
   G_J = jammer antenna gain

and E/N_0 is the ratio of signal energy to noise power per unit bandwidth necessary for reliable detection. The parameter of importance is the average power. The antenna sidelobes are also important. This equation was derived by substituting for kT_0F_n in Eq. (1.7) the jamming noise power per unit bandwidth that would enter the radar receiving-antenna sidelobes. It applies only when the normal receiver noise is negligible compared with the jamming noise.

When the radar is searchlighting a target with a jammer, a mode of operation sometimes called *burnthrough*, the range becomes

$$R^2_{\max} = \frac{P_{av} t_0 G_t}{4\pi} \frac{\sigma}{E/N_0} \frac{B_J}{P_J G_J}$$ (1.9)

The important radar parameters are the average power, the time of observation, and the transmitting-antenna gain. The maximum range is squared rather than raised to the fourth power as in other forms of the radar equation. Note that in neither jamming example does the antenna aperture area enter explicitly. A large aperture collects more signal, but it also collects more jamming noise. The receiver noise figure does not enter because it is assumed that the jamming noise is considerably larger than the receiver noise. Thus in a noisy environment one might not benefit from the effort to design a receiver with the ultimate in sensitivity. The above two examples of jamming radar equations are simplifications. Other variations are possible.

Clutter. When the radar must detect a small target located on the surface of the sea or land, the interfering unwanted clutter echoes can severely limit the detectability of the target. When clutter power dominates receiver noise power, the range equation simply reduces to an expression for the signal-to-clutter ratio. This ratio is equal to the ratio of the target cross section to the clutter cross section. If clutter is distributed more or less uniformly, the clutter echo will depend on the area illuminated by the radar resolution cell. Surface (ground or sea) clutter is described by the ratio of the clutter echo to the area illuminated by the radar. This normalized clutter coefficient is denoted σ^0.

Consider a pulse radar viewing the target and the clutter at low grazing angles. If single-pulse detection is assumed, the signal-to-clutter ratio is

$$\frac{S}{C} = \frac{\sigma}{\sigma^0 R \theta_b (c\tau/2) \sec \phi} \tag{1.10}$$

or

$$R_{max} = \frac{\sigma}{(S/C)_{min}\sigma^0\theta_b(c\tau/2) \sec \phi}$$

where R = range to clutter patch
  θ_b = azimuth beamwidth
  c = velocity of propagation
  τ = pulse width
  ϕ = grazing angle

The clutter patch is assumed to be determined in azimuth by the width of the antenna beam and in the range coordinate by the pulse width. The ratio S/C takes a role similar to the ratio E/N_0 for thermal noise. It must be of sufficient magnitude to achieve reliable detection. The clutter statistics generally differ from the statistics of thermal noise but, as a first guess when no other information is available, the required values of S/C might be taken to be those of E/N_0. It is significant that the range dependence enters linearly rather than as the fourth power. Thus for detection of a target in clutter the radar beam should be narrow and the pulse width should be short. With assumptions other than those above, the important radar parameters for detection of targets in clutter might be different. If n hits are received per scan and if the clutter is correlated from pulse to pulse, no improvement in S/C is obtained as it would be if thermal noise, rather than clutter, were the limitation.

1.3 INFORMATION AVAILABLE FROM THE RADAR ECHO

Although the name *radar* is derived from *radio detection and ranging*, a radar is capable of providing more information about the target than is implied by its name. *Detection* of a target signifies the discovery of its presence. It is possible to consider *detection* independently of the process of information extraction, but it is not often that one is interested in knowing only that a target is

present without knowing something about its location in space and its nature. The extraction of useful target information is therefore an important part of radar operation.

The ability to consider *detection* independent of *information* extraction does not mean that there is no relation between the two. The extraction of information generally requires a matched filter, or its equivalent, for optimum processing. The more information that is known about the target a priori, the more efficient will be the detection. For example, if the target location were known, the antenna could be pointed in the proper direction and energy or time need not be wasted searching empty space. Or, if the relative velocity were known, the receiver could be pretuned to the correct received frequency, negating the need to search the frequency band over which the doppler shift might occur.

The usual radar provides the location of a target in range and angle. The rate of change of target location can also be measured from the change in range and angle with time, from which the track can be established. In many radar applications a *detection* is not said to occur until its track has been established.

A radar with sufficient resolution in one or more coordinates can determine a target's size and shape. Polarization allows a measure of the symmetry of a target. In principle, a radar can also measure the surface roughness of a target and determine something about its dielectric properties.

Range. The ability to determine range by measuring the time for the radar signal to propagate to the target and back is probably the distinguishing and most important characteristic of conventional radar. No other sensor can measure range to the accuracy possible with radar, at such long ranges, and under adverse weather conditions. Surface-based radars can be made to determine the range of an aircraft to an accuracy of a few tens of meters at distances limited only by the line of sight, generally 200 to 250 nmi. Radar has demonstrated its ability to measure interplanetary distances to an accuracy limited only by the accuracy to which the velocity of propagation is known. At more modest distances, the measurement of range can be made with a precision of a few centimeters.

The usual radar waveform for determining range is the short pulse. The shorter the pulse, the more precise can be the range measurement. A short pulse has a wide spectral width (bandwidth). The effect of a short pulse can be obtained with a long pulse whose spectral width has been increased by phase or frequency modulation. When passed through a *matched filter*, the output is a compressed pulse whose duration is approximately the reciprocal of the spectral width of the modulated long pulse. This is called *pulse compression* and allows the resolution of a short (wide-bandwidth) pulse with the energy of a long pulse. A CW waveform with frequency or phase modulation can also provide an accurate range measurement. It is also possible to measure the range of a single target by comparing the phase difference between two or more CW frequencies. Range measurement with CW waveforms has been widely employed, as in aircraft radar altimeters and surveying instruments.

Radial Velocity. From successive measurements of range the rate of change of range, or radial velocity, can be obtained. The doppler frequency shift of the echo signal from a moving target also provides a measure of radial velocity. However, the doppler frequency measurement in many pulse radars is highly ambiguous, thus reducing its utility as a direct measurement of radial velocity.

When it can be used, it is often preferred to successive range measurements since it can achieve a more accurate measurement in a shorter time.

Any measurement of velocity, whether by the rate of change of range or by the doppler frequency shift, requires time. The longer the time of observation, the more accurate can be the measurement of velocity. (A longer observation time also can increase the signal-to-noise ratio, another factor that results in increased accuracy.) Although the doppler frequency shift is used in some applications to measure radial velocity (as, for example, in such diverse applications as the police speed meter and satellite surveillance radars), it is more widely employed as the basis for sorting moving targets from unwanted stationary clutter echoes, as in MTI, AMTI (airborne MTI), pulse doppler, and CW radars.

Angular Direction.   The direction of a target is determined by sensing the angle at which the returning wavefront arrives at the radar. This is usually accomplished with a directive antenna, i.e., one with a narrow radiation pattern. The direction in which the antenna points when the received signal is a maximum indicates the direction of the target. This, as well as other methods for measuring angle, assumes that the atmosphere does not perturb the straight-line propagation of the electromagnetic waves.

The direction of the incident waveform can also be determined by measuring the phase difference between two separated receiving antennas, as with an interferometer. Phase-comparison monopulse also is based on the phase measurement of signals in two separated antennas. The amplitude-comparison monopulse determines the angle of arrival by comparing the amplitudes of the signals received in two squinted beams generated by a single antenna.

The accuracy of the angle of arrival depends on the extent of the antenna aperture. The wider the antenna, the narrower the beamwidth and the better the accuracy.

The angle of arrival, or target direction, is not strictly a radar measurement (as are the range and radial velocity) if a radar measurement is defined as one obtained by comparing the reflected echo signal with the transmitted signal. The determination of angle basically involves only the one-way path. Nevertheless, the angle measurement is an integral part of most surveillance and tracking radars.

Size.   If the radar has sufficient resolution, it can provide a measurement of the target's extent, or size. Since many targets of interest have dimensions of several tens of meters, resolution must be several meters or less. Resolutions of this order can be readily obtained in the range coordinate. With conventional antennas and the usual radar ranges, the angular resolution is considerably poorer than what can be achieved in range. However, target resolution in the cross-range (angle) dimension can be obtained comparable with that obtained in range by the use of resolution in the doppler frequency domain. This requires that there be relative motion between the various parts of the target and the radar. It is the basis for the excellent cross-range resolution obtained in a SAR in which the relative motion between target and radar occurs because of the travel of the aircraft or spacecraft on which the radar is mounted. In an ISAR (inverse synthetic aperature radar) the relative motion is provided by the movement of the target.

Shape.   The size of a target is seldom of interest in itself, but its shape and its size are important for recognizing one type of target from another. A high-resolution radar that obtains the profile of a target in both range and

cross range (as do SAR and ISAR) provides the size and shape of the target. The shape of an object can also be obtained by tomography, in which a two-dimensional image of a three-dimensional object is reconstructed from the measurement of phase and amplitude, at different angles of observation. (The radar might rotate around the fixed object, or the radar can be fixed and the object rotated about its own axis.) Range resolution is not necessary with the coherent tomographic radar method.

As mentioned earlier, comparison of the scattered fields for different polarizations provides a measure of target asymmetry. It should be possible to distinguish targets with different aspect ratios (shapes), as for example, rods from spheres and spheres from aircraft. The complete exploitation of polarization requires the measurement of phase, as well as amplitude of the echo signal at two orthogonal polarizations and a cross-polarization component. Such measurements (which define the polarization matrix) should allow in principle the recognition of one class of target from another, but in practice it is not easy to do.

One characteristic of target *shape* is its surface roughness. This measurement can be of particular interest for echoes from the ground and the sea. Rough targets scatter the incident electromagnetic energy diffusely; smooth targets scatter specularly. By observing the nature of the backscatter as a function of the incident angle it should be possible to determine whether a surface is smooth or rough. Surface roughness is a relative measure and depends on the wavelength of the illuminating signal. A surface that appears rough at one wavelength might appear smooth when illuminated with longer-wavelength radiation. Thus another method for determining surface roughness is by varying the frequency of the illuminating radiation and observing the transition from specular to diffuse scatter. A direct method for determining roughness is to observe the scatter from the object with a resolution that can resolve the roughness scale.

Other Target Measurements.   Just as the radial velocity can be determined from the *temporal* doppler frequency shift, it is possible to measure the tangential (cross-range) component of velocity. This can be obtained from the analogous *spatial* doppler frequency shift that expands or compresses the apparent antenna radiation pattern (just as the radial component of velocity can expand or compress the time waveform of a radar signal reflected from a moving target to produce a temporal doppler frequency shift). A measurement of tangential velocity requires a wide-baseline antenna, such as an interferometer. The measurement of tangential velocity has not seen application because the required baseline is often too wide for practical purposes.

It is also possible to note the change of a complex target's radial projection from the change of the received-signal amplitude with time. (A change in the radial projection of a target usually results in a change of the radar cross section.)

Vibrations of the target, rotation of the propellers of an aircraft, or the rotation of a jet engine can induce distinctive modulation on the radar echo which can be detected by a spectral analysis of the radar echo signal.

1.4   RADAR FREQUENCIES

There are no fundamental bounds on radar frequency. Any device that detects and locates a target by radiating electromagnetic energy and utilizes the echo scattered from a target can be classed as a radar, no matter what its frequency.

Radars have been operated at frequencies from a few megahertz to the ultraviolet region of the spectrum. The basic principles are the same at any frequency, but the practical implementation is widely different. In practice, most radars operate at microwave frequencies, but there are notable exceptions.

Radar engineers use letter designations, as shown in Table 1.1, to denote the general frequency band at which a radar operates. These letter bands are universally used in radar. They have been officially accepted as a standard by the Institute of Electrical and Electronics Engineers (IEEE) and have been recognized by the U.S. Department of Defense. Attempts have been made in the past to subdivide the spectrum into other letter bands (as for waveguides and for ECM operations), but the letter bands in Table 1.1 are the only ones that should be used for radar.

The original code letters (P, L, S, X, and K) were introduced during World War II for purposes of secrecy. After the need for secrecy no longer existed, these designations remained. Others were later added as new regions of the spectrum were utilized for radar application. (The nomenclature *P band* is no longer in use. It has been replaced with *UHF*.)

Letter bands are a convenient way to designate the general frequency range of a radar. They serve an important purpose for military applications since they can describe the frequency band of operation without using the exact frequencies at which the radar operates. The exact frequencies over which a radar operates should be used in addition to or instead of the letter bands whenever proper to do so.

TABLE 1.1 Standard Radar-Frequency Letter-Band Nomenclature*

Band designation	Nominal frequency range	Specific frequency ranges for radar based on ITU assignments for Region 2
HF	3 MHz–30 MHz	
VHF	30 MHz–300 MHz	138 MHz–144 MHz
		216 MHz–225 MHz
UHF	300 MHz–1000 MHz	420 MHz–450 MHz
		890 MHz–942 MHz
L	1000 MHz–2000 MHz	1215 MHz–1400 MHz
S	2000 MHz–4000 MHz	2300 MHz–2500 MHz
		2700 MHz–3700 MHz
C	4000 MHz–8000 MHz	5250 MHz–5925 MHz
X	8000 MHz–12,000 MHz	8500 MHz–10,680 MHz
K_u	12.0 GHz–18 GHz	13.4 GHz–14.0 GHz
		15.7 GHz–17.7 GHz
K	18 GHz–27 GHz	24.05 GHz–24.25 GHz
K_a	27 GHz–40 GHz	33.4 GHz–36.0 GHz
V	40 GHz–75 GHz	59 GHz–64 GHz
W	75 GHz–110 GHz	76 GHz–81 GHz
		92 GHz–100 GHz
mm	110 GHz–300 GHz	126 GHz–142 GHz
		144 GHz–149 GHz
		231 GHz–235 GHz
		238 GHz–248 GHz

*From IEEE Standard 521-1984.

The International Telecommunications Union (ITU) assigns specific frequency bands for radiolocation (radar) use. These are listed in the third column of Table 1.1. They apply to ITU Region 2, which encompasses North and South America. Slight differences exist in the other two ITU regions. Although L band, for example, is shown in the second column of the table as extending from 1000 to 2000 MHz, in practice an L-band radar would be expected to be found somewhere between 1215 and 1400 MHz, the frequency band actually assigned by the ITU.

Each frequency band has its own particular characteristics that make it better for certain applications than for others. In the following, the characteristics of the various portions of the electromagnetic spectrum at which radars have been or could be operated are described. The divisions between the frequency regions are not as sharp in practice as the precise nature of the nomenclature.

HF (3 to 30 MHz). Although the first operational radars installed by the British just prior to World War II were in this frequency band, it has many disadvantages for radar applications. Large antennas are required to achieve narrow beamwidths, the natural ambient noise level is high, the available bandwidths are narrow, and this portion of the electromagnetic spectrum is widely used and restrictively narrow. In addition, the long wavelength means that many targets of interest might be in the Rayleigh region, where the dimensions of the target are small compared with the wavelength; hence, the radar cross section of targets small in size compared with the (HF) wavelength might be lower than the cross section at microwave frequencies.

The British used this frequency band, even though it had disadvantages, because it was the highest frequency at which reliable, readily available high-power components were then available. Ranges of 200 mi were obtained against aircraft. These were the radars that provided detection of hostile aircraft during the battle of Britain and were credited with allowing the limited British fighter resources to be effectively used against the attacking bomber aircraft. They did the job that was required.

Electromagnetic waves at HF have the important property of being refracted by the ionosphere so as to return to the earth at ranges from about 500 to 2000 nmi, depending on the actual condition of the ionosphere. This allows the over-the-horizon detection of aircraft and other targets. The long over-the-horizon ranges that are possible make the HF region of the spectrum quite attractive for the radar observation of areas (such as the ocean) not practical with conventional microwave radar.

VHF (30 to 300 MHz). Most of the early radars developed in the 1930s were in this frequency band. Radar technology at these frequencies represented a daring venture that pushed to the edge of technology known in the thirties. These early radars served quite well the needs of the time and firmly established the utility of radar.

Like the HF region, the VHF (very high frequency) region is crowded, bandwidths are narrow, external noise can be high, and beamwidths are broad. However, the necessary technology is easier and cheaper to achieve than at microwave frequencies. High power and large antennas are readily practical. The stable transmitters and oscillators required for good MTI are easier to achieve than at higher frequencies, and there is relative freedom from the blind speeds that limit the effectiveness of MTI as the frequency is increased. Reflections from rain are not a problem. With horizontal polarization over a good reflecting sur-

face (such as the sea), the constructive interference between the direct wave and the wave reflected from the surface can result in a substantial increase in the maximum range against aircraft (almost twice the free-space range). However, a consequence of this increase in range due to constructive interference is that the accompanying destructive interference results in nulls in the coverage at other elevation angles and less energy at low angles. It is a good frequency for lower-cost radars and for long-range radars such as those for the detection of satellites. It is also the frequency region where it is theoretically difficult to reduce the radar cross section of most types of airborne targets.

In spite of its many attractive features, there have not been many applications of radar in this frequency range because its limitations do not always counterbalance its advantages.

UHF (300 to 1000 MHz). Much of what has been said regarding VHF applies to UHF. However, natural external noise is much less of a problem, and beamwidths are narrower than at VHF. Weather effects usually are not a bother. With a suitably large antenna, it is a good frequency for reliable long-range surveillance radar, especially for extraterrestrial targets such as spacecraft and ballistic missiles. It is well suited for AEW (airborne early warning), e.g., airborne radar that uses AMTI for the detection of aircraft. Solid-state transmitters can generate high power at UHF as well as offer the advantages of maintainability and wide bandwidth.

L Band (1.0 to 2.0 GHz). This is the preferred frequency band for land-based long-range air surveillance radars, such as the 200-nmi radars used for en route air traffic control [designated ARSR by the U.S. Federal Aviation Administration (FAA)]. It is possible to achieve good MTI performance at these frequencies and to obtain high power with narrow-beamwidth antennas. External noise is low. Military 3D radars can be found at L band, but they also are at S band. L band is also suitable for large radars that must detect extraterrestrial targets at long range.

S Band (2.0 to 4.0 GHz). Air surveillance radars can be of long range at S band, but long range usually is more difficult to achieve than at lower frequencies. The blind speeds that occur with MTI radar are more numerous as the frequency increases, thus making MTI less capable. The echo from rain can significantly reduce the range of S-band radars. However, it is the preferred frequency band for long-range weather radars that must make accurate estimates of rainfall rate. It is also a good frequency for medium-range air surveillance applications such as the airport surveillance radar (ASR) found at air terminals. The narrower beamwidths at this frequency can provide good angular accuracy and resolution and make it easier to reduce the effects of hostile main-beam jamming that might be encountered by military radars. Military 3D radars and height finding radars are also found at this frequency because of the narrower elevation beamwidths that can be obtained at the higher frequencies. Long-range airborne air surveillance pulse doppler radars, such as AWACS (Airborne Warning and Control System) also operate in this band.

Generally, frequencies lower than S band are well suited for air surveillance (detection and low-data-rate tracking of many aircraft within a large volume). Frequencies above S band are better for information gathering, such as high-data-rate precision tracking and the recognition of individual targets. If a single fre-

quency must be used for both air surveillance and precision tracking, as in military air defense systems based on phased array multifunction radar, a suitable compromise might be S band.

C Band (4.0 to 8.0 GHz). This band lies between the S and X bands and can be described as a compromise between the two. It is difficult, however, to achieve long-range air surveillance radars at this or higher frequencies. It is the frequency where one can find long-range precision instrumentation radars used for the accurate tracking of missiles. This frequency band has also been used for multifunction phased array air defense radars and for medium-range weather radars.

X Band (8 to 12.5 GHz). This is a popular frequency band for military weapon control (tracking) radar and for civil applications. Shipboard navigation and piloting, weather avoidance, doppler navigation, and the police speed meter are all found at X band. Radars at this frequency are generally of convenient size and are thus of interest for applications where mobility and light weight are important and long range is not. It is advantageous for information gathering as in high-resolution radar because of the wide bandwidth that makes it possible to generate short pulses (or wideband pulse compression) and the narrow beamwidths that can be obtained with relatively small-size antennas. An X-band radar may be small enough to hold in one's hand or as large as the MIT Lincoln Laboratory Haystack Hill radar with its 120-ft-diameter antenna and average radiated power of about 500 kW. Rain, however, can be debilitating to X-band radar.

K_u, K, and K_a Bands (12.5 to 40 GHz). The original K-band radars developed during World War II were centered at a wavelength of 1.25 cm (24 GHz). This proved to be a poor choice since it is too close to the resonance wavelength of water vapor (22.2 GHz), where absorption can reduce the range of a radar. Later this band was subdivided into two bands on either side of the water-vapor absorption frequency. The lower frequency band was designated K_u, and the upper band was designated K_a. These frequencies are of interest because of the wide bandwidths and the narrow beamwidths that can be achieved with small apertures. However, it is difficult to generate and radiate high power. Limitations due to rain clutter and attenuation are increasingly difficult at the higher frequencies. Thus not many radar applications are found at these frequencies. However, the airport surface detection radar for the location and control of ground traffic at airports is at K_u band because of the need for high resolution. The disadvantages that characterize this band are not important in this particular application because of the short range.

Millimeter Wavelengths (above 40 GHz). Although the wavelength of K_a band is about 8.5 millimeters (a frequency of 35 GHz), the technology of K_a-band radar is more like that of microwaves than that of millimeter waves and is seldom considered to be representative of the millimeter-wave region. Millimeter-wave radar, therefore, is taken to be the frequency region from 40 to 300 GHz. The exceptionally high attenuation caused by the atmospheric oxygen absorption line at 60 GHz precludes serious applications in the vicinity of this frequency within the atmosphere. Therefore, the 94-GHz-frequency region (3-mm wavelength) is generally what is thought of as a "typical" frequency representative of millimeter radar.

The millimeter-wave region above 40 GHz has been further subdivided into

letter bands in the IEEE Standard, as shown in Table 1.1. Although there has been much interest in the millimeter portion of the electromagnetic spectrum, there have been no operational radars above K_a band. High-power sensitive receivers and low-loss transmission lines are difficult to obtain at millimeter wavelengths, but such problems are not basic. The major reason for the limited utility of this frequency region is the high attenuation that occurs even in the "clear" atmosphere. The so-called propagation window at 94 GHz is actually of greater attenuation than the attenuation at the water-vapor absorption line at 22.2 GHz. The millimeter-wave region is more likely to be of interest for operation in space, where there is no atmospheric attenuation. It might also be considered for short-range applications within the atmosphere where the total attenuation is not large and can be tolerated.

Laser Frequencies. Coherent power of reasonable magnitude and efficiency, along with narrow directive beams, can be obtained from lasers in the infrared, optical, and ultraviolet region of the spectrum. The good angular resolution and range resolution possible with lasers make them attractive for target information-gathering applications, such as precision ranging and imaging. They have had application in military range finders and in distance measurement for surveying. They have been considered for use from space for measuring profiles of atmospheric temperature, water vapor, and ozone, as well as measuring cloud height and tropospheric winds. Lasers are not suitable for wide-area surveillance because of their relatively small physical aperture area. A serious limitation of the laser is its inability to operate effectively in rain, clouds, or fog.

1.5 RADAR NOMENCLATURE

Military electronic equipment, including radar, is designated by the Joint Electronics Type Designation System (JETDS), formerly known as the Joint Army-Navy Nomenclature System (AN System), as described in Military Standard MIL-STD-196D. The letter portion of the designation consists of the letters AN, a slant bar, and three additional letters appropriately selected to indicate where the equipment is installed, the type of equipment, and the purpose of the equipment. Table 1.2 lists the equipment indicator letters. Following the three letters are a dash and a numeral. The numeral is assigned in sequence for that particular combination of letters. Thus the designation AN/SPS-49 is for a shipboard surveillance radar. The number 49 identifies the particular equipment and indicates it is the forty-ninth in the SPS category to which a JETDS designation has been assigned. A suffix letter (A, B, C, etc.) follows the original designation for each modification where interchangeability has been maintained. A change in the power input voltage, phase, or frequency is identified by the addition of the letters X, Y, or Z to the basic nomenclature. When the designation is followed by a dash, the letter T, and a number, the equipment is designed for training. The letter V in parentheses added to the designation indicates variable systems (those whose functions may be varied through the addition or deletion of sets, groups, units, or combinations thereof). Experimental and developmental systems sometimes are assigned special indicators enclosed in parentheses, immediately following the regular designation, to identify the development organization; for example, (XB) indicates the Naval Research Laboratory, and (XW) indicates the

Rome Air Development Center. Empty parentheses, commonly called "bow-legs," are used for a developmental or series "generic" assignment.

TABLE 1.2 JETDS Equipment Indicators*

Installation (first letter)	Type of equipment (second letter)	Purpose (third letter)
A Piloted aircraft	A Invisible light, heat radiation	A Auxiliary assembly
B Underwater mobile, submarine	C Carrier	B Bombing
D Pilotless carrier	D Radiac (radioactive detection, indication, and computation devices)	C Communications (receiving and transmitting)
F Fixed ground		
G General ground use		
K Amphibious		D Direction finder reconnaissance and/or surveillance
M Ground, mobile	E Laser	
P Portable	G Telegraph or teletype	E Ejection and/or release
S Water	I Interphone and public address	G Fire control or search-light directing
T Ground, transportable		
U General utility	J Electromechanical or inertial wire-covered	H Recording and/or reproducing (graphic meteorological)
V Ground, vehicular		
W Water surface and underwater combination	K Telemetering	
	L Countermeasures	K Computing
	M Meteorological	M Maintenance and/or test assemblies (including tools)
Z Piloted and pilotless airborne vehicle combination	N Sound in air	
	P Radar	
	Q Sonar and underwater sound	N Navigational aids (including altimeters, beacons, compasses, racons, depth sounding, approach and landing)
	R Radio	
	S Special types, magnetic, etc., or combinations of types	
	T Telephone (wire)	Q Special or combination of purposes
	V Visual and visible light	
	W Armament (peculiar to armament, not otherwise covered)	R Receiving, passive detecting
		S Detecting and/or range and bearing, search
	X Facsimile or television	
	Y Data processing	T Transmitting
		W Automatic flight or remote control
		X Identification and recognition
		Y Surveillance (search, detect, and multiple-target tracking) and control (both fire control and air control)

In the first column of Table 1.2, the installation letter M is used for equipment installed and operated from a vehicle whose sole function is to house and transport the equipment. The letter T is used for ground equipment that is normally moved from place to place and is not capable of operation while being trans-

ported. The letter V is used for equipment installed in a vehicle designed for functions other than carrying electronic equipment (such as a tank). The letter G is used for equipment capable of being used in two or more different types of ground installations. Equipment specifically designed to operate while being carried by a person is designated by the installation letter P. The letter U implies use in a combination of two or more general installation classes, such as ground, aircraft, and ship. The letter Z is for equipment in a combination of airborne installations, such as aircraft, drones, and guided missiles.

The equipment indicator letter (second column of Table 1.2) that designates radar is the letter P; but it is also used for beacons which function with a radar, electronic recognition and identification systems, and pulse-type navigation equipment.

Canadian, Australian, New Zealand, and United Kingdom electronic equipment can also be covered by the JETDS designations. For example, a block of numbers from 500 to 599 and 2500 to 2599 is reserved for Canadian use.

The radars used in the air traffic control system of the FAA utilize the following nomenclature:

ASR   airport surveillance radar

ARSR   air route surveillance radar

ASDE   airport surface detection equipment

TDWR   terminal doppler weather radar

The numeral following the letter designation indicates the particular radar model of that type.

Weather radars in use by the U.S. National Weather Service employ the designation WSR, which is not associated with the JETDS nomenclature. The number following the designation indicates the year in which the radar was put into service. When a letter follows the number, it indicates the letter-band designation. Thus, the WSR-74C is a C-band weather radar introduced in 1974.

GENERAL BOOKS ON RADAR

1. Skolnik, M. I.: "Introduction to Radar Systems," 2d ed., McGraw-Hill Book Co., New York, 1980.

2. Skolnik, M. I.: "Radar Applications," IEEE Press, New York, 1988.

3. Barton, D. K.: "Modern Radar System Analysis," Artech House, Norwood, Mass., 1988.

4. Barton, D. K.: "Radar System Analysis," originally published by Prentice-Hall in 1964 and republished by Artech House, Norwood, Mass., in 1977.

5. Nathanson, F.: "Radar Signal Processing and the Environment," McGraw-Hill Book Co., New York, 1969.

6. Brookner, E.: "Aspects of Modern Radar," Artech House, Norwood, Mass., 1988.

7. Brookner, E.: "Radar Technology," Artech House, Norwood, Mass., 1977.

8. Levanon, N.: "Radar Principles," John Wiley & Sons, New York, 1988.

9. Eaves, J. L., and E. K. Reedy: "Principles of Modern Radar," Van Nostrand Reinhold Company, 1987.

10. Berkowitz, R.: "Modern Radar," John Wiley & Sons, New York, 1965.
11. Reintjes, J. F., and G. T. Coate: "Principles of Radar," McGraw-Hill Book Co., New York, 1952.
12. Ridenour, L. N.: "Radar System Engineering," McGraw-Hill Book Co., New York, 1947.

CHAPTER 2
PREDICTION OF RADAR RANGE

Lamont V. Blake
Electronics Consultant

2.1 INTRODUCTION

The basic physics governing the prediction of radar maximum detection range, for a specified target under free-space conditions with detection limited by thermal noise, has been well understood since the earliest days of radar. The term *free space* implies (in the present context) that a spherical region of space, centered at the radar and extending to considerably beyond the target, is empty except for the radar and the target. (*Considerably* as used here can be precisely defined for specific radars, but a general definition would be lengthy and not very useful.) It also implies that the only radar-frequency electromagnetic waves detectable within this region, other than those emanating from the radar itself, are from natural thermal and quasi-thermal noise sources, as described in Sec. 2.5.

Although this condition is never fully realized, it is approximated for some radar situations. Under many non-free-space conditions and with radically non-thermal forms of background noise, the prediction problem is considerably more complicated. Complications not considered in early analyses also result from modification of the signal and noise relationship by the receiving-system circuitry (*signal processing*).

In this chapter the free-space equation will be presented, basic signal processing will be discussed, and some of the most important non-free-space environments will be considered. The effect of some common nonthermal types of noise will be considered. Although it will not be feasible to consider all the possible types of radar situations, the methods to be described will indicate the general nature of the necessary procedures for environments and conditions not specifically treated here. Some of the specialized types of radar, for which special analyses are required, are described in later chapters of this handbook.

Definitions. The radar range equation contains many parameters of the radar system and its environment, and some of them have definitions that are interdependent. As will be discussed in Sec. 2.3, some definitions contain an element of arbitrariness, and it is common for different authors to employ different definitions of some of the range-equation factors. Of course, when

generally accepted definitions do exist, they should be observed. But even more important, although some arbitrariness may be permissible for individual definitions, once a particular definition has been adopted for one of the range-equation factors, it will be found that definition of one or more of the other factors is no longer arbitrary.

As an example, for pulsed radar the definitions of pulse power and pulse length are highly arbitrary individually, but once a definition for either one of them has been adopted, the definition of the other is determined by the constraint that their product *must* equal the *pulse energy*. In this chapter, a set of definitions that are believed to conform to such rules of consistency, as well as to definitions adopted by standards organizations, will be presented.

Conventions. Because of the wide variability of propagation-path and other range-equation factors, certain *conventions* are necessary for predicting range under standard conditions when specific values of those factors are not known. A convention is a generally accepted *standard assumption*, which may never be encountered exactly in practice but which falls within the range of conditions that will be encountered, preferably somewhere near the middle of the range. An example is the conventional geophysical assumption, for calculating certain earth environment effects that depend on the earth's curvature, that the earth is a perfect sphere of radius 6370 km. The importance of conventions is that they provide a common basis for comparison of competing radar systems. To the extent that they are fairly representative of typical conditions, they also allow prediction of a realistic detection range. Commonly accepted conventions will be used in this chapter, and where needed conventions do not exist, appropriate ones will be suggested.

Range Prediction Philosophy. It is apparent from the foregoing discussion that a range prediction based on conventional assumptions will not necessarily be confirmed exactly by experimental results. This conclusion is further warranted by the statistical nature of the "noise" which is usually the limiting factor in the signal detection process. In other words, even if all the environmental factors were precisely known, a range prediction would not be likely to be verified exactly by the result of a single experiment. A statistical prediction refers to the average result of many trials. Therefore, radar range prediction is not an exact science. (In fact, the lesson of quantum mechanics seems to be that there is no such thing as an exact science in the strict sense.)

Nevertheless, calculations to predict radar range are useful. However inexact they may be on an absolute basis, they permit meaningful comparisons of the expected relative performance of competing system designs, and they indicate the relative range performance change to be expected if the radar parameters or environmental conditions are changed. They are therefore a powerful tool for the system designer. The predicted range is a figure of merit for a proposed radar system. It is not necessarily a complete one, since other factors such as target-position-measurement accuracy, data rate, reliability, serviceability, size, weight, and cost may also be important. Despite the inexactness of range predictions in the absolute sense, the error can be made small enough that the calculated range is a good indication of the performance to be expected under average environmental conditions. Section 2.10 is a more detailed discussion of prediction accuracy.

Attempts to evaluate range prediction factors accurately, to better than perhaps 1 dB, are sometimes disparaged on the grounds that some factors are un-

likely to be known with accuracy in operational situations and that hence it is useless to seek better accuracy for any factor. Although there is some basis for this viewpoint, the overall accuracy will be unnecessarily degraded if the accuracy of all the factors in the equation is deliberately reduced. Therefore it is recommended that range predictions be based on as careful an evaluation of all the factors as is possible. A goal of 0.1 dB accuracy is perhaps reasonable, although admittedly it may be impossible to evaluate all the factors in the equation with that degree of precision.

Historical Notes. Possibly the first comprehensive treatise on radar maximum-range prediction was that of Omberg and Norton,[1] published first as a U.S. Army Signal Corps report in 1943. It presents a fairly detailed range equation and contains information on evaluating some of the more problematical factors, such as multipath interference and minimum detectable signal, within the limitations of the then-available knowledge. The signal detection process was assumed to be based on visual observation of a cathode-ray-tube display. The antenna was assumed to "searchlight" the target. Statistical aspects of signal detection were not considered.

D. O. North,[2] in a classical report published with a military security classification in 1943, outlined the basic theory of a statistical treatment of signal detection. (This report was republished in *Proceedings of the IEEE*, but not until 1963.) He introduced the concepts that are now called *probability of detection* and *false-alarm probability*, and he clearly delineated the role of integration in the detection of pulse signals. This report also introduced the concept of the *matched filter*, a contribution for which it had achieved some recognition prior to 1963. But except for the matched-filter concept, its contributions to signal detection theory were virtually unrecognized by radar engineers generally until the report was republished 20 years after its first appearance.

In a famous report[3] first published in 1948 and republished in *IRE Transactions on Information Theory* in 1960, J. I. Marcum extensively developed the statistical theory of detection with the aid of machine computation, referencing North's report. He computed probabilities of detection as a function of a range parameter related to signal-to-noise ratio, for various numbers of pulses integrated and for various values of a false-alarm parameter which he designated *false-alarm number*. He employed this type of computation to study the effects of various amounts and kinds of integration, different detector (demodulator) types, losses incurred by "collapsing" one spatial coordinate on the radar display, and various other effects. His results are presented as curves for probability of detection as a function of the ratio of the actual range to that at which the signal-to-noise ratio is unity, on the assumption that the received-signal power is inversely proportional to the fourth power of the range. Since this proportionality holds only for a target in free space, application of Marcum's results is sometimes complicated by this mode of presentation.

Marcum considered only *steady* signals (target cross section not varying during the period of observation), and most of his results assume the use of a square-law detector. Robertson[4] has published exceptionally detailed and useful steady-signal results applicable to the linear-rectifier detector, which is the type of detector almost universally used. (The square-law-detector results are also useful because they differ very little from the linear-detector results.) Swerling extended Marcum's work to include the case of fluctuating signals.[5] His report was republished in *IRE Transactions on Information Theory* in 1960. Fehlner[6] recomputed Marcum's and Swerling's results and presented them in the more useful form of

curves with the signal-to-noise power ratio as the abscissa. The fluctuating-signal problem has subsequently been further treated by Kaplan,[7] Schwartz,[8] Heidbreder and Mitchell,[9] Bates,[10] and others.

Hall[11] published in 1956 a comprehensive paper on radar range prediction in which the concepts of probability of detection, false-alarm probability, the relative effects of predetection and postdetection integration, and the effects of scanning the antenna beam were considered. The range equation was formulated in terms of an ideal (matched-filter) utilization of the available received-signal power, with loss factors to account for departures from the ideal.

Blake[12] published an updating of the subject in 1961, applying recent advances in system-noise-temperature calculation, atmospheric absorption, plotting of coverage diagrams based on a realistic atmospheric refractive-index model, and multipath-interference calculation. This work was followed by Naval Research Laboratory (NRL) reports[13] and a book[14] in which further details were presented.

Contributions to the subject of range prediction have also been made by many others, far too numerous to mention by name. Only the major contributions can be recognized in this brief history. Special mention should be made, however, of the many contributions in two volumes (13 and 24) of the MIT Radiation Laboratory Series, edited by Kerr[15] and by Lawson and Uhlenbeck.[16] Much use is made in this chapter of results originally published in those volumes.

2.2   RANGE EQUATIONS

Radar Transmission Equation.   The following equation, in the form given in Kerr,[15] is called the *transmission equation* for monostatic radar (one in which the transmitter and receiver are colocated):

$$\frac{P_r}{P_t} = \frac{G_t G_r \sigma \lambda^2 F_t^2 F_r^2}{(4\pi)^3 R^4} \tag{2.1}$$

where P_r = received-signal power (at antenna terminals)
$\quad P_t$ = transmitted-signal power (at antenna terminals)
$\quad G_t$ = transmitting-antenna power gain
$\quad G_r$ = receiving-antenna power gain
$\quad \sigma$ = radar target cross section
$\quad \lambda$ = wavelength
$\quad F_t$ = pattern propagation factor for transmitting-antenna-to-target path
$\quad F_r$ = pattern propagation factor for target-to-receiving-antenna path
$\quad R$ = radar-to-target distance (range)

This equation is not identical to Kerr's; he assumes that the same antenna is used for transmission and reception, so that $G_t G_r$ becomes G^2 and $F_t^2 F_r^2$ becomes F^4. The only factors in the equation that require explanation are the pattern propagation factors F_t and F_r. The factor F_t is defined as the ratio, at the target position, of the field strength E to that which would exist at the same distance from the radar in free space and in the antenna beam maximum-gain direction, E_0. The factor F_r is analogously defined. These factors account for the possibility that the target is not in the beam maxima (G_t and G_r are the gains in the

maxima) and for any propagation gain or loss that would not occur in free space. The most common of these effects are absorption, diffraction and shadowing, certain types of refraction effects, and multipath interference.

For a target in free space and in the maxima of both the transmit and receive antenna patterns, $F_t = F_r = 1$. Detailed definitions of these and other range-equation factors are given in Secs. 2.3 to 2.7.

Maximum-Range Equation. Equation (2.1) is not a range equation as it stands, although it can be rewritten in the form

$$R = \left[\frac{P_t G_t G_r \sigma \lambda^2 F_t^2 F_r^2}{(4\pi)^3 P_r} \right]^{1/4} \tag{2.2}$$

This equation states that R is the range at which the received-echo power will be P_r if the transmitted power is P_t, target size σ, and so forth. It becomes a maximum-range equation by the simple expedient of attaching subscripts to P_r and R so that they become $P_{r,\text{min}}$ and R_{max}. That is, when the value of P_r in Eq. (2.2) is the minimum detectable value, the corresponding range is the maximum range of the radar.

However, this is a very rudimentary maximum-range equation, of limited usefulness. A first step toward a more useful equation is replacement of P_r by a more readily evaluated expression. This is done by first defining the signal-to-noise power ratio:

$$S/N = P_r/P_n \tag{2.3}$$

where P_n is the power level of the noise in the receiving system, which determines the minimum value of P_r that can be detected. This noise power, in turn, can be expressed in terms of a receiving-system noise temperature T_s:

$$P_n = kT_s B_n \tag{2.4}$$

where k is Boltzmann's constant (1.38×10^{-23} Ws/K) and B_n is the noise bandwidth of the receiver predetection filter, hertz. (These quantities are defined more completely in Secs. 2.3 and 2.5.[17]) Therefore,

$$P_r = (S/N) kT_s B_n \tag{2.5}$$

This expression can now be substituted for P_r in Eq. (2.2).

A further convenient modification is to redefine P_t as the transmitter power at the terminals of the transmitter, rather than [as in Eq. (2.1)] the usually somewhat smaller power that is actually delivered to the antenna terminals because of loss in the transmission line. This redefinition is desirable because when radar system designers or manufacturers specify a transmitter power, the actual transmitter output power is usually meant.

With this changed definition, P_t must be replaced by P_t/L_t, where L_t is a *loss factor* defined as the ratio of the transmitter output power to the power actually delivered to the antenna. (Therefore, $L_t \geqq 1$.)

It will later prove convenient to introduce additional loss factors similarly related to other factors in the range equation. These loss factors are multiplicative;

that is, if there are, for example, three loss factors L_1, L_2, and L_3, they can be represented by a single *system loss factor* $L = L_1L_2L_3$. The resulting maximum-range equation is

$$R_{max} = \left[\frac{P_tG_tG_r\sigma\lambda^2F_t^2F_r^2}{(4\pi)^3(S/N)_{min}kT_sB_nL}\right]^{1/4} \tag{2.6}$$

The quantities $(S/N)_{min}$ and T_s as here defined are to be evaluated at the antenna terminals, and that fact detracts from the utility of this form of the equation. As thus defined, $(S/N)_{min}$ is not independent of B_n, and the dependence is difficult to take into account in this formulation. If that dependence were ignored, this equation would imply that R_{max} is an inverse function of B_n; i.e., if all the other range-equation factors were held constant, R_{max} could be made as large as desired simply by making B_n sufficiently small. This is well known to be untrue. To remedy this difficulty, several factors must be considered. It is convenient to do this in terms of a particular transmitted waveform.

Pulse Radar Equation. Equation (2.6) does not specify the nature of the transmitted signal; it can be CW (continuous-wave), amplitude- or frequency-modulated, or pulsed. It is advantageous to modify this equation for the specific case of pulse radars and in so doing to remove the "bandwidth" difficulty encountered in using Eq. (2.6). Pulse radars are of course the most common type. As will be shown, although the equation thus modified will ostensibly be restricted to pulse radars, it can in fact be applied to other types of radar by appropriate reinterpretation of certain parameters.

D. O. North[2] demonstrated that the detectable signal-to-noise ratio $(S/N)_{min}$ will have its smallest possible value when the receiver bandwidth B_n has a particular (optimum) value and that this optimum value of B_n is inversely proportional to the pulse length τ. This implies that an equation can be written with pulse length in the numerator rather than with bandwidth in the denominator. North also showed that signal detectability is improved by *integrating* successive signal and noise samples in the receiver and that the detectability is a function of the total integrated signal energy. (The integration process is discussed in Sec. 2.4.) Finally, he showed that *when the receiver filter is matched to the pulse waveform*, the ratio of the received-pulse energy to the noise power spectral density at the output of the receiver filter is maximized *and is equal to the signal-to-noise power ratio at the antenna terminals*. The term *matched* in this context means, partially, that the filter bandwidth is optimum. The full meaning is that the filter transfer function is equal to the complex conjugate of the pulse spectrum.

Detectability Factor. An equation based on these facts can be derived by utilizing a parameter called *detectability factor*, defined by the Institute of Electrical and Electronics Engineers (IEEE)[18] as follows: "In pulsed radar, the ratio of single-pulse signal energy to noise power per unit bandwidth that provides stated probabilities of detection and false alarm, measured in the intermediate-frequency amplifier and using an intermediate-frequency filter matched to the single pulse, followed by optimum video integration." Deferring

for the moment discussion of the meaning of some aspects of this definition, it can be expressed mathematically as follows:*

$$D_0 = E_r/N_0 = P_r\tau/kT_s \tag{2.7}$$

where D_0 is the detectability factor, E_r is the received-pulse energy, and N_0 is the noise power per unit bandwidth, both measured at the output of the receiver filter (i.e., at the demodulator input terminals).

The next step in this reformulation of the range equation is to define a bandwidth correction factor C_B, to allow for the possibility that the receiver filter bandwidth B_n may not be optimum. This factor is defined by the following relationship:

$$(S/N)_{min}B_n = (S/N)_{min(0)}B_{n,\,opt}C_B = D_0B_{n,\,opt}C_B \tag{2.8}$$

where $B_{n,opt}$ is the optimum value of B_n. The factor C_B has been named the *bandwidth correction factor* because it was originally defined in terms of bandwidth optimization, but in actuality it is a *filter mismatch factor*, in the North matched-filter sense. As Eq. (2.8) implies, $C_B \cong 1$. Evaluation of C_B is discussed in Sec. 2.3.

The quantity $(S/N)_{min(0)}$ in Eq. (2.8) is the optimum-bandwidth (matched-filter) value of $(S/N)_{min}$, which North showed to be equal to D_0. It is this fact that allows the range equation to be written, as desired, in terms of the signal-to-noise ratio at the detector input terminals (filter output) rather than the ratio at the antenna terminals.

North deduced that $B_{n,opt} = 1/\tau$ exactly. As will be discussed later, some radar detection experiments with human observers have subsequently suggested that the constant of proportionality may not be exactly unity. However, North's analysis is theoretically correct for pulses of rectangular shape and for the definition to be given in Sec. 2.3 for the noise bandwidth B_n. For pulses of other shapes the pulse-length–bandwidth relationship is subject to the particular definition used for the pulse length. That definition is not an issue, of course, when the pulse shape is rectangular.

Based on that result, the range equation can be written with pulse length in the numerator by means of the following equivalence, in terms of the parameters of Eq. (2.8):

$$(S/N)_{min}B_n = D_0C_B/\tau \tag{2.9}$$

The expression of the left-hand side of Eq. (2.9), where it occurs in the denominator of Eq. (2.6), can now be replaced by the expression of the right-hand side. The result is the desired pulse radar form of the range equation:

$$R_{max} = \left[\frac{P_t\tau G_t G_r \sigma\lambda^2 F_t^2 F_r^2}{(4\pi)^3 kT_s D_0 C_B L}\right]^{1/4} \tag{2.10}$$

---

*In some of the literature it is stated that the matched-filter output signal-to-noise ratio is $2E_r/N_0$. That statement is based on defining peak signal power as the instantaneous value occurring not only at the peak of the output-pulse waveform but also at the peak of an RF cycle, where the instantaneous power is theoretically twice the average power. North's definition, based on the signal power averaged over an RF cycle, is consistent with the definition of noise power as the average over both the RF cycles and the random noise fluctuations.

A primary advantage of this formulation of the equation is that standard curves for the parameter D_0, as a function of the number of pulses integrated, are available, with the probabilities of detection and false alarm as parameters (Sec. 2.4). Calculation of these curves is necessarily done in terms of D_0, the signal-to-noise ratio at the demodulator input terminals.

The emphasis of this equation on the significance of the pulse *energy* (the product $P_t\tau$ in the numerator) is valuable to the system designer. It also provides a simple answer to the question of which pulse length to use in the range equation when the radar employs *pulse compression*, in which a coded pulse waveform of relatively long duration is transmitted and then "compressed" to a short pulse upon reception. The correct answer is deduced from the fact that the product $P_t\tau$ must equal the transmitted pulse energy. Therefore if the pulse power P_t is the power of the long (uncompressed) transmitted pulse, then τ must be the duration of that pulse.

A further advantage of this form of the equation, or more specifically of the definition of the detectability factor, is the indicated dependence of the radar detection range on the *integration* of successive pulses, if any, that takes place in the receiving system. Integration is discussed in Sec. 2.4.

Finally, as was mentioned earlier, this formulation of the range equation, although derived specifically in terms of pulse radar parameters, can be applied to CW radars and to radars that utilize forms of signal modulation other than pulses. Its application to these other radar types is accomplished by redefining the parameters τ and D_0. Details of this procedure are presented in Ref. 14, Chaps. 2 and 9.

Probabilistic Notation. It has been mentioned (Sec. 2.1) that the radar signal detection process is basically probabilistic or statistical in nature. This results from the nature of the noise voltage that is always present in the receiver circuits. This voltage is randomly varying or fluctuating, and when it is intermixed with a radar echo signal, it becomes impossible to tell with certainty whether a momentary increase of the receiver output is due to a signal or to a chance noise fluctuation. However, it is possible to define probabilities for these two possibilities and to discuss the detection process in terms of them in a quantitative manner. The probability that the signal, when present, will be detected is called the *probability of detection, P_d*, and the probability that a noise fluctuation will be mistaken for a signal is called the *false-alarm probability, P_{fa}*.

The notations R_{max}, $P_{r,min}$, and $(S/N)_{min}$ can then be replaced by more precise notation, using subscripts to denote the applicable values of P_d and P_{fa}. However, the *fa* subscript is ordinarily suppressed, though implied. Thus R_{50} can denote the range for 0.5 (i.e., 50 percent) probability of detection and some separately specified false-alarm probability.

If the target cross section σ fluctuates, this fluctuation will alter the signal-plus-noise statistics. As mentioned in Sec. 2.1, this problem has been analyzed by Swerling[5] and others.[6-10] Curves have been calculated that allow determining the appropriate value of D_0 for the fluctuating-signal case, for various probabilities of detection and false alarm (Sec. 2.4).

Automatic Detection. Detection* is said to be *automatic* if the decision concerning the presence or absence of a received signal is made by a purely

---

*A note on various meanings of the words *detect, detector*, and *detection* is desirable here. In radio usage, a detector has come to mean either a frequency converter (e.g., a superheterodyne first detector) or a demodulator (often the "second detector" of a superheterodyne receiver, which is usually a linear rectifier). Then, *detection* means the waveform modification produced by such a device. An *automatic*

physical device, without direct human intervention. Such a device, described by North,[2] establishes a threshold voltage level (for example, by means of a biased diode). If the processed (e.g., integrated) receiver output exceeds the threshold (as evidenced by diode current flow), some mechanism is actuated to indicate this fact in an unequivocal fashion. The mechanism could be the lighting of a light, the ringing of a bell, or more generally the setting of a bit to 1 in a binary data channel wherein a 0 corresponds to no signal. Various additional consequences may then automatically ensue. The analysis of radar detection can thus be regarded as a problem in statistical decision theory.

Bistatic Radar Equation. The foregoing equations assume that the transmitting and receiving antennas are at the same location (monostatic radar). A bistatic radar (Chap. 25) is one for which the two antennas are widely separated, so that the distance and/or the direction from the transmitting antenna to the target are not necessarily the same as the distance and/or direction from the receiving antenna to the target. Moreover, since the signal reflected from the target to the receiving antenna is not directly backscattered, as it is for monostatic radar, the target cross section is not usually the same (for a given target viewed in a given aspect by the transmitting antenna). A *bistatic radar cross section* σ_b is defined to apply for this situation. The symbol σ in the preceding equations implies a monostatic cross section. Range equations for bistatic radar are obtained from the foregoing monostatic equations by replacing the range R and the target cross section σ by the corresponding bistatic quantities. The bistatic equivalent of R is $\sqrt{R_t R_r}$, where R_t is the distance from the transmitting antenna to the target and R_r is the distance from the target to the receiving antenna.

Equations in Practical Units. The equations that have been given are valid when a consistent system of units is used, such as the rationalized meter-kilogram-second (mks) system. In many applications, however, it is convenient or necessary to employ "mixed" units. Moreover, it is usually more convenient to express the wavelength λ in terms of the equivalent frequency in megahertz. It is also desirable to combine all the numerical factors and the various unit-conversion factors into a single numerical constant. For a particular system of mixed units, the following equation is obtained from Eq. (2.10):

$$R_{\max} = 129.2 \left[\frac{P_{t(kW)} \tau_{\mu s} G_t G_r \sigma F_t^2 F_r^2}{f_{MHz}^2 T_s D_0 C_B L} \right]^{1/4} \qquad (2.11)$$

The subscript notation $R_{\max}$ is now meant to imply the range corresponding to specified detection and false-alarm probabilities. For this equation, the range is given in international nautical miles. (One international nautical mile is exactly 1852 m.) The target cross section σ is in square meters, transmitter power P_t in kilowatts, pulse length τ in microseconds, frequency f in megahertz, and system noise temperature T_s in kelvins. (All other quantities are dimensionless.)

If the range is desired in units other than nautical miles (all other units remaining the same), in place of the factor 129.2 the following numerical constants should be used in Eq. (2.11):

---

detector, however, is a decision-making device—for example, a device that replaces the human observer of a cathode-ray-tube display—and in that context *detection* is the making of a positive decision. In this chapter the meaning will ordinarily be evident from the context. Where confusion might otherwise result, the term *detection-decision device* may be used to denote an automatic detector.

Range units	Constant, Eq. (2.11)
Statute miles	148.7
Kilometers	239.3
Thousands of yards	261.7
Thousands of feet	785.0

A decibel-logarithmic form of the range equation is sometimes useful. An equation of that type, corresponding to Eq. (2.11), is readily obtained as the algebraic sum of the logarithms of the terms of that equation (with appropriate multipliers for the decibel format and for exponents), since it involves only multiplication, division, and exponentiation.

2.3 DEFINITION AND EVALUATION OF RANGE FACTORS

There is an element of arbitrariness in the definition of most of the factors of the radar range equation, and for some of them more than one definition is in common use. Since the definitions in these cases are arbitrary, one definition is in principle as good as another. However, once a definition has been chosen for one factor, there is no longer freedom of choice for one or more of the others. The factors are interdependent, and mutual compatibility is essential. A set of definitions that are believed to be mutually compatible will be given here. Also, information needed for evaluating these factors will be given insofar as is practicable. Certain range-equation factors that present special problems will be considered at greater length in subsequent sections of the chapter.

Transmitter Power and Pulse Length. The radar transmission equation, from which all the subsequent range equations are derived, is an equation for the dimensionless ratio P_t/P_r. Consequently, the most basic requirement on the definition of P_t is that it agree with the definition of P_r. For a CW radar, the power (averaged over an RF cycle) is constant, and there is no definition problem. For a pulse radar, both P_t and P_r are usually defined as the *pulse power*, which is the *average power during the pulse*. More precisely,

$$P_t = \frac{1}{\tau} \int_{T/2}^{-T/2} W(t)\, dt \qquad (2.12)$$

where $W(t)$ is the instantaneous power (a function of time, t). The definition of $W(t)$, however, excludes "spikes," "tails," and any other transients that are not useful for radar detection. The time interval T is the pulse period ($= 1/\text{PRF}$, where PRF is the pulse repetition frequency in pulses per second). Because of the exclusion of nonuseful portions of the waveform (as it exists at the transmitter output terminals), P_t as thus defined may be called the *effective* pulse power. It is often referred to as the *peak power*. However, peak power more properly signifies the power level at the peak of the pulse waveform (averaged over an RF cycle), and *pulse power* is more appropriate.

In the transmission equation, Eq. (2.1), P_t and P_r are the transmitted and received powers at the antenna terminals. As was mentioned in Sec. 2.2, P_t is now defined at the *transmitter output* terminals, and any loss between these terminals and the antenna input terminals must be expressed as a loss factor L_t.

The pulse power P_t and the pulse length τ must be defined so that their product is the pulse energy. Any definition of τ will produce this result if the same definition is used in Eq. (2.12). The customary definition, and the one recommended here, is the time duration between the half-power points of the envelope of the RF pulse (0.707-V points). For some purposes, such as analyzing the range resolution or accuracy, this arbitrary definition of the pulse length is not permissible. But the half-power definition is customary and acceptable for use in the range equation.

The range equation can be written with the product $P_t\tau$ replaced by the pulse energy E_t. The more detailed notation is used here because, for ordinary pulse radars, P_t and τ are usually given explicitly and E_t is not. However, the use of E_t in the equation does have the advantage of avoiding the problems of defining P_t and τ; and it is especially useful when complicated waveforms are transmitted.

If *coherent integration for a fixed integration time* is assumed, the equation can also be written with the transmitted *average* power in the numerator. For simple pulse radars, the average power is the product of pulse power, pulse length, and pulse repetition frequency. In this average-power formulation, the average power $\bar{P}_t$ is multiplied by the integration time t_i (assumed to be long compared with the interpulse period) to obtain the transmitted energy. Then the value of D_0 used is that which would apply if detection were based on observation of a single pulse. (See Sec. 2.4, Fig. 2.3.) The average-power formulation is especially useful for CW or pulse doppler radars.

Antenna Gain, Efficiency, and Loss Factor. The gains G_t and G_r are defined as the power gains of the antennas *in the maximum-gain direction*. If a target of interest is at an elevation angle not in the beam maxima, that fact is accounted for by the pattern propagation factors F_t and F_r, discussed in Sec. 2.6. The maximum power gain of an antenna is equal to its directivity (maximum directive gain) multiplied by its radiation efficiency.[19] The directivity D is defined in terms of the electric-field-strength pattern $E(\theta,\phi)$ by the expression

$$D = \frac{4\pi E^2_{max}}{\int_0^{2\pi} \int_0^\pi E^2(\theta,\phi)\, \sin\,\theta\, d\theta\, d\phi} \qquad (2.13)$$

where θ and ϕ are the angles of a spherical-coordinate system whose origin is at the antenna and E_{max} is the value of $E(\theta,\phi)$ in the maximum-gain direction.

The radiation efficiency of the transmitting antenna is the ratio of the power input at the antenna terminals to the power actually radiated (including minor-lobe radiation). In terms of the receiving antenna, the equivalent quantity is the ratio of the total signal power extracted from the incident field by the antenna, with a matched-load impedance, to the signal power actually delivered to a matched load. The reciprocal of the radiation efficiency is the antenna loss factor L_a, which plays a part in the calculation of antenna noise temperature (Sec. 2.5).

Measured antenna gains are usually power gains, whereas gains calculated

from pattern measurements or theory are directive gains. If the antenna gain figures supplied for use in the range equations of this chapter are of the latter type, they must be converted to power gains by dividing them by the appropriate loss factor. For many simple antennas the ohmic losses are negligible, and in those cases the power gain and the directive gain are virtually equal. However, this is by no means a safe assumption in the absence of specific knowledge. Array antennas in particular are likely to have significant ohmic losses in waveguides or coaxial lines used to distribute the power among the radiating elements.

If separate transmitting and receiving antennas are used and if their maximum gains occur at different elevation angles (this is a possible though not a common situation), appropriate correction can be made by means of the pattern factors $f_t(\theta)$ and $f_r(\theta)$, contained in the pattern propagation factors F_t and F_r (Sec. 2.6).

Antenna Beamwidth. This property of the antenna does not appear explicitly in the range equations, but it affects the range calculation through its effect on the number of pulses integrated when the antenna scans. The conventional definition is the angular width of the beam between the half-power points of the pattern. *Pattern* is used here in the usual antenna sense, for one-way transmission. It is not the two-way pattern of the radar echo signal from a stationary target as the antenna scans past it.

If a radar target, as viewed from the radar antenna, has an angular dimension that is appreciable compared with the beamwidth, the target cross section becomes a function of the beamwidth (see Sec. 2.8). For computing an effective value of σ in this case, in principle a special definition of beamwidth is needed (Ref. 15, p. 483). For practical work, however, the error that results from using the half-power beamwidth in this application is usually acceptable.

Target Cross Section. The definition of *radar target cross section* that applies for use in the foregoing radar range equations is given in Chap. 11, and the reader is referred to that chapter for a detailed discussion of the subject. Here mention will be made of a few aspects of the definitions that are of particular significance to the range prediction problem.

Targets can be classified as either *point targets* or *distributed targets*. A point target is one for which (1) the maximum transverse separation of significant scattering elements is small compared with the length of the arc intercepted by the antenna beam at the target range and (2) the maximum radial separation of scattering elements is small compared with the range extent of the pulse. At distance R from the antenna, the transverse dimension of the antenna beam is R times the angular beamwidth in radians. The range extent of the pulse is $c\tau/2$, where c is the speed of wave propagation in free space, 3×10^5 km/s, and τ is the pulse duration in seconds. Most of the targets for which range prediction is ordinarily of interest are point targets, e.g., aircraft at appreciable distances from the radar.

However, range predictions for distributed targets are sometimes wanted. The moon, for example, is a distributed target if the radar beamwidth is comparable to or less than 0.5° or if the pulse length is less than about 11.6 ms. A rainstorm is another example of a distributed target. Often, distributed targets are of interest because echoes from them (called *radar clutter*) tend to mask the echoes from the point targets whose detection is desired (see Sec. 2.8). Echoes from rain may be regarded as clutter when they interfere with detection of aircraft or other point targets, but they are themselves the signals of prime interest for weather radar.

The radar range equation is derived initially for a point target, and when that equation or the subsequent equations derived from it are used to predict the detection range for distributed targets, complications arise. In many cases, however, the point-target equation can be used for distributed targets by employing a suitable "effective" value of σ (Sec. 2.8).

The cross section of any nonspherical target is a function of the aspect angle from which it is viewed by the radar. It may also be a function of the polarization of the radar electromagnetic field. Therefore, in order to be wholly meaningful, a radar range prediction for a specific target, such as an aircraft, must stipulate the target aspect angle assumed and the polarization employed. Ordinarily, the nose aspect of an aircraft (approaching target) is of principal interest. The commonly used polarizations are horizontal, vertical, and circular. Tabulations of radar cross-section measurements of aircraft sometimes give nose, tail, and broadside values.

If the values are obtained from *dynamic* (moving-target) measurements, they are usually time averages of fluctuating values; otherwise they are *static* values for a particular aspect. Since the *instantaneous* cross section of a target is a function of the aspect angle, targets that are in motion involving random changes of aspect have cross sections that fluctuate randomly with time, as was mentioned in Sec. 2.2. This fluctuation must be taken into account in the calculation of probability of detection, as will be discussed in Sec. 2.4. When σ fluctuates, the value to be used in the range equation as formulated here is the time average.

Because of the wide variation of cross-section values of real targets, the range performance of a radar system is often stated for a particular target–cross-section assumption. A favorite value for many applications is 1 m^2. This represents the approximate cross section of a small aircraft, nose aspect, although the range for different "small" aircraft may be from less than 0.1 m^2 to more than 10 m^2. Radars are often performance-tested by using a metallic sphere, sometimes carried aloft by a free balloon, as the target because the cross section of a sphere can be accurately calculated and it does not vary with the aspect angle or the polarization.

A special definition problem arises when the target is large enough to be nonuniformly illuminated by the radar. A ship, for example, may be tall enough so that the pattern propagation factor F has different values from the waterline to the top of the mast. This matter is discussed in Ref. 15, p. 472 ff.

Wavelength (Frequency).   There is ordinarily no problem in definition or evaluation (measurement) of the frequency to be used in the radar range equation. However, some radars may use a very large transmission bandwidth, or they may change frequency on a pulse-to-pulse basis, so that a question can exist as to the frequency value to be used for predicting range. Also, the presence of f (or λ) in the range equations makes it clear that the range can be frequency-dependent, but the exact nature of the frequency dependence is not always obvious because other factors in the range equation are sometimes implicitly frequency-dependent. Therefore an analysis of how the range depends on frequency can be rather complicated, and the answer depends partly on what factors are regarded as frequency-dependent and which ones are held constant as the frequency is changed. For example, most antennas have gain that is strongly frequency-dependent, but some antenna types are virtually frequency-independent over a fairly wide frequency band.

Bandwidth and Matching Factors. The frequency-response width (bandwidth) of the receiver selective circuits appears explicitly in Eqs. (2.4) to (2.6), but it is an implicit factor in the other range equations as well, through the factor C_B. From Eq. (2.4) it is clear that B_n directly affects the noise level in the receiver output. In general it also affects the signal, but not necessarily in the same manner as the noise is affected, because the signal spectrum is not usually uniform. There is a value of B_n that optimizes the output signal-to-noise ratio, as indicated by Eq. (2.8), and this optimum bandwidth is inversely proportional to the pulse length τ. (This statement applies to pulse compression radars as well as to others if τ is, *in this context*, the *compressed* pulse length, since it is the compressed pulse that is amplified in the receiver. However, as has been emphasized in Sec. 2.2, in the numerator of the range equation the *uncompressed* pulse length must be used along with the actual radiated pulse power, P_t.)

Since the range equation (2.6) and those subsequently derived from it incorporate the assumption of Eq. (2.4) (namely, that the noise output power of the receiver is equal to $kT_sB_nG_0$), the definition of B_n—the *noise bandwidth*—must conform to that assumption. The resulting correct definition, due to North,[20] is

$$B_n = \frac{1}{G_0}\int_0^{\infty} G(f)df \qquad (2.14)$$

where G_0 is the gain at the nominal radar frequency and $G(f)$ is the frequency–power gain characteristic of the receiver predetection circuits, from antenna to detector.

The definition specifies $G(f)$ to be the response characteristic of the predetection circuits only. That is because for maximum postdetection signal-to-noise ratio the video bandwidth should be equal to at least half of the predetection bandwidth; and if it is of this width or wider, its exact width has little or no effect on signal detectability (Ref. 16, p. 211 ff.).

It is common practice to describe receiver bandwidth as the value between half-power points of the frequency-response curve. Fortunately, this value is usually very close to the true noise bandwidth, although the exact relationship of the two bandwidths depends on the particular shape of the frequency-response curve (Ref. 16, p. 177).

The bandwidth correction factor C_B in Eqs. (2.10) and (2.11) accounts for the fact that if B_n is not the optimum value, a value of signal-to-noise ratio larger than the optimum-bandwidth value D_0 is required. Therefore $C_B \geqq 1$. From data obtained in signal detection experiments during World War II at the Naval Research Laboratory, Haeff[21] devised the following empirical expression:

$$C_B = \frac{B_n\tau}{4\alpha}\left[1 + \frac{\alpha}{B_n\tau}\right]^2 \qquad (2.15)$$

where B_n is the noise bandwidth, τ is the pulse length, and α is the product of τ and $B_{n,\text{opt}}$ (optimum bandwidth). Figure 2.1 is a plot of Haeff's equation.

Actually, Haeff deduced from his experiments, as did North from theoretical analysis, that $B_{n,\text{opt}} = 1/\tau$; that is, $\alpha = 1$, for rectangular-shaped pulses. How-
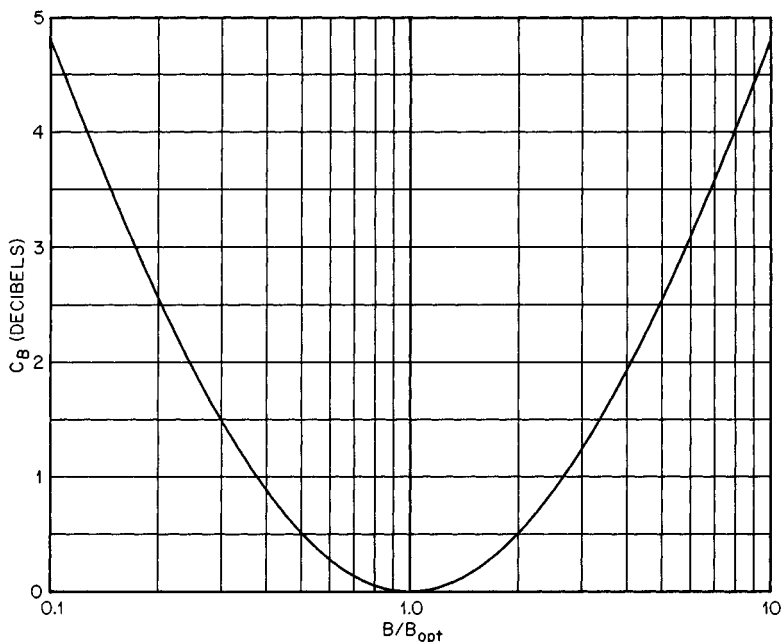
FIG. 2.1 Bandwidth correction factor C_B as a function of bandwidth B_n relative to optimum bandwidth B_{opt}; plotted from Haeff's empirical formula, Eq. (2.15).

ever, on the basis of experiments at the Massachusetts Institute of Technology (MIT) Radiation Laboratory conducted somewhat later (also with rectangular pulses), it was concluded (Ref. 16, p. 202) that $\alpha = 1.2$ for detection of signals by visual observation of cathode-ray-tube displays. Figure 2.2 is a plot of the Radiation Laboratory experimental results. The value 1.2 has subsequently been widely used for determining $B_{n,opt}$ in radar design and for computing C_B in radar range prediction.[11,12] However, North* has suggested that the $\alpha = 1.2$ figure may be based on a misinterpretation of the Radiation Laboratory data. Also, it has been noted† that the number of actual data points in Fig. 2.2 may be too few on which to base a good estimate of the optimum. Consequently, it is possible that the value of α for human observation of visual displays is much closer to 1 than was deduced by Lawson and Uhlenbeck. Fortunately, the minima of the curves (Fig. 2.2) are very broad and flat, and therefore the exact value of α does not make much difference for the usual range of values of $B_n\tau$.

The interpretation of C_B as a factor that accounts only for nonoptimum *width* of the predetection filter is permissible for simple pulse shapes and approximate results, but in principle it must also account for the complete amplitude-phase characteristic of the filter: its departure from a matched-filter characteristic. The

---

*In a private communication to the author in 1963.
†By M. I. Skolnik, editor of this handbook, in his review of this chapter.
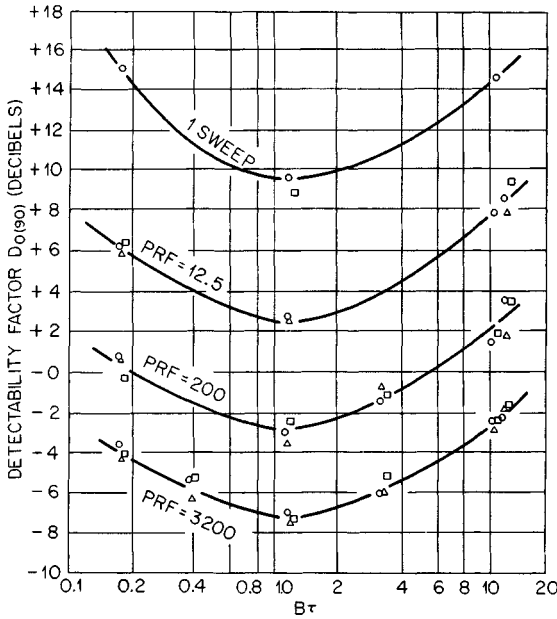
FIG. 2.2 Experimental results showing the effect of bandwidth (parameter $B\tau$) on 90 percent probability detectability factor $D_{0(90)}$, with pulse repetition frequency (PRF) as a parameter. The experiments were performed during World War II at the MIT Radiation Laboratory. *(From Ref. 15, Fig. 8.7.)*

matched-filter condition as stated by North[2] is that the receiver transfer characteristic must be the complex conjugate of the spectrum of the echo at the receiving-antenna terminals.

2.4 MINIMUM DETECTABLE SIGNAL-TO-NOISE RATIO

In Sec. 2.3, factors in the range equations were defined, and some information on how to evaluate them in typical cases was given. However, several very important factors were not covered because they are of sufficient importance to warrant more extensive treatment in separate sections. In this section and in Secs. 2.5 to 2.7, these additional factors will be discussed.

The quantities $P_{r,\text{min}}$, $(S/N)_{\text{min}}$, and D_0 are all related, as indicated in the development of Eqs. (2.3) to (2.9). Determination of the appropriate numbers to use for these quantities in their respective equations is a basic problem of radar range prediction. As will be seen, one of the problems is to define the meaning of *detectable*.

Integration of Signals. Detection of radar echo signals is usually (with some exceptions) accomplished by first integrating (e.g., adding) a sequence of received pulses and basing the detection decision on the resultant integrated signal voltage. Integrators that perform this operation will of course necessarily add noise as well as signals, but it is demonstrable that the ratio of the added signal voltages to the added noise voltages will be greater than the preintegration signal-to-noise ratio. Stated otherwise, the detectable signal-to-noise ratio evaluated ahead of an integrator will be smaller than when detection is performed by using single pulses.

There are many different methods of accomplishing integration. One method is the use of a feedback-loop delay line with a delay time equal to the interpulse period, so that signals (and noise) separated by exactly one pulse period will be directly added. Integration also occurs for visual detection by human radar operators if the phosphor of the cathode-ray-tube radar display (such as a PPI) has sufficient luminous persistence. In recent years, integration methods based on digital circuitry have become practical and are now perhaps the method of choice in many if not most cases.

The benefit of integration is a function of the number of pulses integrated. If integration is performed in the predetection stages of a receiver, ideally the addition of M equal-amplitude phase-coherent signal pulses will result in an output (integrated) pulse of voltage M times the single-pulse voltage. Adding M noise pulses, however, will result in an integrated noise pulse whose rms voltage is only $\sqrt{M}$ times as great as that of a single noise pulse if (as is true of ordinary receiver noise and many other types of noise) the added noise pulses are not phase-coherent. Therefore the signal-to-noise *voltage*-ratio improvement is, ideally, $M/\sqrt{M} = \sqrt{M}$. Consequently the signal-to-noise *power*-ratio improvement, and the reduction of the single-pulse minimum-detectable signal-to-noise power ratio, is equal to M.

Integration can also be performed after detection. In fact postdetection integration is used more commonly than is predetection integration, for reasons that will be explained, but the analysis of the resulting improvement is then much more complicated. After detection, the signals and the noise cannot be regarded as totally separate entities; the nonlinear process of detection produces an inseparable combination of signal and noise, so that one must then consider the comparison of signal-plus-noise to noise. As will be shown, the improvement that results from this type of integration is usually not as great as with ideal predetection integration of the same number of pulses. Nevertheless, postdetection integration produces worthwhile improvement. Moreover, "ideal" predetection integration is virtually unachievable because the echo fluctuation from most moving targets severely reduces the degree of phase coherence of successive received pulses. With rapidly fluctuating signals, in fact, postdetection integration will provide greater detectability improvement than does predetection integration, as discussed later in this section under the heading "Predetection Integration."

Number of Pulses Integrated. The number of pulses integrated is usually determined by the scanning speed of the antenna beam in conjunction with the antenna beamwidth in the plane of the scanning. The following equation can be used for calculating the number of pulses received between half-power-beam-width points for an azimuth-scanning radar:

$$M = \frac{\phi \, \overline{\text{PRF}}}{6 \, \overline{\text{RPM}} \cos \theta_e} \qquad (2.16)$$

where ϕ is the azimuth beamwidth, $\overline{\text{PRF}}$ is the radar pulse repetition frequency in hertz, $\overline{\text{RPM}}$ is the azimuthal scan rate in revolutions per minute, and θ_e is the target elevation angle. This formula strictly applies only if $\phi/\cos\theta_e$, the "effective" azimuth beamwidth, is less than 360°. (At values of θ_e for which $\phi/\cos\theta_e$ is greater than 360°, the number of pulses computed from this formula will obviously be meaningless. Practically, it is suggested that it be applied only for elevation angles such that $\phi/\cos\theta_e$ is less than about 90°.) This formula is based on the properties of spherical geometry. The formula also assumes that the beam maximum is tilted upward at the angle θ_e, but it can be applied with negligible error if θ_e is only approximately equal to the beam tilt angle.

The formula for the number of pulses within the half-power beamwidth for an azimuth- and elevation-scanning radar (which can be applied with minor modification to a radar scanning simultaneously in any two orthogonal angular directions) is

$$M = \frac{\phi\theta\ \overline{\text{PRF}}}{6\omega_v t_v \overline{\text{RPM}}\cos\theta_e} \qquad (2.17)$$

where ϕ and θ are the azimuth (horizontal-plane) and elevation (vertical-plane) beamwidths in *degrees*, θ_e is the target elevation angle, ω_v is the vertical scanning speed in degrees per second, and t_v is the vertical-scan period in seconds (including dead time if any). This formula should also be restricted to elevation angles for which $\phi/\cos\theta_e$ is less than about 90°. Here M is a function of the target elevation angle not only explicitly but also implicitly in that ω_v may be a function of θ_e.

Some modern radars, especially those capable of scanning by electronic means—i.e., without mechanical motion of the antenna—employ *step scanning*. In this method, the antenna beam is pointed in a fixed direction while a programmed number of pulses is radiated in that direction. Then the beam is shifted to a new direction, and the process is repeated. The number of pulses integrated in this scanning method is thus determined by the programming and not by the beamwidth. Also, the integrated pulses are then all of the same amplitude (except for the effect of target fluctuation), and so there is no *pattern loss* of the type described in Sec. 2.7. There is, however, a *statistical* loss if the target direction and the antenna beam maximum do not always coincide when the pulses are radiated.

Evaluation of Probabilities. As was mentioned in Sec. 2.2, if a threshold device is employed to make a decision as to the presence or absence of a signal in a background of noise, its performance can be described in terms of two probabilities: (1) the *probability of detection, P_d*, and (2) the *false-alarm probability, P_{fa}*. The threshold device is characterized by a value of receiver output voltage V_t (the *threshold*, equivalent to Marcum's[3] bias level), which, if exceeded, results in the decision report that a signal is present. If the threshold voltage is not exceeded at a particular instant, the detector reports "no signal."

There is always a definite probability that the threshold voltage will be exceeded when in fact no signal is present. The statistics of thermal random-noise voltage are such that there is a usually small but nonzero probability that it can attain a value at least equal to the saturation level of the receiver. (In the mathematical theory of thermal noise, there is a nonzero probability that it can attain any finite value, however large.) The probability that V_t is exceeded when no signal is present is the false-alarm probability. It is calculated from the equation

$$P_{fa} = \int_{V_t}^{\infty} p_n(v)dv \qquad (2.18)$$

where $p_n(v)$ is the probability density function of the noise. The probability of detection is given by the same expression, with the probability density function that of the signal-noise combination (usually called *signal-plus-noise*, but the "addition" is not necessarily linear):

$$P_d = \int_{V_t}^{\infty} p_{sn}(v)dv \qquad (2.19)$$

The signal-plus-noise probability density function $p_{sn}(v)$ depends on the signal-to-noise *ratio* as well as on the signal and noise statistics. Also, both p_n and p_{sn} are functions of the rectification law of the receiver detector and of any postdetection processing or circuit nonlinearities. Primarily, however, p_{sn} and therefore the probability of detection are functions of the signal-to-noise ratio. From Eq. (2.19), the variation of P_d with S/N can be determined. As would logically be assumed, it is a monotonic-increasing function of S/N for a given value of V_t. Similarly, the variation of P_{fa} as a function of V_t can be found from Eq. (2.18); it is a monotonic-decreasing function.

The method of applying these concepts to the prediction of radar range consists of four steps: (1) decide on a value of false-alarm probability that is acceptable (the typical procedure for making this decision will be described); (2) for this value of P_{fa}, find the required value of threshold voltage V_t, through Eq. (2.18); (3) decide on a desired value of P_d (in different circumstances, values ranging from below 0.5 up to as high as perhaps 0.99 may be selected); and (4) for this value of P_d and for the value of V_t found in step 2 find the required signal-to-noise ratio through Eq. (2.19). This requires evaluating the function $p_{sn}(v)$, taking into account the number of pulses integrated. Iteration is required, in this procedure, to find the value of D_0 corresponding to a specified probability of detection and number of pulses integrated. The value of D_0 thus found is the value to be used in the range equation [e.g., Eqs. (2.10) and (2.11)].

The process of finding the required value of D_0 for use in the range equation is greatly facilitated by curves that relate the number of pulses integrated to D_0 with P_{fa} and P_d as parameters. Many such curves have been published, and some representative ones are given as Figs. 2.3 through 2.7. The principal difficulty in computing them is determination of the probability density functions $p_n(v)$ and $p_{sn}(v)$ and in performing the requisite integrations. North[2] gives the exact functions that apply for single-pulse detection with a linear rectifier as detector and the approximations that apply when many pulses are integrated. The density functions appropriate to other situations, e.g., square-law detection and fluctuation of signals, are given by various authors.[3-10]

The decision as to the acceptable level of false-alarm probability is usually made in terms of a concept called *false-alarm time*, which will here be defined as the average time between false alarms. Other definitions are possible; Marcum[3] defines it as the time for which the probability of at least one false alarm is 0.5. However, the average time between false alarms seems a more practically useful concept. With it, for example, one can compute the average number of false alarms that will occur per hour, day, year, etc. With this definition, the false-alarm time is given by
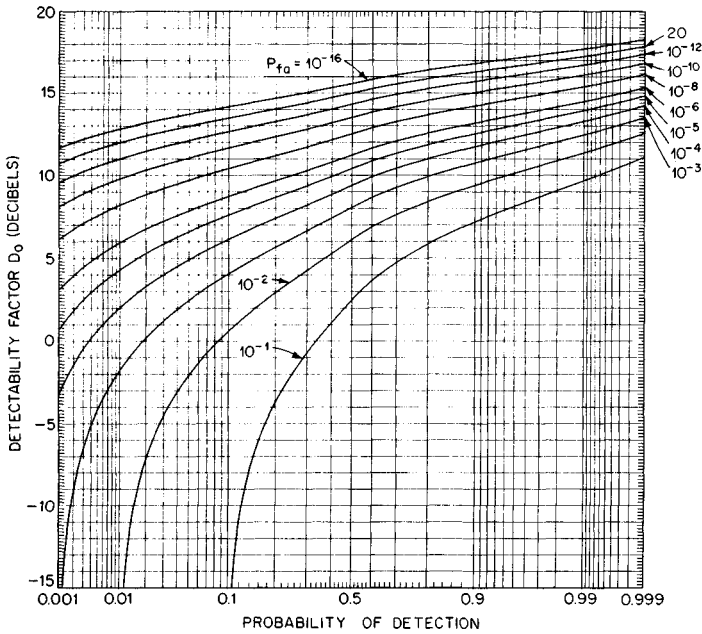
FIG. 2.3 Required signal-to-noise ratio (detectability factor) for a single-pulse, linear-detector, nonfluctuating target as a function of probability of detection with false-alarm probability (P_{fa}) as a parameter. *(From Ref. 13.)*

$$t_{fa} = \frac{M\,\tau}{P_{fa}} \qquad (2.20)$$

where M is the number of pulses integrated and τ is the pulse duration.

This formula assumes that the integrator output is sampled at time intervals equal to τ. If range gates are employed and M pulses are integrated, if the ON time of the gate t_g is equal to or greater than the pulse length τ, and if there is some fraction of the time δ when no gates are open (*dead time*, e.g., just before, during, and after the occurrence of the transmitter pulse), then the formula is

$$t_{fa} = \frac{M\,t_g}{P_{fa}(1-\delta)} \qquad (2.21)$$

These false-alarm-time formulas assume that the receiver predetection noise bandwidth B_n is equal to or greater than the reciprocal of the pulse length and that the postdetection (video) bandwidth is equal to or greater than $0.5\,B_n$ (as it usually is). These assumptions, usually met, amount to assuming that values of the noise voltage separated by the pulse duration are statistically independent; this independence occurs for times separated by $1/B_n$, sometimes called the Nyquist interval. Since ordinarily $B_n = 1/\tau$ and $t_g = \tau$, $1/B_n$ is sometimes used in place of τ or t_g in the false-alarm-time equations.

Marcum's false-alarm number n' is related to the false-alarm probability by the equation
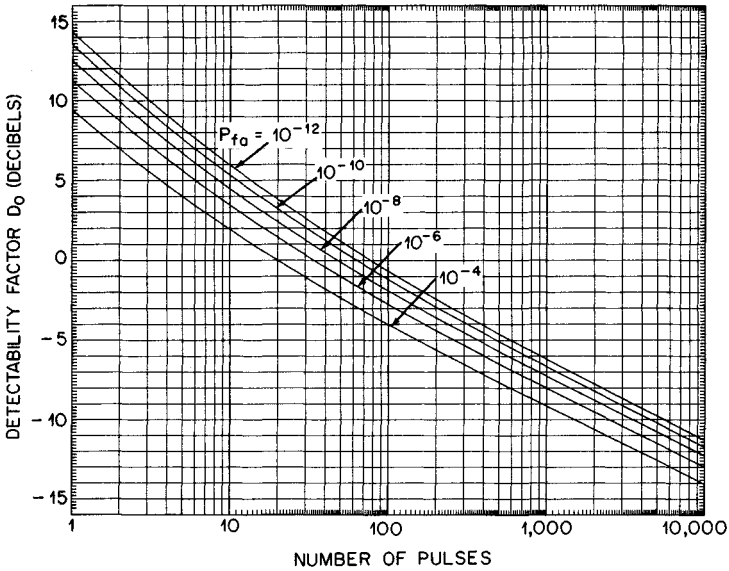
FIG. 2.4 Required signal-to-noise ratio (detectability factor) as a function of number of pulses noncoherently integrated, linear detector, nonfluctuating target, and 0.5 probability of detection. (*From Ref. 13.*)

$$1 - (1 - P_{fa})^{n'} = 0.5 \qquad (2.22a)$$

For the usual large values of n' that are of interest, a highly accurate approximate solution of this equation for P_{fa} is

$$P_{fa} = \frac{\log_e 0.5}{n'} = \frac{0.6931}{n'} \qquad (2.22b)$$

Fluctuating Target Cross Section. In general, the effect of fluctuation is to require higher signal-to-noise ratios for high probability of detection and lower values for low probability of detection than those required with nonfluctuating signals. Swerling has considered four cases, which differ in the assumed rate of fluctuation and the assumed statistical distribution of the cross section. The two assumed rates are (1) a relatively slow fluctuation, such that the values of σ for successive scans of the radar beam past the target are statistically independent but remain virtually constant from one pulse to the next, and (2) a relatively fast fluctuation, such that the values of σ are independent from pulse to pulse within one beamwidth of the scan (i.e., during the integration time).

The first of the two assumed distributions for the received-signal voltage is of the Rayleigh form,* which means that the target cross section σ has a probability density function given by

---

*The *Rayleigh* density function for a voltage v is

$$p(v) = \frac{2v}{r^2} e^{-v^2/r^2}$$

where r is the rms value of v.
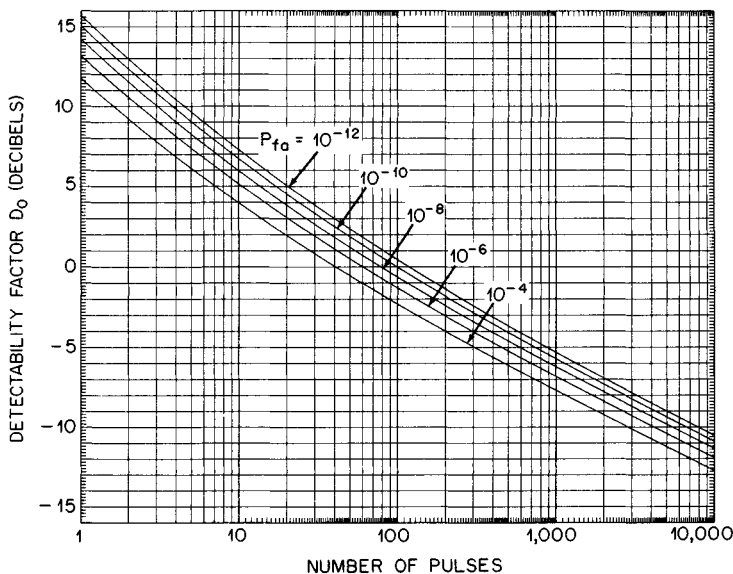
FIG. 2.5 Required signal-to-noise ratio (detectability factor) as a function of number of pulses noncoherently integrated, linear detector, nonfluctuating target, and 0.9 probability of detection. (*From Ref. 13.*)

$$p(\sigma) = \frac{1}{\overline{\sigma}}\, e^{-\sigma/\overline{\sigma}} \tag{2.23}$$

where $\overline{\sigma}$ is the average cross section. (This is a *negative-exponential* density function, but a target having this distribution is called a *Rayleigh target* because this distribution of σ produces a received signal *voltage* which is Rayleigh-distributed.) The second assumed cross-section density function is

$$p(\sigma) = \frac{4\sigma}{\overline{\sigma}^2}\, e^{-2\sigma/\overline{\sigma}} \tag{2.24}$$

The first distribution, Eq. (2.23), is observed when the target consists of many independent scattering elements of which no single one or few predominate. Many aircraft have approximately this characteristic at microwave frequencies, and large complicated targets are usually of this nature. (This result is predicted, for such targets, by the central limit theorem of probability theory.) The second distribution, Eq. (2.24), corresponds to that of a target having one main scattering element that predominates together with many smaller independent scattering elements. In summary, the cases considered by Swerling are as follows:

Case 1   Eq. (2.23), slow fluctuation
Case 2   Eq. (2.23), fast fluctuation
Case 3   Eq. (2.24), slow fluctuation
Case 4   Eq. (2.24), fast fluctuation
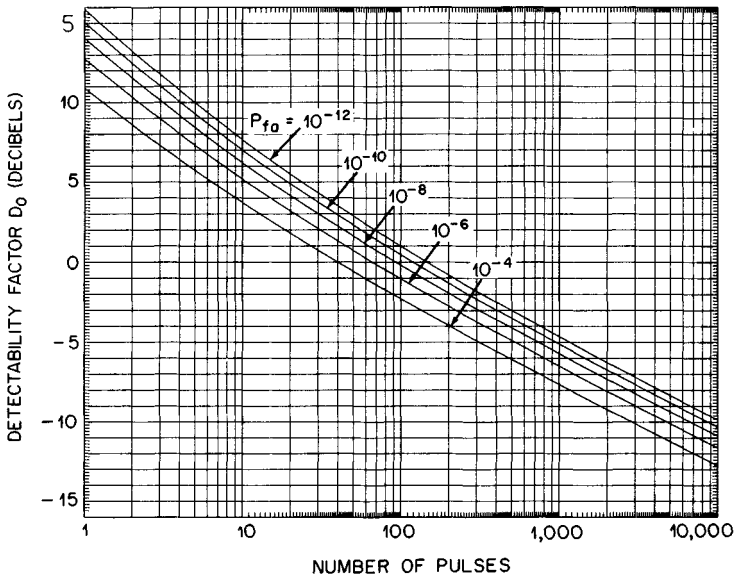
FIG. 2.6 Required signal-to-noise ratio (detectability factor) as a function of number of pulses noncoherently integrated, square-law detector, Swerling Case 1 fluctuating target, and 0.5 probability of detection. (*From Ref. 13.*)

The distribution of Eq. (2.24) is sometimes assumed for a small, rigid streamlined aircraft at the lower radar frequencies (e.g., below 1 GHz). Subsequent to Swerling's work, it has been found that many targets of the non-Rayleigh type are better represented by the so-called log-normal distribution, and analyses have been made for this case.[9]

Swerling's Case 1 is the one most often assumed when range prediction is to be made for a nonspecific fluctuating target. Results for this case are presented in Figs. 2.6 and 2.7. Curves for the other fluctuation cases and for additional values of detection probability are given in Refs. 13 and 14.

Detector Laws.   A linear detector is a rectifier which has the rectification characteristic

$$I_o = \alpha V_i \quad\quad V_i \geq 0 \left.\right\}$$
$$I_o = 0 \quad\quad\quad V_i < 0 \quad\quad\quad\quad (2.25)$$

where I_o is the instantaneous output current, V_i is the instantaneous input voltage, and α is a positive constant. Typical diodes approximate this law if V_i is larger than some very small value (e.g., a few millivolts). Such a diode is ordinarily used as the second detector of a superheterodyne radar receiver. Also, appreciable RF and IF gain usually precedes the second detector, so that the voltage applied to it is usually large enough (typically, an appreciable fraction of a volt) to ensure this "linear" type of operation.

A square-law detector is one that has the nonlinear characteristic

$$I_o = \alpha V_i^2 \quad\quad\quad\quad\quad\quad (2.26)$$
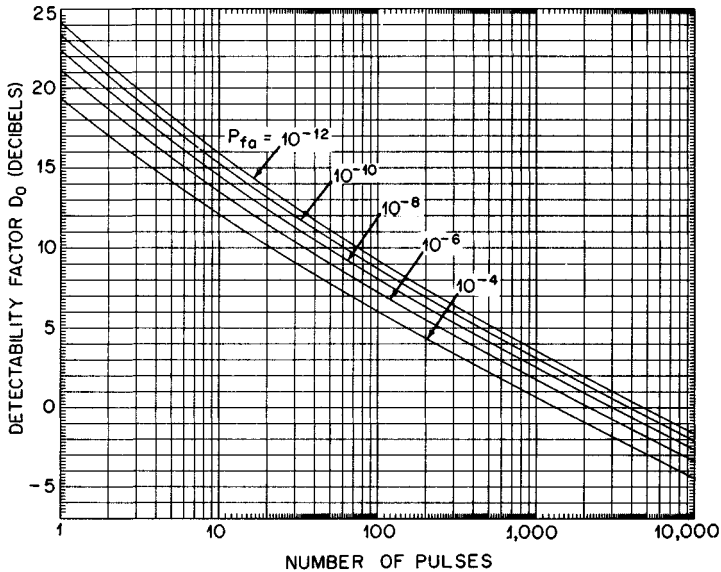
FIG. 2.7 Required signal-to-noise ratio (detectability factor) as a function of number of pulses noncoherently integrated, square-law detector, Swerling Case 1 fluctuating target, and 0.9 probability of detection. (*From Ref. 13.*)

Marcum[3] showed that a square-law detector is very slightly superior to a linear detector when many pulses are integrated, by about 0.2 dB. For a few pulses integrated, 10 or less, a linear detector is slightly superior—again, by about 0.2 dB or less. The mathematical analysis of probability of detection is somewhat more tractable when a square-law detector is assumed; this is probably its principal advantage.

The complexity of this matter is further compounded by the fact that, because of the statistics of the signal-noise superposition, in a *linear rectifier* there is a square-law relationship between the signal input voltage and the signal-plus-noise output voltage for small signal-to-noise ratios. This relationship becomes linear for large signal-to-noise ratios, as shown by Bennett,[22] North,[2] and Rice.[23] Because of this effect, it is sometimes erroneously thought that such a detector becomes square-law for small signal-to-noise *ratios*. But it is the input signal-plus-noise *voltage* V_i, and not the signal-to-noise ratio, that determines whether a diode rectifier is a linear or a square-law detector.

Curves for Visual Detection. The curves of Figs. 2.3 to 2.7 apply when the detection decision is based on an automatic threshold device as described. It is reasonable to suppose, however, that a human observer of a cathode-ray-tube display makes decisions in an analogous manner. That is, the equivalent of a threshold voltage (which would be a luminosity level for the PPI-scope type of display and a "pip-height" level for the A-scope display) exists somewhere in the observer's eye-brain system. This threshold, resulting in a particular false-alarm probability, is probably related to the observer's experience and personality: his or her innate cautiousness or daring. The probability of detection probably depends not only on the signal-to-noise ratio in relation to the threshold but on the observer's visual-mental acuity, alertness or fatigue,

and experience. Consequently, curves calculated for an automatic threshold decision device cannot be assumed to apply accurately to the performance of a human observer of a cathode-ray-tube display. But such an assumption does not give grossly erroneous results, and it is justifiable when experimental human-observer data are not available or are of questionable accuracy.

Curves based on actual experiments with human observers, analogous to those of Figs. 2.4 through 2.7, are given in Ref. 14, Chap. 2, along with further discussion of visual detection.

Other Detection Methods. The discussion and results that have been presented have assumed perfect postdetection (video) integration of pulses prior to decision by an automatic threshold device. The noise statistics have been implicitly assumed to be those of ordinary receiver noise of quasi-uniform spectral density and (before detection) of gaussian probability density. A great many other detection procedures and signal-noise statistics are possible. Some of them are discussed in Ref. 14, Chap. 2.*

Predetection Integration. The results depicted in Figs. 2.3 through 2.7 apply for perfect *postdetection* (video) integration of a specified number of pulses. It was shown by North[2] that under ideal conditions *predetection* integration results in the smallest possible detectability factor and that for ideal predetection integration of M pulses the following relation holds:

$$D_0(M) = D_0(1)/M \qquad (2.27)$$

That is, the minimum detectable signal-to-noise power ratio at the demodulator input terminals is improved, relative to single-pulse detection, by a factor exactly equal to the number of pulses integrated, M. For perfect postdetection integration, the improvement factor is generally less than M, asymptotically approaching $M^{1/2}$ as M becomes indefinitely large.

An exception occurs in the $M < 10$ region for fast-fluctuating targets and high probabilities of detection. For those circumstances the postdetection integration improvement factor can actually exceed M, and under the same circumstances predetection integration yields little or no improvement. The result of adding successive fast-fluctuating signals before detection, with virtually uncorrelated phases, is practically the same as that of adding noise voltages. Therefore there is virtually no integration improvement.

Predetection integration is also called *coherent integration*, because of its dependence on phase coherence of the integrated pulses, and postdetection integration is called *noncoherent integration*.

When integration is not perfect, as is always the case practically, if the value of D_0 used in the range equation is based on perfect integration, an imperfect-integration loss factor or factors must be included in the system loss factor L, as discussed in Sec. 2.7.

Although the full benefits of predetection integration are realizable only for nonfluctuating targets, some benefit can be achieved by predetection integration of a moderate number of *slowly* fluctuating targets. For such targets, the phase fluctuation from pulse to pulse is small. This type of integration is being employed to an increasing extent in modern systems when the utmost sensitivity is important and when fast fluctuation is not expected.

---

*Chapter written by Lowell W. Brooks, Technology Service Corporation, Salida, Colo.

Since radial target motion produces a frequency shift of the received-echo signal (doppler effect) which is proportional to the target's radial velocity, this shift must be taken into account if predetection integration is used. This is done in *pulse doppler radar* (Chap. 17).

Some radar systems that integrate many pulses utilize a combination of coherent and noncoherent integration when the phase stability of the received pulses is sufficient for some coherent integration but not great enough to allow coherent integration of the entire pulse train during the antenna on-target dwell time. If the total number of received pulses is N and M of them (with $M < N$) are coherently integrated and if the coherent integrator is followed by a noncoherent integrator, then (assuming an appropriate implementation and ideal integrations) the detectability factor will be

$$D_{0(M,N)} = D_0(N/M)/M \qquad (2.28)$$

where $D_{0(M,N)}$ means the detectability factor for the assumed combination of coherent and noncoherent integration and $D_0(N/M)$ is the detectability factor for noncoherent integration of N/M pulses with no coherent integration (e.g., a value read from curves such as those of Figs. 2.4 through 2.7). As an example, if a train of $N = 24$ pulses is received and each set of $M = 8$ pulses is predetection- (coherently) integrated and if the predetection integrator is followed by a postdetection (noncoherent) integrator, the integration process produces at best a combined detectability-factor improvement corresponding to that of coherent integration of 8 pulses and noncoherent integration of 3 pulses.

2.5 SYSTEM NOISE TEMPERATURE

The concept of a *noise temperature* is derived from Nyquist's theorem,[24] which states that if a resistive circuit element is at temperature T (kelvins) there will be generated in it an open-circuit thermal-noise voltage given by

$$V_n = \sqrt{4kTRB} \qquad \text{volts} \qquad (2.29)$$

where k is Boltzmann's constant (1.38054×10^{-23} Ws/K), R is the resistance in ohms, and B is the bandwidth, in hertz, within which the voltage is measured (that is, the passband of an infinite-impedance voltmeter). The absence of the frequency in this expression implies that the noise is white—that the spectrum is uniform and extends to infinitely high frequency. But this also implies infinite energy, an obvious impossibility, indicating that Eq. (2.29) is an approximation. A more exact expression, which has frequency dependence, must be used if the ratio f/T exceeds about 10^8, where f is the frequency in hertz and T is the kelvin temperature of the resistor. Thus Eq. (2.29) is sufficiently accurate at a frequency of 30 GHz if the temperature is at least 300 K. The more accurate equations are given in Ref. 14 and in radio astronomy texts.

Available Power, Gain, and Loss. As thus defined, V_n is the open-circuit voltage at the resistor terminals. If an external impedance-matched load of resistance $R_L = R$ is connected, the noise power delivered to it will be

$$P_n = kTB_n \qquad (2.30)$$

which does not depend on the value of R. This is of course also an approximation, but it is quite accurate at ordinary radar frequencies and temperatures. This matched-load power is called the *available* power.[17]

The concepts of *available power, available gain,* and its reciprocal, *available loss,* are assumed in all noise-temperature and noise-factor equations. These and other noise-temperature concepts are explained fully in Refs. 14, 17, and 25. Briefly, available power at an output port is that which would be delivered to a load that matches (in the complex-conjugate sense) the impedance of the source. Available gain of a two-port transducer or cascade of transducers is the ratio of the available power at the output port to that available from the source connected to the input port, with the stipulation that the available output power be measured with the actual input source (not necessarily impedance-matched) connected.

Noise Temperature. The usual noise that exists in a radar receiving system is partly of thermal origin and partly from other noise-generating processes. Most of these other processes produce noise which, within typical receiver bandwidths, has the same spectral and probabilistic nature as does thermal noise. Therefore it can all be lumped together and regarded as thermal noise. This is done, and the available-power level P_n is described by assigning to the noise a semifictitious "noise temperature" T_n, which is

$$T_n = P_n/(kB_n) \qquad (2.31)$$

This is of course simply an inversion of Eq. (2.30), except that T in Eq. (2.30) refers to an actual (thermodynamic) temperature. The temperature defined by Eq. (2.31) is semifictitious because of the nonthermal origin of some of the noise. When this temperature represents the available-noise-power output of the entire receiving system, it is commonly called the system noise temperature or operating noise temperature,[17] and it is then used to calculate the system noise power and signal-to-noise ratio, as in Eqs. (2.4) to (2.6).

The Referral Concept. A receiving system can be represented as a cascade of two-port transducers, preceded by a source (the antenna) and terminated by a load. [However, in the discussion of system noise temperature, only those parts of the receiver that precede the detector (demodulator) are of significance, for the noise level at that point determines the signal-to-noise ratio for signal-detection-calculation purposes.]

Noise may arise at any and all points in this cascade, so that the noise level changes from point to point. The important quantity is the *output* noise power P_{no}. For purposes of signal-to-noise calculation, however, it is convenient to *refer* this output noise to the system input terminals. This is done by defining the *system noise temperature* T_s so that it satisfies the relation

$$kT_sB_n = P_{no}/G_0 \qquad (2.32)$$

where G_0 is the overall-system available gain and B_n is the noise bandwidth of the system [Eq. (2.14)]. The output power P_{no} is thus "referred" to the system input (the antenna terminals), and T_s is actually the system *input* noise temperature. The product kT_sB_n is thus the system output noise power referred to the antenna terminals.

Each two-port transducer of the receiving-system cascade can be regarded as

having its own effective input noise temperature T_e, representing its intrinsic available output noise power referred to its own input terminals. Here *intrinsic* means the power that the transducer would generate with a *noise-free* input termination of the same impedance as the actual input termination. Transducer output power is referred to the input terminals by dividing the output power by the available gain of the transducer.

For an N-transducer cascade, the system input noise temperature (with the antenna terminals considered to be the system input terminals) is then given by

$$T_s = T_a + \sum_{i=1}^{N} \frac{T_{e(i)}}{G_i} \tag{2.33}$$

Here T_a is the antenna noise temperature, representing the available noise power at the antenna terminals, and G_i is the available gain of the system between its input terminals and the input terminals of the ith cascaded component. (By this definition $G_1 = 1$ always.)

To illustrate these principles concretely, this formula will here be applied to a two-transducer cascade representing a typical receiving system (Fig. 2.8). The first transducer is the transmission line that connects the antenna to the receiver input terminals, and the second transducer is the predetection portion of the receiver itself. (As mentioned above, for purposes of signal-noise analysis subsequent portions of the receiver are not considered.) If desired, a many-transducer receiving system could be further broken down, with a preamplifier and possibly other units considered as separate elements of the cascade.

For this system, if the receiving-transmission-line noise temperature is represented by T_r and its loss factor is L_r ($= 1/G_2$) and if the receiver effective input noise temperature is T_e, Eq. (2.33) becomes

$$T_s = T_a + T_r + L_r T_e \tag{2.34}$$

It now remains to discuss evaluation of T_a, T_r, L_r, and T_e.

Antenna Noise Temperature. Antenna noise is the result of (1) noise in the form of electromagnetic waves received by the antenna from external radiating sources and (2) thermal noise generated in the ohmic components (resistive conductors and imperfect insulators) of the antenna structure. The product kT_aB_n is the noise power available at the antenna terminals within the receiver bandwidth.

This noise temperature is dependent in a somewhat complicated way on the noise temperatures of various radiating sources within the receiving-antenna pattern, including its sidelobes and backlobes. The concept of noise temperature of a radiating source is based on Planck's law or on the Rayleigh-Jeans approximation to it, analogous to the relationship of resistor noise temperature to Nyquist's theorem.
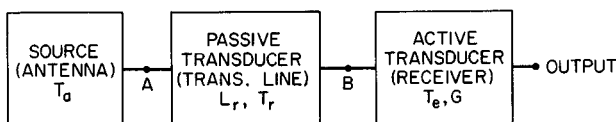
FIG. 2.8   Block diagram of a cascade receiving system.

The antenna noise temperature is not dependent on the antenna gain and beamwidth when a uniform-temperature source fills the beam. If the noise sources within the beam are of different temperatures, the resulting antenna temperature will be a solid-angle-weighted average of the source temperatures. The noise temperatures of most of the radiating sources that an antenna "sees" are frequency-dependent; therefore antenna temperature is a function of frequency. That is, antenna noise is not truly "white," but within any typical receiver passband it is virtually white.

In the microwave region, it is also a function of the antenna beam elevation angle, because in this region most of the "sky noise" is the result of atmospheric radiation. This radiation is related to atmospheric absorption, which is greater at low angles where the antenna beam sees a thicker slice of the lossy atmosphere than it does at higher angles.

Curves of antenna temperature for a lossless antenna are shown in Fig. 2.9, calculated for typical conditions.[14,25]
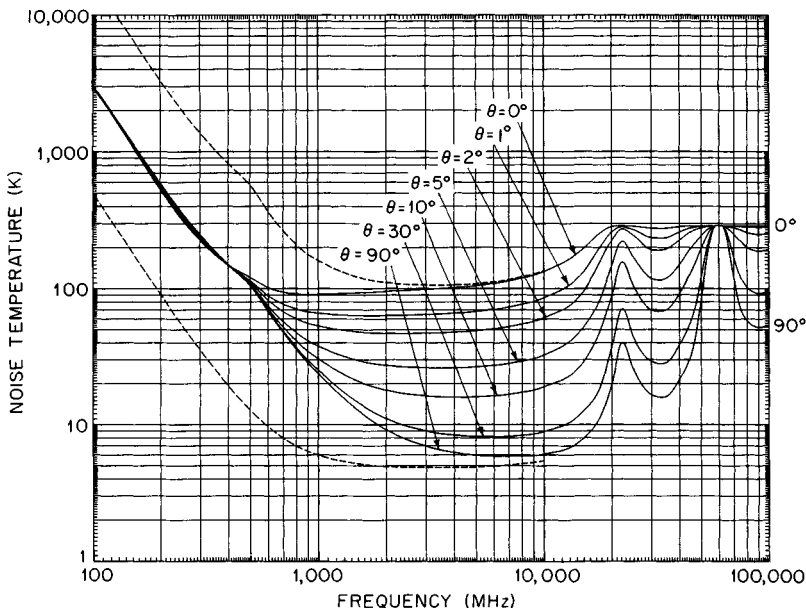
FIG. 2.9 Noise temperature of an idealized antenna (lossless, no earth-directed sidelobes) located at the earth's surface, as a function of frequency, for a number of beam elevation angles. Solid curves are for geometric-mean galactic temperature, sun noise 10 times quiet level, sun in unity-gain sidelobe, cool temperate-zone troposphere, 2.7 K cosmic blackbody radiation, and zero ground noise. The upper dashed curve is for maximum galactic noise (center of galaxy, narrow-beam antenna), sun noise 100 times quiet level, and zero elevation angle; other factors are the same as for the solid curves. The lower dashed curve is for minimum galactic noise, zero sun noise, 90° elevation angle. (The bump in the curves at about 500 MHz is due to the sun-noise characteristic. The curves for low elevation angles lie below those for high angles at frequencies below 400 MHz because of reduction of galactic noise by atmospheric absorption. The maxima at 22.2 and 60 GHz are due to water-vapor and oxygen absorption resonances; see Fig. 2.19.) (*From Ref. 13.*)

The curves of Fig. 2.9 apply to a lossless antenna that has no part of its pattern directed toward a warm earth. The lossless condition means that the curves represent only the noise received from external radiating sources. Therefore any thermal noise generated in the antenna must be added to the noise represented by these curves. In most practical cases, a ground noise-temperature component must also be added because part of the total antenna pattern is directed toward the ground. (This will be true because of sidelobes and backlobes even if the main beam is pointed upward.) But then also the sky-noise component given by Fig. 2.9 must be reduced somewhat because part of the total antenna pattern is not then directed at the sky. The reduction factor is $(1 - T_{ag}/T_{tg})$, where T_{ag} is the ground noise-temperature contribution to the total antenna temperature and T_{tg} is the effective noise temperature of the ground.

If α is the fraction of the solid-angle antenna power pattern subtended by the earth, then $T_{ag} = \alpha T_{tg}$. If the earth is perfectly absorptive (a thermodynamic blackbody), its effective noise temperature may be assumed to be approximately 290 K. A suggested conventional value for T_{ag} is 36 K, which would result if a 290 K earth were viewed over a π-steradian solid angle by sidelobes and backlobes averaging 0.5 gain (-3 dB). These sidelobes are typical of a "good" radar antenna but not one of the ultralow-noise variety.

Moreover, some practical antennas have appreciable ohmic loss, expressed by the loss factor L_a (Sec. 2.3). An additional thermal-noise contribution of amount $T_{ta}(1 - 1/L_a)$ then results, where T_{ta} is the thermal temperature of the lossy material of the antenna. However, the noise from external sources is then also reduced by the factor $1/L_a$. The total correction to the temperature values given by Fig. 2.9, to account for both ground-noise contribution and antenna loss, is then given by the following formula:

$$T_a = \frac{T_a'(1 - T_{ag}/T_{tg}) + T_{ag}}{L_a} + T_{ta}(1 - 1/L_a) \qquad (2.35a)$$

where T_a' is the temperature given by Fig. 2.9. For $T_{ag} = 36$ K and $T_{tg} = T_{ta} = 290$ K, this becomes

$$T_a = \frac{0.876\, T_a' - 254}{L_a} + 290 \qquad (2.35b)$$

and if $L_a = 1$ (lossless antenna), it further simplifies to

$$T_a = 0.876\, T_a' + 36 \qquad (2.35c)$$

Transmission-Line Noise Temperature. Dicke[26] has shown that if a passive transducer of noise bandwidth B_n connected in a cascade system is at a thermal temperature T_t and if its available loss factor is L, the thermal-noise power available at its *output* terminals is

$$P_{no} = kT_t B_n(1 - 1/L) \qquad (2.36)$$

A transmission line is a passive transducer. From Eq. (2.36) together with Eq. (2.31) and the definition of *input* temperature, it is deduced that the input noise temperature of a receiving transmission line of thermal temperature T_{tr} and loss factor L_r is

$$T_r = T_{tr}(L_r - 1) \tag{2.37}$$

(In this referral operation, multiplication by loss factor is equivalent to division by gain.) The receiving-transmission-line loss factor L_r is defined in terms of a CW signal received at the nominal radar frequency by the antenna. It is the ratio of the signal power available at the antenna terminals to that available at the receiver input terminals (points A and B, Fig. 2.8). A suggested conventional value for T_{tr} is 290 K.

Receiver Noise Temperature. The effective input noise temperature of the receiver T_e may sometimes be given directly by the manufacturer or the designer. In other cases, the *noise figure* F_n may be given. The relationship between the noise figure and the effective input noise temperature of the receiver or, in fact, of any transducer is given by[17]

$$T_e = T_0(F_n - 1) \tag{2.38}$$

where T_0 is, by convention, 290 K. In this formula F_n is a power ratio, not the decibel value that is usually given.

This formula is applicable to a *single-response* receiver (one for which a single RF input frequency corresponds to only one output or IF frequency and vice versa). Methods of computing noise temperatures when a double- or multiple-response receiver is used (e.g., for a superheterodyne receiver without preselection) are described in Refs. 17 and 25. Single-response receivers are ordinarily used in radar systems.

It is worth mentioning a point that has been well emphasized in the specialized literature of radio noise but is nevertheless easily overlooked. A receiver noise-temperature or noise-figure rating applies when a particular terminating impedance is connected at the receiver input. If this impedance changes, the noise temperature changes. Therefore, in principle, when a noise-temperature rating is quoted for a receiver, the source impedance should be specified, especially since the optimum (lowest) noise temperature does not necessarily occur when impedances are matched. However, when a receiver noise temperature is quoted without this impedance specification, it is presumable that the optimum source impedance is implied.

2.6 PATTERN PROPAGATION FACTOR

The pattern propagation factors F_t and F_r in the range equation account for the facts that (1) the target may not be in the beam maximum of the vertical-plane antenna pattern and (2) non-free-space wave propagation may occur. This single factor, rather than two separate factors, is designed to account for both of those effects. This is necessary because they become inextricably intertwined in the calculation of multipath interference, which is the most important non-free-space effect.

As will be seen, this effect can result in very considerable increase or decrease of the radar detection range compared with the free-space range. In this chapter, the basic ideas of pattern-propagation-factor calculation and some typical multipath-interference results will be presented. Additional details are given in Ref. 14, Chap. 6, and in Ref. 15.

CHAPTER 3
RECEIVERS

John W. Taylor, Jr.
Westinghouse Electric Corporation

3.1 THE CONFIGURATION OF A RADAR RECEIVER

The function of a radar receiver is to amplify the echoes of the radar transmission and to filter them in a manner that will provide the maximum discrimination between desired echoes and undesired interference. The interference comprises not only the noise generated in the radar receiver but also energy received from galactic sources, neighboring radars and communication equipment, and possibly jammers. The portion of the radar's own radiated energy that is scattered by undesired targets (such as rain, snow, birds, insects, atmospheric perturbations, and chaff) may also be classed as interference. Where airborne radars are used for altimeters or mapping, other aircraft are undesired targets, and the ground is the desired target. More commonly, radars are intended for detection of aircraft, ships, surface vehicles, or personnel, and the reflection from sea or ground is classified as clutter interference.

The boundaries of the radar receiver must be defined arbitrarily. This chapter will consider those elements shown in Fig. 3.1 as the receiver. The input signal is derived from the duplexer, which permits a single antenna to be shared between transmitter and receiver. Some radar antennas include low-noise amplifiers prior to forming the receive beams; although these are generally considered to be antenna rather than receiver elements, they will be discussed in this chapter.

The receiver filters the signal to separate desired echoes from interference in many ways, but some elements of such processing are covered by other chapters because of the depth of treatment required:

Chapter 14 describes continuous-wave (CW) and FM-CW radars; the discussion here will be confined to receivers for pulse radars, the dominant form. Low-pulse-repetition-frequency (PRF) pulse radars transmit a burst of energy and listen for echoes between transmissions. Their outstanding advantage is that neither leakage from the transmitter nor the very strong echoes from close-range clutter occur at the same instant as reception of weak echoes from long-range targets. The delay of the pulsed echo also provides an instantaneous measurement of range.

Chapters 15 to 17 relate to discrimination of desired targets from interference on the basis of velocity or the change in phase from one pulse to the next; the
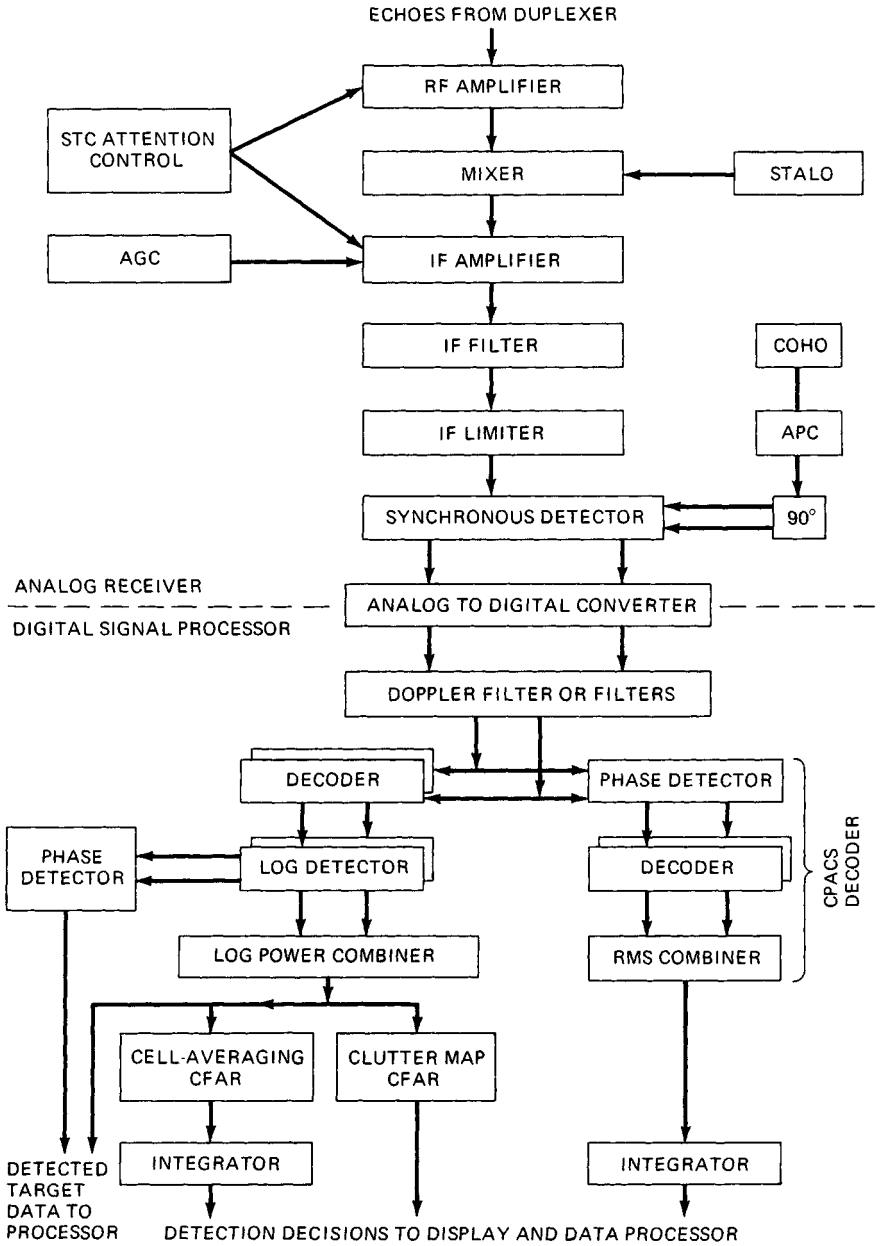
FIG. 3.1   General configuration of a radar receiver.

receiver discussed here serves only to provide the individual pulse signals in proper form for such doppler filtering. Section 3.12 will discuss data distortions in the synchronous detector or analog-to-digital (A/D) converter which affect these doppler filters.

Chapter 10 deals with pulse compression, and only brief mention of its ability to aid in the discrimination process will be included here. Physically, decoding such echoes may occur as part of the intermediate-frequency (IF) filter, typically with surface acoustic-wave (SAW) devices or with digital correlators, either preceding or following doppler filtering. In Fig. 3.1, pulse compression is provided by the cascade combination of IF filter and digital decoders (correlators) after doppler filtering.

Similarly, Chaps. 18 and 20 describe tracking radars and height finding radars, but the peculiarities of the receivers required to perform these functions will be mentioned briefly.

The purpose of Fig. 3.1 is to illustrate the usual sequence of processing functions that may occur in any radar receiver and the variety of possible outputs, although no radar receiver will include all these functions or provide all these outputs.

Virtually all radar receivers operate on the superheterodyne principle shown in Fig. 3.1. The echo, after modest amplification, is shifted to an intermediate frequency by mixing with a local-oscillator (LO) frequency. More than one conversion step may be necessary to reach the final IF, generally between 0.1 and 100 MHz, without encountering serious image- or spurious-frequency problems in the mixing process. Not only is amplification at IF less costly and more stable than at microwave frequency, but the wider percentage bandwidth occupied by the desired echo simplifies the filtering operation. In addition, the superheterodyne receiver can vary the LO frequency to follow any desired tuning variation of the transmitter without disturbing the filtering at IF. These advantages have been sufficiently powerful that competitive forms of receivers have virtually disappeared; only the superheterodyne receiver will be discussed in any detail.

Other receiver types include the superregenerative, crystal video, and tuned radio frequency (TRF). The superregenerative receiver is sometimes employed in radar-beacon applications because a single tube may function as both transmitter and receiver and because simplicity and compactness are more important than superior sensitivity. The crystal video receiver also is simple but of poor sensitivity. The TRF receiver uses only RF and video amplification; although its noise temperature may be low, its sensitivity is poor because optimum-bandwidth filtering of the usual radar echo spectrum is generally impractical to achieve. Only for radars that radiate a relatively wide-percentage-bandwidth signal is filtering practical.

3.2 NOISE AND DYNAMIC-RANGE CONSIDERATIONS

Receivers generate internal noise which masks weak echoes being received from the radar transmissions. This noise is one of the fundamental limitations on the radar range, and for this reason the subject has been treated in Sec. 2.5. The analysis of radar sensitivity is facilitated if the noise contribution of each element of

the system is expressed as a *noise temperature* rather than as a *noise factor* or *noise figure*; these terms are defined and interrelated in Sec. 2.5.

The noise temperature of the radar receiver has been reduced to the point that it no longer represents a dominant influence in choosing between available alternatives. It is a paradox that a noise parameter is usually the first characteristic specified for a radar receiver, yet few radars employ the lowest-noise receiver available because such a choice represents too great a sacrifice in some other characteristic.

Cost is rarely a consideration in rejecting a lower-noise alternative. A reduction of requirements for antenna gain or transmitter power invariably produces cost savings far in excess of any added cost of a lower-noise receiver. More vital performance characteristics generally dictate the choice of receiver front end:

1. Dynamic range and susceptibility to overload
2. Instantaneous bandwidth and tuning range
3. Phase and amplitude stability
4. Cooling requirements

A direct compromise must be made between the noise temperature and the dynamic range of a receiver. The introduction of an RF amplifier in front of the mixer necessarily involves raising the system noise level at the mixer to make the noise contribution of the mixer itself insignificant. Even if the RF amplifier itself has more than adequate dynamic range, the mixer dynamic range has been compromised, as indicated below:

Ratio of front-end noise to mixer noise	6 dB	10 dB	13.3 dB
Sacrifice in mixer dynamic range	7 dB	10.4 dB	13.5 dB
Degradation of system noise temperature by mixer noise	1 dB	0.4 dB	0.2 dB

Definitions.   Section 2.5 defined the noise parameters of a receiver in simple terms. Dynamic range, which represents the range of signal strength over which the receiver will perform as expected, is more difficult to define. It requires the specification of three parameters:

1. *Minimum signal of interest:* This is usually defined as the input signal that produces unity signal-to-noise ratio (*SNR*) at the receiver output. Occasionally, a minimum-detectable-signal definition is employed.

2. *Allowable deviation from expected characteristic:* The maximum signal is one that will cause some deviation from expected performance. Linear receivers usually specify a 1 dB decrease in incremental gain (the slope of the output-versus-input curve). Limiting or logarithmic receivers must define an allowable error in their outputs. Gain-controlled receivers must distinguish between instantaneous dynamic range and that achieved partly as a result of programmed gain variation.

3. *Type of signal:* Three types of signals are of general interest in determining dynamic-range requirements: distributed targets, point targets, and wideband-noise jamming. If the radar employs a phase-coded signal, the elements of the receiver preceding the decoder will not restrict the dynamic range of a point target as severely as they will distributed clutter; the bandwidth-time product of the coded pulse indicates the added dynamic range that the decoder will extract from point targets. Conversely, if the radar incorporates an excessively wide-

bandwidth RF amplifier, its dynamic range to wideband-noise interference may be severely restricted.

When low-noise amplifiers (LNAs) are included in the antenna, prior to forming the receive beams, the sidelobe levels achieved are dependent upon the degree to which gain and phase characteristics are similar in all LNAs. Dynamic range has an exaggerated importance in such configurations because matching nonlinear characteristics is impractical. The effect of strong interference [mountain clutter, other radar pulses, or electronic countermeasures (ECM)], entering through the sidelobes, will be exaggerated if it exceeds the dynamic range of the LNAs because sidelobes will be degraded. The LNAs are wideband devices, vulnerable to interference over the entire radar operating band and often outside this band; although off-frequency interference is filtered in subsequent stages of the receiver, strong interference signals can cause clutter echoes in the LNA to be distorted, degrading the effectiveness of doppler filtering and creating false alarms. This phenomenon is difficult to isolate as the cause of false alarms in such radars owing to the nonrepetitive character of many sources of interference.

Evaluation. A thorough evaluation of all elements of the receiver is necessary to prevent unanticipated degradation of noise temperature or dynamic range. Inadequate dynamic range makes the radar receiver vulnerable to interference, which can cause saturation or overload, masking or hiding the desired echoes. A tabular format for such a computation (a typical example of which is shown as Table 3.1) will permit those components that contribute significant noise or restrict the dynamic range to be quickly identified. "Typical" values are included in the table for purposes of illustration.

One caution is required in using Table 3.1. The dynamic range of each component is computed by comparing the maximum signal and system noise levels at the output of each component. The assumption inherent in this method is that all filtering (bandwidth reduction and decoding) by this component is accomplished prior to any saturation. It is important to treat those stages of the receiver that provide significant filtering as separate elements; if multiple stages are lumped into a single filter, this assumption may be grossly in error.

3.3 BANDWIDTH CONSIDERATIONS

Definitions. The instantaneous bandwidth of a component is the frequency band over which the component can simultaneously amplify two or more signals to within a specified gain (and sometimes phase) tolerance. The tuning range is the frequency band over which the component may operate without degrading the specified performance if suitable electrical or mechanical controls are adjusted.

Important Characteristics. The environment in which a radar must operate includes many sources of electromagnetic radiation, which can mask the relatively weak echoes from its own transmission. The susceptibility to such interference is determined by the ability of the receiver to suppress the interfering frequencies if the sources have narrow bandwidth or to recover quickly if they are more like impulses in character. One must be concerned with the response of the receiver in both frequency and time domains.

Generally, the critical response is determined in the IF portion of the receiver; this will be discussed in Sec. 3.7. However, one cannot ignore the RF portion of the receiver merely by making it have wide bandwidth. Section 3.2 discussed

TABLE 3.1   Noise and Dynamic-Range Characteristics

		Antenna	Trans-line	RF amplifier	Mixer	Filter	Log detector
Noise temperature of component	K	/////	/////	520	1300	300	24K
Gain of component *	dB	/////	-1.0	25	-6	15	
Total gain to input	dB	/////	/////	-1.0	24	18	33
Noise-temperature contribution referred to antenna	System 838 K	80	75	660	6	5	12
	29.3 dB K						
Overall RX bandwidth	63.0 dBHz						
	92.3						
Boltzmann's constant	-198.6						
Narrowband noise level [†]	-106.3 dBm	-106	-107	-82	-88	73	(-73)
Maximum signal capability [†]	dBm	/////	/////	-5	-16	+5	(+7)
Dynamic range to distributed targets [†]	dB	/////	/////	77	72	78	(80)
Bandwidth x time of point target [†]	dB	11	11	11	11	0	0
Dynamic range to point target [†]	dB	/////	/////	88	83	78	(80)
Bandwidth of receiver [†]	MHz	/////	/////	200	100	2	2
Ratio to overall receiver bandwidth [†]		/////	/////	100	50	1	1
Wideband-noise vulnerability [†]	dB	/////	/////	20	17	0	0
Dynamic range to wideband noise [†]	dB	/////	/////	57	55	78	(80)

* CW output-CW input on the center frequency, not coded pulse.

[†] At the output terminal of the designated component except where indicated by parentheses (at the input terminal of a nonlinear device).

how excessively wide bandwidth can penalize dynamic range if the interference is wideband noise. Even more likely is an out-of-band source of strong interference (TV station or microwave communication link) which, if allowed to reach this point, can either overload the mixer or be converted to IF by one of the spurious responses of the mixer.

Ideal mixers in a superheterodyne receiver act as multipliers, producing an output proportional to the product of the two input signals. Except for the effect of nonlinearities and unbalance, these mixers produce only two output frequencies, equal to the sum and the difference of the two input frequencies. Product mixers, although common at intermediate frequencies, are not generally available for RF

conversion down to IF, and diode mixers are most commonly employed. The frequency-conversion properties of the diode are produced by its nonlinear characteristics. If its characteristic is defined by a power series, only the square-law term produces the desired conversion. The other terms produce spurious products, which represent an unwanted ability to convert off-frequency signals to the IF of the receiver. The efficiency of conversion of these unwanted frequencies, except for the image frequency, is sufficiently poor that the system noise temperature is not significantly degraded, but the mixer is vulnerable to strong out-of-band interference. The best radar receiver is one with the narrowest RF instantaneous bandwidth commensurate with the radiated spectrum and hardware limitations, and with good frequency and impulse responses.

A wide tuning range provides a flexibility to escape interference, but if the interference is intentional (jamming), change in frequency on a pulse-to-pulse basis may be required. Such frequency agility can be achieved by using switchable microwave filters or electronically tuned yttrium iron garnet (YIG) filters to restrict the instantaneous bandwidth. Each involves some insertion loss, another sacrifice in noise temperature to achieve more vital objectives.

3.4 RECEIVER FRONT END

Configuration. The radar *front end* consists of a bandpass filter or bandpass amplifier followed by a downconverter. The radar frequency is downconverted to an intermediate frequency, where filters with suitable bandpass characteristics are physically realizable. The mixer itself and the preceding circuits are generally relatively broadband. Tuning of the receiver, between the limits set by the preselector or mixer bandwidth, is accomplished by changing the LO frequency.

Effect of Characteristics on Performance. Noncoherent pulse radar performance is affected by front-end characteristics in three ways. Noise introduced by the front end restricts the maximum range. Front-end saturation on strong signals may limit the minimum range of the system or the ability to handle strong interference. Finally, the front-end spurious characteristic affects the susceptibility of off-frequency interference.

Coherent radar performance is even more affected by spurious mixer characteristics. Range and velocity accuracy is degraded in the pulse doppler radar; stationary-target cancellation is impaired in MTI (moving-target indication) radar; and range sidelobes are raised in high-resolution pulse compression systems.

Spurious Distortion of Radiated Spectrum. It is a surprise to many radar engineers that components of the radar receiver can cause degradation of the radiated transmitter spectrum, generating harmonics of the carrier frequency or spurious doppler spectra, both of which are often required to be 50 dB or more below the carrier. Harmonics can create interference in other electronic equipment, and their maximum levels are specified by the National Telecommunications and Information Administration (NTIA) and MIL-STD-469. Spurious doppler spectra levels are dictated by requirements to suppress clutter interference through doppler filtering.

Harmonics are generated by any component which is nonlinear at the power level created by the transmitter and which passes those harmonics to the an-

tenna. Gaseous or diode receiver-protectors are designed to be nonlinear during the transmitted pulse and reflect the incident energy back toward the antenna. Isolators or circulators are often employed to absorb most of the reflected fundamental, but they are generally much less effective at the harmonics. Moreover, these ferrite devices are nonlinear in themselves and can generate harmonics.

Harmonic filters are included in most radars but often are improperly located to perform adequately. It is useless to locate the harmonic filter between the transmitter and the duplexer if the latter generates unacceptable harmonic levels itself; the filter must be located between the antenna and the duplexer.

Spurious doppler spectra are created by any process which does not reoccur precisely on each transmitted pulse. Gaseous receiver-protectors ionize under transmitter power levels, but there is some small statistical variation in the initiation of ionization on the leading edge of the pulse and in its subsequent development. In radars demanding high clutter suppression (in excess of 50 dB), it has sometimes been found necessary to prevent this variable reflected power from being radiated by use of both a circulator and an isolator in the receive path.

Spurious Responses of Mixers

Mathematical Mixer Model. The power-series representation of the mixer is perhaps the most useful in predicting the various spurious effects that are often noted. The current i flowing in a nonlinear resistance may be represented by a power series in the voltage V across the resistor terminals:

$$i = a_0 + a_1V + a_2V^2 + a_3V^3 + \cdots + a_nV^n \tag{3.1}$$

The voltage applied to the mixer is the sum of the LO voltage $V_1e^{j\omega_1 t}$ and the signal voltage $V_2e^{j\omega_2 t}$:

$$V = V_1e^{j\omega_1 t} + v_2e^{j\omega_2 t} \tag{3.2}$$

When V from Eq. (3.2) is substituted into Eq. (3.1) and the indicated operations performed, the spectral characteristics are predicted.

Mixer Spurious-Effects Chart. The results of these calculations have been tabulated in several forms to show the system designer at a glance which combinations of input frequencies and bandwidths are free of strong low-order spurious components. The most useful form of the mixer chart[1] is shown in Fig. 3.2. The heavy line shows the variation of normalized output frequency $(H - L)/H$ with normalized input frequency L/H. This response is caused by the first-order mixer product $H - L$, which originates mainly from the square-law term in the power-series representation. All other lines on the chart define spurious effects arising from the cubic and higher-order terms in the power series. To simplify use of the chart, the higher input frequency is designated by H and the lower input frequency by L.

Seven particularly useful regions have been outlined on the chart. Use of the chart is illustrated by means of the region marked A, which represents the widest available spurious-free bandwidth centered at $L/H = 0.63$. The available RF passband is from 0.61 to 0.65, and the corresponding IF passband is from 0.35 to 0.39. However, spurious IF frequencies of 0.34 ($4H - 6L$) and 0.4 ($3H - 4L$) are generated at the extremes of the RF passband. Any extension of the instantaneous RF bandwidth will produce overlapping IF frequencies, a condition that is not corrected by IF filtering. The $4H - 6L$ and $3H - 4L$ spurious frequencies,
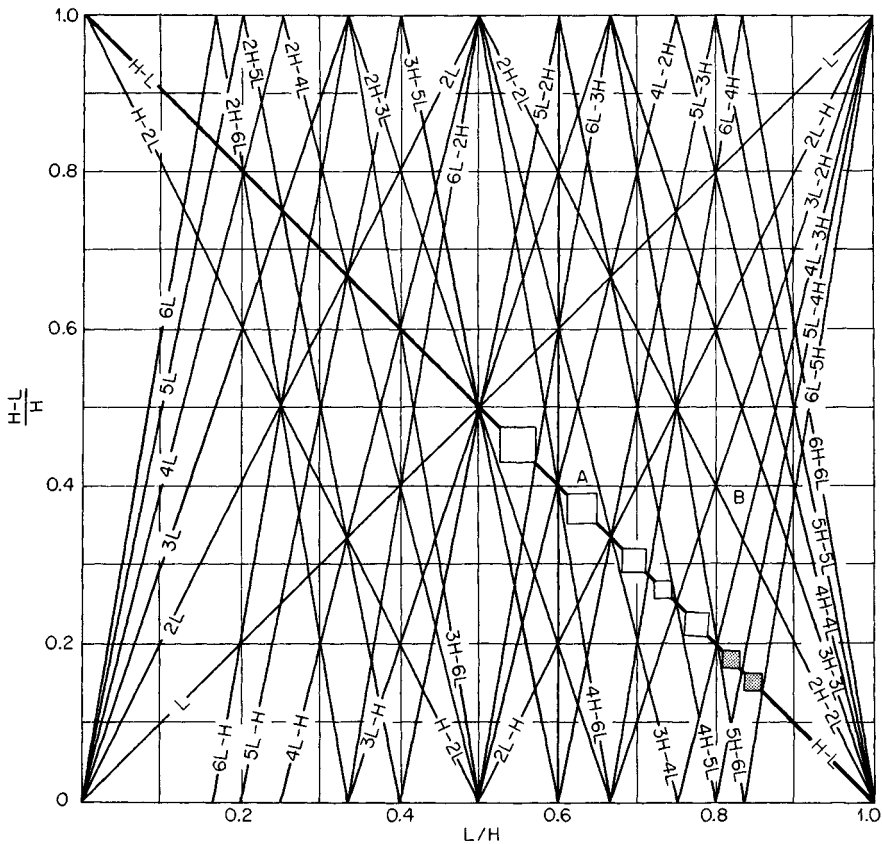
FIG. 3.2 Downconverter spurious-effects chart. H = high input frequency; L = low input frequency.

like all spurious IF frequencies, arise from cubic or higher-order terms in the power-series model of the mixer.

The available spurious-free bandwidth in any of the designated regions is roughly 10 percent of the center frequency or $(H - L)/10H$. Thus receivers requiring a wide bandwidth should use a high IF frequency centered in one of these regions. For IF frequencies below $(H - L)/H = 0.14$ the spurious frequencies originate from extremely high-order terms in the power-series model and are consequently so low in amplitude that they can usually be ignored. For this reason, single-conversion receivers generally provide better suppression of spurious responses than double-conversion receivers. The rationale for a choice of double conversion should always be validated.

The spurious-effects chart also demonstrates spurious input responses. One of the stronger of these occurs at point B, where the $2H - 2L$ product causes a mixer output in the IF passband with an input frequency at 0.815. All the products of the form $N(H - L)$ produce potentially troublesome spurious responses. These frequencies must be filtered at RF to prevent their reaching the mixer.

A spurious input response not predicted by the chart occurs when two or more off-frequency input signals produce by intermodulation a third frequency that lies within the RF passband. This effect is caused by quartic and higher-order even terms in the series. Its effect will be noted, for example, when

$$\frac{2H - L_1 - L_2}{H} = \frac{H - L}{H} \tag{3.3}$$

Intermodulation is reduced in some mixer designs by forward-biasing the mixer diodes to reduce the higher-order curvature.

The Balanced Mixer. The mixer model and the spurious-effects chart predict the spectral characteristics of a single-ended mixer. In the balanced-mixer configuration these characteristics are modified by symmetry. The two most common forms of balanced-mixer configuration are shown by Fig. 3.3a and b.

The configuration of Fig. 3.3a suppresses all spurious IF frequencies and spurious RF responses derived from even harmonics of the *signal* frequency. For the case where the subtraction is not obtained by a time delay, the LO frequency and *all* its harmonics are suppressed at the signal input port. Also of importance, noise sidebands of the LO which are converted to IF frequency are suppressed at the mixer IF port.

The configuration of Fig. 3.3b suppresses all spurious IF frequencies and spurious responses derived from even harmonics of the LO frequency. For the case where the RF phase shift is not obtained by a time delay, the LO frequency and
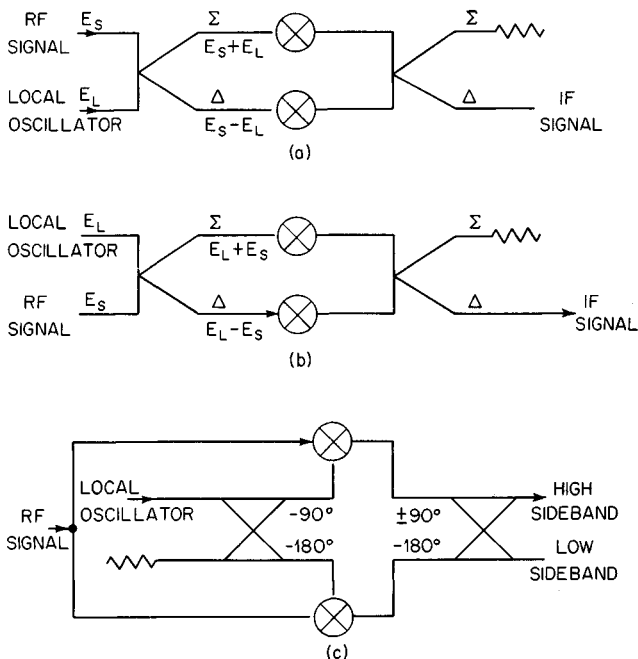
FIG. 3.3 (a) Balanced mixer with an inverted signal. (b) Balanced mixer with an inverted LO. (c) Image-reject mixer.

its *odd* harmonics are suppressed at the signal input port. Noise sidebands of the LO converted to IF *are not* rejected by this configuration, however.

Image-Reject Mixer. The single-ended mixer has two input responses which are derived from the square-law term in the power series. The responses occur at points above and below the LO frequency where the frequency separation equals the IF. The unused response, known as the image, is suppressed by the image-reject or single-sideband mixer shown in Fig. 3.3*c*. The RF hybrid produces a 90° phase differential between the LO inputs to the two mixers (which may be balanced mixers). The effect of this phase differential on the IF outputs of the mixers is a +90° shift in one sideband and a −90° shift in the other. The IF hybrid, adding or subtracting another 90° differential, causes the high-sideband signals to add at one output port and to subtract at the other. Where wide bandwidths are involved, the IF hybrid is of the all-pass type.

Characteristics of Amplifiers and Mixers

Noise Temperature. The most frequently cited figure of merit for a mixer or amplifier is its noise figure. However, the concept of noise temperature has proved more useful. Chapter 2 defines the usage of these parameters in determining the detectability of signals in a noise background.

Dynamic Range. A second useful figure of merit of the front-end device is the dynamic range from rms noise to the signal level that causes 1 dB compression in dynamic gain. Since the rms noise is dependent on the IF bandwidth, the effective dynamic range decreases with increasing IF bandwidth.

The balanced-diode mixer exhibits the largest dynamic range for a given IF bandwidth. However, the dynamic range of the mixer preceded by a low-noise amplifier will be reduced in proportion to the gain of the amplifier. Thus noise performance and dynamic range cannot be simultaneously optimized. A solution to this problem may come in the form of an active converter.[2,3]

3.5 LOCAL OSCILLATORS

Functions of the Local Oscillator. The superheterodyne receiver utilizes one or more local oscillators and mixers to convert the echo to an intermediate frequency that is convenient for filtering and processing operations. The receiver can be tuned by changing the first LO frequency without disturbing the IF section of the receiver. Subsequent shifts in intermediate frequency are often accomplished within the receiver by additional LOs, generally of fixed frequency.

Pulse-amplifier transmitters also use these same LOs to generate the radar carrier with the required offset from the first local oscillator. Pulsed oscillator transmitters, with their independent "carrier" frequency, use automatic frequency control (AFC) to maintain the correct frequency separation between the carrier and first LO frequencies.

In many early radars, the only function of the local oscillators was conversion of the echo frequency to the correct intermediate frequency. The majority of modern radar systems, however, coherently process a series of echoes from a target. The local oscillators act essentially as a timing standard by which the echo delay is measured to extract range information, accurate to within a small fraction of a wavelength. The processing demands a high degree of phase stability

throughout the radar. Although these processing techniques are described elsewhere (Chaps. 15 to 17 and 21), they determine the basic stability requirements of the receiver.

The first local oscillator, generally referred to as a stable local oscillator (stalo), has a greater effect on processing performance than the transmitter. The final local oscillator, generally referred to as a coherent local oscillator (coho), is often utilized for introducing phase corrections which compensate for radar platform motion or transmitter phase variations.

Stalo Instability. The stability requirements of the stalo are generally defined in terms of a tolerable phase-modulation spectrum. Sources of unwanted modulation are mechanical or acoustic vibration from fans and motors, power supply ripple, and spurious frequencies and noise generated in the stalo. In general, the tolerable phase deviation decreases with increasing modulation frequency because the doppler filter is less efficient in suppressing the effects. In a radar having two-pulse MTI, there is a linear relationship between the tolerable phase deviation and the period of the modulation. Their ratio is the allowable FM (frequency modulation) or *short-term frequency stability* sometimes encountered in the literature. This parameter does not adequately define the phase-stability requirements for pulse doppler or MTI radars where more than two pulses are coherently processed.

The phase-modulation spectrum of the stalo may be measured and converted into the MTI improvement factor limitation, which is dependent on range to the clutter and the characteristics of the two cascaded filters in the radar receiver. This conversion process involves three steps, described below.

It should be noted that some spectrum analyzers do not distinguish between frequencies below the desired stalo frequency and those above; their response is the sum of the power in the two sidebands at each designated modulation frequency. This is of no consequence in MTI radars which have equal response to positive and negative doppler frequencies. In radars using doppler filters unsymmetrical about zero doppler, it is necessary to assume that the stalo spectrum measured is symmetrical, generally a valid assumption. The examples shown subsequently employ measured data from this type of double-sideband (DSB) spectrum analyzer. If a single-sideband (SSB) spectrum analyzer is available, positive and negative modulation-frequency components can be measured separately and analyzed without any assumption of symmetry. It is essential that the measured data be defined as SSB or DSB, since there is a 3 dB difference in the two forms of data.

Range Dependence. Most modern radars use the stalo to generate the transmitted pulse as well as to shift the frequency of the received echoes. The transmitters are power amplifiers (traveling-wave tubes, klystrons, twystrons, crossed-field amplifiers, solid-state amplifiers, etc.) rather than oscillators (magnetron, etc.). It is this double use of the stalo that introduces a dependence on range of the clutter and exaggerates the effect of certain unintentional phase-modulation components by 6 dB, the critical frequencies being those which change phase by odd multiples of 180° during the time period between transmission and reception of the clutter echo from a specified range. At these critical frequencies, a maximum positive phase deviation on transmission changes to a maximum negative deviation at the time of reception, doubling the undesired phase modulation of the echo at IF.

Figure 3.4 shows this range-dependent filter characteristic, which may be expressed mathematically as

$$dB = 10 \log 4 \sin^2 (2\pi f_m R/c) = 10 \log 4 \sin^2 (\pi f_m t) \qquad (3.4)$$

where f_m = modulation frequency, Hz
 R = range, m
 c = propagation velocity, 3×10^8 m/s
 t = time delay = $2R/c$

A short time delay can tolerate much higher disturbance at low modulation frequencies, as illustrated by the two cases in Fig. 3.4. Consequently, stalo stability needs to be computed for several time delays.

CURVE	DELAY (μs)
1	15
2	1200
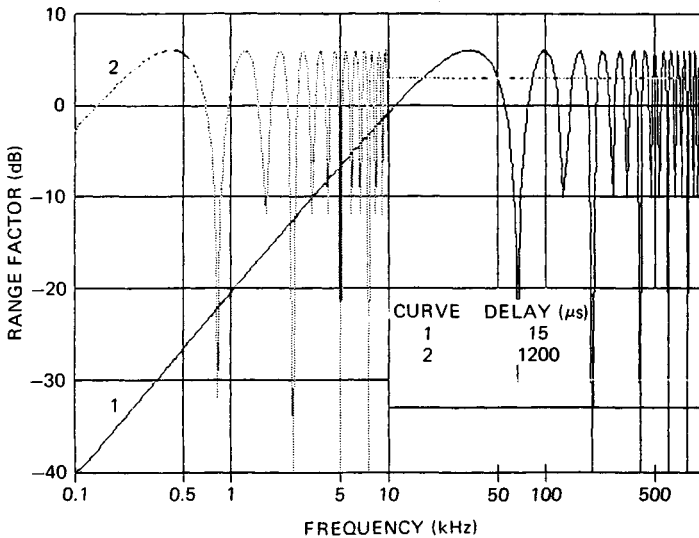
FIG. 3.4 Effect of range delay on clutter cancellation.

For example, stalo phase modulation caused by power supply ripple at 120 Hz creates nearly equivalent phase modulation of a clutter echo from nearly 100-nmi range (delay of 1200 μs resulting in 0 dB range factor). Phase modulation of a clutter echo with a range delay of 15 μs is about 38 dB less than the stalo phase modulation because the stalo phase has changed only slightly in this short time interval; the phase added by the stalo to the transmitted pulse is nearly the same as the phase subtracted from the received echo in the mixer.

Adding the decibel values of the measured stalo spectrum and the range-dependent effect at each modulation frequency provides the spectrum of undesired doppler modulation at the output of the mixer.

Receiver Filtering.   Subsequent stages of the radar receiver have responses which are functions of the doppler modulation frequency; so the output residue spectrum can be obtained by adding the decibel responses of these filters to the preceding spectrum at the mixer.

The receiver contains two cascaded filters: an optimum-bandwidth filter at IF and a doppler filter, generally implemented digitally in modern radars. The example illustrated in Fig. 3.5 includes a gaussian filter at IF with a 3 dB bandwidth of 1.6 MHz and a four-pulse MTI with variable interpulse periods and time-
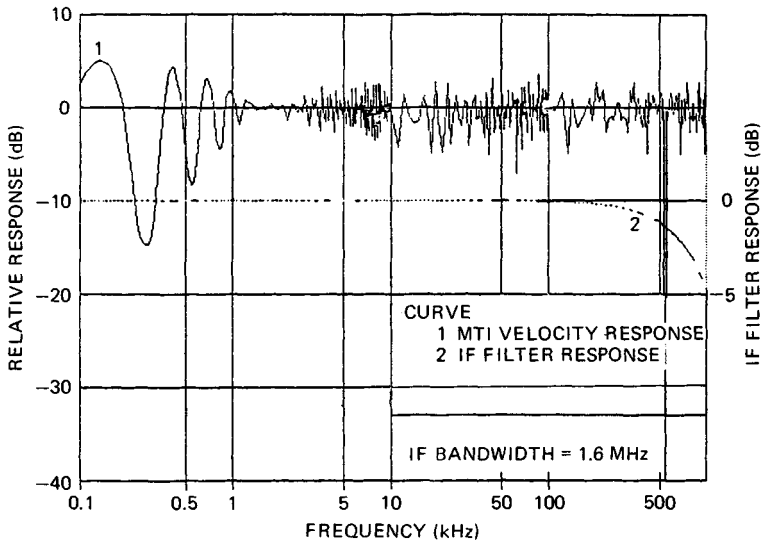
FIG. 3.5   Frequency responses of radar receiver.

varying weights. This MTI exaggerates certain stalo modulation frequencies by as much as 5 dB more than average. Note that for this analysis to be correct the MTI velocity response must be scaled to have 0 dB gain to noise, not to an optimum doppler frequency.

Integration of Residue Power.   The MTI improvement factor limitation of the stalo may be expressed as the ratio of the stalo power to the total power of the echo modulation spectrum it creates at the output of the cascaded filters (Figs. 3.4 and 3.5).

Figure 3.6 shows an example of the measured modulation spectrum of a stalo (curve 1) and the effect of 15-μs delay on the clutter residue (curve 2). A computer program can be utilized to alter the measured spectral data, using the filters of Fig. 3.4 and 3.5, and to integrate the total power in the doppler residue spectrum, with the exception of modulation frequencies below 100 Hz, which cannot be measured. The result (51.8 dB) is the MTI improvement factor limitation due to stalo instability.

Figure 3.7 shows the same measured stalo modulation spectrum, subjected to a range delay of 1200 μs. The clutter residue spectrum contains more power at the lower modulation frequencies than Fig. 3.6, but the MTI residue is increased by only 1 dB. Long-range clutter is suppressed to nearly the same degree as short-range clutter.

If the radar utilizes more than one doppler filter, the effect of stalo instability should be calculated for each individually. If an individual filter's doppler response is unsymmetrical, residues from positive and negative doppler bands must be computed separately and added in power.

It should be noted that many textbooks analyze only a simple two-pulse MTI, and the resulting equations for the limitation on MTI improvement factor cannot be employed for more sophisticated doppler filters. This important fact is generally overlooked by the casual reader.
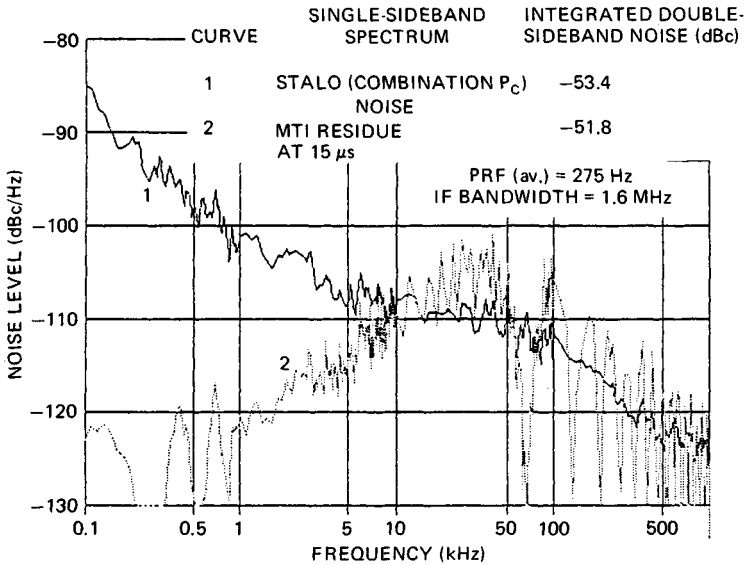
FIG. 3.6 Effect of 15-μs range delay on stalo MTI limitation.
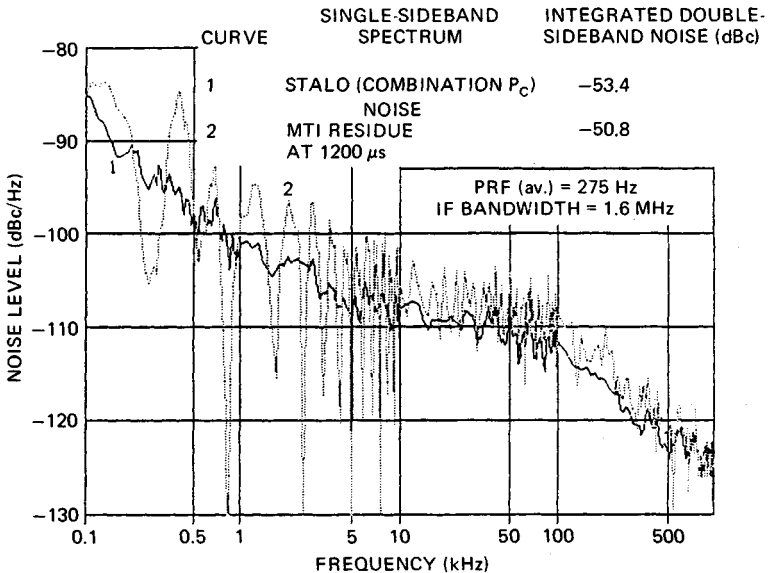
FIG. 3.7 Effect of 1200-μs range delay on stalo MTI limitation.

It should be noted that most textbook analyses assume that the stalo instability is due either to a single modulation frequency or to a combination of white gaussian-noise modulations. Rarely are these assumptions valid for real stalos; so a different method of analysis must be employed.

The computer analysis described measures the colored modulation spectrum of the stalo, modifies it in conformance with the range-dependent effect and receiver filters of the radar, and integrates the output residue power.

It is a valid procedure for determining stalo stability that requires no assumptions, but the resulting value must be considered merely a figure of merit to compare *different* stalos for a *given* application. The same stalo would have a different stability value in another application with different receiver filters.

Coho and Timing Instability. In modern radars using pulsed amplifier transmitters, the coho is rarely a significant contributor to receiver instability. However, in older radars using pulsed oscillator transmitters, the coho must compensate for the random phase of each individual transmitter pulse, and imperfect compensation results in clutter residue at the output of the doppler filters. Compensation is possible only for echoes from the most recent transmission; echoes from prior transmissions (multiple-time-around clutter) cannot be suppressed by doppler filtering in pulsed oscillator radars, and for this reason alone such radars are no longer popular. Those readers interested in the methods employed to compensate the coho in these older radars are referred to Sec. 5.5 of the 1970 edition of this handbook.

When the radar is on a moving platform or when the clutter is moving rain or sea, the frequency of the coho is sometimes varied to try to compensate for this motion, shifting the clutter spectrum back to zero doppler. The servo which accomplishes this task, if properly designed, will introduce insignificant instability under ideal environmental conditions (solely clutter echoes and receiver noise typical of laboratory tests), but the effect of strong moving targets and pulsed interference from other radars can sometimes be serious, shifting the coho frequency from the proper compensation value.

Timing signals for the transmitter and A/D converter are usually generated from the coho, and timing jitter can cause clutter attenuation to be degraded. However, the effect of timing jitter is too complex to predict accurately; so it is rarely measured separately.

Total Radar Instability. The primary sources of radar instability are usually the stalo and the transmitter. If the doppler spectra of these two components are available, either through measurements or through predictions based on similar devices, the convolution of the two-way stalo spectrum (modified by the range-dependent effect) and the transmitter doppler spectrum provides an estimate of the spectrum of echoes from stable clutter, which is then modified by the two receiver filters and integrated to obtain the residue power caused by these two contributors. This power can be larger than the sum of the residues created by each contributor alone. These procedures are employed to diagnose the source of radar instability in an existing radar or to predict the performance of a radar in the design stage.

Measurement of total radar instability can be conducted with the radar antenna searchlighting a stable point clutter reflector which produces an echo close to (but below) the dynamic-range limit of the receiver and doppler filter. Suitable clutter sources are difficult to find at many radar sites, and interruption of rotation of the antenna to conduct such a test may be unacceptable at others; in this case, a microwave delay line can be employed to feed a delayed sample of the transmitter pulse into the receiver. All sources of instability are included in this single measurement except for any contributors outside the delay-line loop. It is important to recognize that timing jitter does not produce equal impact on all

parts of the echo pulse and generally has minimal effect on the center of the pulse; so it is essential to collect data samples at a multiplicity of points across the echo, including leading and trailing edges. The total radar instability is the ratio of the sum of the multiplicity of residue powers at the output of the doppler filter to the sum of the powers at its input, divided by the ratio of receiver noise at these locations. Stability is the inverse of this ratio; both are generally expressed in decibels.

In radars with phase-coded transmission and pulse compression receivers, residue may be significant in the range sidelobe region as well as in the compressed pulse, caused by phase modulation during the long transmitted pulse rather than solely from pulse to pulse. Measurement of stability of such radars must employ a very large number of data points to obtain an answer valid for clutter distributed in range.

Radar instability produces predominantly phase modulation of the echoes, and the scanning antenna produces predominantly amplitude modulation; so the combined effect is the sum of the residue powers produced by each individually.

3.6 GAIN-CONTROLLED AMPLIFIERS

Sensitivity Time Control (STC). The search radar detects echoes of widely differing amplitudes, typically so great that the dynamic range of any fixed-gain receiver will be exceeded. Differences in echo strength are caused by differences in radar cross sections, in meteorological conditions, and in range. The effect of range on radar echo strength overshadows the other causes, however.

The radar echo power received from a reflective object varies inversely with the fourth power of the range or propagation time of the radar energy. The effect of range on signal strength impairs the measurement of target size. Yet determination of target size is needed to discriminate against radar echoes from insects, atmospheric anomalies, or birds (which in some cases have radar cross sections only slightly less than that of a jet fighter). Also, many radar receivers exhibit objectionable characteristics when signals exceed the available dynamic range. These effects are prevented by a technique known as sensitivity time control, which causes the radar receiver sensitivity to vary with time in such a way that the amplified radar echo strength is independent of range.

Search radars often employ a cosecant-squared antenna pattern whose gain diminishes with increasing elevation angle. The pattern restricts the power at high elevation angles, since an aircraft at a high elevation angle is necessarily at close range and little power is required for detection. At the high elevation angles, however, the echo power becomes independent of range and varies instead with the inverse fourth power of the altitude. An STC characteristic that is correct for the low-angle radar echoes restricts high-angle coverage. This incompatibility of the STC requirements at the elevation extremes severely limits the usefulness of STC.

The restriction on STC imposed by the cosecant-squared antenna pattern can be reduced by a more realistic radar design philosophy. It is recognized that the antenna must radiate more energy at the high angles than is provided by the cosecant-squared pattern. There are two reasons for this. First, high-angle coverage is limited by clutter from the stronger low-altitude section of the beam rather than from system noise. Second, ECM reduces both the maximum range

and the altitude coverage of the radar. Of these, the loss of altitude coverage is the more serious. Both factors have caused the cosecant-squared pattern to be abandoned in favor of one that directs more energy upward.

The advent of stacked-beam radars, which achieve their coverage pattern by use of multiple beams, has liberated STC from the restrictions of the antenna pattern. In these systems, there is one receiver channel for each beam, and STC may be applied to the receiver channels independently. Consequently, the upper-beam receivers may be allowed to reach maximum sensitivity at short ranges, whereas the lower-beam receiver reaches maximum sensitivity only at long range.

Most modern radars generate STC waveforms digitally. The digital commands may be used directly by digital attenuators or converted to voltage or current for control of diode attenuators or variable-gain amplifiers. Digital control permits calibration of each attenuation to determine the difference between the actual attenuation and the command, by injecting a test pulse during *dead time*. This is essential in monopulse receivers which compare the echo amplitudes received in two or more beams simultaneously to accurately determine the target's position in azimuth or elevation. Accurate measurements depend on compensation for any difference in gains of the monopulse receivers.

Readers interested in the methods of generating the analog STC waveforms used in older radars may find descriptions of various methods in Sec. 5.6 of the 1970 edition of this handbook.

Clutter Map Automatic Gain Control.   In some radars, mountain clutter can create echoes which would exceed the dynamic range of the subsequent stages of the receiver (A/D converter, etc.) if the STC attenuation at that range allows detection of small aircraft. The spatial area occupied by such clutter is typically a very small fraction of the radar coverage; so AGC is sometimes considered as an alternative to either boosting the STC curve (a performance penalty affecting detectability of small aircraft in areas of weaker clutter or no clutter) or increasing the number of bits of the A/D converter and subsequent processing (an economic penalty).

Clutter map AGC is controlled by a digital map which measures the mean amplitude of the strongest clutter in each map cell of many scans and adds attenuation where necessary to keep the mean amplitude well below saturation. One disadvantage of clutter map AGC is that it degrades detectability of small aircraft over clutter which, in the absence of AGC, would be well below saturation. The scan-to-scan fluctuation of clutter requires a 6 to 10 dB safety margin between the maximum mean level controlled and saturation. Another problem is the vulnerability of the map to pulsed interference from other radars.

Clutter map AGC can serious degrade other critical signal-processing functions, and the following fundamental incompatibilities prevent its successful application to many types of radars:

- Suppression of clutter by doppler filtering is degraded by change of attenuation from one interpulse period to the next.

- Control of false alarms in distributed clutter (rain, sea) can be degraded by change of attenuation from one range sample to the next (see Sec. 3.13).

- Time sidelobes of compressed pulses in radars which transmit coded waveforms are degraded by attenuation variation in range prior to compression. Gradual STC variations can be tolerated, but not large step changes.

Automatic Noise-level Control. AGC is widely employed to maintain a desired level of receiver noise at the A/D converter. As will be described in Sec. 3.11, too little noise relative to the quantization increment of the A/D converter causes a loss in sensitivity; too much noise means a sacrifice of dynamic range. Samples of noise are taken at long range, often beyond the instrumented range of the radar (in dead time), to control the gain by means of a slow-reaction servo. If the radar has RF STC prior to any amplification, it can achieve meaningful dead time by switching in full attenuation; this minimizes external interference with minimal (and predictable) effect on system noise temperature. Most radars employ amplifiers prior to STC; so they cannot attenuate external interference without affecting the noise level which they desire to sense, and the servo must be designed to tolerate pulses from other radars and echoes from rainstorms or mountains at extreme range. This interference occasionally can be of high amplitude but generally has a low duty cycle during a 360° scan; so the preferred servo is one which increments a counter when any sample is below the desired median noise level and decrements the counter when the sample is above that level, independently of how great the deviation is. The most significant bits of the counter control the gain, and the number of bits of lesser significance in the counter dictates the sluggishness of the servo.

3.7 FILTERING

Filtering of the Entire Radar System. The filter provides the principal means by which the receiver discriminates between desired echoes and interference of many types. It may approximate either of two forms: a matched filter, which is a passive network whose frequency response is the complex conjugate of the transmitted spectrum, or a correlation mixer, an active device which compares the received signals with a delayed replica of the transmitted signal. Receiver filters are assumed to have no memory from one transmission to the next; their response is to a single transmission.

Actually, most radars direct a multiplicity of pulse transmissions at a target before the antenna beam is moved to a different direction, and the multiplicity of echoes received is combined in some fashion. The echoes may be processed by an integrator, which is analogous to a matched filter in that ideally its impulse response should match the echo modulation produced by the scanning antenna. The echoes may be applied directly to a PPI, with the viewer visually integrating the dots in an arc which he or she associates with the antenna beamwidth. Various doppler processes (including MTI) may be applied to separate desired from undesired targets. From the radar system standpoint, these are all filtering functions, but they are treated in other chapters of this handbook. The receiver filtering to be discussed here is that associated with separating a single pulse from interference, although the subsequent problem of filtering the train of echoes from a single target dictates the stability of the receiver filter.

At some point in the radar receiver, a detector produces an output voltage which is some function of the envelope of the IF signal. If it provides a linear function, it is termed an envelope detector; logarithmic detectors will be described in Sec. 3.8. The response of a linear detector to weak signals which do not greatly exceed noise level has been extensively analyzed.[4] Various pairs of frequency components of input noise, which may be far removed from the spec-

trum of the desired echo, can intermodulate to produce a beat-frequency component at the detector output that is within the desired band. Similarly, the noise intermodulation smears some of the signal energy outside the desired band. As a consequence, filtering after envelope detection is less efficient than filtering prior to detection. All postdetection circuitry should have several times the bandwidth of the echo, and predetection filtering should be optimized, as will be described.

Definitions. The reader is cautioned that there are no universally accepted definitions of the terms *pulse duration* and *spectral bandwidth* of the transmitted signal, *impulse response* and *bandwidth* of filters, or the equivalent antenna parameters, *beamwidth* and *spectral bandwidth* caused by scanning. These terms should always be used with clarifying adjectives to define their meaning.

Energy Definitions. For detection of radar echoes against a noise background, the only fundamental parameters are the energy content of the transmitted signal, of the receiver noise, and of the echoes received as the antenna scans past the target. These energy parameters define the width of a rectangular function that has the same peak response and same energy content as the real function. Their only purpose is to relate the peak value of the function to the more vital energy content.

Of these parameters, only the energy of the pulse is easy to measure (average power/PRF), and this may be employed in the radar range equation directly, without distinguishing peak power and "energy" pulse width. The noise or energy bandwidth of a receiver is often employed in theoretical analyses but rarely stated in the tabulation of radar parameters; bandwidth need not even be included in the radar range equation if the receiver approximates a matched filter.

3 dB Definitions. In the interest of making possible direct measurement of parameters from oscilloscope waveforms or pen recordings of these functions, it has been customary to utilize widths measured either at half-power (3 dB) or half-voltage (6 dB) points. For functions that resemble a gaussian pattern, the 3 dB width is a close approximation to the energy width; the receiver bandpass generally fulfills this condition sufficiently to make the 3 dB bandwidth meaningful. Transmitter pulse shapes and spectra generally deviate significantly from gaussian.

6 dB Definitions. Although antenna beamwidths (and number of echoes received) are often specified between the 3 dB points, this actually represents a 6 dB definition of the echo response as the radar antenna scans past the target; in those radars whose transmitting and receiving beams are not identical, the 6 dB points of the two-way pattern are usually specified. Most definitions of pulse shapes include voltage parameters, with rise and fall times being represented by 10 and 90 percent points and pulse duration by 50 percent (6 dB) points. Likewise, filter bandpass characteristics are often defined by their widths at the 6 and 60 dB points. The 6 dB definitions will be the dominant definitions employed in this chapter.

Entirely aside from custom, there are several valid arguments favoring the use of 6 dB parameters. As indicated in Table 3.2, the optimum bandwidth–time product for detection of a pulse in white gaussian noise, with each defined at the 6 dB points, does not deviate significantly from unity for most practical functions. The 3 dB or energy definitions yield widely variable optimum bandwidths, dependent upon the shape of the pulse and the bandpass of the filter; there can be no quick estimation of optimum bandwidth if these parameters are utilized.

Distributed clutter, rain or chaff, is often a more serious interference with target detection than noise. In passing through an optimum-bandwidth filter and optimum-bandwidth integrator, the echo is stretched in both range and angle; the clutter spectrum, being the product of the transmitted spectrum and the bandpass of the receiver, is narrower than either. As a result, the 6 dB two-way beamwidth and the 6 dB pulse width closely approximate the extent of the radar cell from which the "optimum" receiver accumulates clutter energy.

To summarize the general utility of 6 dB definitions: (1) The range equation for detection in noise need not include peak power, pulse width, or receiver bandwidth; only the efficiency of the integrator requires a definition of the number of pulses being received, and a 6 dB echo definition is universally employed. (2) The optimum bandwidth is close to the inverse of the echo duration if both are 6 dB definitions; this applies to both the receiver filter and the integrator. (3) The energy of the interference from clutter, rain or chaff, that is accepted by an approximately matched receiver is well defined by the 6 dB pulse duration and 6 dB two-way beamwidth.

Approximations to Matched Filters. The most efficient filter for discriminating between white gaussian noise and the desired echoes is a matched filter, a passive network whose frequency response is the complex conjugate of the transmitted spectrum. It can process echoes from all ranges. The correlation mixer, an active device which compares the received signals with a delayed replica of the transmitted signal, is mathematically equivalent to a matched filter, but it is responsive only to echoes from one specific range; consequently its use in radar systems is more limited.

Table 3.2 illustrates the degree of sacrifice in detectability that results in approximating a matched filter, either to simplify the hardware or to achieve better filtering of other forms of interference. The optimum bandwidths of these filters

TABLE 3.2 Approximations to Matched Filters

Transmitted pulse shape	Receiver filter	Optimum bandwidth–time product			Mismatch loss, dB
		6 dB	3 dB	Energy	
Gaussian	Gaussian bandpass	0.88	0.44	0.50	0
Gaussian	Rectangular bandpass	1.05	0.74	0.79	0.51
Rectangular	Gaussian bandpass	1.05	0.74	0.7	0.51
Rectangular	5 synchronously tuned stages	0.97	0.67	0.76	0.51
Rectangular	2 synchronously tuned stages	0.95	0.61	0.75	0.56
Rectangular	Single-pole filter	0.70	0.40	0.63	0.88
Rectangular	Rectangular bandpass	1.37	1.37	1.37	0.85
Phase-coded					
Biphase	Gaussian	1.05	0.74	0.79	0.51
Quadriphase	Gaussian	1.01	0.53	0.5	0.09

also are tabulated in terms of product of the filter bandwidth and pulse duration. Typically, the bandwidth may deviate 30 to 50 percent from the optimum value before the detectability is degraded by an additional 0.5 dB. This rather broad "optimum" region is centered near unity bandwidth-time product for virtually all practical filters if one uses the 6 dB definitions.

Sometimes the bandwidth of a radar receiver is in excess of the optimum to allow for some offset between the echo spectrum and the filter bandpass, caused by target velocity and receiver tuning tolerances. Although this makes the radar more susceptible to off-frequency narrowband interference (Fig. 3.8), it reduces the time required to recover from impulse interference (Fig. 3.9). These figures also illustrate that, to provide good suppression of both forms of interference, the shape of the filter bandpass characteristic is even more important than its bandwidth. Rectangular bandpass or impulse responses should be avoided; the closer one approximates a gaussian filter, the better the skirts in both frequency and time domains.

In the case of phase-coded transmissions, the duration of the subpulses is the time parameter. This is equivalent to the spacing of the subpulses with one exception: the quadriphase code[5,6] employs half-cosine subpulses with 6 dB width equal to four-thirds of subpulse spacing. One of the virtues of the quadriphase code is the unusually low mismatch loss, owing to the fact that the impulse response of the gaussian filter is an excellent approximation to the subpulse shape. A digital correlator in a later stage of the receiver completes the matched filter.

Filtering Problems Associated with Mixer Spurious Responses. The approximation of a matched filter is generally most easily accomplished at some
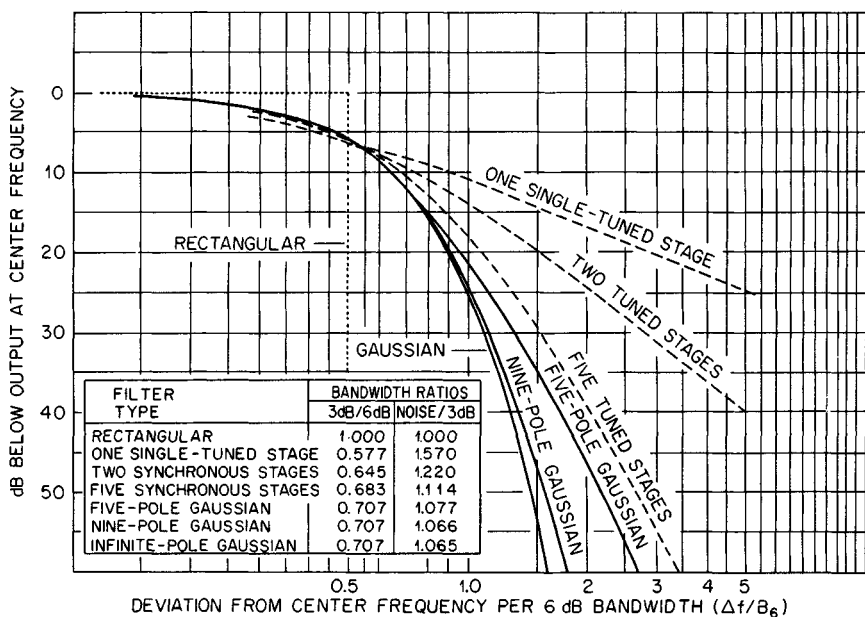
FIG. 3.8   Bandpass characteristics of filters.
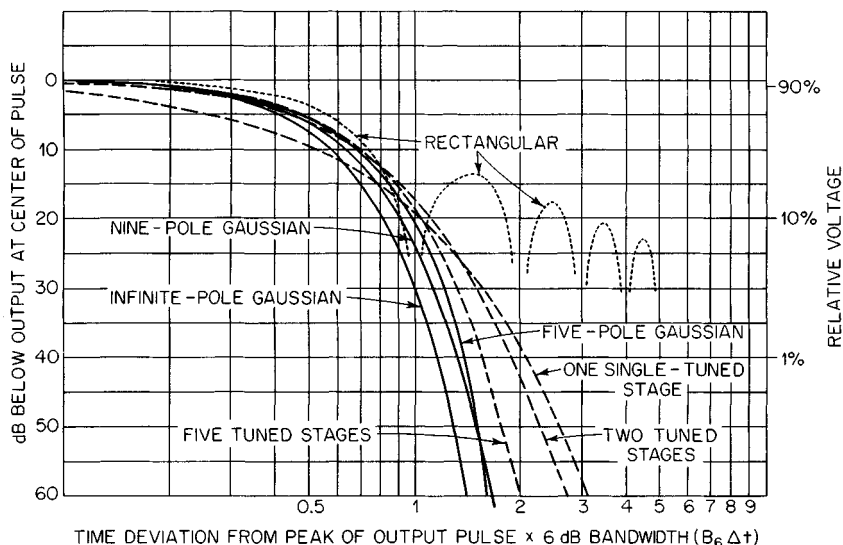
FIG. 3.9 Impulse characteristics of filters.

frequency other than that radiated by the radar. The optimum filtering frequency is a function of the bandwidth of the echo and the characteristics of the filter components. Consequently, it is necessary for the radar receiver to translate the frequency of the echo to that of the filter, in one or more steps, using local oscillators and mixers.

Section 3.4 described how spurious responses are generated in the mixing process. Unwanted interference signals can be translated to the desired intermediate frequency even though they are well separated from the echo frequency at the input to the mixer. The ability of the radar to suppress such unwanted interference is dependent upon the filtering preceding the mixer as well as on the quality of the mixer itself.

The image frequency is the most serious of the spurious-response bands, but an image-rejection mixer can readily suppress these signals by 20 dB. A filter can further attenuate image-frequency signals before they reach the mixer unless the ratio of input to output frequencies of the mixer exceeds the loaded Q of the available filters. This image-suppression problem is the reason why some receivers do not translate from the echo frequency directly to the final intermediate frequency in a single step.

The other spurious products of a mixer generally become more serious if the ratio of input to output frequencies of the downconverter is less than 10. The spurious-effects chart (Fig. 3.2) shows that there are certain choices of frequency ratio that provide spurious-free frequency bands, approximately 10 percent of the intermediate frequency in width. By the use of a high first IF, one can eliminate the image problem and provide a wide tuning band free of spurious effects. Filtering prior to the mixer remains important, however, because the neighboring spurious responses are of relatively low order and may produce strong outputs from the mixer.

In addition to external sources of interference, the radar designer must be concerned with internal signal sources. MTI and pulse doppler radars are particularly

susceptible to any such internal oscillators that are not coherent, i.e., that do not have the same phase for each pulse transmission. The effect of the spurious signal then is different for each echo, and the ability to reject clutter is degraded.

A truly coherent radar generates all frequencies, including its interpulse periods, from a single stable oscillator. Not only all the desired frequencies but also all the internally generated spurious signals are coherent, and they do not affect the clutter rejection.

More commonly, MTI and pulse doppler radars are pseudo-coherent, as illustrated in Fig. 3.10. The coho is the master reference for the phase detector and may be the clock from which the interpulse periods are determined. The coho also is employed in generating the transmitter frequency, offset from a noncoherent local oscillator. Neither the local oscillator nor the transmitter is co-
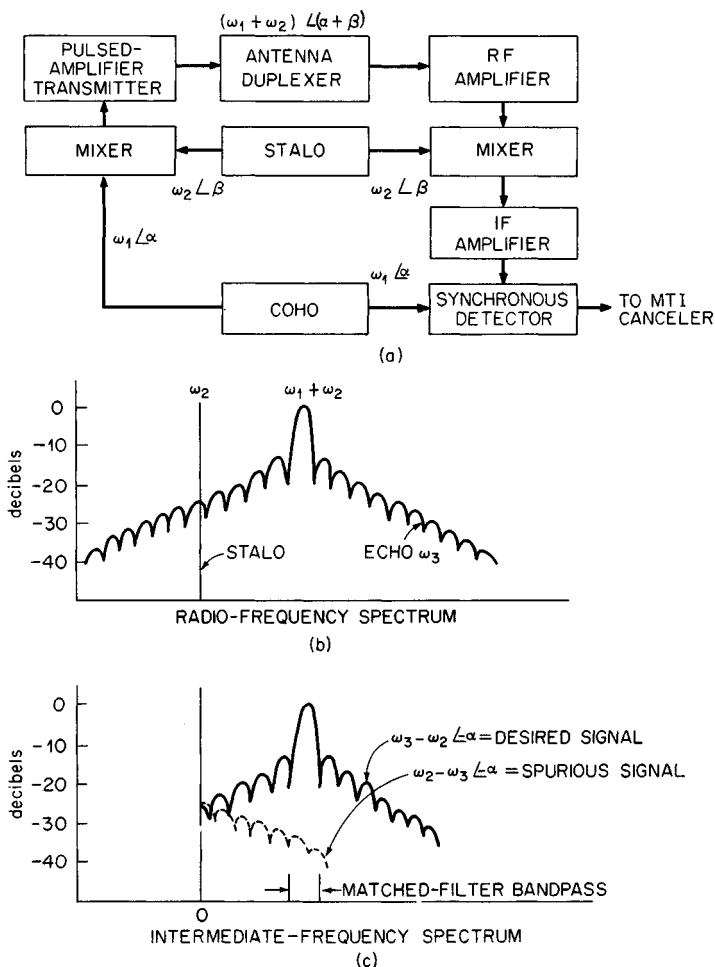
FIG. 3.10   Pseudo-coherent radar. (*a*) Block diagram. (*b*) RF spectrum of the echo. (*c*) IF spectrum of the echo.

herent; their phases are different for each pulse transmitted. Only in the IF portion of the receiver is the desired echo from a stationary target coherent (α = constant, where α is the phase of the echo when the transmitter is turned on); virtually all the spurious outputs of the mixer are noncoherent and produce a fluctuating signal at IF.

The only spurious component that may be coherent in the pseudo-coherent radar is the image frequency. As illustrated in Fig. 3.10, a pulsed spectrum can have fairly wide skirts and can overlap into the image band. The folded spectrum at intermediate frequency has two components within the bandpass of the matched filter, and if the percentage bandwidth at IF is large, the undesired image component can be significant. Only when it is coherent is a degradation of clutter cancellation impossible.

Consider a pseudo-coherent radar in which the interpulse periods are generated from an independent source. Now the phase of the coho (α) changes from pulse to pulse, and although the phase of the desired echo from a stationary target changes equally, the phase of the spurious image echo changes oppositely. The cancellation of the stationary-target echo is limited to the level of the spurious echo.

It is clear that the overall filtering capability of the radar, its ability to enhance the desired echoes and suppress undesired interference, may be degraded by the spurious responses of the various mixing stages. Particularly susceptible to degradation are MTI and pulse doppler radars, which may not provide the expected improvement in the rejection of clutter if the coho is not at the same phase condition each time that the transmitter pulses. All radars are susceptible to off-frequency interference, which, unless it is filtered before reaching the mixer, may create a detectable output in the desired IF band.

The ability of MTI or pulse doppler processing to suppress clutter may be degraded if the receiver filter is not perfectly stable. The receiver's transfer characteristic (gain, time delay, and bandpass or impulse response) must be constant so that its effect on each echo pulse is identical.

The mixer spurious responses just discussed and the stalo and coho problems of Sec. 3.5 represent only the most likely sources of instability to be encountered. Other elements of the receiver require attention to avoid instability problems. Vibration or power supply ripple can cause gain and phase modulation, particularly in RF amplifiers. Such modulation will degrade clutter attenuation unless the ripple frequency is a harmonic of the PRF.

3.8 LOGARITHMIC DEVICES

Characteristics

Accuracy.   Logarithmic devices and IF amplifiers are devices whose output is proportional to the logarithm of the envelope of the IF input. They often approximate the logarithmic characteristic by multiplicity of linear segments. Normally linear segments of equal *length ratio* and varying slope are joined to give a best fit to a logarithmic curve. Each segment will be correct at two points and will have a maximum error at the ends and center. The magnitude of the error[7] increases with the length ratio of the segment. Figure 3.11 shows how this error changes with segment length ratio (also called gain per segment). In practice, the joints between the "linear" segments are not abrupt, and the best fit to the logarithmic curve may have less than the theoretical error.

Logarithmic detectors and amplifiers are frequently designed with adjustments in each stage. This allows for adjustment of the slope and/or length of the segment for a better fit. A precision exponentially decaying IF waveform from a test set is applied to the unit under test. The unit is adjusted for a linearly decaying output, which indicates the correct adjustment.

Dynamic Range. The dynamic range of a logarithmic detector or amplifier is dependent on the number of linear segments N and on the length ratio G of the segments:

$$\text{Dynamic range} = 20N \log G \qquad (3.5)$$

A well-designed logarithmic detector may have a dynamic range of 80 dB derived from nine stages with an error as low as ±0.2 dB.

Bandwidth. The bandwidth of a logarithmic detector or amplifier generally varies with signal level. For this reason the logarithmic device is usually designed with excess bandwidth and is preceded by filters which establish the receiver bandwidth. However, the large signal bandwidth of the logarithmic device itself may be measured by using the method indicated in Fig. 3.12. The input voltage is increased from V_i to $V_i + 3$ or 6 dB, changing the operating point from A to B. The frequency is now changed in both directions to find those two frequencies that place the operating point at C. The same result also may be obtained by changing the frequency until the output is reduced by $3S$ or $6S$, where S is the volts-per-decibel characteristic. This places the operating point at C, which lies on the same curve as C. Because of the dependence of bandwidth on signal level, if possible the logarithmic amplifier should be aligned with signals below the threshold and its pulse response used as a criterion of performance.

The pulse response is measured with a pulsed IF signal having much faster rise and fall times than those of the logarithmic device being tested. The rise time is the time required for the output to rise from the $-20S$ to the $-S$ point, and the fall time is the time required for the output to fall from the $-S$ point to the $-20S$ point. Because of the logarithmic characteristic, the fall time tends to be a straight line and to exceed the rise time.
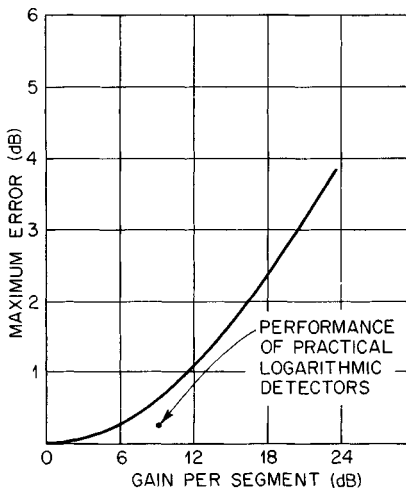
FIG. 3.11 Error in the approximation of a logarithmic curve by linear approximation.
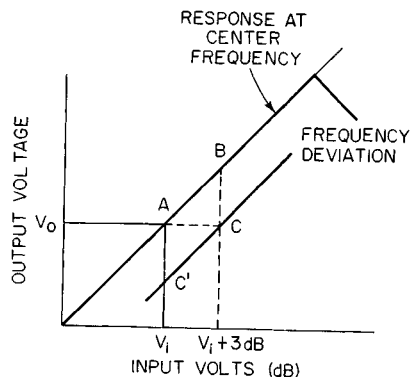
3.12 Method of measuring logarithmic-amplifier bandwidth.

Analog Logarithmic Devices

Logarithmic Detector. A well-known form of the logarithmic detector uses successive detection,[6] wherein the detected outputs of N similar limiter stages are summed as shown by Fig. 3.13. If each stage has a small signal gain G and a limited output level E, the intersections of the approximating segments fall on a curve described by

$$E_0(M) = n\left[E\frac{\log E_i(M)G^{N+1}}{\log G} + E\left(\frac{1}{G^{M-1}} + \cdots + \frac{1}{G^2} + \frac{1}{G}\right)\right] \qquad (3.6)$$

where n is the detector efficiency and $E_i(M)$ represents the particular input levels that correspond to the intersections of the line segments,

$$E_i(M) = \frac{E}{G^M} \qquad (3.7)$$

(a)

(b)
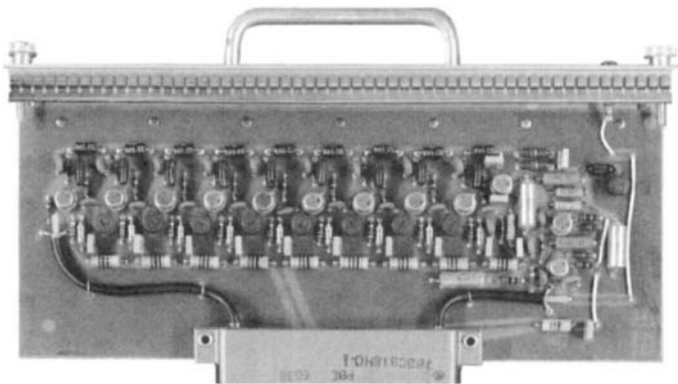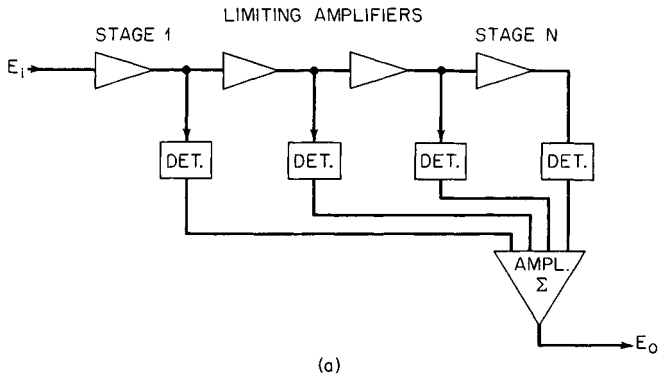
FIG. 3.13 (a) Logarithmic detector. (b) Nine-stage logarithmic detector. Gain adjustment for each stage is shown below the transistor.

The independent variable M is the order of the stage at incipient saturation and will take on only integer values from 1 to N.

A nonlogarithmic term in the form of a power series in $1/G$ has only a minor effect on logarithmic accuracy. The successive difference in this term as M is varied from 2 to N is E/G^{M-1}. The overall tendency of this term, therefore, is to produce an offset in the output with only a minor loss in logarithmic accuracy at the highest signal levels.

A typical logarithmic detector may have an accuracy of ± 4 dB, a dynamic range of 80 dB, and a bandwidth of 5 to 10 MHz at 30 MHz, derived from nine stages (Fig. 3.13b). However, Rubin[9] describes a four-stage successive detection design having a bandwidth of 640 MHz centered at 800 MHz. Pulse response time of 2.5 ns is claimed.

Logarithmic Amplifier.   A logarithmic IF amplifier may be implemented with N identical cascaded dual-gain stages. In this case a precise logarithmic response results if each amplifier has a threshold E_T below which the gain is a fixed value G and above which the incremental gain is unity. The intersections of the approximating segments fall on a curve described by

$$E_0(M) = E_T G \frac{\log E_i(M) G^N}{\log G} \tag{3.8}$$

where $E_i(M)$ represents the input levels corresponding to the intersection of the line segments:

$$E_i(M) = E_T G^{1-M} \tag{3.9}$$

The independent variable M is the order of the stage that is at incipient saturation; M will have integer values between 1 and N. In the case of the dual-gain logarithmic amplifier, all intersections of the line segments fall on a logarithmic curve.

A typical logarithmic amplifier may have a dynamic range of 80 dB, derived from nine stages, with an overall bandwidth of 5 MHz or more. Typical accuracy is $\pm \frac{1}{4}$ dB over 70 dB dynamic range and ± 1 dB over the full dynamic range.

Details of a typical stage are shown in Fig. 3.14, and the voltage characteristics in Fig. 3.15. The stage consists of an attenuator bridged by a series-diode limiter and followed by an amplifier. In the absence of an input voltage, the current divides equally between the limiter diodes. The thresholds are reached when diode $D2$ carries either all or none of the current. If the voltage drops across the diodes are neglected, this occurs when

$$E_i = E_T = \pm \frac{V}{2}\left(\frac{R}{R + R_s}\right) \tag{3.10}$$

Signals above the threshold in magnitude are attenuated by a factor $R/(R + R_f)$, the reciprocal of the amplifier gain. The incremental gain in this region is therefore unity. The blocking capacitors are used because of the dc offset voltage that is characteristic of this form of limiter.

Digital Logarithm.   The trend toward digital processing requires mention of a piecewise linear digital approximation which may be accomplished after analog-to-digital conversion and digital doppler filtering to suppress clutter
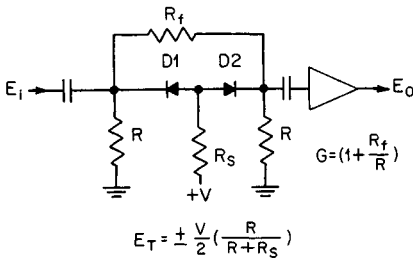
FIG. 3.14 Dual-gain IF stage.
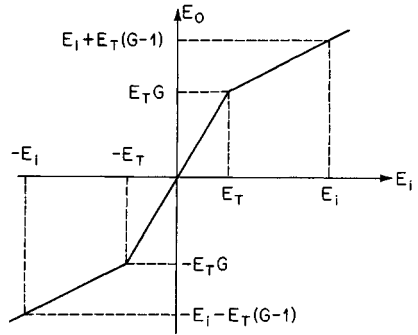
FIG. 3.15 Dual-gain-stage voltage characteristic.

interference prior to any nonlinear operation. A digital word in the power-of-2 binary format may be written

$$E = 2^{N-M}\left(b_{N+1-M} + \frac{b_{N-M}}{2} + \frac{b_{N-1-M}}{4} + \cdots + \frac{b_0}{2^{N-1}}\right) \qquad (3.11)$$

where M is the place beyond which all coefficients to the left are zero. Note that M has essentially the same significance as in previous sections. The logarithm of E to the base 2 is

$$\log_2 E = N - M + \log_2\left(1 + \frac{b_{N-M}}{2} + \frac{b_{N-1-M}}{4} + \cdots + \frac{b_0}{2^{N-1}}\right) \qquad (3.12)$$

$$\log_2 E \approx N - M + \left(\frac{b_{N-M}}{2} + \frac{b_{N-1-M}}{4} + \cdots + \frac{b_0}{2^{N-1}}\right) \qquad (3.13)$$

The whole number $(N - M)$ becomes the characteristic, and the series, a fractional number, is a line approximation of the mantissa. The approximation is accurate to ± 0.25 dB if the mantissa contains at least 4 bits. The accuracy may be improved to any desired degree by use of a programmable read-only memory (PROM) to convert the linear approximation to the true mantissa.

Digital Log Power Combiner. In log format fewer bits of data need to be stored or manipulated, and many arithmetic computations are simplified. For example, the rms combination of the voltages (I and Q) is complicated in linear format, and approximations are generally employed which introduce some error. In log format the process is simple and more accurate:

$$\log_2 I^2 = 2 \log_2 |I| \qquad \log_2 Q^2 = 2 \log |Q| \qquad (3.14)$$

$$\log_2 (I^2 + Q^2) = \log_2 I^2 + \log_2 (1 + Q^2/I^2) \qquad (3.15)$$

The latter term of Eq. (3.15) is the output of a PROM, using $\log_2 I^2 - \log_2 Q^2$ as the address.

The size of the PROM can be substantially reduced by comparing the two variables to determine the larger (L) and smaller (S):

$$\log_2 (I^2 + Q^2) = \log_2 L + \log_2 (1 + S^2/L^2) \tag{3.16}$$

The latter term of Eq. (3.16) requires fewer bits than that of Eq. (3.15) because its maximum value is 3 dB. The PROM address also is unipolar and may be limited at a ratio where the PROM output drops to zero.

The log power combiner makes power (square-law) integration feasible. A number of variables may be weighted by addition of logarithmic scaling factors and successively accumulated, using the log power combiner to combine each with the prior partial power summation. Square-law integration of multiple echoes from a target provides better sensitivity than prior methods, but this was impossible with analog signal processing and costly when using conventional digital processing.

3.9 IF LIMITERS

Applications. When signals are received that saturate some stage of the radar receiver which is not expressly designed to cope with such a situation, the distortions of operating conditions can persist for some time after the signal disappears. Video stages are most vulnerable and take longer to recover than IF stages; so it is customary to include a limiter in the last IF stage, designed to quickly regain normal operating conditions immediately following the disappearance of a limiting signal. The limiter may be set either to prevent saturation of any subsequent stage or to allow saturation of the A/D converter, a device which is usually designed to cope with modest overload conditions.

The IF phase detector described in Sec. 3.10 requires a limiter to create an output dependent on phase and independent of amplitude. It is employed in phase-lock servos and phase-monopulse receivers.

An IF limiter is sometimes employed prior to doppler filtering to control the false-alarm rate when the clutter echo is stronger than the filter can suppress below noise level. This was widely used in early two-pulse MTI, but it has drastic impact on the performance of the more complex doppler filters of modern radars. It is only compatible with phase-discrimination constant false-alarm rate (CFAR; Sec. 3.13), but it serves a useful purpose in radars utilizing this CFAR process after doppler filtering.

Characteristics. The limiter is a circuit or combination of like circuits whose output is constant over a wide range of input signal amplitudes. The output waveform from a bandpass limiter is sinusoidal, whereas the output waveform from a broadband limiter approaches a square wave.

There are three basic characteristics of limiters whose relative importance depends upon the application. They are performance in the presence of noise, amplitude uniformity, and phase uniformity. When the input signal varies over a sufficiently wide range, all these characteristics become significant. Amplitude uniformity and phase uniformity are dependent largely on the design of the limiter and are a direct measure of its quality.

Noise. Limiter performance in the presence of noise is characterized by a failure to limit signals buried in noise and by an output signal-to-noise ratio that

CHAPTER 4
TRANSMITTERS

T. A. Weil

Equipment Division
Raytheon Company

4.1 INTRODUCTION

The Transmitter as Part of a Pulsed Radar System. Figure 4.1 shows a block diagram of a typical pulsed radar system. Of these dozen blocks, the public news media generally show only the antenna and displays. The rest of the blocks are "unsung heroes," but they are equally important to the system and can be equally interesting from a design standpoint.
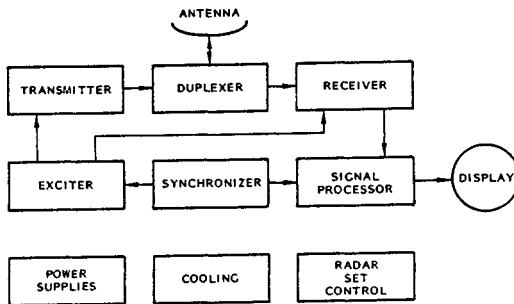
FIG. 4.1 Block diagram of a typical radar system.

The transmitter is usually a large fraction of radar system cost, size, weight, and design effort, and it typically requires a major share of system prime power and maintenance. It generally ends up being a big box that sits in the corner of the radar equipment room, hums to itself, and has a big sign on it that says, "Danger, High Voltage"; so most people prefer to keep away from it. Its insides tend to look peculiar, more like a brewery than a TV set or a computer. This chapter will try to explain why transmitters have to be what they are and hopefully will make them appear a little less peculiar to the reader.

Why So Much Power? Transmitters are big, heavy, and costly and draw so much prime power because they are required to generate so much RF power output; and that requirement, in turn, comes from the radar system design tradeoffs.

The useful range of a search radar varies as the fourth root of the product of average RF power, antenna aperture area (which determines antenna gain), and the time allowed to scan the required solid angle of coverage (which limits how long the signal in each direction can be collected and integrated to improve signal-to-noise ratio):

$$R^4 \propto P \times A \times T \tag{4.1}$$

The range varies as the fourth root of power because both the outgoing transmitted power density and the returning echo energy density from the target become diluted as the square of the distance traveled. Trying to increase range by increasing transmitter power is costly: a 16-fold increase in power is needed to double the range. Conversely, negotiating a reduced range requirement can produce remarkable savings in system cost.

Power-aperture product as a measure of radar performance is fundamental. It is so fundamental that it was explicitly mentioned in the SALT-I Treaty as the basis for limiting the capabilities of antiballistic missile (ABM) radars.

Receiver sensitivity is not shown as a factor in Eq. (4.1) because thermal noise sets a definite limit on receiver sensitivity, and this simplified range equation merely assumes that the receiver is always made as sensitive as possible.

Since average transmitter power is only one of the factors in the range equation and is so costly, why does power usually end up being so high? Wouldn't it be better to use less power and to make up for it with more aperture or more scan time? The flaw in this argument is that increasing the antenna aperture increases its cost quickly because its weight, structural complexity, dimensional tolerance problems, and pedestal requirements grow rapidly with antenna size. The only other factor, scan time, is usually set by some definite system operational requirement: to look at all aircraft within 100 mi every 4 s, for example, to permit prompt recognition of changes in aircraft direction of travel; so scan time is usually not flexible (which probably explains why everyone talks about the "power-aperture product" of a radar rather than its "power-aperture–scan-time product").

It obviously would not make sense for a radar to have a huge, costly antenna and a tiny, inexpensive transmitter, or vice versa, because doubling the tiny part would allow cutting the huge part in half, which would clearly reduce total system cost. Thus, minimizing total system cost requires a reasonable *balance* between the costs of these two subsystems. The result, for any nontrivial radar task, is that significant transmitter power is always demanded by the system designers.

The same result occurs when the system design is based on a required range coverage in the presence of standoff jammers (rather than just thermal noise).

For detection of a target carrying a self-screening jammer, the range equation becomes

$$R^2 \propto (P_r \times A_r) / (P_j \times A_j) \tag{4.2}$$

where P_r and A_r are the power and aperture of the radar and P_j and A_j are those of the jammer. The result is very similar: power and aperture are still the driving

factors, and a balanced system design again results in significant transmitter power.

The inescapable conclusion is that "It's watts up front that count." The desire to attain maximum radar performance capability thus means, more often than not, that *both* the antenna size and the transmitter power are pushed to the maximum affordable.

Pushing the transmitter design to the maximum affordable power is not without its problems, however. Historically, this pressure has often led to problems in development time, unexpected costs, and other risks, especially when a new RF tube had to be developed for the application. The AN/FPN-10 L-band beacon radar system development, for example, was never completed because the tube vendor was unable to make the magnetron stable enough over the wide range of duty cycle. The ballistic missile early-warning system (BMEWS) radar development was in similar danger until a second (backup) tube development contract was placed that used integral vacuum cavities rather than external cavities for the high-average-power klystron development.[1] Even a "successful" RF tube development may end up with a design that is marginal in arcing rate and/or in cooling design, leading to reliability problems, excessive maintenance and logistics costs, and an unhappy customer.

As a result of the risks of pushing RF tube developments to (or unwittingly beyond) the state of the art, and especially if the desired power is known to be beyond the capabilities of a single tube, it becomes attractive to use more than a single RF tube and to combine their RF outputs; this turns out to be a very practical approach, as will be discussed later (Sec. 4.5). This ability to combine, readily and reliably, is also what makes solid-state transmitters practical, since individual solid-state RF devices have much lower power-handling capability than single RF tubes. Combining a few RF tubes to obtain a needed high-power level adds complexity to a transmitter, of course; but, on the other hand, combining a *large* number of RF devices, as must typically be done for solid-state transmitters, leads to certain advantages, such as graceful degradation and improved reliability, as described in Chap. 5.

Why Pulsed? Radar transmitters would be much less complex and costly if they could simply operate CW (continuous-wave) like broadcast stations. Having to generate very high pulsed RF power leads to much higher operating voltages (both dc and RF), energy storage problems, and the necessity for high-power switching devices. Some RF devices, like Class C amplifiers (tube or solid-state), are *self-pulsing* and draw current only when RF drive is applied, but most microwave tubes require some type of pulse modulator (Sec. 4.8) so they won't waste power and so they won't generate interfering noise during the receiving period between pulses.

Basically, pulsing is used because it's hard to hear while you're talking (not everyone at meetings seems to understand that point). In a radar system, if the transmitter is always on, it is very hard to keep the transmitter from interfering with the receiver that is trying to hear faint echoes from distant targets. CW radars have been made to work by using separate transmit and receive antennas to isolate the receiver from the transmitter. When the two antennas cannot be widely enough separated to reduce transmitter leakage into the receiver below the receiver noise level (such as when both antennas have to be on the same vehicle), the residual transmitter leakage can be reduced by *feedthrough nulling*, which works by using negative feedback at the receiver input to cancel whatever transmitter carrier signal may appear there. The feedback loop must be selective

enough, however, to cancel only the carrier, since signals offset from the carrier include the desired target doppler signals. As a result, a fundamental limitation in CW radar system sensitivity is that leakage into the receiver of transmitter noise sidebands (resulting from imperfect transmitter stability) sets a limit below which small moving-target signals cannot be seen; the maximum range capability of CW radars is often limited by this factor.

A pure CW radar system can detect moving targets by their doppler offset, but no range information is obtained. The normal solution to that problem is to use an FM–CW system, in which the transmitted frequency is swept (usually linearly versus time) so that both range and doppler information can be extracted by proper interpretation of the received signals;[2] the frequency of the echo determines how long ago the signal was transmitted and thus the range to the target. Nevertheless, one fundamental limitation in such CW radars is that weak echoes from distant targets must compete with strong echoes from short-range clutter. This requires superb clutter cancellation, which in turn is limited by transmitter instabilities (which produce noise sidebands). In other words, strong short-range clutter effectively adds more transmitter leakage into the receiver.

In pulsed radar systems, short-range and long-range echoes arrive at different times, and receiver sensitivity can be adjusted accordingly with STC (sensitivity time control). Note that high-PRF pulse doppler systems, which also receive signals from multiple ranges simultaneously, suffer the same type of limitation as CW radars, so long-range pulse radars seldom use a continuous pulse doppler waveform. However, much of the same benefit of wide unambiguous doppler coverage can be obtained by a compromise waveform called *burst*, in which a finite group of high-PRF pulses is transmitted; the duration of the burst is made short enough to avoid making long-range target echoes compete with short-range clutter echoes.

A further disadvantage of CW radar is that it requires two antennas, which effectively "wastes" 3 dB of range-equation performance that could be gained if that total aperture area were combined into a single antenna and used for both transmit and receive. Pulsed radar does exactly that; it shares a single antenna for both the transmitter and the receiver by using a *duplexer*,[2] as shown in Fig. 4.2.

A gas-tube duplexer (Fig. 4.2a) uses the presence of high power during transmit to fire the gas-filled T/R (transmit/receive) "tubes," which are actually just sections of transmission line filled with a low-breakdown-voltage gas, to direct the transmitter power to the antenna. The T/R tubes recover (deionize) quickly after the transmitted pulse, which then allows received signals to flow to the receiver. A limiter is also used, as shown, to pro-
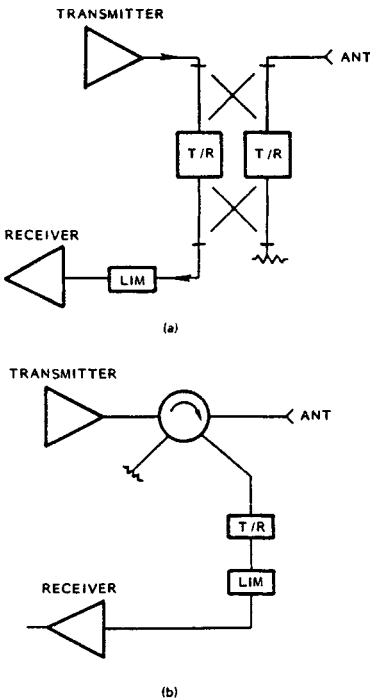
FIG. 4.2   Duplexers. (a) Gas-tube duplexer. (b) Ferrite circulator duplexer.

tect the receiver from power leakage through the T/R tubes during transmit. The limiter also protects the receiver from signals from nearby radars that may not be strong enough to fire the T/R tubes but could be large enough to hurt the receiver.

A ferrite duplexer (Fig. 4.2*b*) uses a ferrite circulator,[3] instead of T/R tubes, to send the transmitter power to the antenna and the received signals to the receiver. However, in this case reflected power from the antenna voltage–standing-wave ratio (VSWR) during transmit also is directed to the receiver, so a T/R tube and limiter are still required to protect the receiver during transmit.

In either case, the duplexer accomplishes the objective of letting the transmitter and the receiver share a single antenna in pulsed radar systems.

4.2 MAGNETRON TRANSMITTERS

Historically, the invention of the microwave magnetron during World War II made pulsed radar practical, and early radar systems were undoubtedly tailored around what magnetrons could do. The 5J26, for example, has been used in search radars for over 40 years. It operates at L band and is mechanically tunable from 1250 to 1350 MHz. It is typically operated at 500-kW peak power with 1-μs pulse duration and 1000 pulses per second (pps), or 2 μs and 500 pps, either of which is 0.001 duty cycle and provides 500 W of average RF power. Its 40 percent efficiency is typical for magnetrons. The 1- to 2-μs pulse duration provides 150- to 300-m range resolution and is "convenient" for magnetrons, which simply oscillate at the resonant frequency of their mechanical cavities and are subject to frequency instabilities that would be unacceptable compared with the narrower signal bandwidth of longer pulse widths.

Magnetron transmitters are well described in the literature.[2] They readily produce high peak power; and they are quite small, simple, and low in cost. Pulsed magnetrons vary from a 1-in[3], 1-kW peak-power beacon magnetron up to several megawatts peak and several kW average power, and CW magnetrons have been made up to 25 kW for industrial heating. All commercial marine radars have used magnetrons.

Magnetron transmitters have been widely used for moving-target indication (MTI) operation, typically allowing 30 to 40 dB of clutter cancellation. It is remarkable that magnetrons are stable enough for MTI operation at all, considering that it requires the self-excited magnetron to repeat its frequency, pulse to pulse, within about 0.00002 percent. The starting RF phase, however, is arbitrary on each pulse as the magnetron starts to oscillate, so a locked coho (coherent oscillator)[4] or an equivalent (which measures phase on transmit and corrects in a signal processor on receive) must always be used. The high-voltage power supply (HVPS) and pulse modulator must provide very stable (repeatable) pulsing to the magnetron, as well, in order not to spoil MTI performance. Modulation of magnetron frequency by microphonics, from ambient vibration, has also been a limiting factor in some cases.

Automatic frequency control (AFC) is typically used to keep the receiver tuned to the transmitter as the magnetron slowly drifts with ambient temperature and self-heating. The AFC can be applied to the magnetron instead to keep it operating on an assigned frequency, within the accuracy limits of its tuning mechanism.

Limitations. In spite of their wide capabilities, magnetrons may not be suitable for various reasons:

1. If precise frequency control is needed, better than can be achieved through the magnetron tuner after allowing for backlash, warmup drift, pushing, pulling, etc.

2. If precise frequency jumping is required, or frequency jumping within a pulse or within a pulse group.

3. If the best possible stability is required. Magnetrons are not stable enough to be suitable for very long pulses (e.g., 100 μs), and starting jitter limits their use at very short pulses (e.g., 0.1 μs), especially at high power and lower frequency bands.

4. If coherence is required from pulse to pulse for second-time-around clutter cancellation, etc. Injection locking has been tried but requires too much power to be attractive. For the same reason, combining the power outputs of magnetrons has not been attractive.

5. If coded or shaped pulses are required. A range of only a few decibels of pulse shaping is feasible with a magnetron, and even then frequency pushing may prevent obtaining the desired benefits.

6. If lowest possible spurious power levels are required. Magnetrons cannot provide a very pure spectrum but instead produce considerable electromagnetic interference (EMI) across a bandwidth much wider than their signal bandwidth (coaxial magnetrons are somewhat better in this respect).

Magnetron Features.   Where a magnetron is suitable, it can be obtained with features that have broadened considerably since its early days.

Tuners.   High-power magnetrons can be mechanically tuned over a 5 to 10 percent frequency range routinely, and in some cases as much as 25 percent.

Rotary Tuning.   The rotary-tuned ("spin-tuned") magnetron was developed around 1960.[5,6] A slotted disk is suspended above the anode cavities as shown in Fig. 4.3 and, when rotated, alternately provides inductive and capacitive loading of the cavities to raise and lower the frequency. Very fast tuning rates are feasible because each revolution of the tuner disk tunes the tube across the band and back a number of times equal to the number of cavities around the anode. The disk is mounted on bearings inside the vacuum (developed first for rotating-anode x-ray tubes) and is magnetically coupled to a shaft outside the vacuum. At 1800 r/min, a tube with 10 cavities tunes across the band 300 times per second. By ensuring that the modulator pulse rate is not synchronous with the tuning rate, the transmitted frequency will vary from pulse to pulse in a regular pattern as the PRF beats with the tuning rate. Irregular (pseudorandom) jumping of the frequency can be obtained by varying the modulator PRF or by varying the motor speed rapidly. First-order tracking information for the receiver local oscillator (LO) is obtained from an
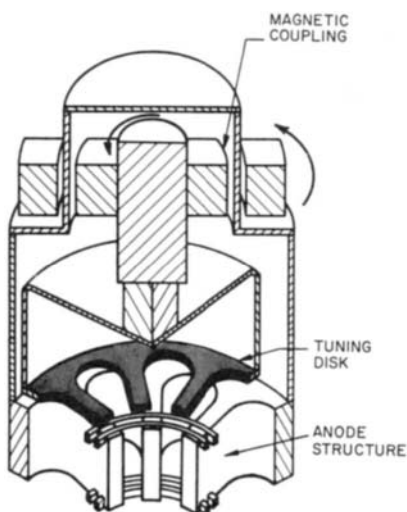
FIG. 4.3   Magnetron rotary tuner. (*Courtesy of Raytheon Company.*)

internal transducer, usually capacitive, on the same shaft as the tuning disk.

The use of rotary tuners involves some penalties besides higher cost and weight. Less average power output is feasible than for tubes with conventional tuners, since cooling the rotary tuner is more difficult. Precise band-edge tuning is not assured; since the entire tuning range is always covered on each cycle and since system operation outside the assigned band is usually not permissible, tolerances on tuning range must be absorbed within the band. When used for MTI (with the tuner stopped), stability is less good than with other tuners.

Stabilized Magnetrons. The most common form of stabilized magnetron is the coaxial magnetron, in which a high-Q annular cavity is intimately coupled to the anode vanes inside the inner cylinder, as shown in Fig. 4.4. At the higher frequencies (above X band) an inside-out version, called an inverted coaxial magnetron, as shown in Fig. 4.5, is more suitable because the cavity becomes very small and the normal construction would leave inadequate room for the cathode and the anode structure.
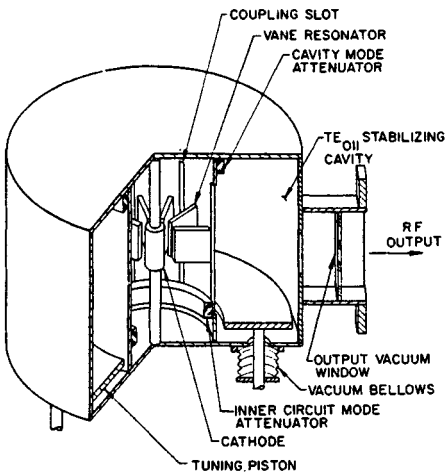
FIG. 4.4 Coaxial magnetron. (*From Ref. 8.*)
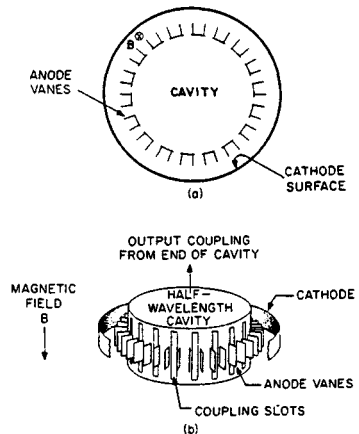
FIG. 4.5 Inverted coaxial magnetron. (*a*) Simplified cross section. (*b*) Simplified perspective.

These techniques[5,7,8] permit an increase in stability by a factor of 3 to 10 in pushing figure and pulling figure (defined later in this section). This is of most importance at high frequency (X and K bands), where the effects of pushing and pulling are more significant compared with the bandwidth occupied by typical system pulse widths. Stabilization is also most practical at these frequencies because a high-Q cavity is then of acceptable size. MTI performance may be improved over conventional magnetrons because the pulse-to-pulse and intrapulse frequency stability is better. However, the expected benefit may not be realizable unless jitter and noise during the starting time of each pulse are low enough, and these characteristics vary considerably among the different types of stabilized magnetrons.

Common Problems. The classic problems of magnetron operation still exist, but they now tend to be better understood, specified, and controlled. Briefly, the most common problems are as follows:

1. *Sparking:* Especially when a magnetron is first started, it is normal for anode-to-cathode arcing to occur on a small percentage of the pulses. Sometimes this also applies to moding and/or misfiring. The modulator must tolerate this for brief periods without tripping off and must deliver normal output immediately following sparking.

2. *Moding:* If other possible operating-mode conditions exist too close to the normal-mode current level, stable operation is difficult to achieve. Starting in the proper mode requires the proper rate of rise of magnetron cathode voltage, within limits that depend on the tube starting time and the closeness of other modes. Too fast a rate of rise may even result in failure to start at all. Since starting time is roughly $4Q_L/f_o$, where Q_L is the loaded Q, it is more difficult and inefficient to operate high-power, low-frequency magnetrons at short pulse lengths. Techniques to minimize this problem include *despiking*, usually a simple series-*RC* network to slow down the modulator voltage rate of rise, or a *pulse bender*, which uses a diode and a parallel-*RC* network to load down only the last portion of the rate of rise.[9] On the other hand, too slow a voltage rate of rise (or too slow a fall time at the end of the pulse) can also excite a lower-current mode if that tube has one.[10]

3. *Noise rings:* Excessive inverse voltage following the pulse, or even a small forward "postpulse" of voltage applied to the magnetron, may make it produce sufficient noise to interfere with short-range target echoes. The term *noise ring* is used because this noise occurs at a constant delay after the transmitted pulse and produces a circle on a plan position indicator (PPI). This can also occur if the pulse voltage on the magnetron does not fall fast enough after the pulse.[11]

4. *Spurious RF output:* In addition to their desired output power, magnetrons generate significant amounts of spurious noise. The kinds and amounts are similar to those listed under crossed-field amplifiers (CFAs) in Table 4.2 below and discussed in Secs. 4.3 and 4.4, but the resonant nature of the magnetron tends to suppress noise that is far from the operating frequency, except for harmonics.

5. *RF leakage out of the cathode stem:* Typically, an S-band tube may radiate significant VHF and UHF energy as well as fundamental and harmonics out of its cathode stem. This effect varies greatly among different magnetrons, and when it occurs, it also varies greatly with lead arrangements, filament voltage, magnetic field, etc. Although it is preferable to eliminate cathode-stem leakage within the tube, it has sometimes been successfully trapped, absorbed, or tolerated outside the tube.

6. *Drift:* Magnetron frequency varies with ambient temperature (of the cooling air or water) according to the temperature coefficient of its cavities, and it may also vary significantly during warmup. Even during continuous operation, a change in tuner setting may result in drifting again after the change if cavity or tuner heating varies with tuner setting. Temperature-compensated designs are available in some cases.

7. *Pushing:* The amount by which a magnetron's frequency varies with changes in anode current is called its *pushing figure*,[10] and the resulting pulse-to-pulse and intrapulse frequency changes must be kept within system requirements by proper modulator design.

8. *Pulling:* The amount by which a magnetron's frequency varies as the phase of a mismatched load is varied is called its *pulling figure*.[10] Thanks to the ready availability of ferrite isolators, pulling is seldom a problem in modern radar

transmitters. For the same reason, long-line effect[12] is a problem of the past, since isolators readily reduce the mismatch seen by the magnetron to a value low enough to guarantee freedom from frequency skipping.

9. *Life:* Although some magnetrons have short wear-out life, many others have short life because of mishandling by inexperienced personnel. Dramatic increases in average life have been obtained by improved handling procedures and proper operator training (Sec. 4.4).[13,14]

10. *Tuner life:* Because of cost and size tradeoffs, tube life may be limited by the finite fatigue life of the bellows required to allow actuating tuners inside the vacuum. Tuners that operate outside the vacuum must still have adequate gear and bearing design if they are not to limit tube life; in particular, backlash may be a limit.

4.3 AMPLIFIER CHAIN TRANSMITTERS

It was the limitations of magnetrons that eventually pushed radars to use amplifier chain transmitters, which are more capable but also more complex. The key difference is that the transmitter signal is generated at low level, as precisely as desired, and is amplified all the way from there to the required peak power level. As shown in Figs. 15.44 and 15.45, the change in the system block diagram is small, just the direction of two arrows and the change from oscillator to amplifier for the high-power RF source; but there is a *huge* difference in the hardware required to implement the amplifier chain system, including many stages of RF amplification, each with its own power supplies, modulator, and controls; and *all* these stages must be stable to achieve good system MTI performance (Chap. 15).

Oscillator versus Amplifier. Amplifier chain systems can readily achieve full coherence from pulse to pulse and can provide all the features that pulsed oscillator systems (usually magnetrons) cannot provide: coded pulses, true frequency agility, and combining and arraying. The price is higher system complexity and cost. Thus, "oscillator versus amplifier" is one of the basic choices that must be made early in radar system design. Some of the factors entering into this choice are given below.

Accuracy and Stability of Carrier Frequency. In an oscillator-type transmitter, the RF power tube determines its own operating frequency, as opposed to having it determined by a separate low-power stable oscillator. The frequency may thus be affected by tube warmup drift, temperature drift, pushing, pulling, tuner backlash, and calibration error. In an amplifier chain type of transmitter, the frequency accuracy is essentially equal to that of its low-level stable crystal (or other) oscillator. Furthermore, the frequency of the amplifier chain can be changed instantaneously by electronic switching among several oscillators, at a rate faster than that of any mechanical tuner.

Coherence. An amplifier chain system can generate its LO (local oscillator) and coho [coherent intermediate-frequency (IF) oscillator] signals with precision, whereas an oscillator-type transmitter requires manual tuning or an automatic frequency control (AFC) servosystem to tune the LO to the correct frequency. Since an oscillator-type transmitter starts each pulse at an arbitrary phase angle with respect to the coho and LO, coho locking must be provided; in an amplifier system, coho locking is inherent in the signal generation process. Furthermore,

since phase coherence can be maintained over a train of pulses in an amplifier-type transmitter, second-time-around clutter can also be canceled, whereas in an oscillator-type system, second-time-around clutter will be noise-modulated by the random starting phase of the oscillator tube. Amplifier chains also allow full coherence, in which the PRF, IF, and RF frequencies are all locked together; this is sometimes necessary to keep PRF harmonics out of IF doppler bands.

Instabilities.   As discussed in Secs. 4.6 and 15.11, different kinds of instabilities are associated with a pulsed oscillator system and a pulsed amplifier chain. For the oscillator, pulse-to-pulse *frequency* stability depends on HVPS ripple, and intrapulse *frequency* changes depend on modulator droop and ringing. Tolerable limits are shown in Table 15.4, but these limits may be loosened if coho locking is based on an effective average of the transmitter frequency during the pulse length. For an amplifier chain, pulse-to-pulse *phase* stability depends on HVPS ripple, and intrapulse *phase* variations depend on droop and ringing; tolerable limits are also shown in Table 15.4.

An interesting compromise is also feasible: if a locked coho is used with an amplifier chain, then pulse-to-pulse phase variations in the chain are not significant (except on second-time-around clutter). This technique is particularly convenient when a CFA power booster is added to an existing pulsed oscillator MTI radar system; simply by changing the point at which the RF sample is taken for locking the existing coho, the added CFA is not required to have tight pulse-to-pulse phase stability.

A digital equivalent of the locked coho has also been used. The phase of the transmitter is simply measured on each pulse, and the proper correction is made on the received signals in the signal processor. Like the locked coho, this technique is not effective on second-time-around targets.

Amplifier Chains: Special Considerations.   The decision to use an amplifier chain, most often for coherence and agility, introduces many complications, some of which are noted here.

Timing.   Because modulator rise times differ, triggers to each amplifier stage must usually be separately adjusted to provide proper synchronization without excessive wasted beam energy. In CFA chains, allowance must also be made for the pulse-width shrinkage that occurs because of the necessary overlap of RF drive, as noted in Sec. 4.4.

Isolation.   Each intermediate stage of a chain must see proper load match even if the following stage has high VSWR input, as in a typical broadband klystron, or even if it has significant reverse-directed power coming back from it, as is the case with CFAs. This reverse-directed power results from mismatch at the CFA output that sends power back through the low-loss structure of the CFA. For example, a load with 1.5:1 VSWR reflects power 14 dB down. At certain frequencies, this reflected power will combine with reflections inside the tube and may typically return to the input of the CFA at a power level only 8 dB down from full output power. This amount of reverse-directed power is 2 dB *greater* than the RF input power arriving at that point even if the CFA has only 10 dB of gain. Although this does not seem to interfere with normal CFA operation, it does require an isolator at the CFA input with 16 dB isolation, in this case, just to bring the VSWR seen by the previous stage down to 1.5:1.

Matching.   RF tubes used in amplifier chains are often more "fussy" about the match they see than oscillator tubes. Because good isolators are now generally available, improved amplifier ratings are sometimes available if the tube is guaranteed to see a good match, such as 1.1:1. Furthermore, CFAs and traveling-wave tubes (TWTs) generally require that the match they see be controlled over

a much wider range than the specified operating frequency band to ensure that the amplifier tube will remain stable.

Signal-to-Noise Ratio. The noise power output of an amplifier tube may be significant. When several tubes are connected as a chain, the output signal-to-noise ratio cannot be better than that of the worst stage. For this reason the input stage, especially, must be checked to see if it has an adequately low noise figure; otherwise, it may prevent the entire chain from achieving a satisfactory signal-to-noise ratio. For example, a low-level TWT with 0.5-mW RF signal input and 35 dB noise figure will limit signal-to-noise ratio of the amplifier chain to 74 dB in a 1-MHz bandwidth. Conventional CFAs have higher noise levels than linear-beam tubes, and their signal-to-noise ratio is typically only 55 dB in a 1-MHz bandwidth; the low-noise CFA, however, can be 70 dB or better (Sec. 4.4).

Leveling. In a multistage chain of linear-beam tubes, the performance of each tube depends in part on the performance of the tubes preceding it. In particular, power flatness (constant power output across the frequency band) requires careful specification of flatness for each stage in the presence of a suitable allowance of nonflatness of the stage preceding it. For example, the *saturated gain* of a tube may be constant across the band, and yet the power output may vary considerably across the band with constant RF drive. Saturated gain is measured by varying the drive at each frequency until the point of maximum power output is found; at that point, saturated gain is the ratio of RF power output divided by RF power input. Unless the saturated power output is constant over the band, the saturated gain bears little relationship to power flatness across the band with constant RF drive. Nor does flat small-signal gain indicate power flatness at large-signal conditions. Therefore, it is usually best to specify that the tube be tested in a way that will ensure proper performance in the system, including adequate tolerances on the RF drive.

Naturally, the transmitter gain and leveling plan must cover all losses and tolerances of components between the stages as well as the tube tolerances. It is also feasible to consider passive frequency-shaping networks to compensate for known deviations from flatness in the RF tube characteristics.

In CFA chains, leveling is far simpler because excess drive power is harmless (it just feeds through and adds to the output),[15] and it is only necessary to ensure that there is always adequate drive power.

Stability Budgets. In a multistage chain, each stage must have better stability than the overall requirement on the transmitter, since the contributions of all stages may add. They may add randomly or directly, or in certain cases they may be arranged to cancel, depending on the nature and source of the instabilities. Normally it is necessary to subdivide the transmitter stability requirement into several smaller numbers that are then allocated to each stage according to *degree of difficulty*. Such *stability budgets* are usually required for pulse-to-pulse variations, for intrapulse variations, and sometimes for phase linearity. Jitter is usually dependent primarily on a single stage and is therefore usually not budgeted among stages.

RF Leakage. A typical amplifier chain may have 90 dB of gain at the transmitter frequency in one shielded room or one location. In order to keep the chain from oscillating, leakage from the output of the chain back to its input must clearly be at least 90 dB down. However, a more stringent requirement is that RF leakage into the input stage of the chain must be kept below the desired level of MTI *purity* with respect to the signal level at that point, since the leakage path might conceivably be modulated by fan blades, cabinet vibration, etc. The leakage feedback will also affect pulse compression sidelobe levels. Since a typical

level of purity desired for MTI or pulse compression might be 50 dB, this leads to an isolation requirement of 140 dB from chain output to input. Since typical waveguide joints and coaxial-cable connectors may have leakage levels of the order of −60 dB, 140 dB of isolation can be difficult. Other contributors to amplifier chain RF leakage problems often include collector seals on linear-beam tubes and cathode stems on CFAs. Successful amplifier chain design therefore requires conscious and careful control of RF leakage.

Reliability. The complexity of transmitter amplifier chains often makes it difficult to achieve the desired reliability. Solutions usually involve the use of redundant stages or a whole redundant chain, and many combinations of switching are feasible. Careful analysis and restraint are usually necessary; otherwise, the complexity and cost of fault monitoring and automatic switching very quickly grow out of bounds. Appropriate design for acceptable reliability involves trading off various serial and redundant transmitter chain and switching alternatives, but such system-reliability calculations are beyond the scope of this handbook.

RF Amplifiers. Successful amplifier chain transmitter design depends upon the availability of suitable RF amplifier devices or the feasibility of developing them. Since solid-state transmitters are covered in Chap. 5, we will limit this chapter to discussions of RF tubes for radar systems, as described in the next section.

4.4 RF AMPLIFIER TUBES

Until the mid-1970s, radar transmitters used only vacuum tubes of one kind or another for microwave power generation. The earliest systems all used magnetrons, as has been noted, and amplifier chain system development had to await the development of suitable high-power pulsed-amplifier tubes. Although many varieties were developed, the successful kinds were klystrons, TWTs, and CFAs. Triodes and tetrodes[2] have also been used in radars at frequencies below 600 MHz.

Klystrons and TWTs are called *linear-beam tubes* because the direction of the dc electric field that accelerates the electron beam coincides with the axis of the magnetic field that focuses and confines the beam. This is in contrast to *crossed-field tubes*, such as magnetrons and CFAs, in which the electric and magnetic fields are at right angles to each other.

Since there are a number of excellent references that describe the theory and operation of RF amplifier tubes,[2-4,16-18] this discussion will be limited primarily to system considerations in selecting and using microwave amplifier tubes in radar transmitters.

Crossed-Field Amplifiers (CFAs). High efficiency, small size, and relatively low-voltage operation make CFAs especially attractive for lightweight systems for transportable or airborne use, from UHF to K band. Having relatively low gain, CFAs are generally used only in the one or two highest-power stages of an amplifier chain, where they may offer an advantage in efficiency, operating voltage, size, and/or weight compared with linear-beam tubes. The output-stage CFA is usually preceded by a medium-power TWT that provides most of the chain gain. CFAs have also been used to boost the power output of previously existing radar systems.

The dominant types of CFAs are all reentrant, distributed emission

CFAs.[4,16-18] The high-gain CFA[19] was not developed until 1987, but it is very attractive, both because it requires less drive power and because the presence of RF drive at the cathode results in lower noise levels.

Backward-wave CFAs were developed first and were applied first (the amplitron).[15] In backward-wave devices, the voltage required for a given peak current is essentially proportional to frequency, but this can be accommodated by the inherent contant-power characteristic of a line-type modulator or by a hard-tube modulator operated in the constant-current region. The constant-current switch tube also helps to regulate CFA current against HVPS capacitor-bank-voltage droop. Forward-wave CFAs, which were developed later, operate at nearly constant voltage across their frequency band and can therefore be considered for *dc operation*, which requires only a control electrode (see below) instead of a full-power pulse modulator.
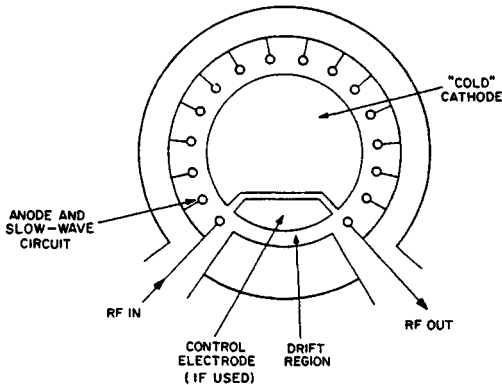
FIG. 4.6 Drift region and control electrode in a reentrant CFA.

Some CFAs have cold cathodes and are started by applying RF drive. The RF drive power must be applied in time to permit the tube to start drawing current before the cathode voltage pulse overshoots the proper operating voltage. However, even when there is a drift region in the CFA, as shown in Fig. 4.6, the tube will not stop when RF drive is removed; the reentrant electrons still carry enough energy that secondary emission from the cathode is maintained, and the tube will oscillate near a band edge or generate broadband noise until the cathode voltage pulse ends. In addition, once operation has been started by RF drive, back bombardment heats the cathode, and on following pulses the cathode current may start from thermionic emission even before RF drive is applied. Since this would also produce noise output, it is customary to make the RF drive pulse straddle the modulator voltage pulse to prevent this. Allowance must be made for the resulting pulse-width shrinkage in an amplifier chain with one or more CFAs. The output pulse will also have "pedestals" owing to feedthrough of the wider RF drive pulse length, as shown in Fig. 4.7.

A control electrode[20,21] usually consists of a segment of the cathode structure in the drift region, as shown in Fig. 4.6. The control electrode is pulsed positive

POWER
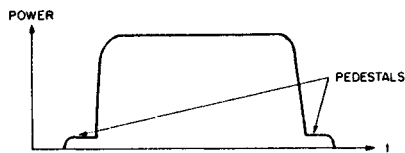
FIG. 4.7 Pedestals on a CFA RF output pulse.

(with respect to the cathode) at the end of the RF pulse to collect the electrons passing through the drift region and thereby make the tube turn off even though high voltage is still applied.[22] Turnoff control electrodes in CFAs thus make possible dc operation, which eliminates the high-power modulator. In dc operation, the high voltage is continuously present between anode and cathode, and the current is turned on by applying RF drive and turned off by pulsing the control electrode. To prevent the tube from starting without RF drive, the cathode must be kept cool enough to prevent thermionic emission. The control electrode requires only a short, medium-power pulse, typically one-third of the anode voltage and one-third of the anode peak current. The greatly reduced modulator requirements for dc operation make it practical to use more complex pulse coding. However, some energy is dissipated on the control electrode each time it is pulsed, and since it is an insulated electrode, it is difficult to cool. Control-electrode heating can therefore be a limitation on the maximum PRF that may be used.

In practice, dc operation has seldom been used[23] because it requires a much larger capacitor bank to limit droop (as opposed to using a switch tube in the constant-current mode) and because an arc in a dc-operated CFA requires crowbarring (Sec. 4.9), which interrupts operation for a few seconds instead of only for a single pulse; it has not been possible to make CFAs arc-free. Problems have also occurred as a result of adjacent radars injecting enough RF energy into the system antenna and back into the transmitter to turn on a dc-operated CFA at the wrong times.

The low insertion loss of CFAs from RF input to RF output without modulator voltage applied permits convenient programming of CFA amplifier chain power output in steps.[24] For example, in an amplifier chain transmitter with two CFAs preceded by a TWT, three power-output levels can be selected simply by choosing which modulators to pulse. Power programming, also called *feedthrough operation*, is especially useful in 3D radar applications, since it permits conserving average power by reducing the peak power output at high scan angles.

The low insertion loss of a CFA also allows power reflected at the output to be passed back through the tube to its input; in many cases the reflected output power coming back out of the input may even exceed the incoming drive power. A properly rated isolator[3,25] is thus a necessity between stages of a CFA chain.

Certain additional problems long identified with magnetron operation are also common to operation of CFAs. For details, see paragraphs on sparking, moding, noise rings, spurious RF output, and RF leakage, discussed under "Common Problems" in magnetrons (Sec. 4.2). One difference is that because RF drive is present during the voltage rise time, many (but not all) cathode-pulsed CFAs allow a much faster voltage rate of rise than magnetrons. For the same reason (i.e., RF drive is already present) there is little starting-time delay in the desired CFA operating mode; but the π-mode oscillation has a finite starting time and will produce little energy if the voltage passes quickly enough through the range in which it can occur. In a dc-operated CFA, the π-mode oscillation should not occur at all because the cathode voltage is at full value all the time.

Klystrons. The multicavity klystron has always been known for its high-

gain and high-power capability. However, its bandwidth during the 1950s tended to be 1 percent or less, with wider ranges being covered by mechanical tuning of the cavities; *gang tuning* (tuning all cavities at once in response to rotation of a single tuning knob or motor drive) is often used. Although a tradeoff between klystron gain and bandwidth was always known to be feasible, the stagger tuning of a klystron is far more complex than that of an IF strip. The overall frequency response of a klystron contains intermediate gain products as well as the total product of the individual cavity responses; certain tuning combinations produce excessive harmonic output; and broadband small-signal gain does not ensure broadband saturated gain. Klystron bandwidth capability increases with power level[16] because the stronger beam provides heavier loading of the cavities.

Modern digital computers made it possible to determine improved cavity-tuning arrangements, and klystron bandwidth improved greatly; 8 percent bandwidth 3 dB down) has been obtained with fixed cavity settings, and even 11 percent in rare cases (Varian VA-812C). The achievement of this bandwidth in klystrons also depended partly on improvements in beam perveance, but, more important, it required progress in output-cavity design, because the *power bandwidth* can be no better than the ability of the output cavity alone to extract the energy from the beam, regardless of the gain or drive power available preceding it. Single-cavity output circuits are therefore replaced in broadband klystrons by double-tuned and triple-tuned cavities, sometimes called an *extended-interaction circuit*,[26,27] which uses more than one interaction gap to extract energy from the beam, as shown in Fig. 4.8. This technique of grouping cavities was later extended to the prior cavities as well, and by discovering that the cavities in each group need not be coupled to each other, the *clustered-cavity klystron*[28] has achieved as much as 20 percent bandwidth. Although more complex and expensive than a normal klystron, the clustered-cavity klystron is still less complex and costly than a comparable TWT or Twystron.
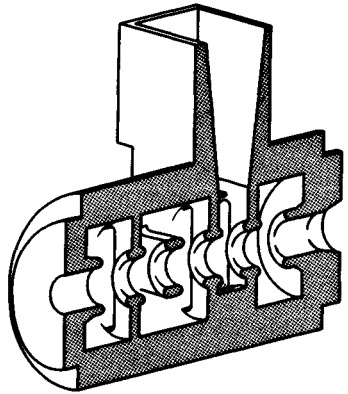
FIG. 4.8 Extended-interaction output circuit. (*From Ref. 26.*)

Traveling-Wave Tubes (TWTs). The low-power helix TWT is still the king of bandwidth. Because it has virtually constant phase velocity at all frequencies, the helix permits TWTs to have bandwidths in excess of an octave. However, the helix TWT has not been used in high-power radars because high power requires a high-voltage beam, and the resulting electron velocity is too fast to synchronize with the low velocity of the RF wave on a helix slow-wave circuit. The limit of helix tubes is about 10 kV and a peak RF power output of a few kilowatts. For higher power levels, other kinds of slow-wave circuits with a higher RF velocity must be used, and the bandpass characteristics of those circuits can lead to band-edge oscillation problems. Furthermore, both forward waves and backward waves may propagate on the RF structures, leading to the possibility of backward-wave oscillations. Depending on the circuit used, other kinds of oscillations can also occur. For
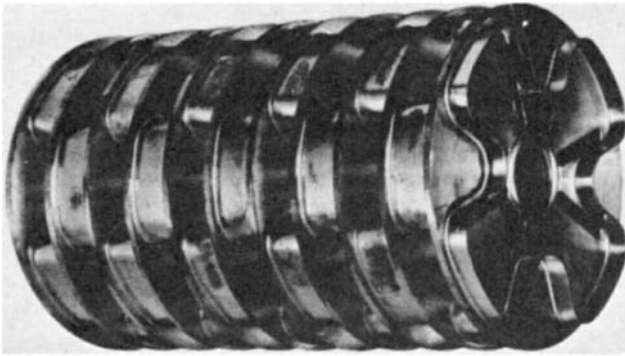
FIG. 4.9 Cloverleaf slow-wave circuit. (*Courtesy of Varian Associates.*)

these reasons, high-power TWT development lagged that of klystrons and is still more expensive. By 1963, however, Varian had produced multimegawatt pulsed TWTs using the cloverleaf circuit, as shown in Fig. 4.9,[29] a structure that can be made heavy and rugged enough to handle power comparable with that of klystrons.

Slow-wave structures for high-power TWTs include helix-derived structures (contrawound helix, or ring-bar circuit) and coupled-cavity circuits, of which the cloverleaf is one example, and the ladder network.[30] Below 100 kW, ring-bar circuits usually have broader bandwidth and higher efficiency than coupled-cavity circuits. Above 200 kW or even below that if average power is a limitation, coupled-cavity circuits dominate.[31]

If a TWT using a coupled-cavity circuit is cathode-pulsed, there is an instant during the rise and fall of voltage when the beam velocity becomes synchronous with the cutoff frequency (π mode) of the RF circuit, and the tube usually oscillates. These oscillations at the leading and trailing edges of the RF output pulse have a characteristic appearance that has given them the name *rabbit ears*, as shown in Fig. 4.10. Only in rare cases has it been possible to suppress these oscillations completely. However, since this particular oscillation depends on electron velocity, which in turn depends on beam voltage, the problem is avoided by the use of mod-anode or grid pulsing (described later in this section). In this case, it is only necessary to be sure not to let the modulator begin pulsing the beam current during turn-on of the HVPS until the voltage is safely above the oscillation range, which is typically somewhere between 60 and 80 percent of full operating voltage.
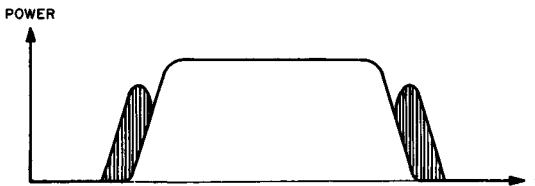
FIG. 4.10 Rabbit-ear oscillations on the envelope of the RF output from a cathode-pulsed TWT amplifier.

Discontinuities, called *severs*, are necessary in the slow-wave structure of high-power TWTs to prevent oscillation due to reflections at input and output of the structure. Although oscillation could also be prevented by distributing loss along the structure, this would result in lower efficiency, which is unattractive in high-power tubes. Typically, a sever is used for each 15 to 30 dB of tube gain. At each sever, the modulated beam carries the signal forward, while the power traveling in the slow-wave circuit at that point is dissipated in the sever loads; thus reverse-directed power is stopped at each sever. The sever loads may be placed external to the tube to reduce dissipation within the RF structure itself.

TWTs tend to be less efficient than klystrons because of the necessity for loading the structure for stability and because relatively high RF power is present in an appreciable fraction of the entire structure. One important technique for improving the efficiency of high-power TWTs is called *velocity tapering*. This technique consists of tapering the length of the last few circuit sections of the slow-wave circuit to take into account the slowing down of the beam as the energy is extracted from it. Velocity tapering permits extracting more of the energy from the beam and significantly improves the power-bandwidth performance of the tube.[31] Nevertheless, high-power TWTs generally show an appreciable falloff of power output toward the band edges, so that the rated bandwidth depends very much on how much power falloff can be tolerated by the system.

To improve the efficiency of TWTs (or klystrons), the use of depressed collectors[16,32] has been developed to a remarkably successful degree. The use of multiple collector sections at intermediate voltages allows catching each spent electron at a voltage near optimum. Up to 10 collector sections have been used in some communications tubes, but 3 sections, as shown in Fig. 4.11, are more typical for modern high-power TWTs for radar systems. The several different voltages needed for the depressed collectors add complexity to the HVPS, but fortunately these voltages need not be as well regulated as the main beam voltage.

Twystrons. In 1963 Varian assembled a hybrid tube consisting of klystron cavities in all but the output section, while a cloverleaf traveling-wave circuit was used for the output section. The purpose was to produce a more efficient version of the VA-125 broadband S-band TWT, based on the more effective beam-bunching action of the cavities. The result was not only slightly higher efficiency but also a significant improvement in bandwidth as a result of the flexibility in tuning of the cavities combined with the broad power-bandwidth capability of the TWT output section. To compensate for the inherent droop in gain of the TWT output section at the edges of the band, the klystron cavities were purposely tuned to boost the gain at those frequencies. Because it is part klystron and part TWT, Varian named the hybrid tube a Twystron.[33] A 3 dB bandwidth of 14 percent has been demonstrated in the VA-145, or a 1 dB bandwidth of 12 percent; 48 percent efficiency has been shown at midband with 41 dB gain. Although more complex and expensive than most klystrons, the Twystron appears capable of equally high power with broader bandwidth than all but perhaps the clustered-cavity klystron.

RF Tube Selection. Table 4.1 summarizes the main differences among the leading RF tube types. The factors that most often dominate in tube selection are cost, bandwidth, spurious noise level, control electrodes, gain, size, voltage, and availability (not shown in the table). Linear-beam tubes are quieter and higher-gain and can be grid-pulsed, but CFAs are smaller, lighter, lower-voltage, and less costly. Sometimes one of these factors is overwhelmingly important, and the transmitter designer is forced to accommodate all the less
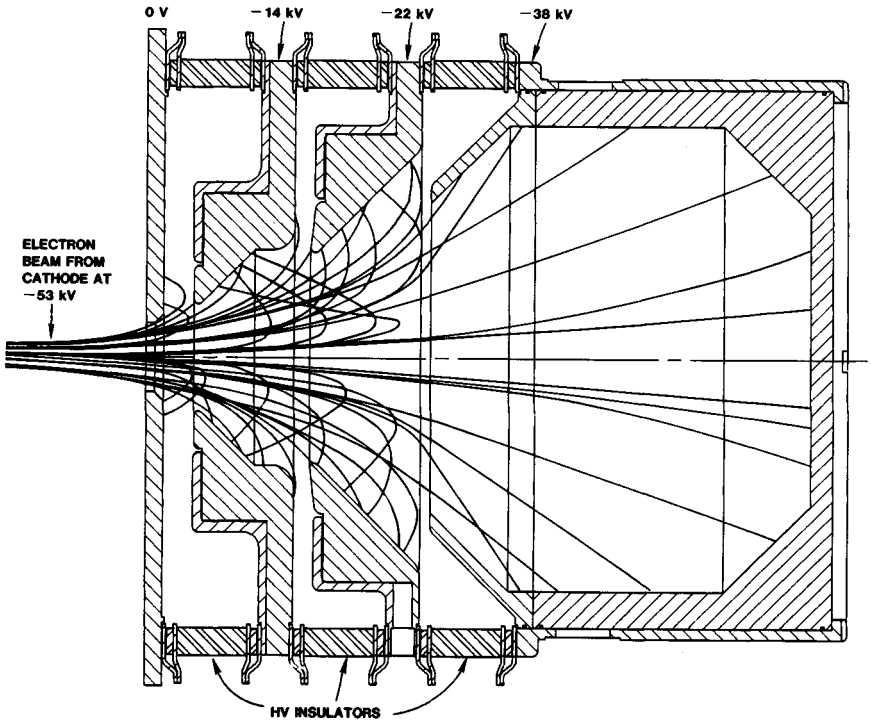
FIG. 4.11 Multiple depressed collectors. (*Courtesy of Hughes Aircraft Electron Dynamics Division.*)

significant disadvantages. Often, several choices are feasible, and tradeoffs are then considered among system cost, schedule, and performance. Further comments on Table 4.1 are given below.

Voltage.   Voltage affects the size and cost of the HVPS and modulator as well as x-ray severity.

Gain.   Gain strongly affects the number of stages and therefore the complexity (parts count, control, fault monitoring, and maintenance) required in an amplifier chain.

Bandwidth.   The bandwidth listed here is the instantaneous bandwidth of the tube, i.e., without tuning adjustments. The tube bandwidth must be compatible with system requirements; in turn, system bandwidth tends to be based on available or presumed tube capabilities.

X-rays.   X-rays affect transmitter weight because of the shielding required to protect personnel (and sensitive semiconductors).

Efficiency.   This strongly affects transmitter weight, cost, and cooling requirements as well as prime power. The numbers shown do *not* include heater power, solenoid power, or cooling-system power, which may be significant.

Ion Pump.   Residual outgassing in a microwave tube can spoil its vacuum and cause RF or dc breakdown. A Vac-Ion pump (trademark of Varian Associates) can be used to maintain a good vacuum, even during storage, and to indicate the quality of the vacuum. Most crossed-field tubes don't require ion pumps because they will pump themselves down when operated.

TABLE 4.1 High-Power Pulsed Amplifiers Compared for Same Frequency and Peak and Average Power Output

	Linear-beam tubes		Crossed-field tubes*	
	Klystron	TWT	Conventional	High-gain
Voltage	High voltage (1 MW requires approx. 90 kV)		Low voltage (1 MW requires approx. 40 kV)	
Gain	30–70 dB		8–15 dB	15–30 dB
Bandwidth	1–8%†	10–15%	10–15%	
X-rays	Severe, but lead is reliable		Not usually a problem	
Efficiency				
Basic	15–30%		35–45%	
With depressed collectors	40–60%		N.A.	
Ion pump	Required on large tubes		Self-pumping	
Weight	Higher		Lower	
Size	Larger		Smaller	
Cost	Medium	High	Medium	
Spurious noise‡	−90 dB		−55 dB	−70 dB
Spurious modes (typical)	None	π mode during rise and fall if cathode-pulsed; none if mod-anode- or grid-pulsed	π mode during rise and fall if cathode-pulsed; full-power noise output if turned on without RF drive	
Usable dynamic range	40–80 dB		A few decibels	
Control electrode	None, or mod-anode, or grid		None, or turnoff electrode	
Magnetic field	PPM up to 1 MW at S band; solenoid otherwise; barrel magnet rare (SLAC); none for ESFK		Permanent magnet	
Dynamic/static impedance	0.8		0.05–0.2	
Phase-modulation sensitivity	5–40° per 1% $\Delta E/E$		0.5–3.0° per 1% $\Delta I/I$	

*Distributed emission, reentrant, circular.
†Clustered-cavity klystron can achieve 10 to 15 percent bandwidth at higher cost.
‡In 1-MHz bandwidth.

Weight. The weights of the tubes alone are compared in Table 4.1; however, the solenoid required for high-power linear-beam tubes and their higher voltages usually make linear-beam-tube transmitters considerably heavier than CFA transmitters.

Size. The same comments apply as noted under "Weight."

Cost. The costs listed here refer only to tube costs. Also see comments under "Control Electrodes," below. The comments listed apply roughly to both development costs and unit costs.

Spurious Noise. This has become a more significant factor as radar bands become more crowded, receiver sensitivities increase, and electromagnetic compatibility (EMC) receives more attention. Spurious noise should be considered in four parts:

1. *Harmonics:* Both linear-beam tubes and CFAs produce harmonic power output of typically −25 dB (with respect to fundamental power output) at the sec-

ond harmonic, −30 dB at the third harmonic, and considerably less at higher harmonics. Although these figures vary greatly from one tube design to another, there is no strong difference between linear-beam tubes and CFAs in general. If harmonic power output is a problem, excellent high-power microwave filters are practical.

2. *Adjacent-band spurious noise:* Problems of this type may occur from adjacent modes in CFAs or in TWTs, typically appearing a few percent above (in forward-wave tubes) or below (in backward-wave tubes) the desired operating band. The problem may be severe in cathode-pulsed tubes but is avoided in tubes pulsed by a control electrode (dc-operated CFA, or mod-anode or gridded TWT). If present, adjacent-band spurious noise is usually easy to filter out unless it is too close to the desired operating band.

3. *In-band spurious noise:* This is the factor noted in Table 4.1, since it is most serious to system operation and usually cannot be filtered out. In-band spurious noise may interfere with other systems or may prevent achieving desired MTI cancellation or pulse compression sidelobe levels in the tube's own system. In-band spurious noise also sets a limit on the spectrum improvement that may be obtained by pulse shaping (Sec. 4.7). In-band spurious noise may also be degraded by the source of RF drive (Sec. 4.3).

4. *Interpulse noise:* Unlike the above three factors, interpulse noise is the noise produced by an RF tube when it is supposed to be completely off; that is, between pulses. Noise generated at this time is of concern because, in almost all radar system configurations, this noise will feed directly into the radar receiver and may either produce false targets or mask real targets. In cathode-pulsed tubes, the high voltage is removed from the RF tube between pulses, and no interference with receiver operation is normally encountered unless there is excessive modulator fall time or backswing, which can produce *noise rings* (Sec. 4.2). With dc-operated CFAs or with mod-anode or gridded linear-beam tubes, the high voltage remains on the tube between pulses, and serious noise may be generated if even a small amount of beam current is allowed to flow through the tube. Since all dc-operated CFAs use cold cathodes, no current can flow until RF drive is applied. With linear-beam tubes, beam current must be well enough cut off to keep noise output (and amplified input signals) small enough. Despite the nearly 200 dB between typical RF peak power output and typical receiver noise levels, most RF tubes readily meet the interpulse-noise requirements. Problems occurred mostly with older-style intercepting-grid linear-beam tubes because the hot grid may emit and produce residual beam current even if cathode current is cut off.

Spurious Modes.   The spurious modes listed in Table 4.1 are the ones most commonly present. In some cases tubes have been made in which these modes are fully suppressed, whereas in poorly designed tubes other modes may also appear, such as band-edge oscillations, harmonic oscillations, etc.

Usable Dynamic Range.   Dynamic range and linearity may be of importance for pulse shaping, as discussed in Sec. 4.7.

Control Electrodes.   These determine the type of modulator required, which in turn affects transmitter size, weight, complexity, and cost. Control electrodes avoid the need for a full-power cathode-pulse modulator. A *mod-anode* (modulating anode) can be used in any linear-beam tube; it acts like a control grid with

a mu of 1 and is inexpensive and reliable. A high-mu grid is also feasible for all but the very highest power linear-beam tubes, and it greatly simplifies the modulator requirements; but it raises the cost and may lower reliability and life of the tube.

Magnetic Field. Except for a few electrostatically focused klystrons,[34,35] all magnetrons, CFAs, klystrons, and TWTs require a magnetic field to control the path of the electron beam. Virtually all CFAs use permanent magnets. Periodic permanent magnet (PPM) focusing is used on all but very high power linear-beam tubes, which still require solenoids. Use of a solenoid affects transmitter size, weight, efficiency, cost, servicing, and tube protection.

Dynamic Impedance. This indicates how rapidly the tube current changes for a given change in applied voltage (also see Table 4.2). The significance of this factor depends on the type of modulator used, or it may affect the HVPS capacitor bank size required for a given permissible power droop during the pulse length (Sec. 4.8).

Phase-Modulation Sensitivity. This indicates how hard the transmitter designer must work to ensure that system phase-stability requirements will be met (Sec. 4.6). Large as the difference may seem between linear-beam tubes and CFAs in this characteristic, modulation sensitivity is seldom a dominant factor in RF tube selection. However, it does enter into the size, weight, and cost tradeoffs by its effect on HVPS filter size or on modulator complexity.

Historically, it has always been feasible to obtain the extremely low ripple levels desired for MTI systems, limited only by inherent noise levels (including jitter and starting-time noise) in the tubes. For pulse compression systems, the necessary freedom from ringing during the pulse length has been fairly easy to achieve with hard-tube modulators and difficult to achieve with line-type modulators, with either linear-beam tubes or CFAs.[36]

Life. Linear-beam tubes and CFAs have both shown the feasibility of long life (over 40,000 h in some cases), but both have also shown very short life when the wrong tube-design compromises were made, when development problems remained, or when tubes were misapplied or carelessly handled in the field. Attainment of very long life, in the region of 10,000 h or more, requires judicious selection of power ratings, conservative cathode-current density, and conscientious counseling of the marriage between tube and transmitter.

RF Tube Power Capabilities. The peak power capabilities of RF tubes have progressed sufficiently far that the limitation has become breakdown in practical waveguide systems, even with 20 lb/in^2 of SF_6 in the waveguide. Therefore, since the early 1960s there has been a tendency for radar systems to employ increased duty cycle, by the use of techniques such as pulse compression, to achieve higher average power without a further increase in peak power. Although in many cases a single RF tube can produce so much average power that even pure-copper waveguide requires water cooling, the limit in system average power may still be the RF tube. Furthermore, asking for the ultimate power capability in one RF tube has a high risk of leading to an unsuccessful development program or to an unreliable tube even if the development is "successful." From a reliability standpoint as well, multiple smaller tubes are often preferable to a single very large tube. Therefore, it has become quite common for high-power radar systems to use more than a single RF tube, as described in the next section.

CHAPTER 5
SOLID-STATE TRANSMITTERS

Michael T. Borkowski
Raytheon Company

5.1 INTRODUCTION

Solid-state devices have largely superseded vacuum tubes in logic and other low-power circuits and even in some very high power applications such as power supplies and power converters below 1 MHz. The only exception seems to be cathode-ray tubes (CRTs), which are less costly than large plasma displays. In radar transmitters, the transition from high-power klystrons, traveling-wave tubes (TWTs), crossed-field amplifiers (CFAs), and magnetrons to solid-state has been more gradual because the power output of individual solid-state devices is quite limited. However, compared with tubes, solid-state devices offer many advantages:

1. No hot cathodes are required; therefore, there is no warmup delay, no wasted heater power, and virtually no limit on operating life.

2. Device operation occurs at much lower voltages; therefore, power supply voltages are on the order of volts rather than kilovolts. This avoids the need for large spacings, oil filling, or encapsulation, thus saving size and weight and leading to higher reliability of the power supplies as well as of the microwave power amplifiers themselves.

3. Transmitters designed with solid-state devices exhibit improved mean time between failures (MTBF) in comparison with tube-type transmitters. Module MTBFs greater than 100,000 h have been measured.

4. No pulse modulator is required. Solid-state microwave devices for radar generally operate Class-C, which is *self-pulsing* as the RF drive is turned on and off.

5. Graceful degradation of system performance occurs when modules fail. This results because a large number of solid-state devices must be combined to provide the power for a radar transmitter, and they are easily combined in ways that degrade gracefully when individual units fail. Overall power output, in decibels, degrades only as 20 log r, where r is the ratio of operating to total amplifiers.

6. Extremely wide bandwidth can be realized. While high-power microwave radar tubes can achieve 10 to 20 percent bandwidth, solid-state transmitter modules can achieve up to 50 percent bandwidth or more with good efficiency.

7. Flexibility can be realized for phased array applications. For phased array

systems, an active transceiver module can be associated with every antenna element. RF distribution losses that normally occur in a tube-powered system between a point-source tube amplifier and the face of the array are thus eliminated. In addition, phase shifting for beam steering can be implemented at low power levels on the input feed side of an active array module; this avoids the high-power losses of the phase shifters at the radiating elements and raises overall efficiency. Also, peak RF power levels at any point are relatively low since the outputs are combined only in space. Furthermore, amplitude tapering can be accomplished by turning off or attenuating individual active array amplifiers.

The general replacement of high-power microwave tubes by solid-state devices has proceeded more slowly than was once forecast. With hindsight, the reason for this is that it is usually too costly to use solid-state devices to replace a pulsed RF tube directly while operating at the same peak power and duty cycle. This is true because microwave semiconductor devices have much shorter thermal time constants than RF tubes (milliseconds rather than seconds). The result is that a microwave transistor that is capable of perhaps 50-W average power cannot handle much more than 100 to 200 W of peak power without overheating during the pulse. The short pulse lengths and low duty cycles typical of older tube-type radars would thus make very inefficient use of the average power capabilities of microwave transistors. For example, to replace the old, well-proven 5J26 L-band magnetron that develops 500 W of average RF power at 0.1 percent (typical) duty cycle would require 2500 to 5000 of the 50-W transistors just mentioned. In other words, microwave transistors are much more cost-effective when the required radar system average power can be provided by a lower peak power at a higher duty cycle. As a result, there have been relatively few direct replacements of older low-duty-cycle transmitters by solid-state transmitters; the AN/SPS-40 is an interesting exception to this rule and will be discussed later. For new radar systems, the system designers have been motivated by these considerations to choose as high a duty cycle as possible, both to reduce the peak power required and to permit using solid-state devices at a reasonable cost. With a 10 percent duty cycle, for example, the 500-W average power mentioned earlier in the paragraph could be provided by only 25 to 50 of the 50-W transistors.

The decision to use a high transmitter duty cycle, however, has significant impact on the rest of the radar system. Operation at a high duty cycle generally requires the use of pulse compression to provide the desired unambiguous range coverage together with reasonably small range resolution. Other consequences follow in turn: the wide transmitted pulse used with pulse compression blinds the radar at short ranges, so a "fill-in" pulse must also be transmitted and processed. To prevent points of strong clutter from masking small moving targets, the signal processor must achieve low pulse compression time sidelobes and high clutter cancellation ratio. As a result, it is much easier to design a solid-state transmitter as part of a new system than it is to retrofit one into an old system that usually does not have all these features.

High-power microwave transistors have been developed more quickly at HF through L band than at higher-frequency bands, so the widest use of solid-state transmitters has been at these lower bands, as shown in Table 5.1. Note, also, that solid-state transmitters at UHF and below have generally been much higher in peak and average power than those at L band.

The use of solid-state does not eliminate all the problems of transmitter design, of course. The RF combining networks must be designed with great care and skill to minimize combining losses in order to keep transmitter efficiency

TABLE 5.1 Fielded Solid-State Transmitters

System	Contractor	Frequency, MHz	Peak power, kW	Duty cycle	Average power, kW	No. of modules	Peak power per module, W	Year fielded
ROTHR	Raytheon	5–30	210	CW	210	84	3000	1986
NAVSPASUR*	Raytheon	218	850	CW	850	2666	320	1986
SPS-40*	Westinghouse	400–450	250	1.6%	4	112	2500	1983
PAVE PAWS†	Raytheon	420–450	600	25.0%	150	1792	340	1978
BMEWS*	Raytheon	420–450	850	30.0%	255	2500	340	1986
TPS-59	GE	1200–1400	54	18.0%	9.7	1080	50	1975
TPS-59‡	GE	1200–1400	54	18.0%	9.7	540	100	1982
SEEK IGLOO	GE	1200–1400	29	18.0%	5.2	292	100	1980
MARTELLO*	Marconi	1250–1350	132	3.75%	5	40	3300	1985
RAMP	Raytheon	1250–1350	28	6.8%	1.9	14	2000	1986
SOWRBALL	Westinghouse	1250–1350	30	4.0%	1.2	72	700	1987

*Solid-state replacements of tube-type transmitters.
†Parameters per array face.
‡Upgraded with 100-W peak power modules.

high. Suitable isolation from excessive voltage–standing-wave ratio (VSWR) must be provided to protect the microwave transistors, and their harmonic power output must be properly filtered to meet MIL-STD-469 and other specifications on RF spectrum quality. Because most microwave transistors operate Class-C, no pulse modulators are required, but Class-C operation makes it more difficult to provide controlled shaping of the RF rise and fall time for spectrum control. Also, just as in tube-type transmitters, energy management is still crucial. Each dc power supply must have a capacitor bank large enough to supply the energy drawn by its solid-state modules during an entire pulse, and each power supply must recharge its capacitor bank smoothly between pulses without drawing an excessive current surge from the power line. While the required power supply is generally not a "catalog" power supply, there are plenty of solid-state devices and circuits available to satisfy these requirements.

As a result of unavoidable losses in combining the outputs of many solid-state devices, it is especially tempting to avoid combining before radiating, since combining in space is essentially lossless. For this reason, many solid-state transmitters consist of modules that feed either rows, columns, or single elements of an array antenna. Especially in the last-named case, it is necessary to build the modules (and probably their power supplies) into the array structure. Furthermore, locating the modules at the antenna avoids the losses of long waveguide runs. Nevertheless, there are cases where building that much equipment weight into the antenna is undesirable, which may force designers to stay with conventional combining schemes. One such case is shipboard radars; the antenna is always mounted as high on the ship as possible, where weight must be minimized to maintain roll stability, and where access for maintenance is extremely difficult.[1]

Because of the large number of individual modules in a typical solid-state transmitter, failure of an individual or a few modules has little effect on overall transmitter performance. However, the module outputs add as voltage vectors,

so that loss of 2 of 10 modules (or 20 percent of 1000 modules) results in a reduction to 80 percent of voltage output, which is 64 percent of power output; but even this is only a 2-dB reduction (the difference between 64 and 80 percent of the power ends up in the combiner loads or in sidelobes if the combining is in space). As a result of this "graceful degradation," overall reliability of solid-state transmitters is very high even if maintenance is delayed until convenient scheduled periods; however, this advantage should not be abused. Consider a case where 20 percent of 1000 modules are allowed to fail before output power falls below requirements, and assume that maintenance occurs at scheduled 3-month intervals. In this case, module MTBF need only be 22,000 h to provide 90 percent confidence that the transmitter will not "fail" in less than 3 months; but the cost of replacement modules and labor would be very unattractive, since nearly 40 percent of the transmitter would have to be replaced every year. Higher MTBFs are thus essential to ensure that the transmitter is not only available but also affordable. Fortunately, solid-state module reliability has proved to be even better than the MIL-HDBK-217 predictions; AN/FPS-115 (PAVE PAWS), for example, actual transceiver module MTBF, including the receiver transmit/receiver (T/R) switches and phase shifters as well as the power amplifiers, has grown to 141,000 h, which is 2.3 times the predicted value. In fact, MTBF for the output power transistors measures better than 1.1 million h.[2]

5.2 SOLID-STATE MICROWAVE POWER GENERATION

Although the RF power-generating capability of single solid-state devices is small with respect to the overall peak and average power requirements of a radar transmitter, solid-state devices can be used quite effectively. Large peak and average powers can be attained by combining the outputs of hundreds or thousands of identical solid-state amplifiers. The power output level from a particular device is a function of the exact operating frequency and the operating conditions, namely, the pulse width and duty cycle, and within normal operating constraints bipolar transistors can provide power outputs in the 50-W through 500-W range. These devices have been used for successful designs in the UHF through L-band frequency ranges, as noted in Table 5.1. Bipolar devices satisfy system requirements of reliability, electrical performance, packaging, cooling, availability, and maintainability. In fact, these devices offer an attractive alternative to tube operation at the lower frequencies.

At higher frequencies, and especially for microwave phased array applications where a physically small module with transmit and receive functions is necessary, the gallium arsenide field-effect transistor (GaAs FET) and its associated batch-processed monolithic microwave integrated circuitry (MMIC) can be used. GaAs FETs are well established as low-noise devices up to 60 GHz;[3] and, with individual cell-combining techniques,[4] GaAs FETs can be used as power amplifiers in the 1- to 20-GHz range. In general, the attribute that makes GaAs an attractive technology is that the GaAs FET can be fully integrated with the passive circuitry that is necessary to provide the biasing, loading, filtering, and switching functions that are necessary for multistage transceiver module designs. As a result of fundamental device limitations, however, this approach is not envisioned

as a cost-effective alternative for module designs that require power outputs exceeding 25 to 30 W.

For the upper end of the solid-state microwave spectrum, i.e., the millimeter-wave range, the single-port microwave diode can be used as a low-power oscillator. Unfortunately, the power output and efficiency of these devices are in general very low; in fact, the efficiency is significantly lower than that of their tube counterparts. However, CW and pulsed power output are attainable up to 300 GHz. Gunn and IMPATT diodes are the devices that offer the most promise for millimeter-wave solid-state operation.

Brief descriptions of these device types and their associated technologies are given in the following subsections.

Microwave Bipolar Power Transistors.   The silicon bipolar power transistor is a common device choice for a solid-state system. At the lower frequencies, especially below 3 GHz, this component provides adequate performance at the lowest cost among competing solid-state technologies. Amplifier design is realizable for frequencies up through S Band, where the tradeoff between device performance and overall system cost begins to reach a point of diminishing returns. The silicon bipolar transistor technology is very mature, and, with the continuing advances in device processing, packaging, and circuit design techniques, manufacturers should be able to continue demonstrating increased levels of power output, bandwidth, and reliability for these transistors. In addition, as a relative figure of merit, the cost per watt of device output power has been decreasing as a result of improvements in processing yields and as a result of increased automated or semiautomated assembly techniques.

Microwave power transistors can be considered complex hybrid circuits and are generally single-chip or multichip devices. For devices with very high power output, several transistor dice are always combined in parallel within a small hermetic ceramic package. In addition, some form of internal impedance prematching circuitry is often included in order to preserve the high intrinsic bandwidth of the semiconductor chip and to make the task of external impedance matching easier. The internal matching also increases the terminal impedances of the packaged device to a level where the component losses of the circuitry external to the transistor become less critical.

The processing and planar layout of these chips is somewhat standardized among manufacturers. Figure 5.1 shows a partial cross section of a typical microwave bipolar power transistor chip. The structure is an NPN silicon device with a vertical diffusion profile; i.e., the collector contact forms the bottom layer of the chip. The P-type base region has been diffused or implanted into the collector, the N-type emitter has been diffused or implanted into the base, and both base and emitter regions are accessible from the top surface of the chip. The collector region consists of an N-doped, low-resistivity epitaxial layer that is grown on a very low resistivity silicon substrate. The characteristics of the epitaxial layer, i.e., thickness and resistivity, can determine the upper limit of performance of the device in terms of ruggedness, efficiency, and saturated power output.

The fundamental limitation on high-frequency transistor performance is the overall collector-to-emitter delay time. If a signal is introduced to either the base or the emitter, four separate regions of attenuation or time delay are encountered: the emitter-base junction capacity charging time, the base transit time, the col-
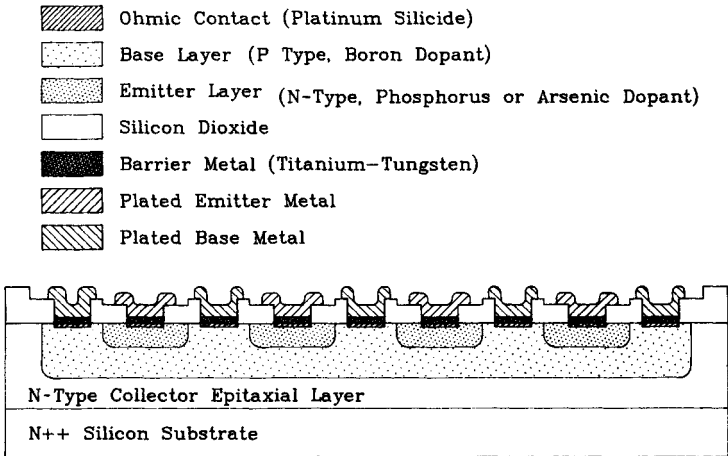
FIG. 5.1 Cross section of a microwave bipolar transistor chip. (*Reprinted with permission from E. D. Ostroff et al., "Solid-State Transmitters," Artech House, Norwood, Mass., 1985.*)

lector depletion-layer transit time, and the collector capacitance-resistance charging time. High-frequency transistor design is concerned with optimizing the physical parameters that contribute to the time-delay components.[5]

For high-power chips, the design challenge is to maintain a uniform high current density over a large emitter area with a minimum temperature rise. High-frequency devices require shallow, narrow, high-resistance base regions under the emitter region. This causes most of the current carried in the device to be crowded along the periphery of the emitter. Thus, in order to maximize the current-handling capability of the device and hence the power output capability of the device, the emitter periphery is maximized. Since the capacitance of the collector-base junction appears as a deleterious parasitic electrical component, the emitter-periphery to base-area ratio, or *Ep/Ba*, is maximized where possible. Generally, higher-frequency devices exhibit higher *Ep/Ba* ratios; and to obtain a high *Ep/Ba* ratio very fine line geometries are required, where the term *geometry* refers to the surface construction details of the transistor.

One limit on the RF power output capability of the transistor is the breakdown voltage of the collector-base junction. Within that limit, the maximum practical level of power output that can be obtained from a single transistor over a given bandwidth is governed by two further limitations: the thermal-dissipation limit of the device and the terminal input or output impedance limit of the device. These latter two limitations are somewhat related by virtue of the physical construction of typical devices.

Active transistor area on the surface of the chip is divided into cells, where the cell size is most often custom-designed for a particular application or range of applications. Pulse width and duty cycle or, as a result, the peak and average dissipated power are the parameters that determine the cell size and arrangement of cells on a chip. As devices become larger and the dissipative heat flux from the top surface of the die to the bottom layer of the transistor increases, the junction temperature increases to the point where the transistor becomes thermally limited.

The ultimate operating junction temperature of the transistor is largely dependent on the transient heating that will be encountered and the layout and area of the individual cells. For devices that are designed to operate for long pulses or CW, an increase in the average power capability of the transistor can be achieved by dividing the active area of a transistor into small, thermally isolated cell areas.

There is a thermal time constant associated with the numerous thermally resistive layers between the transistor junction and the heat sink or cold plate to which the device is attached. This occurs because each layer (silicon, ceramic, transistor flange) not only has a thermal resistance but also exhibits a thermal capacity. Since the overall thermal time constant for a typical L-band power transistor may be on the order of hundreds of microseconds, the tradeoff between peak and average power versus device size can be significant for typical radar pulse widths in the 20- to 1000-μs range. Devices that operate for short-pulse and low-duty-cycle applications, such as DME (distance-measuring equipment), Tacan, and IFF (identification, friend or foe) systems, differ in design from the devices that have been designed for the longer pulse widths and moderate-duty-cycle waveforms that are more typical for surveillance radars. Very high duty cycles or CW operation dictates careful thermal design. An illustration of the thermal-time-constant effect, as it relates to a train of RF pulses, is shown in Fig. 5.2. Table 5.2 illustrates some reported device applications and their general performance characteristics. A photograph of the 115-W UHF transistor used for the PAVE PAWS transmitter is shown in Fig. 5.3, and the schematic, shown in Figure 5.4, may be considered typical for a packaged multichip 100-W L-band transistor.
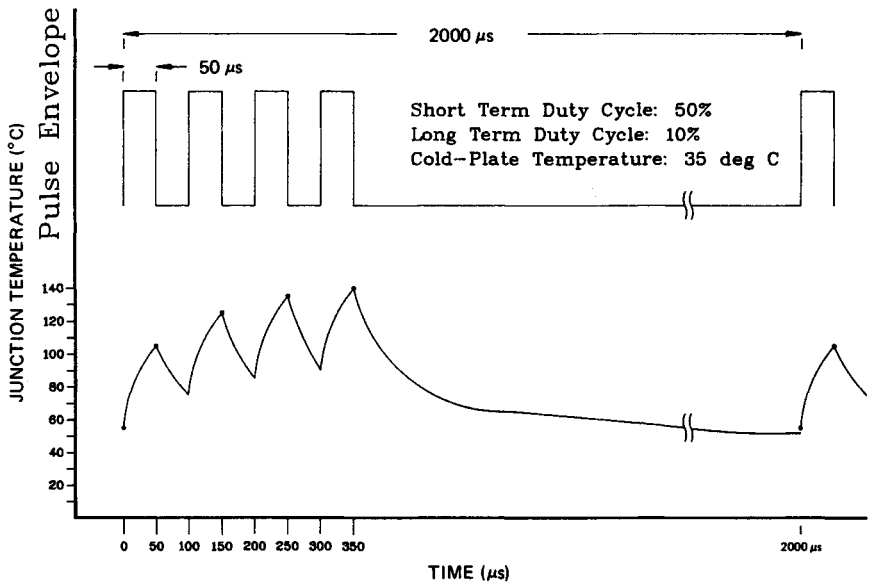
FIG. 5.2 Transient thermal response of a Class-C-biased silicon power transistor for a pulsed RF input.

TABLE 5.2 System Applications for Microwave Power Transistors*

| System | Frequency, MHz | Pulse/duty | Transistor performance | | |
			Peak power, W	Gain, dB	Efficiency, percent
OTH	5–30	CW	130	14.0	60
NAVSPASUR	217	CW	100	9.2	72
AN/SPS-40	400–450	60 μs at 2%	450	8.0	60
PAVE PAWS	420–450	16 ms at 20%	115	8.5	65
BMEWS	420–450	16 ms at 20%	115	8.5	65
AN/TPS-59	1215–1400	2 ms at 20%	55	6.6	52
RAMP	1250–1350	100 μs at 10%	105	7.5	55
MARTELLO S723	1235–1365	150 μs at 4%	275	6.3	40
MATCALS	2700–2900	100 μs at 10%	63	6.5	40
AN/SPS-48	2900–3100	40 μs at 4%	55	5.9	32
AN/TPQ-37	3100–3500	100 μs at 25%	30	5.0	30
HADR	3100–3500	800 μs at 23%	50	5.3	35

*Reprinted with permission from E. D. Ostroff et al., "Solid-State Transmitters," Artech House, Norwood, Mass., 1985.
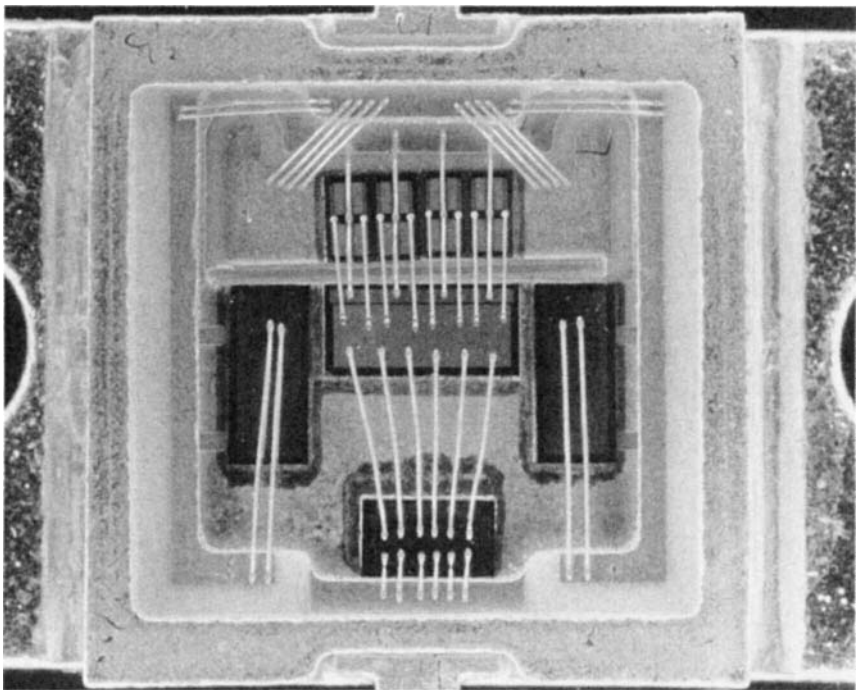
FIG. 5.3 A UHF 115-W power transistor for long pulse and high duty cycle, used in the PAVE PAWS transmitter. (*Photograph courtesy of M/A-COM, PHI.*)
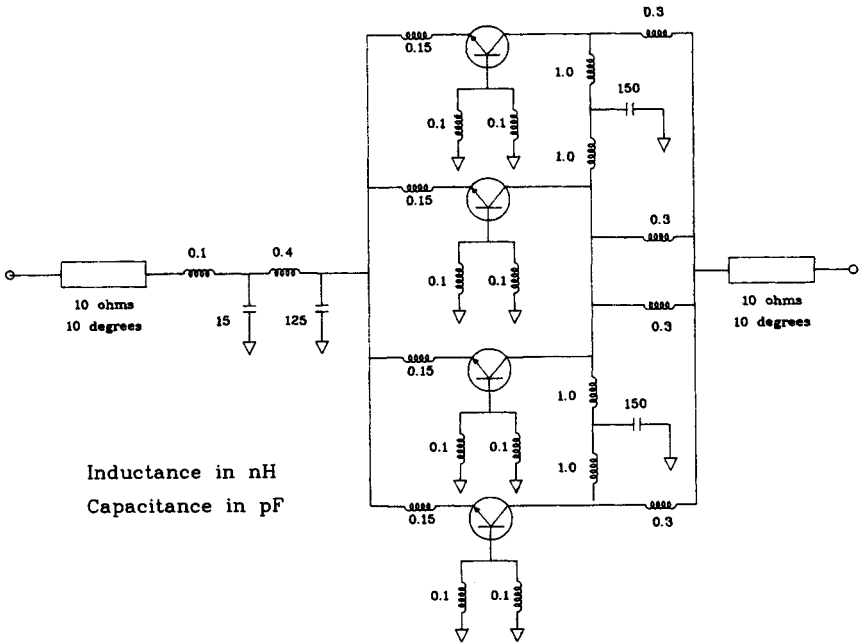
FIG. 5.4 Circuit schematic of an internally input- and output-matched 100-W L-band power transistor.

Microwave Field-Effect Transistors (FETs). Silicon power FETs have demonstrated power output characteristics comparable to the silicon bipolar transistor in the 0.1- to 1.0-GHz frequency range[6]; however, power FETs that are made from gallium arsenide (GaAs) are limited in power output capability, primarily because of the poor thermal conductivity of GaAs and lower typical breakdown voltages; but they are capable of much higher frequency operation than silicon devices.[7] In addition, the GaAs FET can be utilized in a monolithic format, where the remaining passive circuitry of the amplifier is totally integrated on the same substrate with the active device.

GaAs Power FETs. In an FET, the flow of charge carriers between the source and drain electrodes is modulated by one or more gate electrodes, and when the FET is operated as a power device, it may be considered as a simple current switch. Power devices in GaAs are commonly built as metal-semiconductor field-effect transistors (MESFETs), so called because the gate metal is deposited directly onto the semiconductor channel region, forming a Schottky barrier. Power FETs that are fabricated on GaAs exhibit substantially higher frequency performance than similar devices fabricated on silicon because GaAs has a higher electron bulk mobility and a greater maximum electron drift velocity than silicon. In addition, the electron mobility in epitaxial GaAs approaches the bulk value; hence GaAs FETs exhibit lower parasitic series resistances and higher transconductances than silicon FETs with similar geometries. Electrons in GaAs travel at approximately twice the speed that is possible in sil-

icon. In addition, the electron mobility in GaAs is a factor of 3 higher than in silicon. Thus, for comparable geometries, the intrinsic gain for a GaAs device will be substantially higher than for a silicon device.

The cell design and geometry configuration of power GaAs FETs follow design rules similar to those used by designers of silicon bipolar devices. Figure 5.5 shows the cross section of a power GaAs FET that uses air bridge construction and *via holes* to ground the source terminal of the FET. The gate length, seemingly a misnomer because it is shorter than the gate width, is the major parameter that determines device gain and hence operating frequency. Gate width is sometimes also referred to as the periphery. In general, it is desirable to use the largest gate length that permits sufficient gain at the operating frequency. This maximizes processing yield and hence minimizes component cost. Gate lengths for devices in the 1- to 30-GHz range may vary from 2.0 μm to as little as 0.25 μm. While frequency limits can be increased to some degree by decreasing the gate length, increases in power output require greater transistor gate width to support the increased current flow; however, if gate fingers are made too wide, the microwave signal will accumulate a phase shift and will be attenuated while propagating down the gate metal; consequently, the overall gain will be degraded. Total effective increases in gate width can be achieved by paralleling several adjacent gates in order to increase the total channel area per given area. This is similar to increasing emitter periphery per unit base area in the design of bipolar transistors. In addition, the structure must be designed to maintain as high a source-drain breakdown voltage as is possible in order to maximize power output capability. The output power capability of GaAs FETs increases almost linearly with increases in the total gate width, while the power gain decreases slowly with increasing total gate width. The maximum practical total gate width that can be accommodated on a single chip is limited by the following factors:

1. *Yield of the device:* A typical dc processing yield may be 0.995 per 100 μm of gate periphery. A 24-mm chip would therefore have an expected yield of only 30 percent.

2. *Difficulty of impedance matching:* Increases in power output are the result of adjacent channels being connected in parallel. For higher power levels, overall device impedances become lower and lower. For example, the real part of the input impedance for a 24-mm chip would be less than 1 ohm.

3. *Decrease of total power gain:* Uniform current distribution among gate fingers becomes increasingly difficult to manage as the number of paralleled gates increases. Combining inefficiencies result, and the overall device gain decreases.

4. *Physical device size:* The size of a 24-mm periphery chip would be approximately 3000 square mils (75 mils by 40 mils). Larger chip areas decrease the probable assembly yield because assembly difficulty increases with larger chips, greater numbers of wires, and larger packages.

5. *Dissipated power:* The thermal conductivity of GaAs is poor. The dissipated power from larger devices will result in extremely high channel temperatures on the die surface, and reliability will be impacted.

Silicon Power FETs. Silicon bipolar power transistors have been under development far longer than their silicon FET equivalents; consequently, many of the earliest solid-state transmitter designs utilized the bipolar devices. However, the silicon power FET is a viable alternative device type for the amplifier designer. Unlike bipolar transistors, which are minority carrier devices, FETs are

Gate Metal

Ohmic Contact, n+ Implant

Drain Metal

Source Metal

Plated Metal, Source Air Bridge

Plated Metal, Backside Via Holes

N-Type Channel

Semi—Insulating GaAs

Via Hole   Via Hole
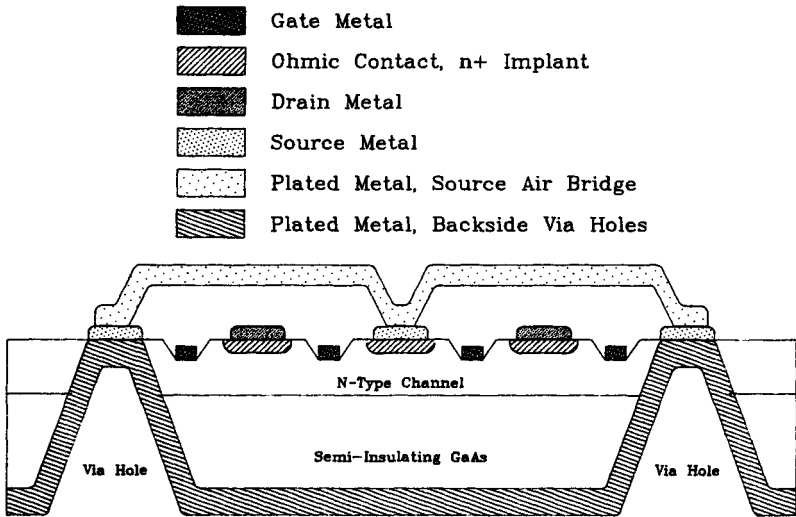
FIG. 5.5 Cross section of a GaAs power MESFET chip.

majority carrier devices and exhibit inherent thermal stability. In contrast, bipolar devices are minority carrier devices, and since minority carriers are thermally generated, bipolar power transistors tend to generate localized hot spots and can become thermally unstable. Resistive ballasting techniques used in both the collector and the emitter leads of a bipolar transistor reduce the intrinsic device gain and efficiency but offset the thermal-instability problem. In the power FET, however, large active cell areas can be combined without using these ballasting techniques and without experiencing the problems of thermal runaway.

In general, there are advantages that silicon FETs exhibit, namely,[8]

1. *Thermal stability:* This results from a negative temperature coefficient of power gain.

2. *Gain control ability:* Pulse shaping can be accomplished by using low power gate bias modulation.

3. *Ease of impedance matching:* This is particularly true for the static induction transistor (SIT), which can operate from dc supply voltages as high as 100 V and hence can provide higher impedance levels than other device types for a given power output level.

Millimeter-Wave Solid-State Power Sources. Solid-state power in the millimeter-wave frequency range is generated from low-power oscillators or negative-resistance amplifiers. The most promising results have been obtained from IMPATT diodes or Gunn diodes. However, TUNNETT (tunnel injection transit time) devices and BARITT (barrier injection transit time) devices are also used. When an extremely short gate length is used, a MESFET construction can be employed at these frequencies, but fundamental limitations on charge carrier velocities and processing tolerances on physically short gate lengths limit the practical operation of MESFET oscillators to below 50 GHz.

IMPATT diodes have been made from silicon, gallium arsenide (GaAs), or

indium phosphide (InP) and operate as millimeter-wave oscillators. Silicon devices offer the most promise because silicon provides the most efficient junction heat removal. Overall performance of the diode depends upon the doping density and the thicknesses of the epitaxial, junction, and interface layers. The level of power output from the devices depends upon whether the device is operated pulsed or CW but can range from 1 W CW at 80 GHz to approximately 20 W pulsed at 80 GHz.[9] In addition to IMPATT diodes, transferred electron devices (more commonly named Gunn diodes) made from GaAs or InP can be used up to about 100 GHz. At the lower part of the millimeter-wave band, CW power levels up to 2 W with 15 percent efficiency and pulsed power levels up to 5 W with 20 percent efficiency have been reported.

5.3 SOLID-STATE MICROWAVE DESIGN

The solution to the radar range equation for most applications invariably requires high peak and average radiated power from the antenna in order to ultimately maintain some minimum signal-to-noise ratio on receive. The impact on the solid-state transmitter designer of the requirement for high radiated power is fundamental: high power must be achieved by combining the outputs of lower-power amplifiers in order to develop the required radiated levels. The amplifier-combining approach generally takes one of two different configurations: space-combined or corporate-combined structures, as shown in Fig. 5.6; however, there are also hybrid approaches in which corporate-combined modules feed
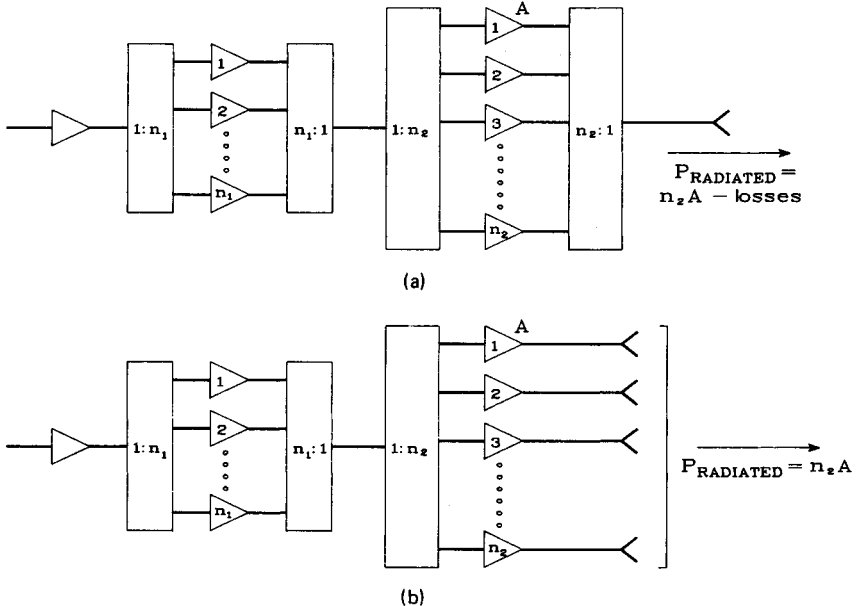
FIG. 5.6 Block diagram of (a) a corporate-combined power amplifier and (b) a space-combined power amplifier.

rows of a space-combined array. The phased array configuration is an example of the space-combined approach wherein each radiating antenna element is fed by an amplifier module and the wavefront is formed in space. An example of the corporate-combined design is a parabolic reflector antenna illuminated from a single feedhorn. The power radiated from the horn has been combined from the outputs of many amplifier modules. Solid-state transmitter designs have been built around each of these generic forms, and the components that are required in the implementation of each share similar characteristics and devices.

High-Power Amplifier Design. In the corporate-combined system, high power levels are generated at a single point by combining the outputs of many low-power amplifiers. The amplifier module is usually partitioned such that the required electrical performance is achieved while the constraints imposed by the mechanical, cooling, maintenance, repair, and reliability disciplines are simultaneously achieved. In general, a module, as shown in Fig. 5.7, consists of a number of identical amplifier stages that are parallel-combined and isolated from one another through the use of microwave combining and isolating techniques. Drive power for this parallel group is obtained from driver or predriver stages, using microwave power dividers. A circulator at the module output port is commonly used to protect the amplifier from the damaging effects of high-load VSWR, most notably from the antenna. Also, ancillary circuitry such as energy storage capacitance, built-in-test (BIT) sensors, or adaptive control components may be included.[10]

Single-Stage Characteristics. Transistors that operate in the HF through S-band frequency ranges are commonly biased either Class-B or Class-C. Class-C
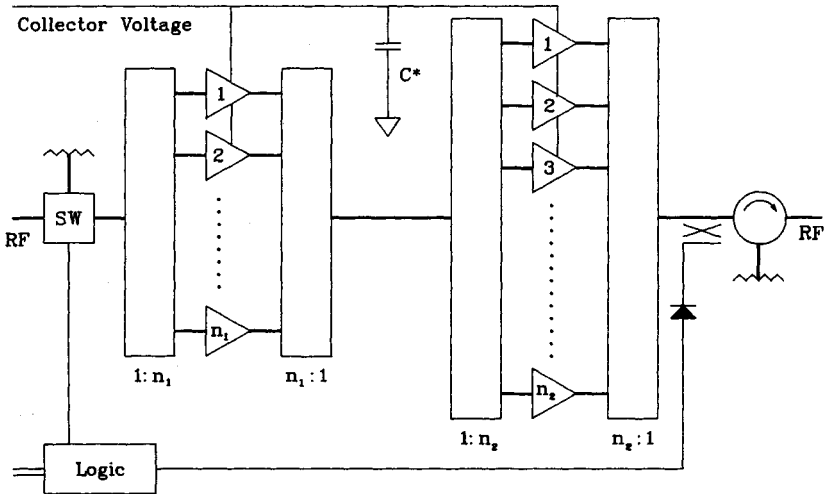
C* : Energy storage capacitance for pulsed amplifiers

$$C = \frac{I \times dt}{dv} \qquad \text{where} \qquad \begin{aligned} &I = \text{peak current} \\ &dt = \text{pulse length} \\ &dv = \text{pulse voltage droop} \end{aligned}$$

FIG. 5.7 Block diagram of a high-power solid-state amplifier module for a corporate-combined transmitter.

operation is the preferred mode since the RF output power of the amplifier is maximized for a given prime power input.[11,12] In general, the base-emitter junction is reverse-biased, and collector current is drawn for less than half of an RF cycle. Collector current is drawn only when the input voltage exceeds the reverse bias across the input and the output voltage is developed across a resonant-tuned load. The net result is high amplifier efficiency. The practical implications of a Class-C-biased amplifier stage are as follows:

1. No quiescent dc current is drawn while the device is not being driven, such as in the radar receive mode. Hence there is no power dissipation in the amplifier while the transmitter is operating in this mode.

2. Only one power supply voltage is necessary for the collector terminal of the transistor. The Class-C operation is a *self-bias*, wherein the transistor draws collector current only when the RF voltage swing on the input exceeds the built-in potential of the emitter-base junction. Additional reverse biasing may be introduced as a result of the voltage drop induced by current flow across parasitic resistance of the base or emitter bias return, and in common-base operation this will result in degraded power gain.

Unlike Class-A linear amplifiers, there are peculiar operating characteristics of the Class-C-biased amplifier that must be recognized in the overall amplifier design. Among these are the RF characteristics of the device as a function of varying RF input drive levels, varying collector voltage supply levels, or varying load impedance. As the RF input drive level of a Class-C-biased device is increased from zero, the dc potential of the reverse-biased base-emitter junction is surpassed and the device begins to draw collector current from the fixed dc supply voltage. The amplifier shows somewhat "linear" transfer characteristics as the drive is increased until the device begins to saturate. Eventually a point of saturated power output capability of the device is attained, and further increases in RF input drive level will actually produce a degraded power output level. At this point the device may also be thermally limited at a device junction. One of the characteristics of this mode of operation is that devices will continue to draw collector current as the amplifier is driven harder; consequently, there exists an optimum operating point with regard to RF drive level. This generally occurs as the transistor is driven approximately 0.5 to 0.75 dB into saturation.

At the chosen operating point and under the conditions of fixed RF input drive and fixed dc supply voltage, amplifiers of this type also exhibit sensitivities of insertion phase and power output to changes in the input drive level, collector voltage, temperature, and load impedance.[13] Some of the common sensitivities of a Class-C-biased amplifier are given in Table 5.3. Although a nominal 50-ohm load impedance may be assumed, the typical loading effect from the microwave power combiners and the circulator will produce variations in the load impedance presented to the transistor stage that may vary by ±50 percent from the nominal level. Depending on the phase of this mismatch, which can vary among adjacent devices, the port-to-port characteristics of an amplifier can vary dramatically. An important facet of Class-C design is that the response of the single-stage amplifier to these external perturbations must be addressed. Proper selection of the nominal load impedance directly affects the power output, gain, insertion phase, efficiency, and peak junction temperature of the single stage. Changes in the port-to-port insertion phase may result in combining inefficiencies among adjacent amplifiers since the RF power that is lost to the fourth-port termination of a microwave combiner, when adjacent amplifiers are combined, is given by

TABLE 5.3   Performance Sensitivities for a Class-C-Biased Amplifier

Parameter	Value
Amplitude sensitivity to RF drive	0.2–0.9 dB/dB*
Amplitude sensitivity to collector voltage	0.2–04 dB/V†
Phase sensitivity to RF drive	10–13°/dB
Phase sensitivity to collector voltage	0.5–1.5°/V
Phase runout	5–20°‡

*Function of the saturation level.
†Function of the collector voltage level.
‡Function of pulse length and transistor geometry.

$$\text{Power lost (dB down)} = 10 \log [(1 + \cos \theta)/2] \qquad (5.1)$$

where θ is the phase difference between amplifiers. Finally, the long-term reliability of the amplifier may be affected by the choice of nominal load impedance since this affects the operating junction temperature of the transistors.

Module Design.   In a very simple sense, the design of an amplifier module consists of matching the power transistors to the proper impedance level and then combining the power levels at these impedances. A typical packaged power transistor has low input and output impedances that must be transformed up to higher level, usually 50 ohms. Thus, the typical amplifier design task must address both low-loss and inexpensive reactive impedance-transforming networks that can provide the proper source and load impedances to the transistor. The common medium for providing this function is a microstrip transmission line. Microstrip is a quasi-TEM mode transmission-line medium that requires photolithographically defined lines on a low-loss, high-quality dielectric substrate. Reactive components that are necessary as impedance-matching elements can be approximated in the microstrip format. An inexpensive reactive matching network can be formed by using an interconnected pattern of microstrip elements. Shunt- and series-connected inductive reactances as well as shunt capacitive reactances are the most easily fabricated and most frequently used matching elements up through 10 GHz.

The outputs of identical single-stage power amplifiers are commonly summed by using splitting and combining techniques that also provide isolation between paralleled amplifiers. It is important to note that isolation is necessary between adjacent combined amplifier stages. Should one device fail, the power combiner must provide a constant load impedance to the remaining device. Half the power of the remaining active device, however, will be dissipated in the isolation resistor of the combiner.

A splitting-combining network must also provide serial isolation among amplifier stages as well as parallel isolation. As a Class-C-biased transistor is pulsed, it passes through its cutoff, linear, and saturation regions. Consequently, the input and output impedances are dynamically varying, and the input impedance changes most dramatically. The input impedance match may change from a near-infinite VSWR in the OFF state to a well-matched condition in the ON state. When amplifier stages are serially cascaded without any means of isolating successive stages, the changing input impedance will appear as a varying load impedance to the previous stage. This may very well send the previous stage into oscillation. Figure 5.8 illustrates splitting-combining configurations that provide the necessary isolation by utilizing reflected signal phase cancellation techniques.
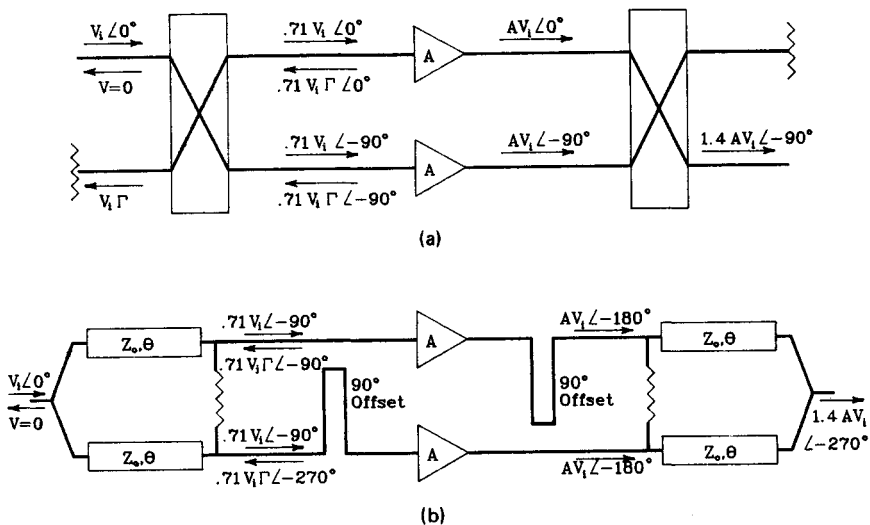
FIG. 5.8 Power amplifier combining configurations that provide minimum input port reflected power. (*a*) Quadrature-coupled amplifier pair. (*b*) Split-T amplifier pair with a 90° offset. (*Reprinted with permission from E. D. Ostroff et al., "Solid-State Transmitters," Artech House, Norwood, Mass., 1985.*)

Phased Array Transceiver Module Design. In contrast to the design of a corporate-combined output, where significant losses can accrue in the combining circuitry, the solid-state phased array system uses individual transceiver modules to feed antenna elements on an array face. Consequently, the phase shifting can be done at a low power level, where dissipated power levels in the phase shifters can be much smaller. The transceiver module, regardless of complexity, has four fundamental functions: (1) to provide gain and power output in the transmit mode, (2) to provide gain and low-noise figure in the receive mode, (3) to switch between transmit and receive states, and (4) to provide phase shift for beam steering in the transmit and receive states. A block diagram of a typical transceiver module is shown in Fig. 5.9.

Microwave Monolithic Integrated Circuits. The use of integrated circuits in transceiver designs has enabled bold new module configurations, and hence phased array systems, to be envisioned. Since some of the more complex functions in the generic transceiver block diagram can be fabricated by using MMIC technology, the components that can be realized through the use of this technology can be employed to create system architectures that are difficult if not impossible to design with other, less integrated technologies. The MMIC design approach utilizes active and passive devices that have been manufactured by using a single process. Active and passive circuit elements are formed on a semi-insulating semiconducter substrate, commonly GaAs, through various deposition schemes. The monolithic approach to circuit design inherently offers several advantages:
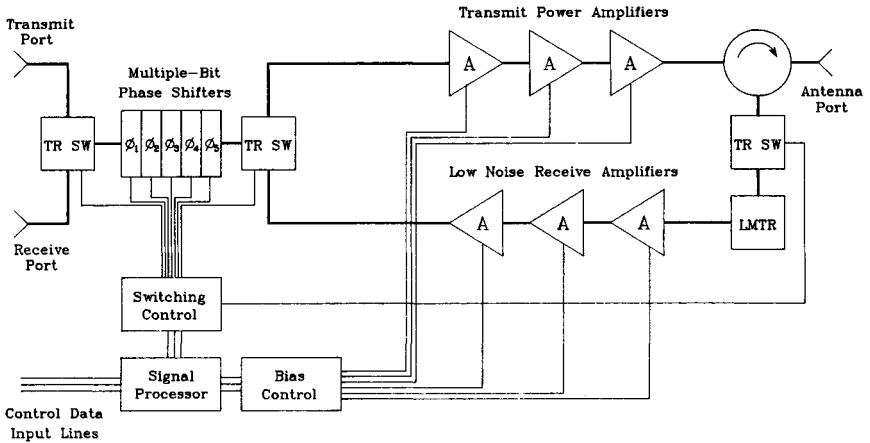
FIG. 5.9 Block diagram of a generic transceiver module for phased array radar.

1. *Low-cost circuitry:* Component assembly is eliminated since complex circuit configurations using both active and passive components are batch-processed on the same substrate.

2. *Increased reliability:* Batch-processed components lead to a reduced number of parts, from the reliability standpoint, and hence to increased reliability.

3. *Increased reproducibility:* Circuitry that is batch-processed or circuits that originate from the same wafer exhibit consistent electrical characteristics from component to component.

4. *Small size and weight:* Integration of active and passive components onto a single chip results in high-density circuitry with multiple functions on a single chip. Overall, the transceiver module can be made much smaller than with discrete components.

However, there are inherent drawbacks to the MMIC approach to component design. The nonrecurring engineering design cost is very high. A typical multistage low-noise amplifier chip design, for example, may easily consume a year's worth of effort before a final design is realized. The cost of wafers and the processing costs in general are very high, such that low-yielding circuit designs may still result in high-cost components. Little circuit trimming can be accommodated in the MMIC approach; consequently, designs must be made tolerant to processing variations or lower performance standards must be accepted for a given design.

The typical processing construction sequence for a GaAs MMIC chip is fairly similar among the GaAs foundries.[14] The active channel region of an FET is delineated by any of several patterning techniques on a semi-insulating GaAs substrate. A combination of deposited dielectric films and metal layers is used to form the passive components and also to interconnect all the elements of the circuit. Standard libraries of circuit elements may include FETs (used as linear amplifiers, low-noise amplifiers, saturating power amplifiers, or switches), resistors, capacitors, inductors, diodes, transmission lines, interconnects, and plated ground vias.

Monolithic circuit elements can be viewed as consuming real estate on a GaAs substrate, and the processing complexity of each step determines the relative yield for individual elements. Although typical processing yields for FET devices may exceed 95 percent per millimeter of gate periphery and greater than 99 percent per picofarad of capacitance, the net yield for a complex circuit may be quite low. For example, a four-stage power amplifier chip that is capable of 3 W of power output at 2 GHz may require more than 9 mm of total gate periphery and may have a total of 75 pF of blocking, bypassing, and matching capacitance. The overall dc yield for this device may be as low as 30 percent when the processing yields for each step are accounted for; high-reliability screening that addresses visually detected imperfections may reduce that yield again by half.

Transceiver Module Performance Characteristics.   The partitioning of transceiver module circuit functions onto GaAs chips usually represents a tradeoff among several design issues, and the resultant circuit configurations represent a compromise among the goals of optimum RF performance, high levels of integration, and fabrication yields that are consistent with processing capabilities of GaAs MMICs. Among the single-chip circuit designs that have been reported from UHF through millimeter-wave frequencies are power amplifiers, low-noise amplifiers, wideband amplifiers, phase shifters, attenuators, T/R switches, and other special function designs.

Component Characteristics.   Performance characteristics for monolithic circuits vary significantly and are the result of processing variations, layout considerations, yield optimization, or circuit complexity. However, the design tradeoffs have resulted in commonly partitioned circuits. Some of the design criteria or characteristics peculiar to amplifiers and other circuits are as follows:

LOW-NOISE AMPLIFIERS.   (1) Multiple-stage linear designs require proper device sizing of successive stages in order to maintain low intermodulation distortion products. (2) Circuit losses on the input, before the first stage, degrade the noise figure of the design; therefore some designs utilize off-chip matching. (3) A low noise-figure requires a bias condition that is close to the pinch-off voltage of the FET. Both gain and noise figure are highly dependent on the pinch-off voltage when the FET is biased close to pinch-off. Since the pinch-off voltage can vary significantly for devices from the same wafer, the bias condition must be chosen carefully. Gain and noise figures are usually traded off against repeatable performance.

POWER AMPLIFIERS.   (1) Total gate periphery is usually at a premium. For high-power design, the load impedance presented to the final device must be carefully chosen such that power output and efficiency are maximized. (2) Losses in the output circuit of the final stage can significantly reduce power output and efficiency. Off-chip matching may be necessary to maximize power output for a given design. (3) GaAs is a poor thermal conductor. Power FET design that addresses thermal management is required. Adequate heat sinking of the chip is mandatory. (4) For efficient multiple-stage designs, it is necessary that the final stage of the amplifier reach saturation before the preceding stages. This must be addressed in the circuit design.

T/R SWITCHING.   (1) For switching applications, the FET design should be chosen such that the ratio of OFF-ON resistance of the FET is kept as high as possible. The channel length largely determines the ON resistance and hence the insertion loss of the device. The tradeoff between short gate length (thus lower processing yield) and insertion loss must be examined. (2) The value of the parasitic drain-source capacitance will affect the OFF-state isolation of the device. This capacitance depends largely on the source-drain spacing of the FET geometry. Critical

applications are usually only the front-end switching configurations in a transceiver module, i.e., before the receive low-noise amplifier or after the transmit amplifier.

PHASE SHIFTERS. (1) Phase-shifter designs generally utilize either a switched-line or a loaded-line circuit configuration, using either distributed transmission-line components or lumped-element equivalent circuits, to achieve multiple-bit phase shifting. Switched-line configurations rely on FET switches to switch lengths of transmission line in and out of the circuit and are typically used for higher frequencies where less chip area is needed. Loaded-line configurations use the switched FET parasitics as circuit elements to introduce the necessary phase changes.

The photograph of a representative MMIC chip, shown in Fig. 5.10, is a 12-W power amplifier chip pair that operates at S band. The final stages of this particular design use FETs with a total of 30-mm gate periphery.

Module Characteristics. Performance data exists for modules that use GaAs MMIC technology in the 1- to 10-GHz frequency range. Animated interest in lightweight, adaptive array applications will continue to push this technology. The reported performance characteristics of transceiver modules are generally a combination of multiple MMIC chip configurations and/or multiple-chip configurations with additional hybrid components to augment the performance of the GaAs components. In addition, the complexity of the transmit/receive functions varies among the module configurations. Data for various modules is enumerated in Table 5.4.[15,16,17] An integrated transceiver module that operates in the X-band frequency range is shown in Fig. 5.11.
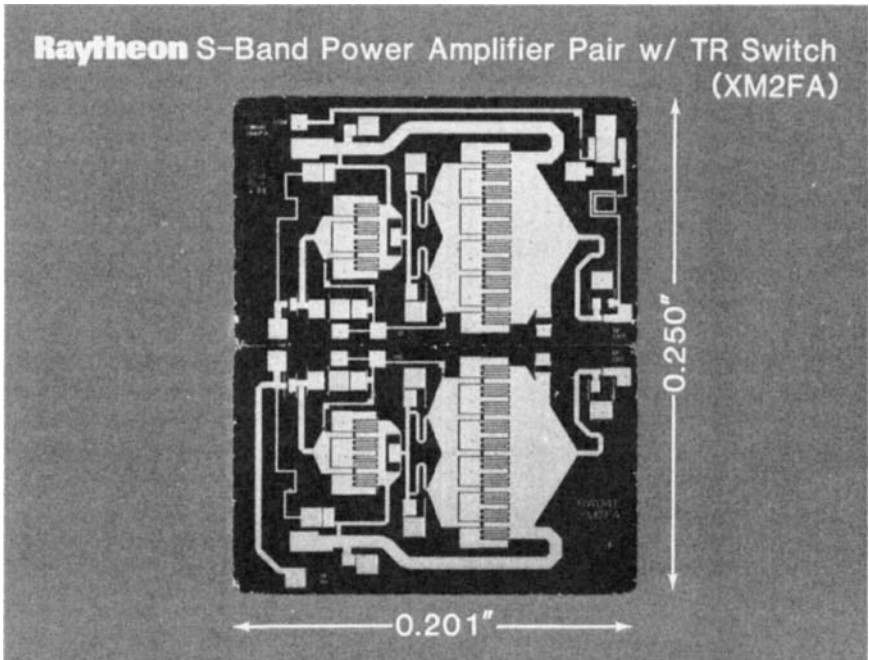
FIG. 5.10 An S-band 12-W GaAs power amplifier MMIC chip pair. (*Photograph courtesy of Raytheon Company.*)

TABLE 5.4 Integrated Transceiver Module Performance Characteristics and Sensitivities

	Performance characteristics								
	Transmit mode			Receive mode					
						rms phase error			
Frequency	RF power, W	Gain, dB	Efficiency, percent	Gain, dB	Noise figure, dB	Gain, dB	Phase, deg	Size, in³	weight, oz
L band*	11	35	30	30	3.0	0.8	5.0	4.0	4.0
S band	10	31	16	25	4.1	0.5	4.0	2.4	2.4
S band	2	23	22	27	3.8	N.A.	4.6	2.9	3.6
S/X band†	2	30	25	...	...	...	...	0.25	N.A.
X band	2.5	30	15	22	4.0	0.6	6.0	0.7	0.7

Performance sensitivities		
Parameter	Transmit	Receive
Gain sensitivity to drain voltage	1 dB/V	1 dB/V
Phase sensitivity to drain voltage	4°/V	2°/V

*Includes hybrid stage on transmit output and receive input.
†Transmit amplifier only.
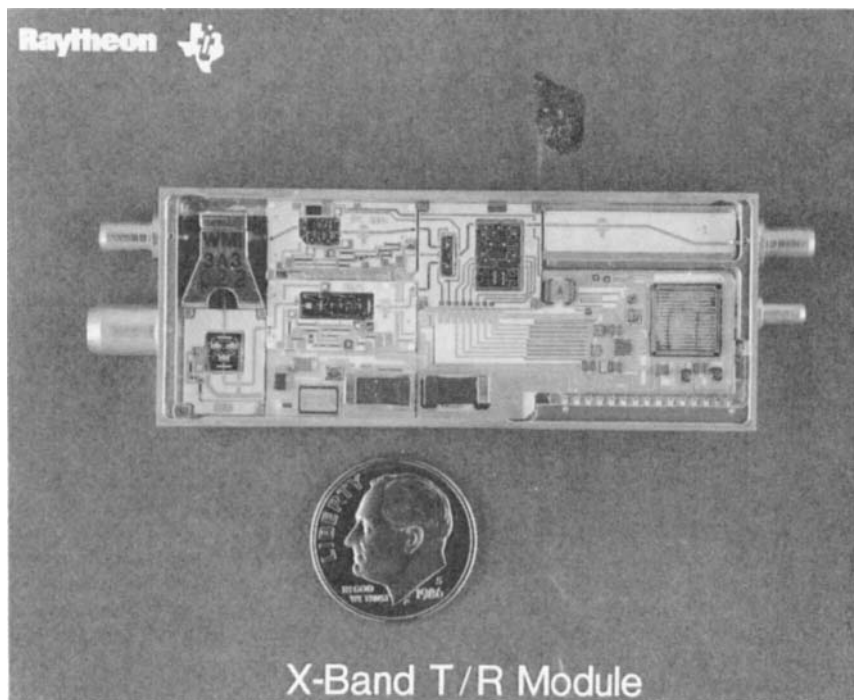
FIG. 5.11 An integrated X-band transceiver module. (*Photograph courtesy of Raytheon Company/Texas Instruments; development work sponsored by RADC, Griffiss Air Force Base.*)

5.4 TRANSMITTER SYSTEM DESIGN

Solid-state amplifiers are usually used in space-combined configurations, corporate-combined configurations, or a hybrid combination of the two. The advantages of solid-state in transmitter applications have already been noted, and some of the performance nuances are noted in the following paragraphs.

Performance Sensitivities

Peak and Average Power Differences. A significant difference between the optimum operating characteristics of solid-state amplifiers and tube amplifiers is that transistors are average-power devices while tubes typically are peak-power devices. The operating characteristics that result from this difference have tremendous impact on the overall system design. The peak power output of microwave transistors is limited by electrical characteristics, while the average power output capability is determined by thermal layout of the transistor. Transistors can be designed to operate for short ($<$ 10 μs), medium (10 to 150 μs), or long ($>$ 150 μs) pulse widths and for duty cycle ranges to CW. Unlike tubes, where the ratio of peak to average power capability can be very high, the peak power output capability of a transistor operated in a short-pulse low-duty mode may only be 2 to 3 times greater than the peak power capability when that device is operated CW. Thus, solid-state operation favors long pulse widths and high duty cycles. For retrofit systems, where high peak power from a short pulse width is required, the acquisition cost of a solid-state transmitter may be prohibitive. The cost of the transmitter may ultimately be reduced if the waveform can be altered to favor higher duty cycle, since the cost in dollars per average watt of the transistor becomes lower as the duty cycle is increased. As long as this cost differential is large enough to offset the added cost in signal processing to accommodate long pulses, a less costly transmitter may be realized.

Amplitude and Phase Sensitivities. Transistor amplifiers that utilize Class-C-biased devices exhibit sensitivity to RF drive level that may degrade the output pulse characteristics. The single-stage amplifier will typically exhibit a very narrow "linear" transfer characteristic; the linear region may exist over only a narrow 1- to 3-dB window. This becomes strikingly critical when several Class-C-biased stages are cascaded in series, as is common in most amplifier configurations. The final tier of output transistors in a serial amplifier chain must be driven into saturation by the preceding stages, and the drive level must be held relatively constant as a function of time and temperature. Since these devices show a narrow operating range, small decreases in the input RF drive level to a multistage amplifier may bring the final tier of devices out of saturation. The net result is an unacceptably degraded output pulse fidelity. One solution has involved a feedback path from a drive-level monitoring point to a variable power supply voltage that maintains the drive level within a specified range for various operating conditions.[10]

In addition to the problems associated with pulse envelope distortions, the phase and amplitude sensitivity of transistor amplifiers to power supply ripple may impact the MTI improvement factor that can be attained. The sensitivities of amplifier stages have already been described. In a multistage amplifier the phase errors due to power supply sensitivity of serially cascaded stages will add, and the limit on MTI improvement factor is

$$I = 20 \log d\theta \qquad (5.2)$$

where $d\theta$ is the magnitude of the insertion phase ripple. The corresponding limit on improvement factor caused by amplifier amplitude instability is

$$I = 20 \log (dA/A) \tag{5.3}$$

where dA and A are the amplitude ripple and the magnitude of the amplitude voltage, respectively.

Time jitter of the RF output pulse envelope can also result from power supply ripple, as a result of Class-C operation of the module, and will also limit the MTI improvement factor. The limitation from this effect is

$$I = 20 \log (dt/T) \tag{5.4}$$

where dt is the time jitter and T is the pulse width. If pulse compression is used, T is still the transmitted-pulse width, not the compressed-pulse width.

In addition, careful design must take into account interactions that can occur as a result of the many cascaded stages of solid-state amplification. These include the following:

1. Phase errors in cascaded stages simply add. However, it may also be possible to arrange them to cancel by proper phasing of power supply ripples for different stages. Similarly, in a stage with N modules in parallel, each with its own high-frequency power-conditioned power supply, the overall phase ripple can usually be assumed to be reduced by a factor equal to the square root of N if the power supply clocks are purposely not synchronized.

2. Because of saturation effects amplitude errors in cascaded stages do not simply add. However, amplitude errors in driving stages will cause drive-induced phase variations in the following stages, as noted above, all of which must be counted.

3. Time jitter in cascaded stages simply adds unless they are arranged to cancel or to be root-sum-squared, as discussed in Par. 1. In addition, amplitude fluctuations in the RF drive will also cause drive-induced jitter, which may even exceed power-supply-ripple-induced jitter, so this factor must be carefully measured.

Spectral Emissions. As a result of Class-C-biased amplifier operation, when a rectangular RF drive pulse is applied to a module, the amplifier will typically show rise and fall times that are on the order of nanoseconds. The output signal spectrum of this pulse shape may not meet spectral emissions requirements, and it may be necessary to slow the rise and fall times. This becomes very difficult when stages are serially cascaded. Because of the highly nonlinear effect described in the preceding subsection, each Class-C stage tends to speed up the rise and fall times of the driving pulse. Consequently, an input pulse shape with slow rise and fall times may be necessary to achieve the desired output-pulse spectral composition.

Control of the rise and fall times is complicated by the necessary use of external bias injection networks. The pulse fall times are generally very fast, on the order of nanoseconds. However, the rise time for a high-power transistor amplifier may be slower, on the order of 100 nanoseconds, and is the result of the reverse biasing that may be encountered by instantaneous current flow in the emitter bias return as the transistor is turning on. This is a design problem for common-base operation, but a common-base configuration is often necessary

since it provides more power gain than common-emitter operation at frequencies above approximately 1 GHz.

Power Combining. To achieve very high levels of output power from a single port, combining the outputs of a large number of modules is required, and therefore a complex combining design is necessary. A power combiner coherently adds together the RF output voltages of individual modules and delivers to a single port the sum total of the modules' output power, minus the losses of the combiner. There are several power combiner-splitter configurations available to the module circuit designer, and all display somewhat varied characteristics.[18] In general, the requirements for a power combiner are:

1. The combiner should have low insertion loss, such that transmitter power output and efficiency are not compromised.

2. The combiner should have RF isolation among ports, such that failed modules do not affect the load impedances or combining efficiency for the remaining functioning modules.

3. The combiner should provide a controlled RF impedance to the amplifier modules, such that the amplifier characteristics are not degraded.

4. The combiner reliability should far exceed the reliability of other transmitter components.

5. The dissipated power capability of the power combiner terminations should be sufficient to accommodate any combination of power amplifier failures.

6. The mechanical packaging of the power combiner should allow modules to be repaired easily. The packaging should also provide short, equal phase and low-loss interconnections between the amplifier modules and the combiner.

High-power combiners may be either reactive or hybrid (or equivalent magic-T) designs. In the hybrid design, any imbalance or difference between the phase and amplitude of the voltages that are being combined is directed to a resistive termination. The net result is that a constant load impedance is presented to the amplifier module under all conditions even when an adjacent module in a combining tier has failed. In a reactive combiner design, any imbalance in power or phase between two input signals results in reflected power and increased VSWR to the module driving it. Power amplifier modules that are not protected from high-load VSWR can be damaged by reflected power from the combiner. In addition, frequency-dependent phase and amplitude ripple may result from this configuration.

Typical RF transmission media that are used in the construction of high-power combiners include coaxial transmission lines, microstrip or stripline transmission lines, or waveguide. The choice of transmission medium is generally a function of many parameters, including peak and average power-handling capability, operating frequency and bandwidth, mechanical packaging constraints, and, of course, the overall loss that can be tolerated. More often than not a combiner design utilizes a hierarchy of cascaded designs to sum the outputs of many modules;[10] however, unique configurations that sum many ports to a single port have been built.[19,20,21]

Solid-State Transmitter Design Examples

AN/TPS-59. The AN/TPS-59 (Sec. 20.1) is a solid-state, L-band, long-range, 3D air defense surveillance radar developed for the Marine Corps by the Elec-

tronic Systems Division of the General Electric Company.[22] The radar is tactically mobile and consists of a trailer-mounted rotating antenna and two radar shelters. The shelters house the digital signal processor, waveform generator, preprocessor, computer, peripherals, and display consoles. One additional shelter is provided for maintenance aids. Radar surveillance coverage encompasses a volume out to 300 nmi and up to 100,000 ft with a 90 percent probability of detection within 200 nmi for a 1-m^2 fluctuating target. The search volume is scanned mechanically in azimuth through 360° and electronically in elevation through 20°. The rotating 15-ft by 30-ft antenna structure houses 54 row transceivers, each of which drives an RF distribution board that feeds one row of 24 dipole antenna elements. The peak power output capability of the system is 54 kW at an average duty cycle of 18 percent.

The row electronics circuitry feeds each of 54 row feed networks. The row electronics consists of transmit preamplifiers, transmit amplifiers, phase shifters, circulator, and logic control. There are ten 100-W transmit amplifier modules in the final output stage of the row electronics circuitry. Each power amplifier consists of two 55-W silicon bipolar power transistors driven by a smaller 25-W device. The 55-W transistors provide a minimum of nearly 7.0-dB gain from 1215 through 1400 MHz from a 28-V dc power supply. All three devices are beryllia-based transistors that are soldered into a metal-ceramic hermetic enclosure, a photograph of which is shown in Fig. 5.12.

Variants of this system were sold by General Electric to the North Atlantic Treaty Organization (NATO) as the GE-592 and to the U.S. Air Force as the AN/
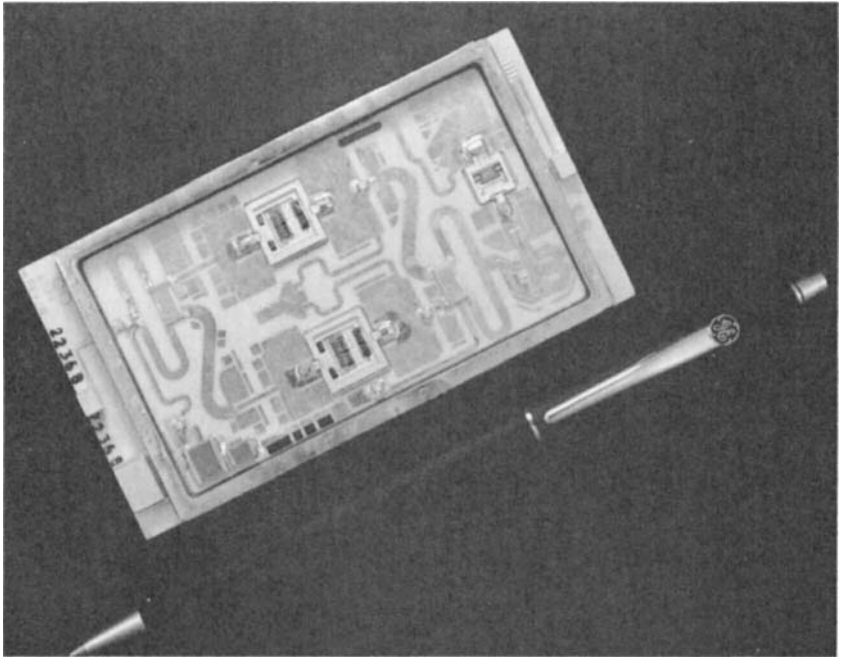
FIG. 5.12 AN/TPS-59 transmitter amplifier module. (*Photograph courtesy of General Electric Company.*)

FPS-117. The AN/FPS-117 radars are operational for the U.S. Air Force in Alaska (SEEK IGLOO Program), northern Canada (North Warning Program), and Berlin. The radar is also in use in Korea, for NATO in Iceland, and for Saudi Arabia as part of the Peace Shield Program. MTBF for these radars has been demonstrated to exceed the 1076-h specification, and more than 750,000 h MTBF has been demonstrated for the 100-W solid-state RF power amplifier modules.

PAVE PAWS. The PAVE PAWS (AN/FPS-115) system is a UHF solid-state active aperture phased array radar that was built for the Electronic Systems Division of the U.S. Air Force by the Equipment Division of the Raytheon Company.[23] The radar is a long-range system with a primary mission to detect and track sea-launched ballistic missiles. The radar uses 1792 active transceiver modules per face to feed dipole antenna elements. Extra elements and a narrow beam are used on receive, and upgrade capability has been included for the future installation of up to 5354 transceiver modules per array face. The peak power output from each face of the baseline system is 600 kW, and the average power output is 150 kW.

Among the 1792 modules per face, groups of 32 transceiver modules are operated as a subarray. In transmit, a high-power array predriver is used to drive 56 subarray driver amplifiers. Each of these power amplifiers provides enough RF drive for all 32 modules in one subarray. In receive, the signal from each of the 56 subarrays is fed into a receive beamforming network.

The transceiver module contains predriver, driver, and final transmit output stages, transmit/receive switching, low-noise amplifiers, limiter, phase shifters, and logic control. The transceiver module block diagram is shown in Fig. 5.13, and a photograph is shown in Fig. 5.14. The transmitter portion of the T/R module contains seven silicon bipolar power transistors, operated Class-C from a +31-V dc power supply. The amplifier is a 1-2-4 configuration, and each of the four final stages delivers 110 W peak for 16-ms pulse widths at duty cycles up to 25 percent. Table 5.5 enumerates some of the salient measured performance
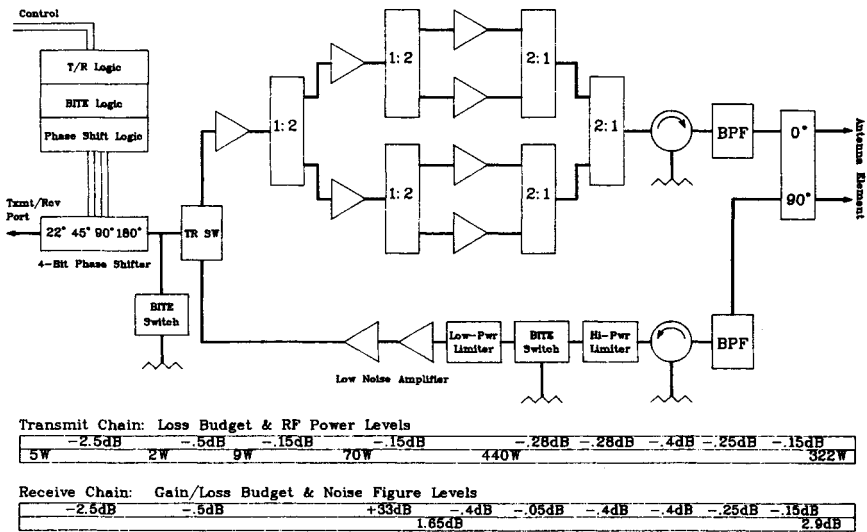
FIG. 5.13 Block diagram of the PAVE PAWS transceiver module.
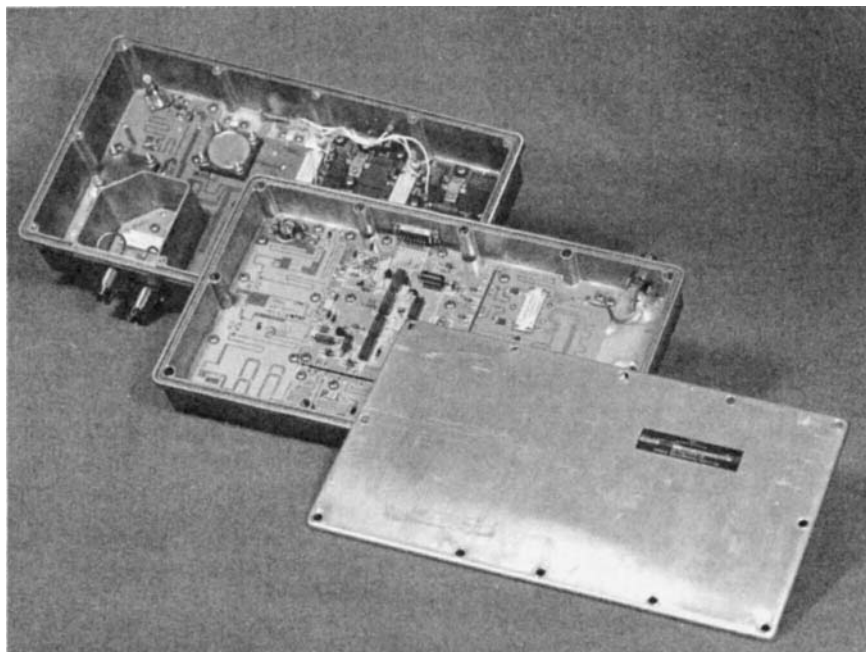
FIG. 5.14 PAVE PAWS transceiver module. *(Photograph courtesy of Raytheon Company.)*

TABLE 5.5 PAVE PAWS Solid-State Transceiver Module Performance*

Parameter	Performance	Specification
Peak power output	330 W	322 W
Output power tracking	0.24 dB rms	0.58 dB rms
Transmit phase tracking	6.7° rms	14° rms
Transmit phase-shifter error	2.52° rms	4.6° rms
Receive phase-shifter error	2.30° rms	4.6° rms
Efficiency	37.9 percent	36 percent
Receive gain	34 dB	27 dB
Receive gain tracking	0.57 dB rms	0.81 dB rms
Receive phase tracking	5.56° rms	10° rms
Noise figure	2.71 dB rms	2.9 dB rms

*Reprinted with permission from D. J. Hoft, Solid-State Transmit/Receive Module for the PAVE PAWS Phase Array Radar, *Microwave J.*, Horizon House, Norwood, Mass., October 1978.

NOTE: Frequency: 420–450 MHz; pulse width: 0.25–16 ms; duty cycle: 0–25 percent.

characteristics of the module. More than 180,000 transistors have been built into more than 25,000 modules.

AN/SPS-40. The AN/SPS-40 was an existing UHF, tube-type, long-range, 2D shipboard search radar system, for which a new solid-state transmitter, built for the Naval Sea Systems Command by the Westinghouse Electric Corporation, replaced the tube transmitter.[10] The existing waveform from the original trans-

mitter was not changed, and the solid-state unit was installed as a direct retrofit. This was not quite as difficult as usual, because the tube-type system already used long pulses and pulse compression, with a duty cycle of nearly 2 percent, which is a lot higher than older 0.1 percent duty cycle systems. Although it may have been desirable to go to a higher duty cycle and lower peak power to make the solid-state retrofit easier, the Navy preferred not to have to modify the rest of the system.

The 250-kW peak power transmitter uses a total of 128 high-power amplifier modules, which, along with power combining, predrivers, drivers, and control circuitry, are housed in three separate cabinets. There are 112 final power output modules arranged in two groups of 56. Each module produces 2500 W peak and 50 W average for a 60-μs pulse width at a 2 percent duty cycle. Drive power for the two banks of final output modules, 17.5 kW, is provided from the combined outputs of 12 more identical modules in the driver group. Predrivers and a redundant preamplifier are used as preceding drive stages.

The power amplifier module consists of 10 identical silicon bipolar power transistors arranged in a 2-driving-8 amplifier configuration to develop more than 2500-W peak power output over the 400- to 450-MHz frequency bandwidth. A photograph of the transmitter module is shown in Fig. 5.15. Each transistor is a 400-W peak-power device that is operated in a balanced push-pull circuit design. By using a push-pull configuration, the circuit designers have alleviated some of the low-impedance-matching problems normally associated with very high power transistors. The RF input drive to the module is 120 W peak and is used to drive two devices. A combined power level of greater than 600 W is split eight ways and drives the eight identical output stages. Losses in the output circulator, final power combining, and the fault detection circuitry reduce the combined power level to 2500 W. Output modules are liquid-cooled for normal operation, but an emergency backup forced-air cooling is provided in the event of a primary-cooling-system failure. The dissipated heat can be tolerated because the system operates at a low duty cycle.
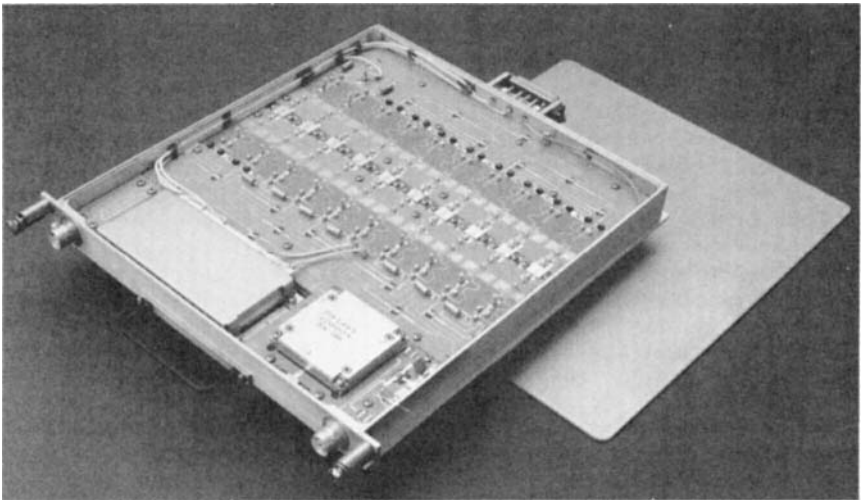
FIG. 5.15 AN/SPS-40 transmitter amplifier module. (*Photograph courtesy of Westinghouse Electric Corporation.*)

The power combining for each output cabinet consists of 56:1 combiners. The reactive power combiner consists of seven groups of 8:1 combiners fabricated in air stripline using 0.5-in ground-plane spacing. The seven outputs are combined by using a single 7:1 air stripline combiner with 1.0-in ground-plane spacing. The 130-kW outputs of the two 56:1 combiners are combined in a single 2:1 isolated hybrid that is manufactured by using 3⅛-in coaxial transmission line. The advertised losses of the 2:1 and 56:1 combiners are 0.1 dB and 0.25 dB, respectively.

Other features of the system include a self-monitoring and self-adjusting driver group power output level. The output level is monitored, and changes to the programmable 24- to 40-V dc power supplies maintain a constant driver group output level as a function of time and temperature.

NAVSPASUR. The solid-state transmitter for the Naval Space Surveillance System (NAVSPASUR) was designed and built by the Equipment Division of the Raytheon Company.[24] The NAVSPASUR is a CW radar, operating at 217 MHz, and is used to provide detection and track data on satellites and other objects as they pass over the continental United States. The solid-state transmitter was procured as a direct replacement for the prior tube-type version. This system produces a very high average power output (850 kW), and with the antenna gain of the system it produces an effective radiated power of over 98 dBW.

There is one main transmitter site at Lake Kickapoo, Texas, with smaller auxiliary sites in Gila River, Arizona, and Jordan Lake, Alabama. The transmitter sites are phased array dipole antennas driven through a coaxial corporate-feed system. The main transmitter site at Lake Kickapoo consists of 2556 antenna elements, each driven by a 300-W solid-state module located directly below the antenna. The most apparent advantages that the solid-state system has over the former tube version are:

1. Much lower dissipation is experienced in the corporate feed because the modules are colocated with the antenna. As a result, the overall site efficiency has been nearly doubled, thus contributing to lower life-cycle costs.

2. A fault-tolerant architecture has led to a system availability of near unity. With a module MTBF of 100,000 h, the maintenance for failed modules can be neglected for nearly 2 years before the transmitter power output degrades by 1 dB. Figure 5.16 illustrates the comparison between the original tube system and the solid-state retrofit.

The solid-state module is a 300-W CW amplifier that uses a 1-driving-4 configuration of silicon bipolar transistors operated common-emitter by using Class-C bias and a 28-V dc power supply. The operating characteristics of the module are delineated in Table 5.6, and Figs. 5.17 and 5.18 are a photograph and a block diagram, respectively. An input power level of 6.2 W is required to drive the module into saturation at the 300-W level. The module dissipates 200 W to the baseplate, which is convection- and radiation-cooled in the outdoor environment. Other module features include automatic fault detection and shutoff, harmonic filtering, and factory-adjustable delay line for module insertion phase matching.

RAMP. The RAMP (Radar Modernization Project) radar system is an L-band system built by the Raytheon Company to replace the earlier primary and secondary surveillance radars used for air traffic control by Canada's Ministry of Transport.[25,26] The primary surveillance radar consists of a rotating reflector, horn-fed by a solid-state transmitter, and redundant receive channels with receiver-exciters and signal processors. The primary surveillance radar operates
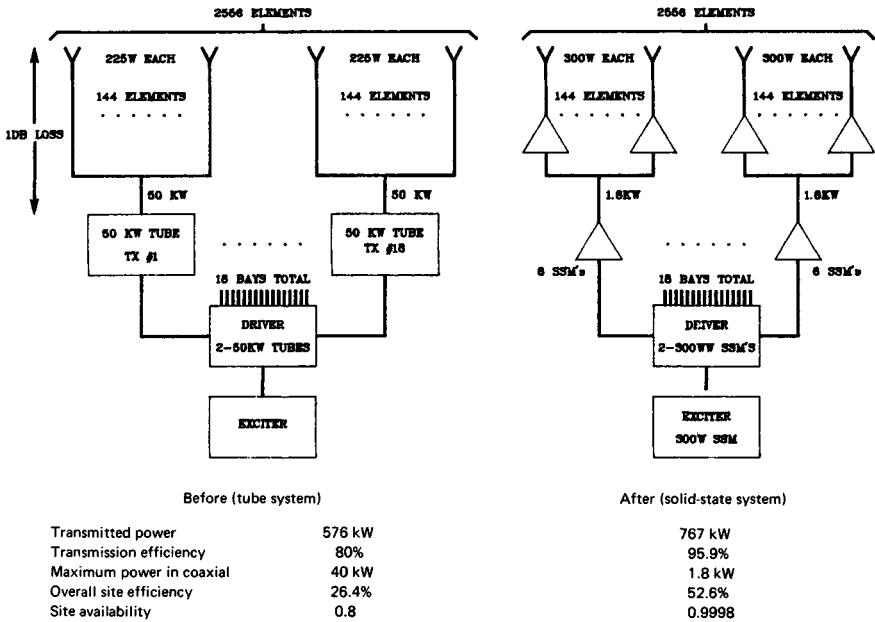
FIG. 5.16 NAVSPASUR transmitter design tradeoffs.

	Before (tube system)	After (solid-state system)
Transmitted power	576 kW	767 kW
Transmission efficiency	80%	95.9%
Maximum power in coaxial	40 kW	1.8 kW
Overall site efficiency	26.4%	52.6%
Site availability	0.8	0.9998

TABLE 5.6   NAVSPASUR Solid-State Power Amplifier Module Performance

Parameter	Performance	Specification
Frequency	216.98 MHz	216.98 MHz
RF power output (CW)	320.0 W	300 W + 0.5 dB
Gain	17.1 dB	16.8 dB
Spurious RF output (near-in)	−75 dBc	−70 dBc maximum
Spurious RF output (far-out)	−85 dBc	− 80 dBc maximum
RF dc efficiency	61.5 percent	58 percent minimum
Input return loss	14 dB	14 dB
Power output similarity (1σ)	0.29 dB	0.5 dB
Phase similarity (1σ)	3.0°	3°
DC voltages	28 V/16.5 A and 8.9 V/0.18 A	28 V/19 A and 8.9 V/0.2 A
Size	21 × 16 × 4.3 in	21 × 26 × 4.3 in
Weight	47 lb	47 lb
Operating ambient temperatures	0–116°F	0–116°F
Pressurization	5 lb/in^2	5 lb/in^2

between 1250 and 1350 MHz with a 25-kW peak power output and provides radar coverage to 80 nmi and to an altitude of 23,000 ft with an 80 percent probability of detection for a 2-m^2 target; with azimuth and range resolution to 2.25° and 0.25 nmi, respectively. The receiver-exciter efficiently utilizes the transmitter solid-state devices with a high-duty-cycle waveform. A double-pulse pair is used in the frequency-agile system, and target returns are processed by a four-pulse moving-target detector. The pulse pair consists of a 1-μs CW pulse that provides cover-
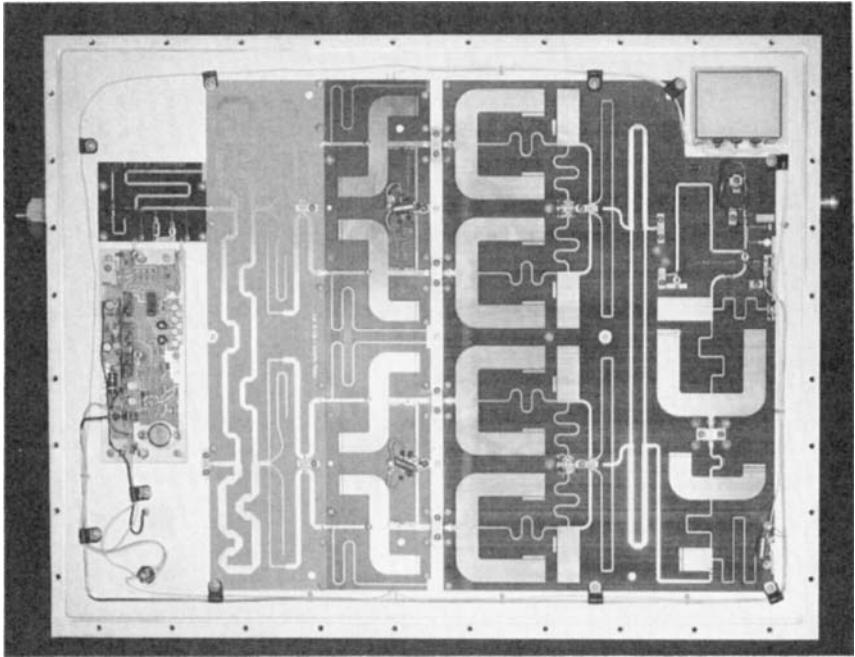
FIG. 5.17 NAVSPASUR transmitter amplifier module. (*Photograph courtesy of Raytheon Company.*)

age to 8 nmi and a 100-μs nonlinear chirp pulse that provides coverage to 80 nmi. The 100-μs pulse is compressed to 1 μs such that high duty cycle is achieved without compromising range resolution.

The transmitter consists of 14 modules, each capable of 2000-W power output, that are combined to produce the greater than 25-kW peak level. Two modules and a 33-V dc power supply make up a single transmitting group. The module consists of a 2-8-32 amplifier configuration of silicon bipolar power transistors. The 32 final output devices and the eight driver devices are 100-W transistors capable of operating up to a 10 percent duty cycle over the 100-MHz bandwidth at collector efficiencies greater than 52 percent. Each module is air-cooled, and the measured efficiency is greater than 25 percent when the module is operating at an 8.2 percent average duty cycle. Module power gain is greater than 16 dB. Figure 5.19 shows a photograph of the 50-lb module and the lineup of power transistors down the center spine of the module. A circulator is used on the output port to protect the 100-W devices from antenna-generated reflections, and control circuitry has been included to switch off modules in the event of cooling-system failure. A 14:1 high-power replicated combiner,[21] built by using a combination of reactive and resistive power-combining techniques in air dielectric stripline, is employed to sum the module outputs to the 25-kW level.
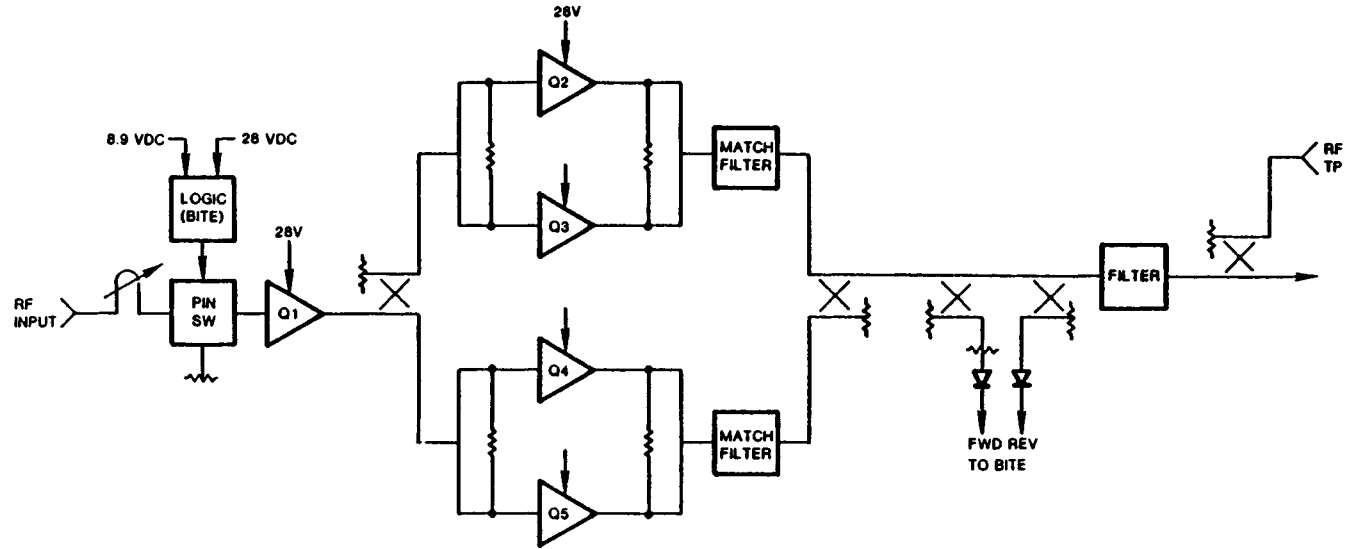
	RF SWITCH	DRIVER	BRANCH-LINE 3-dB DIVIDER		FINALS	BRANCH-LINE COMBINER			BITE MONITORS	PRODUCTION TOLERANCE	
BITE MONITOR				3-dB WILKINSON		3-dB COMBINER/ FILTER (WILKINSON)					
	−0.5	+9.70	−0.30	−0.20	+9.2	−0.06(−0.2)	−0.10		−0.03	−0.60	Gain/loss (dB)
6.2		5.7	52	11.5	(4x94)				344	300	Power (watts)

FIG. 5.18   Block diagram of the NAVSPASUR transmitter amplifier module.
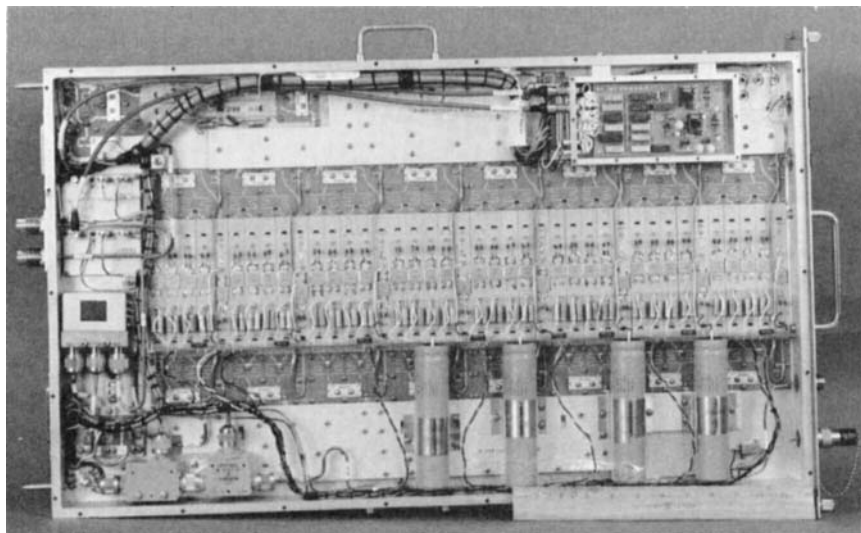
FIG. 5.19 RAMP transmitter amplifier module. (*Photograph courtesy of Raytheon Company.*)

REFERENCES

1. Skolnik, M. I.: The Application of Solid-State RF Transmitters to Navy Radar, *NRL Memo. Rept.* 3074, Naval Research Laboratory, Washington, June 1975.

2. Considine, B.: personal communication, Raytheon Company, January 1988.

3. Watkins, E. T., et al.: A 60 GHz GaAs FET Amplifier, *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 145–147, 1983.

4. Soares, R., J. Graffeuil, and J. Obregon, "Applications of GaAs MESFET's," Artech House, Norwood, Mass., 1983, pp. 157–207.

5. Cooke, H.: Microwave Transistors: Theory and Design, *Proc. IEEE*, vol. 59, pp. 1163–1181, August 1971.

6. Ostroff, E., M. Borkowski, H. Thomas, and J. Curtis: "Solid-State Radar Transmitters," Artech House, Norwood, Mass., 1985.

7. DiLorenzo, J. V., and D. D. Khandelwal: "GaAs FET Principles and Technology," Artech House, Norwood, Mass., 1982, pp. 201–276.

8. Browne, J.: RF Devices Gain Higher Power Levels, *Microwaves & RF*, p. 148, November 1987.

9. Bhartia, P., and I. J. Bahl: "Millimeter Wave Engineering and Applications," John Wiley & Sons, New York, 1984, pp. 75–152.

10. Lee, K., C. Corson, and G. Mols: A 250 kW Solid-State AN/SPS-40 Radar Transmitter, *Microwave J.*, pp. 93–105, July 1983.

11. Krauss, H. L., C. W. Bostian, and F. H. Raab: "Solid State Radio Engineering," John Wiley & Sons, New York, 1980, pp. 394–431.

12. Pitzalis, O., and R. Gilson: Broadband Microwave Class-C Transistor Amplifiers, *IEEE Trans.*, vol. MTT-21, pp. 660–668, November 1973.

13. Lavallee, L. R.: Two-Phased Transistors Shortchange Class-C Amps, *Microwaves*, pp. 48–54, February 1978.

14. Ferry, D. K.: "Gallium Arsenide Technology," Howard W. Sams & Co., Indianapolis, 1985, pp. 189–300.

15. Laighton, D. G.: private communication, Raytheon Company, January 1988.

16. Pierro, J., and R. Clouse: An Ultraminiature 5–10 GHz, 2-Watt Transmit Module for Active Aperture Application, *IEEE MIT Symp. Dig.*, pp. 941–944, June 1987.

17. Green, C. R., et al.: A 2-Watt GaAs TX/RX Module with Integral Control Circuitry for S-Band Phased Array Radars, *IEEE MTT Symp. Dig.*, pp. 933–936, June 1987.

18. Howe, H.: "Stripline Circuit Design," Artech House, Norwood, Mass., 1974, pp. 77–180.

19. Sanders, B. J.: A 110-Way Parallel Plate RF Divider/Combiner Network & Solid-State Module, *Military Microwaves Conf.*, 1980.

20. Beltran, F.: "Multi-Port Radio Frequency Networks for an Antenna Array," U.S. Patent 4,612,548, Raytheon Company, Sept. 16, 1986.

21. O'Shea, R. L.: "Radio Frequency Power/Divider Combiner Networks," U.S. Patent 4,583,061, Raytheon Company.

22. Perkins, W. B.: personal communication, January 1988.

23. Hoft, D. J.: Solid-State Transmit/Receive Module for the PAVE PAWS Phased Array Radar, *Microwave J.*, pp. 33–35, October 1978.

24. Francoeur, A.: private communication, Raytheon Company, January 1988.

25. Dyck, J. D., and H. R. Ward: RAMP's New Primary Surveillance Radar, *Microwave J.*, p. 105, December 1984.

26. Ward, H. R.: The RAMP PSR, A Solid State Surveillance Radar, *IEE Int. Radar Conf.*, London, October 1987.

CHAPTER 6
REFLECTOR ANTENNAS

Helmut E. Schrank
Gary E. Evans
Daniel Davis
Electronic Systems Group
Westinghouse Electric Corporation

6.1 INTRODUCTION

Role of the Antenna. The basic role of the radar antenna is to provide a transducer between the free-space propagation and the guided-wave propagation of electromagnetic waves. The specific function of the antenna during transmission is to concentrate the radiated energy into a shaped directive beam which illuminates the targets in a desired direction. During reception the antenna collects the energy contained in the reflected target echo signals and delivers it to the receiver. Thus the radar antenna is used to fulfill reciprocal but related roles during its transmit and receive modes. In both of these modes or roles, its *primary purpose is to accurately determine the angular direction of the target*. For this purpose, a highly directive (narrow) beamwidth is needed, not only to achieve angular accuracy but also to resolve targets close to one another. This important feature of a radar antenna is expressed quantitatively in terms not only of the beamwidth but also of *transmit gain* and *effective receiving aperture*. These latter two parameters are proportional to one another and are directly related to the detection range and angular accuracy. Many radars are designed to operate at microwave frequencies, where narrow beamwidths can be achieved with antennas of moderate physical size.

The above functional description of radar antennas implies that a single antenna is used for both transmitting and receiving. Although this holds true for most radar systems, there are exceptions: some monostatic radars use separate antennas for the two functions; and, of course, bistatic radars must, by definition, have separate transmit and receive antennas. In this chapter, emphasis will be on the more commonly used single antenna and, in particular, on the widely used reflector antennas. Phased array antennas are covered in Chap. 7.

Beam Scanning and Target Tracking. Because radar antennas typically have directive beams, coverage of wide angular regions requires that the narrow beam be scanned rapidly and repeatedly over that region to assure detection of targets wherever they may appear. This describes the *search* or *surveillance function* of a radar. Some radar systems are designed to follow a target once it has been detected, and this *tracking function* requires a specially designed antenna different from a surveillance radar antenna. In some radar systems,

particularly airborne radars, the antenna is designed to perform both search and track functions.

Height Finding. Most surveillance radars are two-dimensional (2D), measuring only range and azimuth coordinates of targets. In early radar systems, separate height finding antennas with mechanical rocking motion in elevation were used to determine the third coordinate, namely, elevation angle, from which the height of an airborne target could be computed. More recent designs of three-dimensional (3D) radars use a single antenna to measure all three coordinates: for example, an antenna forming a number of stacked beams in elevation in the receive mode and a single broad-coverage elevation beam in the transmit mode. The beams are all equally narrow in azimuth, but the vertically stacked receive beams allow measuring echo amplitudes in two adjacent overlapping beams to determine the elevation angle of the target.

Classification of Antennas.   Radar antennas can be classified into two broad categories, *optical antennas* and *array antennas.* The optical category, as the name implies, comprises antennas based on optical principles and includes two subgroups, namely, *reflector antennas* and *lens antennas.* Reflector antennas are still widely used for radar, whereas lens antennas, although still used in some communication and electronic warfare (EW) applications, are no longer used in modern radar systems. For that reason and to reduce the length of this chapter, lens antennas will not be discussed in detail in this edition of the handbook. However, references on lens antennas from the first edition will be kept in the list at the end of the chapter.

6.2 BASIC PRINCIPLES AND PARAMETERS

This section briefly reviews basic antenna principles with emphasis on definitions of terms useful to a radar system designer. In order to select the best type of antenna for a radar system, the system designer should have a clear understanding of the basic performance features of the wide variety of antenna types from which he or she must choose.[1] This includes the choice between reflector antennas, covered in this chapter, and phased arrays, covered in Chap. 7. Another alternative is a reflector fed by a phased array.

Although the emphasis in this chapter is on reflectors, many of the basic principles discussed in this section apply to all antennas. The three basic parameters that must be considered for any antenna are:

- Gain (and effective aperture)
- Radiation pattern (including beamwidth and sidelobes)
- Impedance (voltage–standing-wave ratio, or VSWR)

Other basic considerations are *reciprocity* and *polarization*, which will be briefly discussed in this section.

Reciprocity.   Most radar systems employ a single antenna for both transmitting and receiving, and most such antennas are reciprocal devices, which means that their performance parameters (gain, pattern, impedance) are identical for the two functions. This reciprocity principle[2] allows the antenna to be considered either as a transmitting or as a receiving device, whichever is

more convenient for the particular discussion. It also allows the antenna to be tested in either role (Sec. 6.10).

Examples of *nonreciprocal* radar antennas are phased arrays using nonreciprocal ferrite components, active arrays with amplifiers in the transmit/receive (T/R) modules, and height finding antennas for 3D (range, azimuth, and elevation) radars. The last-named, typified by the AN/TPS-43[3] radar, uses several overlapping beams stacked in elevation for receiving with a broad elevation beam for transmitting. All beams are equally narrow in the azimuth direction. These nonreciprocal antennas must be tested separately for their transmitting and receiving properties.

Gain, Directivity, and Effective Aperture. The ability of an antenna to concentrate energy in a narrow angular region (a directive beam) is described in terms of antenna gain. Two different but related definitions of antenna gain are *directive gain* and *power gain*. The former is usually called *directivity*, while the latter is often called *gain*. It is important that the distinction between the two be clearly understood.

Directivity (directive gain) is defined as the maximum radiation intensity (watts per steradian) relative to the average intensity, that is,

$$G_D = \frac{\text{maximum radiation intensity}}{\text{average radiation intensity}} = \frac{\text{maximum power per steradian}}{\text{total power radiated}/4\pi} \quad (6.1)$$

This can also be expressed in terms of the maximum radiated-power density (watts per square meter) at a far-field distance R relative to the average density at that same distance:

$$G_D = \frac{\text{maximum power density}}{\text{total power radiated}/4\pi R^2} = \frac{p_{\max}}{P_t/4\pi R^2} \quad (6.2)$$

Thus the definition of directivity is simply how much stronger the actual maximum power density is than it would be if the radiated power were distributed isotropically. Note that this definition does not involve any dissipative losses in the antenna but only the concentration of *radiated power*.

Gain (power gain) does involve antenna losses and is defined in terms of *power accepted* by the antenna at its input port P_0 rather than radiated power P Thus gain is given by

$$G = \frac{\text{maximum power density}}{\text{total power accepted}/4\pi R^2} = \frac{p_{\max}}{P_0/4\pi R^2} \quad (6.3)$$

For realistic (nonideal) antennas, the power radiated P_t is equal to the power accepted P_0 times the *radiation efficiency* factor η of the antenna:

$$P_t = \eta P_0 \quad (6.4)$$

As an example, if a typical antenna has 1.0 dB dissipative losses, $\eta = 0.79$, and it will radiate 79 percent of its input power. The rest, $(1 - \eta)$ or 21 percent, is converted into heat. For reflector antennas, most losses occur in the transmission line leading to the feed and can be made less than 1 dB.

By comparing Eqs. (6.2) and (6.3) with (6.4), the relation between gain and directivity is simply

$$G = \eta G_D \quad (6.5)$$

Thus antenna gain is always less than directivity except for ideal lossless antennas, in which case $\eta = 1.0$ and $G = G_D$.

Approximate Directivity—Beamwidth Relations. An approximate but useful relationship between directivity and antenna beamwidths (see Sec. 2.3) is

$$G_D \simeq \frac{40,000}{B_{az}B_{el}} \tag{6.6}$$

where B_{az} and B_{el} are the principal-plane azimuth and elevation half-power beamwidths (in degrees), respectively. This relationship is equivalent to a 1° by 1° pencil beam having a directivity of 46 dB. From this basic combination, the approximate directivities of other antennas can be quickly derived: for example, a 1° by 2° beam corresponds to a directivity of 43 dB because doubling one beamwidth corresponds to a 3 dB reduction in directivity. Similarly, a 2° by 2° beam has 40 dB, a 1° by 10° beam has 36 dB directivity, and so forth. Each beamwidth change is translated into decibels, and the directivity is adjusted accordingly. This relation does not apply to shaped (e.g., cosecant-squared) beams.

Effective Aperture. The aperture of an antenna is its physical area projected on a plane perpendicular to the main-beam direction. The concept of effective aperture is useful when considering the antenna in its receiving mode. For an ideal (lossless), uniformly illuminated aperture of area A operating at a wavelength λ, the directive gain is given by

$$G_D = 4\pi A/\lambda^2 \tag{6.7}$$

This represents the *maximum available gain* from an aperture A and implies a perfectly flat phase distribution as well as uniform amplitude.

Typical antennas are not uniformly illuminated but have a tapered illumination (maximum in the center of the aperture and less toward the edges) in order to reduce the sidelobes of the pattern. In this case, the directive gain is less than that given by Eq. (6.7):

$$G_D = 4\pi A_e/\lambda^2 \tag{6.8}$$

where A_e is the effective aperture or *capture area* of the antenna, less than the physical aperture A by a factor ρ_a usually called the *aperture efficiency*.

$$A_e = \rho_a A \tag{6.9}$$

This aperture efficiency would better be called *aperture effectiveness* because it does not involve RF power turned into heat; i.e., it is not a dissipative effect but only a measure of how effectively a given aperture is utilized. An antenna with an aperture efficiency of, say, 50 percent ($\rho_a = 0.5$) has a gain 3 dB below the uniformly illuminated aperture gain but does not dissipate half the power involved. The effective aperture represents a smaller, uniformly illuminated aperture having the same gain as that of the actual, nonuniformly illuminated aperture. It is the area which, when multiplied by the incident power density P_i, gives the power received by the antenna:

$$P_r = P_i A_e \tag{6.10}$$

Radiation Patterns. The distribution of electromagnetic energy in three-dimensional angular space, when plotted on a relative (normalized) basis, is called the *antenna radiation pattern*. This distribution can be plotted in various ways, e.g., polar or rectangular, voltage intensity or power density, or power per unit solid angle (radiation intensity). Figure 6.1 shows a typical radiation pattern for a circular-aperture antenna plotted isometrically in terms of the logarithmic power density (vertical dimension in decibels) versus the azimuth and elevation angles in rectangular coordinates. The *main lobe* (or *main beam*) of the pattern is a *pencil beam* (circular cross section) surrounded by *minor lobes*, commonly referred to as *sidelobes*. The angular scales have their origins at the peak of the main lobe, which is generally the *electrical reference axis* of the antenna.

This axis may or may not coincide with the *mechanical axis* of the antenna, i.e., the axis of symmetry, sometimes called the *boresight axis*. If the two do not coincide, which usually happens unintentionally, the angular difference is referred to as a *boresight error* and must be accounted for in the measurement of target directions.

Figure 6.1a shows the three-dimensional nature of all antenna patterns, which require extensive data to plot in this form. This same data can also be plotted in the form of constant-power-level contours, as shown in Fig. 6.1c. These contours are the intersections of a series of horizontal planes through the 3D pattern at various levels and can be quite useful in revealing the distribution of power in angular space.

More frequently, however, 2D plots are sufficient and more convenient to measure and plot. For example, if the intersection of the 3D pattern of Fig. 6.1a with a vertical plane through the peak of the beam and the zero azimuth angle is taken, a 2D slice or "cut" of the pattern results, as shown in Fig. 6.1b. This is called the *principal-plane elevation pattern*. A similar cut by a vertical plane or
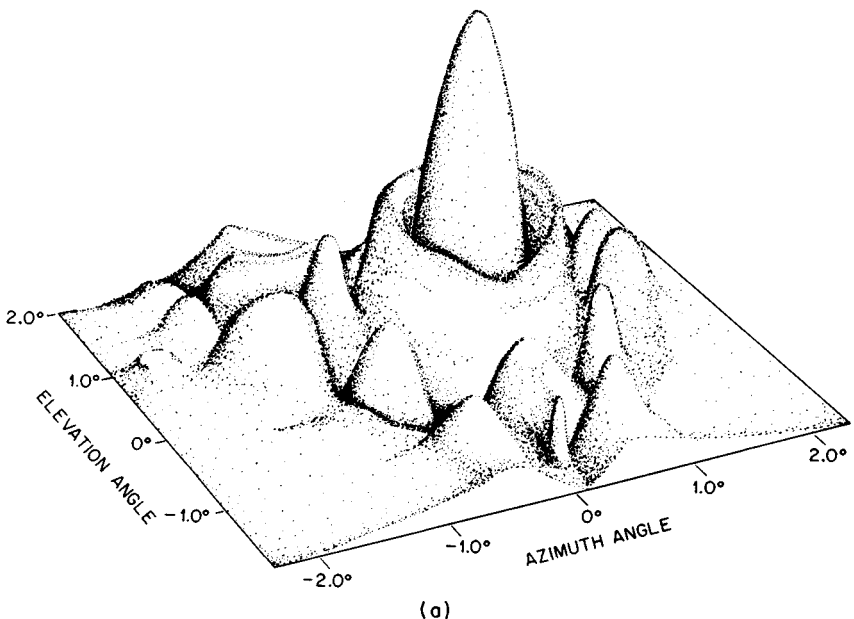

(a)

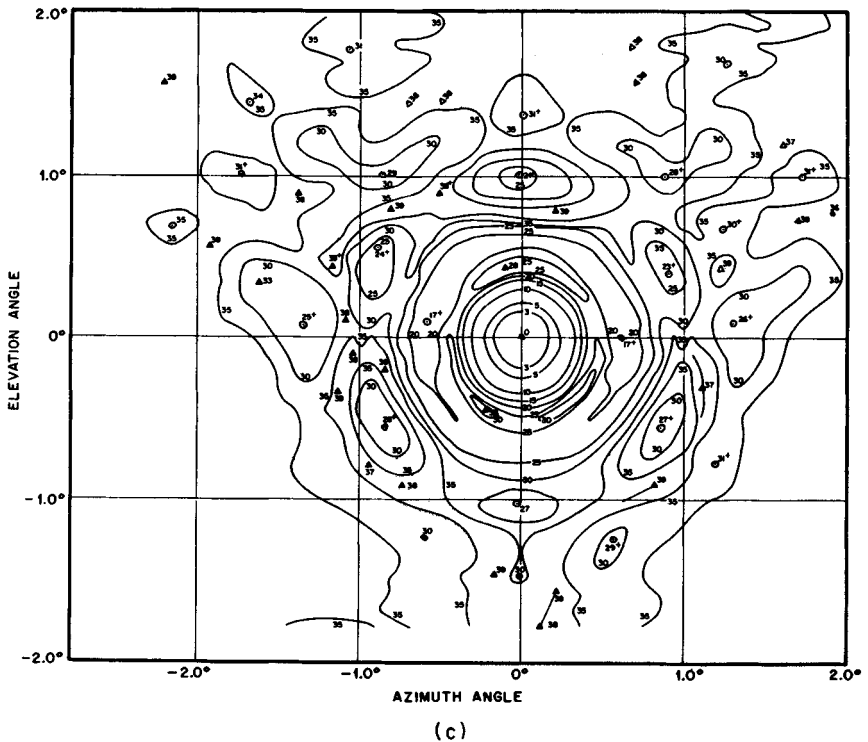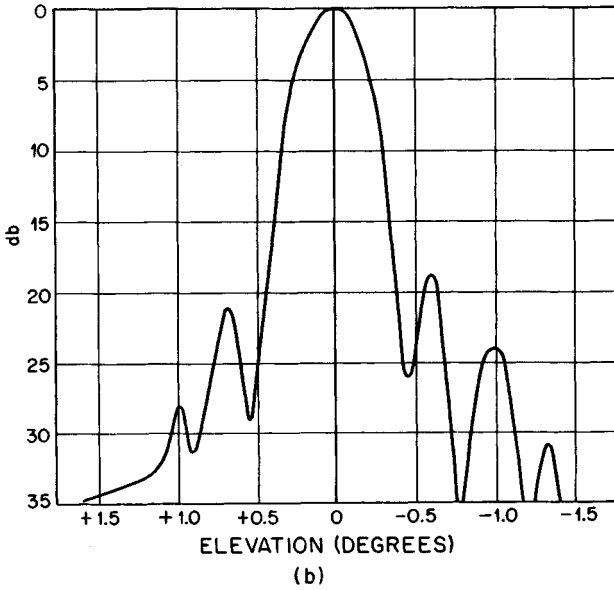FIG. 6.1   Typical pencil-beam pattern. (*a*) Three-dimensional cartesian plot of complete pattern.

(b)

(c)

FIG. 6.1 (*Continued*) (*b*) Principal-plane elevation pattern. (*c*) Contours of constant intensity (isophotes). (*Courtesy of D. D. Howard, Naval Research Laboratory.*)

thogonal to the first one, i.e., containing the peak and the 0° elevation angle, results in the so-called *azimuth pattern*, also a principal-plane cut because it contains the peak of the beam as well as one of the angular coordinate axes.

These principal planes are sometimes called *cardinal planes*. All other vertical planes through the beam peak are called *intercardinal planes*. Sometimes pattern cuts in the ±45° intercardinal planes are measured and plotted, but most often it is sufficient to plot only the azimuth and elevation patterns to describe the pattern performance of an antenna. In other words, it is sufficient (and much less costly) to *sample* the 3D pattern with two planar cuts containing the beam axis.

The terms *azimuth* and *elevation* imply earth-based reference coordinates, which may not always be applicable, particularly to an airborne or space-based (satellite) system. A more generic pair of principal planes for antennas in general are the so-called *E* and *H* plane of a linearly polarized antenna. Here the *E*-plane pattern is a principal plane containing the direction of the electric (*E*-field) vector of the radiation from the antenna, and the *H* plane is orthogonal to it, therefore containing the magnetic (*H*-field) vector direction. These two principal planes can be independent of earth-oriented directions such as azimuth and elevation and are widely used.

It should be noted that sampling 3D antenna patterns is not limited to planar cuts as described above. Sometimes it is meaningful and convenient (from a measuring-technique viewpoint) to take *conical cuts*, i.e., the intersections of the 3D pattern with cones of various angular widths centered on the electrical (or mechanical) axis of the antenna.

The typical 2D pattern shown in Fig. 6.1*b* is plotted in rectangular coordinates, with the vertical axis in decibels. This is by far the most widely used form of plotting patterns because it provides a wide dynamic range of pattern levels with good visibility of the pattern details. However, other forms of plotting-pattern data are also used, as illustrated in Fig. 6.2. This shows four forms of plotting the same sin x/x pattern, including (*a*) a polar plot of relative voltage (intensity), (*b*) a rectangular plot of voltage, (*c*) a rectangular plot of relative power (density), and (*d*) a rectangular plot of logarithmic power (in decibels). The linear voltage and power scales in *a, b,* and *c* leave much to be desired in showing lower-level pattern details, while *d* provides good "visibility" of the entire pattern. Of course, polar patterns can also be plotted by using decibels in the radial dimension, but lower-level details are compressed near the center of the pattern chart and visibility is very poor. Figure 6.2 shows the reason why *rectangular-decibel pattern plots* are most often preferred.

Beamwidth.   One of the main features of an antenna pattern is the *beamwidth* of the main lobe, i.e., its angular extent. However, since the main beam is a continuous function, its width varies from the peak to the nulls (or minima). The most frequently expressed width is the *half-power beamwidth* (HPBW), which occurs at the 0.707-relative-voltage level in Fig. 6.2*a* and *b*, at the 0.5-relative-power level in *c*, and at the 3 dB level in *d*. Sometimes other beamwidths are specified or measured, such as the one-tenth power (10 dB) beamwidth, or the width between nulls, but unless otherwise stated the simple term *beamwidth* implies the half-power (3 dB) width. This half-power width is also usually a measure of the *resolution* of an antenna, so that two identical targets at the same range are said to be resolved in angle if separated by the half-power beamwidth.

The beamwidth of an antenna depends on the size of the antenna aperture as well as on the amplitude and phase distribution across the aperture. For a given distribution, the beamwidth (in a particular planar cut) is inversely proportional to the size of the aperture (in that plane) expressed in wavelengths. That is, the half-power beamwidth is given by

$$\text{HPBW} = K/(D/\lambda) = K\lambda/D \qquad (6.11)$$

where D is the aperture dimension, λ is the free-space wavelength, and K is a proportionality constant known as the *beamwidth factor*. Each amplitude distribution (assuming a linear phase distribution) has a corresponding beamwidth factor, expressed either in radians or in degrees.
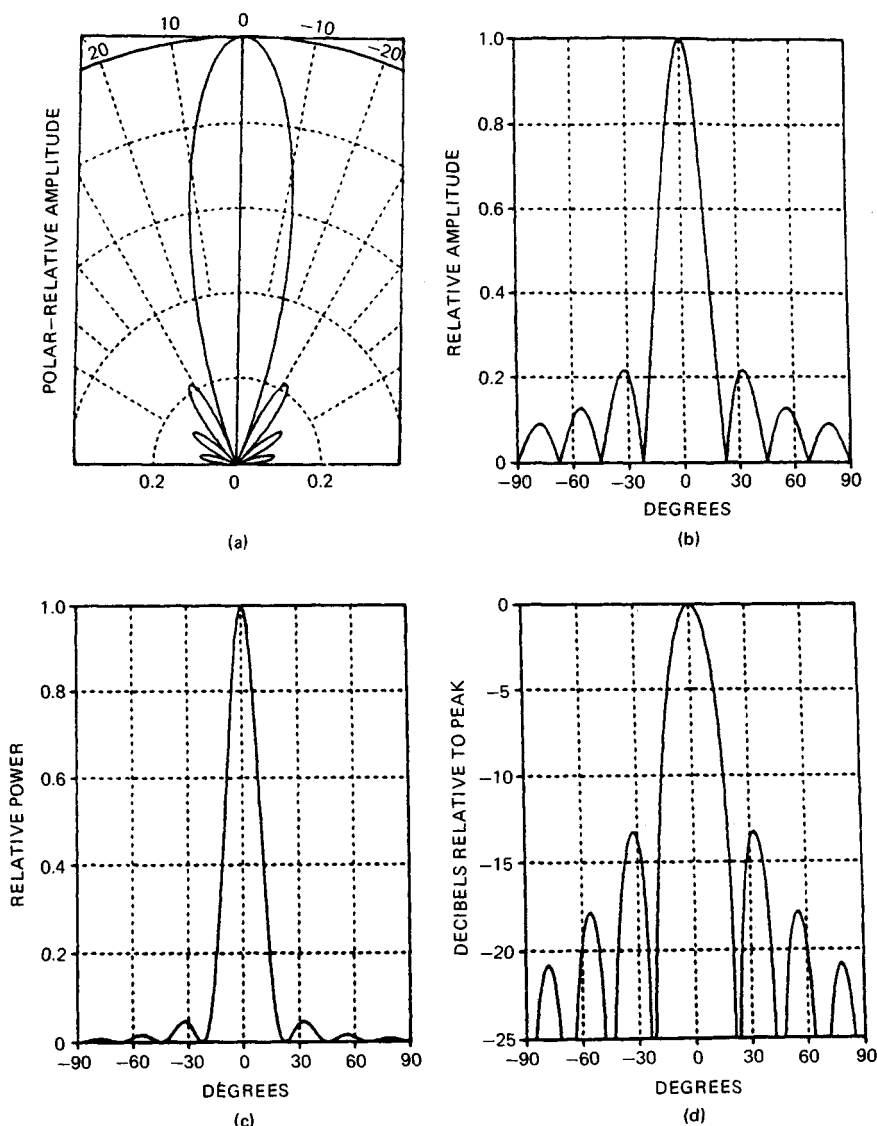

FIG. 6.2   Various representations of the same sin x/x pattern.

Sidelobes. The lobe structure of the antenna radiation pattern outside the major-lobe (main-beam) region usually consists of a large number of *minor lobes*, of which those adjacent to the main beam are *sidelobes*. However, it is common usage to refer to all minor lobes as sidelobes, in which case the adjacent lobes are called *first sidelobes*. The minor lobe approximately 180° from the main lobe is called the *backlobe*.

Sidelobes can be a source of problems for a radar system. In the transmit mode they represent wasted radiated power illuminating directions other than the desired main-beam direction, and in the receive mode they permit energy from undesired directions to enter the system. For example, a radar for detecting low-flying aircraft targets can receive strong ground echoes (*clutter*) through the sidelobes which mask the weaker echoes coming from low radar cross-section targets through the main beam. Also, unintentional interfering signals from friendly sources (electromagnetic interference, or EMI) and/or intentional interference from unfriendly sources (jammers) can enter through the minor lobes. It is therefore often (but not always) desirable to design radar antennas with sidelobes as low as possible (consistent with other considerations) to minimize such problems. (NOTE: There are systems in which the lowest possible sidelobes are not desirable; for example, to minimize main-beam clutter or jamming, it may be better to tolerate higher sidelobes in order to achieve a narrower main-beam null width).

To achieve low sidelobes, antennas must be designed with special tapered amplitude distributions across their apertures. For a required antenna gain this means that a larger antenna aperture is needed. Conversely, for a given size of antenna, lower sidelobes mean less gain and a correspondingly broader beamwidth. The optimum compromise (tradeoff) between sidelobes, gain, and beamwidth is an important consideration for choosing or designing radar antennas. Figure 7.23 of Chap. 7 shows these tradeoff relations for the optimum Taylor amplitude distribution[4,5] being widely used for sidelobe suppression in radar antennas. One set of curves is for rectangular (linear) apertures; the other, for circular Taylor distributions.

The *sidelobe levels* of an antenna pattern can be specified or described in several ways. The most common expression is the *relative sidelobe level*, defined as the peak level of the highest sidelobe relative to the peak level of the main beam. For example, a "-30 dB sidelobe level" means that the peak of the highest sidelobe has an intensity (radiated power density) one one-thousandth (10^{-3} or -30 dB) that of the peak of the main beam. Sidelobe levels can also be quantified in terms of their *absolute level* relative to isotropic. In the above example, if the antenna gain were 35 dB, the absolute level of the -30 dB relative sidelobe is $+5$ dBi, i.e., 5 dB above isotropic. For some radar systems, the peak levels of individual sidelobes are not as important as the *average level* of all the sidelobes. This is particularly true for airborne "down-look" radars like the Airborne Warning and Control System, or AWACS (E-3A), which require very low (ultralow) average sidelobe levels in order to suppress ground clutter. The average level is a *power average* (sometimes referred to as the *rms level*) formed by integrating the power in all minor lobes outside the main lobe and expressing it in decibels relative to isotropic (dBi). For example, if 90 percent of the power radiated is in the main beam, 10 percent (or 0.1) is in all the sidelobes: this corresponds to an average sidelobe level of -10 dBi. If the main beam contains 99 percent of the radiated power, the average sidelobe level is 0.01, or -20 dBi, etc. Ultralow aver-

age sidelobe levels, defined as better than −20 dBi, have been achieved with careful design and manufacturing processes.

One other way to describe sidelobe levels (not often used but sometimes meaningful) is by the *median level*; this is the level such that half of the angular space has sidelobe levels above it and the other half has them below that level.

Polarization. The direction of polarization of an antenna is defined as the direction of the electric-field (*E*-field) vector. Many existing radar antennas are *linearly polarized*, usually either *vertically* or *horizontally*; although these designations imply an earth reference, they are quite common even for airborne or satellite antennas.

Some radars use *circular polarization* in order to detect aircraftlike targets in rain. In that case, the direction of the *E* field varies with time at any fixed observation point, tracing a circular locus once per RF cycle in a fixed plane normal to the direction of propagation. Two senses of circular polarization (CP) are possible, right-hand (RHCP) and left-hand (LHCP). For RHCP, the electric vector appears to rotate in a clockwise direction when viewed as a wave receding from the observation point. LHCP corresponds to counterclockwise rotation. These definitions of RHCP and LHCP can be illustrated with hands, by pointing the thumb in the direction of propagation and curling the fingers in the apparent direction of *E*-vector rotation. By reciprocity, an antenna designed to radiate a particular polarization will also receive the same polarization.

The most general polarization is *elliptical polarization* (EP), which can be thought of as imperfect CP such that the *E* field traces an ellipse instead of a circle. A clear discussion of polarizations can be found in Kraus.[6,7]

Another increasingly important consideration for radar antennas is not only what polarization they radiate (and receive) but how pure their polarization is. For example, a well-designed vertical-polarization antenna may also radiate small amounts of the orthogonal horizontal polarization in some directions (usually off the main-beam axis). Similarly, an antenna designed for RHCP many also radiate some LHCP, which is mathematically orthogonal to RHCP. The desired polarization is referred to as the *main polarization* (COPOL), while the undesired orthogonal polarization is called cross polarization (CROSSPOL). Polarization purity is important in the sidelobe regions as well as in the main-beam region. Some antennas with low COPOL sidelobes may, if not properly designed, have higher CROSSPOL sidelobes, which could cause clutter or jamming problems. A well-designed antenna will have CROSSPOL components at least 20 dB below the COPOL in the main-beam region, and 5 to 10 dB below in the sidelobe regions. Reflecting surfaces near an antenna, such as aircraft wings or a ship superstructure, can affect the polarization purity of the antenna, and their effects should be checked.

6.3 TYPES OF ANTENNAS

Reflector antennas are built in a wide variety of shapes with a corresponding variety of feed systems to illuminate the surface, each suited to its particular application. Figure 6.3 illustrates the most common of these, which are described in some detail in the following subsections. The paraboloid in Fig. 6.3*a* collimates radiation from a feed at the focus into a pencil beam, providing high gain and minimum beamwidth. The parabolic cylinder in Fig. 6.3*b* performs this
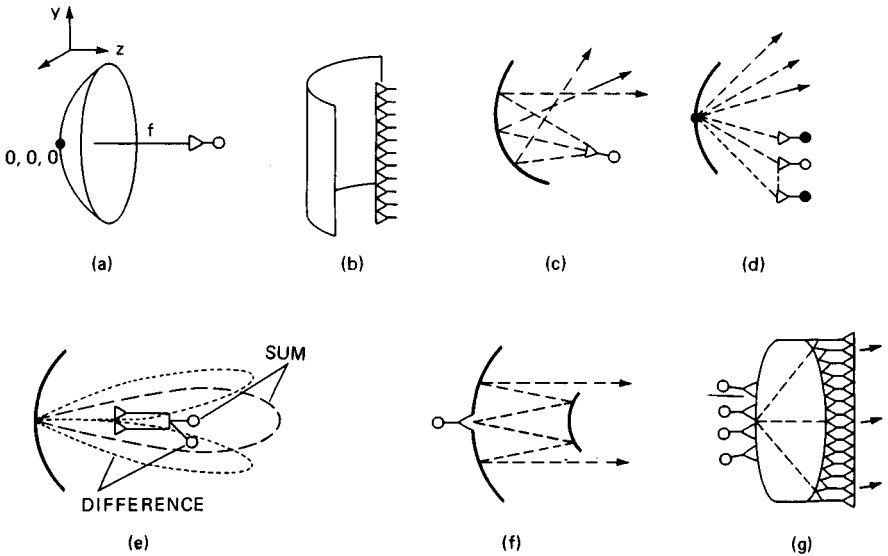
FIG. 6.3   Common reflector antenna types. (*a*) Paraboloid. (*b*) Parabolic cylinder. (*c*) Shaped. (*d*) Stacked beam. (*e*) Monopulse. (*f*) Cassegrain. (*g*) Lens.

collimation in one plane but allows the use of a linear array in the other plane, thereby allowing flexibility in the shaping or steering of the beam in that plane. An alternative method of shaping the beam in one plane is shown in Fig. 6.3*c*, in which the surface itself is no longer a paraboloid. This is a simpler construction, but since only the phase of the wave across the aperture is changed, there is less control over the beam shape than in the parabolic cylinder, whose linear array may be adjusted in amplitude as well.

Very often the radar designer needs multiple beams to provide coverage or to determine angle. Figure 6.3*d* shows that multiple feed locations produce a set of secondary beams at distinct angles. The two limitations on adding feeds are that they become defocused as they necessarily move away from the focal point and that they increasingly block the aperture. An especially common multiple-beam design is the monopulse antenna of Fig. 6.3*e*, used for angle determination on a single pulse, as the name implies. In this instance the second beam is normally a difference beam with its null at the peak of the first beam.

Multiple-reflector systems, typified by the Cassegrain antenna of Fig. 6.3*f*, offer one more degree of flexibility by shaping the primary beam and allowing the feed system to be conveniently located behind the main reflector. The symmetrical arrangement shown has significant blockage, but offset arrangements can readily be envisioned to accomplish more sophisticated goals.

Lenses (Fig. 6.3*g*) are not as popular as they once were, largely because phased arrays are providing many functions that lenses once fulfilled. Primarily they avoid blockage, which can become prohibitive in reflectors with extensive feed systems. A very wide assortment of lens types has been studied.[8-13]

In modern antenna design, combinations and variations of these basic types are widespread, with the goal of minimizing loss and sidelobes while providing the specified beam shapes and locations.
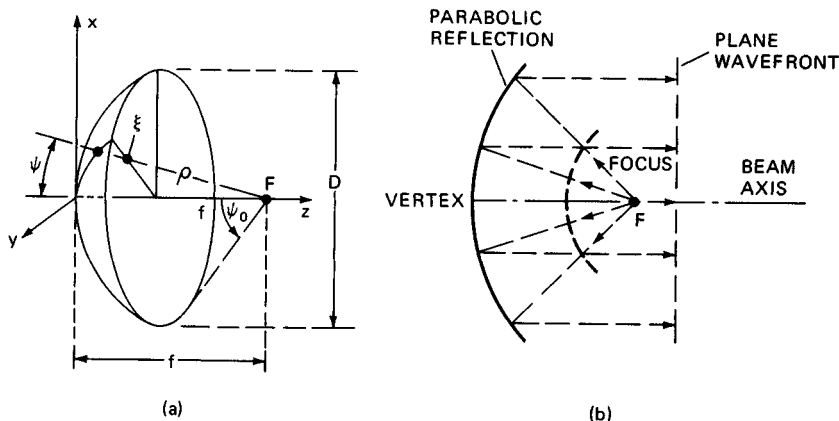
FIG. 6.4 Geometrical representation of a paraboloidal reflector. (*a*) Geometry. (*b*) Operation.

Paraboloidal Reflector Antennas. The theory and design of paraboloidal reflector antennas are extensively discussed in the literature.[2-4,14,15] The basic geometry is that of Fig. 6.4*a*, which assumes a parabolic conducting reflector surface of focal length *f* with a feed at the focus *F*. It can be shown from geometrical optics considerations that a spherical wave emerging from *F* and incident on the reflector is transformed after reflection into a plane wave traveling in the positive *z* direction (Fig. 6.4*b*).

The two coordinate systems that are useful in analysis are shown in Fig. 6.4*a*. In rectangular coordinates (x, y, z), the equation of the paraboloidal surface with a vertex at the origin $(0, 0, 0)$ is

$$z = (x^2 + y^2)/4f \qquad (6.12)$$

In spherical coordinates (ρ, ψ, ξ) with the feed at the origin, the equation of the surface becomes

$$\rho = f \sec^2 \frac{\psi}{2} \qquad (6.13)$$

This coordinate system is useful for designing the pattern of the feed. For example, the angle subtended by the edge of the reflector at the feed can be found from

$$\tan \frac{\psi_0}{2} = D/4f \qquad (6.14)$$

The aperture angle $2\psi_0$ is plotted as a function of f/D in Fig. 6.5. Reflectors with the longer focal lengths, which are flattest and introduce the least distortion of polarization and of off-axis beams, require the narrowest primary beams and therefore the largest feeds. For example, the size of a horn to feed a reflector of $f/D = 1.0$ is approximately 4 times that of a feed for a reflector of $f/D = 0.25$. Most reflectors are chosen to have a focal length between 0.25 and 0.5 times the diameter.
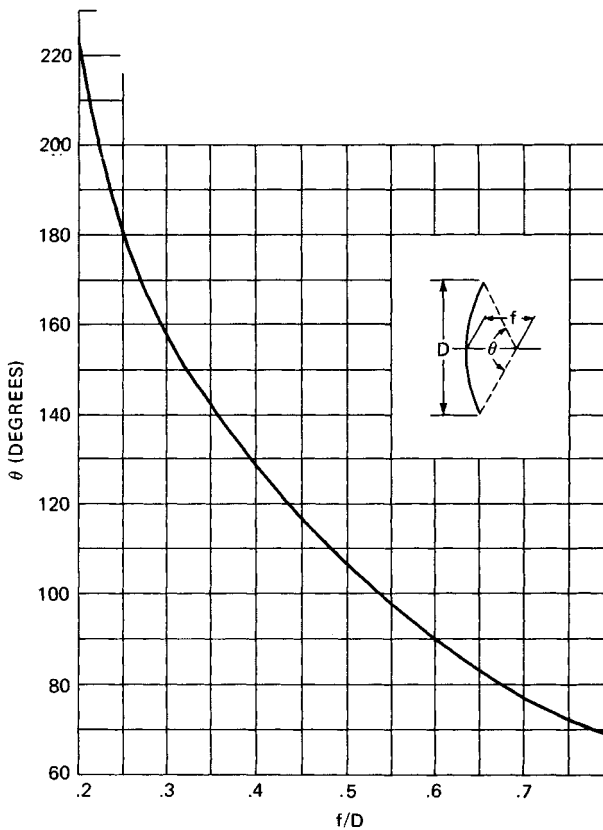
FIG. 6.5   Subtended angle of the edge of a paraboloidal reflector.

When a feed is designed to illuminate a reflector with a particular taper, the distance ρ to the surface must be accounted for, since the power density in the spherical wave falls off as $1/\rho^2$. Thus the level at the edge of the reflector will be lower than at the center by the product of the feed pattern and this "space taper." The latter is given in decibels as

$$\text{Space taper (dB)} = 20 \log \frac{(4f/D)^2}{1 + (4f/D)^2} \qquad (6.15)$$

Equation (6.15) is graphed in Fig. 6.6, showing a significant contribution at the smaller focal lengths. In low-sidelobe applications this amplitude variation may be used in conjunction with the feed pattern to achieve a specific shaping to the skirts of the distribution across the aperture.

Although this reflector is commonly illustrated as round with a central feed point, a variety of reflector outlines are in use, as shown in Fig. 6.7. Often, the
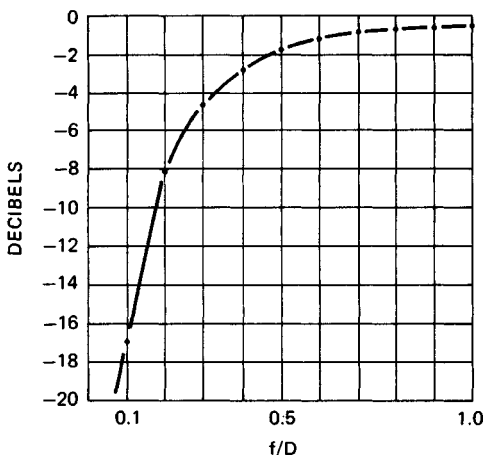
FIG. 6.6 Edge taper due to the spreading of the spherical wave from the feed ("space loss").
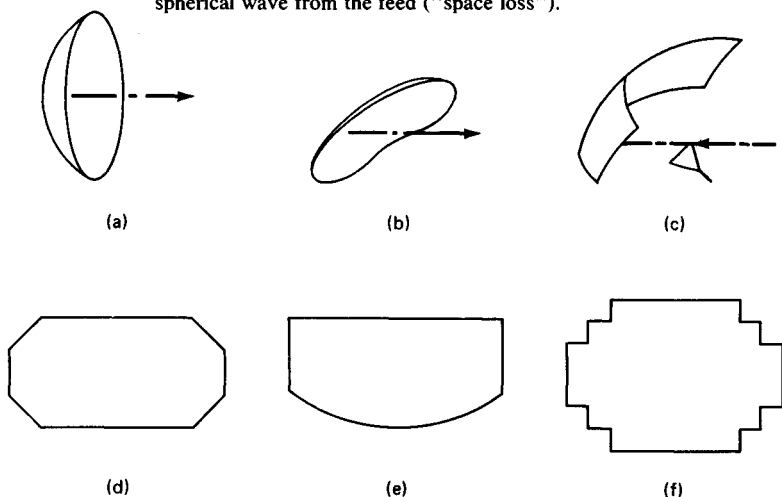
FIG. 6.7 Paraboloidal reflector outlines. (a) Round. (b) Oblong. (c) Offset feed. (d) Mitered corner. (e) Square corner. (f) Stepped corner.

azimuth and elevation beamwidth requirements are quite different, requiring the "orange-peel," or oblong, type of reflector of Fig. 6.7b.

As sidelobe levels are reduced and feed blockage becomes intolerable, offset feeds (Fig. 6.7c) become necessary. The feed is still at the focal point of the portion of the reflector used even though the focal axis no longer intersects the reflector. Feeds for an offset paraboloid must be aimed beyond the center of the reflector area to account for the larger space taper on the side of the dish away from the feed. The result is an unsymmetrical illumination.

The corners of most paraboloidal reflectors are rounded or mitered as in Fig. 6.7d to minimize the area and especially to minimize the torque required to turn

the antenna. The deleted areas have low illumination and therefore least contribution to the gain. However, circular and elliptical outlines produce sidelobes at all angles from the principal planes. If low sidelobes are specified away from the principal planes, it may be necessary to maintain square corners, as shown in Fig. 6.7e.

Parabolic reflectors still serve as a basis for many radar antennas in use today, since they provide the maximum available gain and minimum beamwidths with the simplest and smallest feeds.

Parabolic-Cylinder Antenna.[2,16,17] It is quite common that either the elevation or the azimuth beam must be steerable or shaped while the other is not. A parabolic cylindrical reflector fed by a line source can accomplish this flexibility at a modest cost. The line source feed may assume many different forms ranging from a parallel-plate lens to a slotted waveguide to a phased array using standard designs.[2-4]

The parabolic cylinder has application even where both patterns are fixed in shape. The AN/TPS-63 (Fig. 6.8) is one such example in which elevation beam
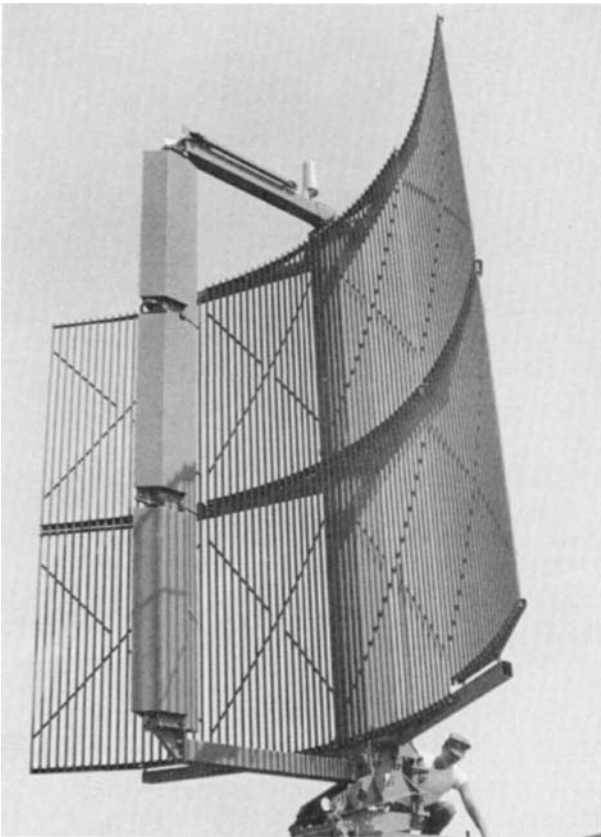
FIG. 6.8 AN/TPS-63 parabolic-cylinder antenna. (*Courtesy Westinghouse Electric Corporation.*)

shaping must incorporate a steep skirt at the horizon to allow operation at low elevation angles without degradation from ground reflection. A vertical array can produce much sharper skirts than a shaped dish of equal height can, since a shaped dish uses part of its height for high-angle coverage. The array can superimpose high and low beams on a common aperture, thereby benefiting from the full height for each.

The basic parabolic cylinder is shown in Fig. 6.9, in which the reflector surface has the contour

$$z = y^2/4f \tag{6.16}$$

The feed is on the focal line $F\text{-}F'$, and a point on the reflector surface is located with respect to the feed center at x and $\rho = f \sec^2 \psi/2$. Many of the guidelines for paraboloids except space taper can be carried over to parabolic cylinders. Since the feed energy diverges on a cylinder instead of a sphere, the power density falls off as ρ rather than ρ^2. Therefore, the space taper of Eq. (6.15) is halved in decibels.
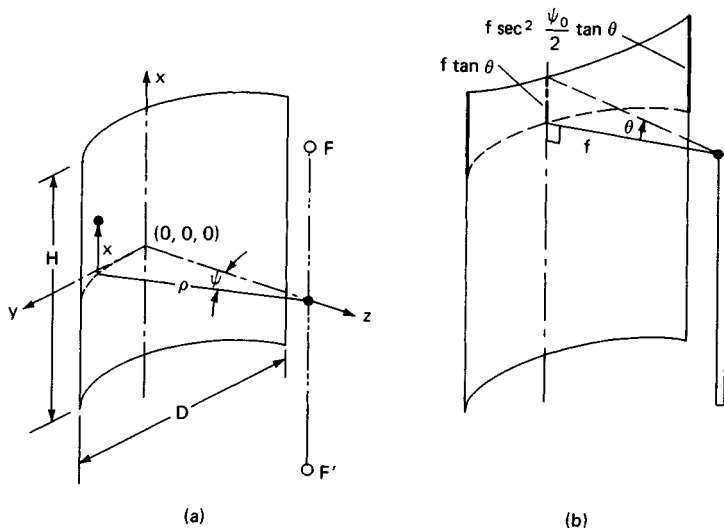
FIG. 6.9 Parabolic cylinder. (*a*) Geometry. (*b*) Extension.

The height or length of the parabolic cylinder must account for the finite beamwidth, shaping, and steering of the linear feed array. As Fig. 6.9 indicates, at angle θ from broadside the primary beam intercepts the reflector at $f \tan \theta$ past the end of the vertex. Since the peak of the primary beam from a steered line source lies on a cone, the corresponding intercepts on the right and left corners of the top of reflector are farther out at $f \sec^2 \psi_0/2 \tan \theta$. For this reason, the corners of a parabolic cylinder are seldom rounded in practice.

Parabolic cylinders suffer from large blockage if they are symmetrical, and they are therefore often built offset. Properly designed, however, a cylinder fed by an offset multiple-element line source can have excellent performance[18] (Fig. 6.10). A variation on this design has the axis of the reflector horizontal, fed with
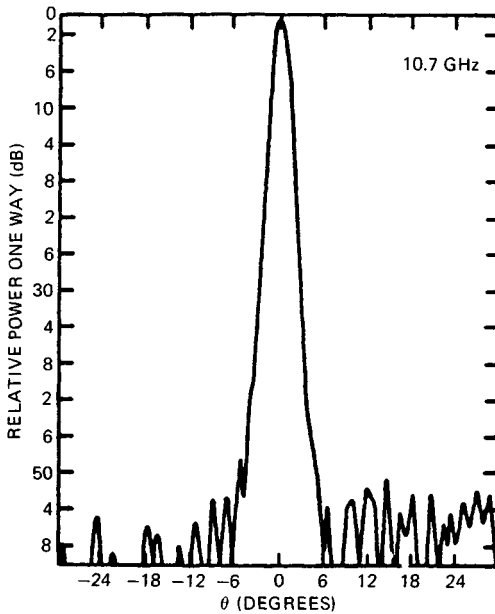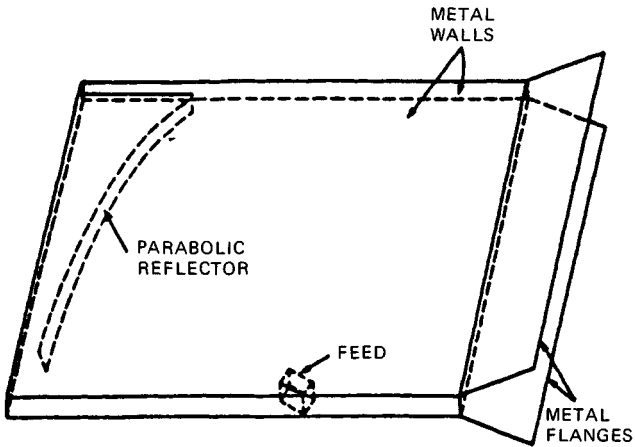
FIG. 6.10 Pillbox structure used to test a low-sidelobe parabolic cylinder and its measured pattern. (*Courtesy Ronald Fante, Rome Air Development Center.*)

a linear array for low-sidelobe azimuth patterns and shaped in height for elevation coverage. It is an economical alternative to a full two-dimensional array.

Shaped Reflectors. Fan beams with a specified shape are required for a variety of reasons. The most common requirement is that the elevation beam

provide coverage to a constant altitude. If secondary effects are ignored and if the transmit and receive beams are identical, this can be obtained with a power radiation pattern proportional to $csc^2\theta$, where θ is the elevation angle.[2,19] In practice, this well-known cosecant-squared pattern has been supplanted by similar but more specific shapes that fit the earth's curvature and account for sensitivity time control (STC).

The simplest way to shape the beam is to shape the reflector, as Fig. 6.11 illustrates. Each portion of the reflector is aimed in a different direction and, to the extent that geometric optics applies, the amplitude at that angle is the integrated sum of the power density from the feed across that portion. Silver[2] describes the procedure to determine the contour for a cosecant-squared beam graphically. However, with modern computers arbitrary beam shapes can be approximated accurately by direct integration of the reflected primary pattern. In so doing, the designer can account for the approximations to whatever accuracy he or she chooses. In particular, the azimuth taper of the primary beam can be included,
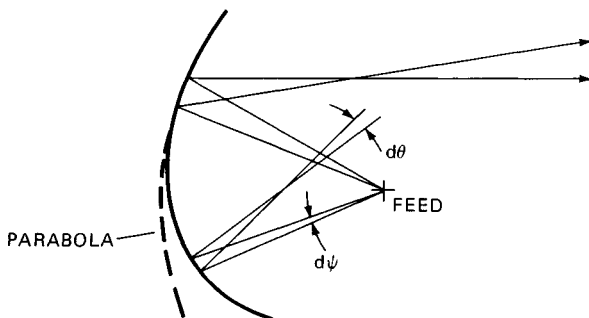
FIG. 6.11   Reflector shaping.

VIEW FROM ELEVATION ϕ
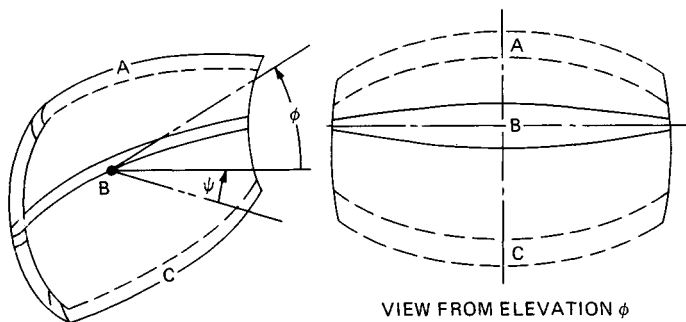
FIG. 6.12   Three-dimensional shaped-reflector design.

and the section of the reflector aimed at elevation θ can be focused in azimuth and have a proper outline when viewed from elevation θ (Fig. 6.12). Without these precautions off-axis sidelobes can be generated by banana-shaped sections.

Most shaped reflectors take advantage of the shaping to place the feed outside the secondary beam. Figure 6.13 shows how blockage can be virtually eliminated

even though the feed appears to be opposite the reflector.

The ASR-9 (Fig. 6.14) typifies shaped reflector antennas designed by these procedures. The elevation shaping, azimuth beam skirts, and sidelobes are closely controlled by the use of the computer-aided design process.

A limitation of shaped reflectors is that a large fraction of the aperture is not used in forming the main beam. If the feed pattern is symmetrical and half of the power is directed to wide

FIG. 6.13   Elimination of blockage.

angles, it follows that the main beam will use half of the aperture and have double the beamwidth. This corresponds to shaping an array pattern with phase only and may represent a severe problem if sharp pattern skirts are required. It can be avoided with extended feeds.

Multiple Beams and Extended Feeds.[19–21]   A feed at the focal point of a parabola forms a beam parallel to the focal axis. Additional feeds displaced from the focal point form additional beams at angles from the axis. This is a powerful capability of the reflector antenna to provide extended coverage with a modest increase in hardware. Each additional beam can have nearly full gain, and adjacent beams can be compared with each other to interpolate angle.

A parabola reflects a spherical wave into a plane wave only when the source is at the focus. With the source off the focus, a phase distortion results that in-
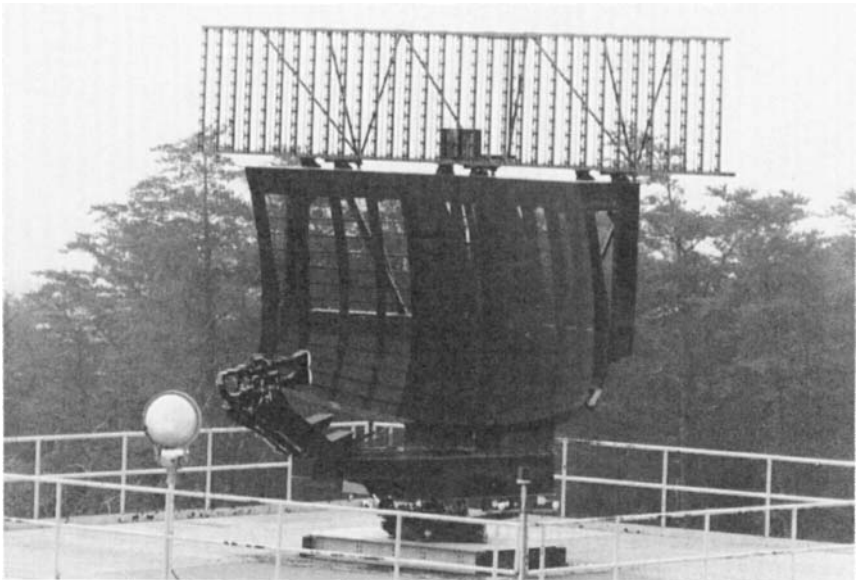
FIG. 6.14   ASR-9 shaped reflector with offset feed and an air traffic control radar beacon system (ATCRBS) array mounted on top. (*Courtesy Westinghouse Electric Corporation.*)
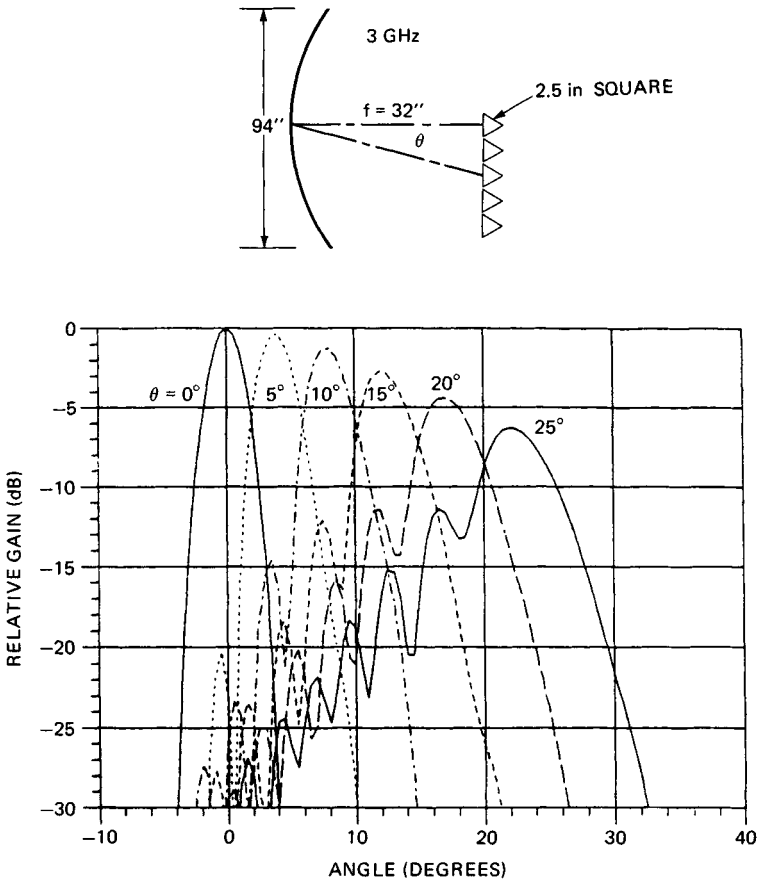
FIG. 6.15 Patterns for off-axis feeds.

creases with the angular displacement in beamwidths and decreases with an increase in the focal length. Figure 6.15 shows the effect of this distortion on the pattern of a typical dish as a feed is moved off axis. A flat dish with a long focal length minimizes the distortions. Progressively illuminating a smaller fraction of the reflector as the feed is displaced accomplishes the same purpose.

Two secondary effects are influential in the design of extended feeds. If an off-axis feed is moved parallel to the focal axis, the region of minimum distortion moves laterally in the reflector. At the same time, if the reflector is a paraboloid of revolution, the focus in the orthogonal plane (usually the azimuth) is altered. For the reflector region directly in front of the displaced feed, it has been found that both planes are improved by moving progressively back from the focal plane. This is clearly illustrated in the side view of the AN/TPS-43 antenna of Fig. 6.16. If that feed is examined carefully, one can also see that off-axis feeds become progressively larger so as to form progressively wider elevation beams that maintain a nearly constant number of beamwidths off axis. This is often made possible by radar coverage requirements having reduced range at wide elevation angles.
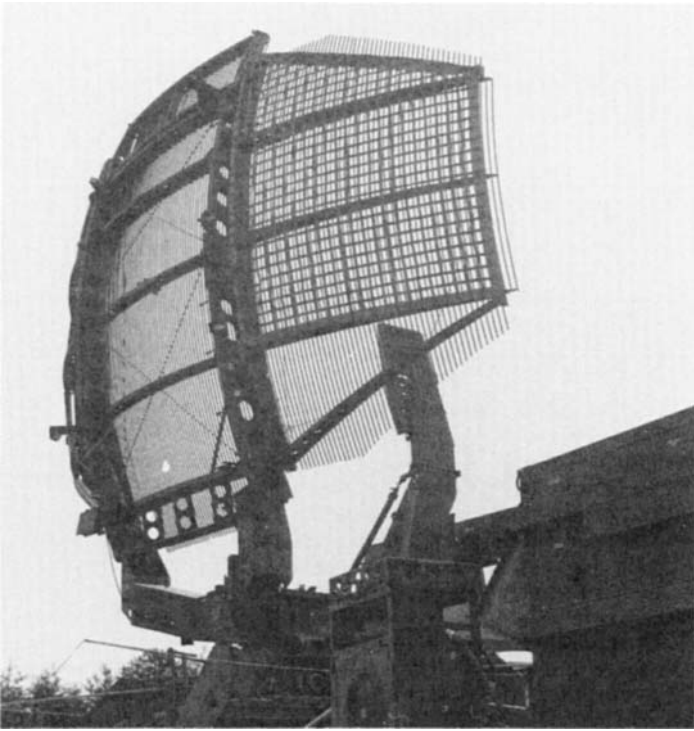
FIG. 6.16   AN/TPS-43 multiple-beam antenna. (*Courtesy Westinghouse Electric Corporation.*)

For some purposes the extended feed is not placed about the focal plane at all. If we consider the reflector as a collector of parallel rays from a range of angles and examine the converging ray paths (Fig. 6.17) it is evident that a region can be found that intercepts most of the energy. A feed array in that region driven with suitable phase and amplitude can therefore efficiently form beams at any of the angles. This ability has been used in various systems as a means of forming agile beams over a limited sector. It has also been used as a means of shaping beams and of forming very low sidelobe illumination functions. One such antenna is illustrated in Fig. 6.18.

Monopulse Feeds.[22-25]   Monopulse is the most common form of multiple-beam antenna, normally used in tracking systems in which a movable antenna keeps the target near the null and measures the mechanical angle, as opposed to a surveillance system having overlapping beams with angles measured from RF difference data.

Two basic monopulse systems, phase comparison and amplitude comparison, are illustrated in Fig. 6.19. The amplitude system is far more prevalent in radar antennas, using the sum of the two feed outputs to form a high-gain, low-sidelobe beam, and the difference to form a precise, deep null at boresight. The sum beam is used on transmit and on receive to detect the target. The difference port provides angle determination. Usually both azimuth and elevation differences are provided.

If a reflector is illuminated with a group of four feed elements, a conflict arises
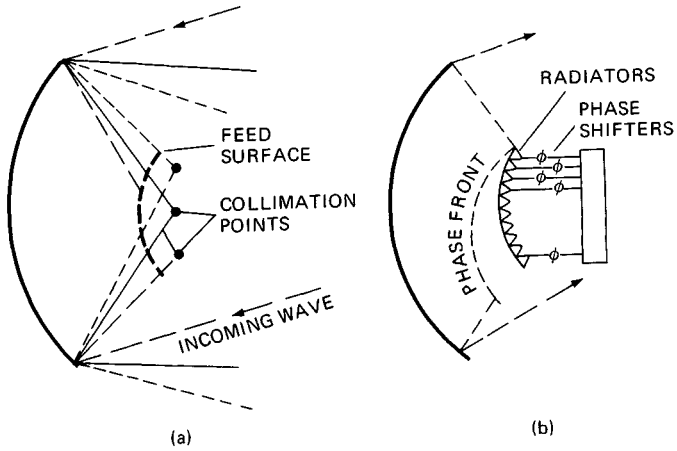
FIG. 6.17 Extended feeds off the focal plane. (*a*) Geometry. (*b*) Feed detail.
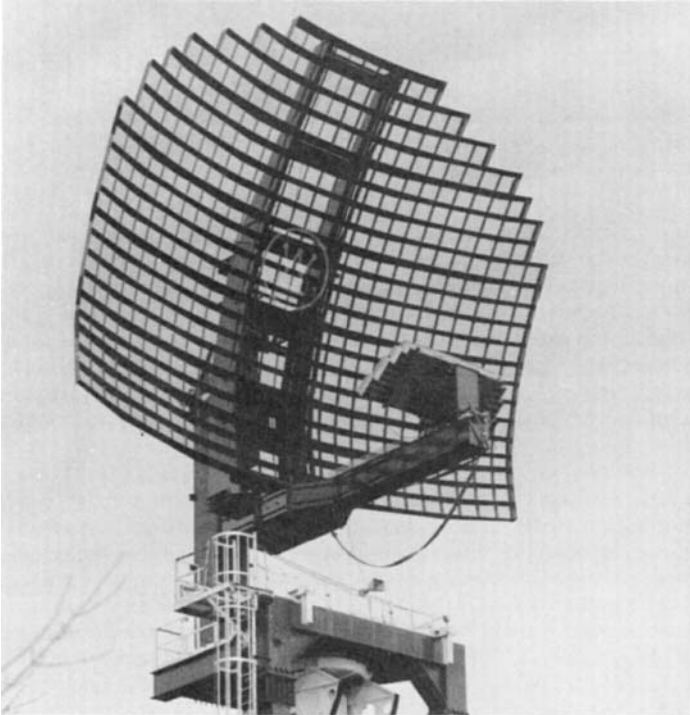
FIG. 6.18 Low-sidelobe reflector using an extended feed. (*Courtesy Westinghouse Electric Corporation.*)
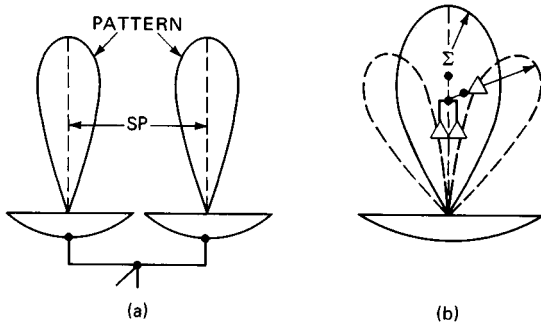
FIG. 6.19 Monopulse antennas. (*a*) Phase. (*b*) Amplitude.
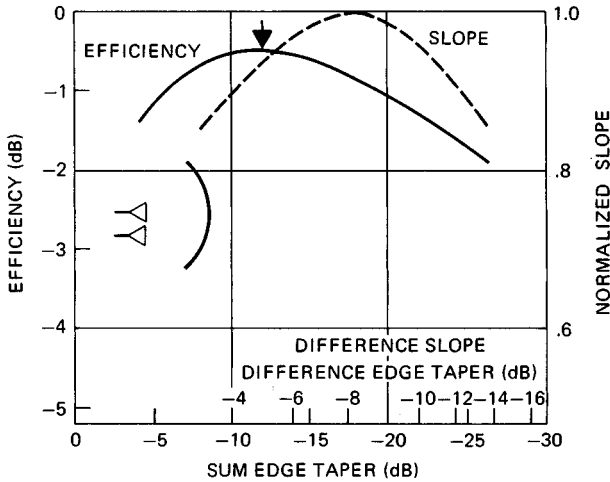
FIG. 6.20 Conflicting taper requirements for sum and difference horn designs (*H* plane illustrated).

between the goals of high sum-beam efficiency and high difference-beam slopes. The former requires a small overall horn size, while the latter requires large individual horns (Fig. 6.20). Numerous methods have been devised to overcome this problem, as well as the associated high-difference sidelobes. In each case the comparator is arranged to use a different set of elements for the sum and difference beams. In some cases this is accomplished with oversized feeds that permit two modes with the sum excitation. Hannan[24] has tabulated results for several configurations, as summarized in Table 6.1.

Multiple-Reflector Antennas.[26-31] Some of the shortcomings of paraboloidal reflectors can be overcome by adding a secondary reflector. The contour of the added reflector determines how the power will be distributed across the primary reflector and thereby gives control over amplitude in addition to phase in the aperture. This can be used to produce very low spillover or to produce a specific low-sidelobe distribution. The secondary reflector may also be used to

TABLE 6.1 Monopulse Feedhorn Performance

| Type of horn | H plane | | E plane | Sidelobes, dB | | Feed shape |
	Efficiency	Slope	Slope	Sum	Difference	
Simple four-horn	0.58	1.2	1.2	19	10	
Two-horn dual-mode	0.75	1.6	1.2	19	10	
Two-horn triple-mode	0.75	1.6	1.2	19	10	
Twelve-horn	0.56	1.7	1.6	19	19	
Four-horn triple-mode	0.75	1.6	1.6	19	19	

relocate the feed close to the source or receiver. By suitable choice of shape, the apparent focal length can be enlarged so that the feed size is convenient, as is sometimes necessary for monopulse operation.

The Cassegrain antenna (Fig. 6.21), derived from telescope designs, is the most common antenna using multiple reflectors. The feed illuminates the hyperboloidal subreflector, which in turn illuminates the paraboloidal main reflector. The feed is placed at one focus of the hyperboloid and the paraboloid focus at the other. A similar antenna is the gregorian, which uses an ellipsoidal subreflector in place of the hyperboloid.

The parameters of the Cassegrain antenna are related by the following expressions:

$$\tan \psi_r/2 = 0.25 D_m/f_m \qquad (6.17)$$

$$1/\tan \psi_v + 1/\tan \psi_r = 2f_s/D_s \qquad (6.18)$$

$$1 - 1/e = 2 L_r/f_c \qquad (6.19)$$

where the eccentricity e of the hyperboloid is given by

$$e = \sin [(\psi_v + \psi_r)/2]/\sin [(\psi_v - \psi_r)/2] \qquad (6.20)$$

The equivalent-paraboloid[4,26] concept is a convenient method of analyzing the radiation characteristics in which the same feed is assumed to illuminate a virtual reflector of equal diameter set behind the subreflector. The equation

$$f_e = D_m/(4 \tan \psi_r/2) \qquad (6.21)$$

defines the equivalent focal length, and the magnification m is given by

$$m = f_e/f_m = (e + 1)/(e - 1) \qquad (6.22)$$

Thus the feed may be designed to produce suitable illumination within subtended angles $\pm\psi_r$ for the longer focal length. Typical aperture efficiency can be better than 50 to 60 percent.

Aperture blocking can be large for symmetrical Cassegrain antennas. It may
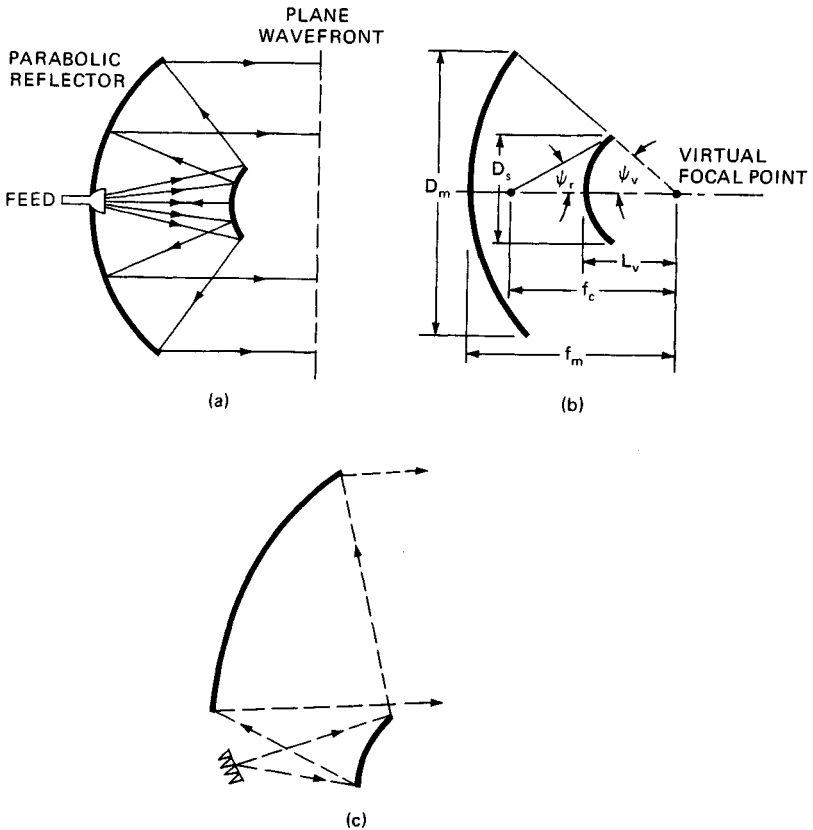
FIG. 6.21 The Cassegrain reflector antenna. (*a*) Schematic diagram. (*b*) Geometry. (*c*) Offset dual reflector.

be minimized by choosing the diameter of the subreflector equal to that of the feed.[26] This occurs when

$$D_s' \simeq \sqrt{2f_m\lambda/k} \qquad (6.23)$$

where k is the ratio of the feed-aperture diameter to its effective blocking diameter. Ordinarily k is slightly less than 1. If the system allows, blocking can be reduced significantly by using a polarization-twist reflector and a subreflector made of parallel wires.[3,26] The subreflector is transparent to the secondary beam with its twisted polarization.

In the general dual-reflector case, blockage can be eliminated by offsetting both the feed and the subreflector (Fig. 3.21c). With blockage and supporting struts and spillover virtually eliminated, this is a candidate for very low sidelobes.[32] It can be used in conjunction with an extended feed to provide multiple or steerable beams.[33]

Special-Purpose Reflectors. Several types of antennas are occasionally used for special purposes. One such antenna is the spherical reflector,[34] which can be scanned over very wide angles with a small but fixed phase error known as spherical aberration. The basis of this antenna is that, over small regions, a spherical surface viewed from a point halfway between the center of the circle and the surface is nearly parabolic. If the feed is moved circumferentially at constant radius $R/2$, where R = the radius of the circular reflector surface, the secondary beam can be steered over whatever angular extent the reflector size permits. In fact, 360° of azimuth steering may be accomplished if the feed polarization is tilted 45° and the reflector is formed of conducting strips parallel to the polarization. The reflected wave is polarized at right angles to the strips on the opposite side. This antenna is known as a Helisphere.[35]

If the scanning is in azimuth only, the height dimension of the reflector may be parabolic for perfect elevation focus. This is the parabolic torus,[3,4] which has been used in fixed radar installations.

6.4 FEEDS

Because most radar systems operate at microwave frequencies (L band and higher), feeds for reflector antennas are typically some form of flared waveguide horn. At lower frequencies (L band and lower) dipole feeds are sometimes used, particularly in the form of a linear array of dipoles to feed a parabolic-cylinder reflector. Other feed types used in some cases include waveguide slots, troughs, and open-ended waveguides, but the flared waveguide horns are most widely used.

Paraboloidal reflectors (in the receive mode) convert incoming plane waves into spherical phase fronts centered at the focus. For this reason, feeds must be point-source radiators; i.e., they must radiate spherical phase fronts (in the transmit mode) if the desired directive antenna pattern is to be achieved. Other characteristics that a feed must provide include the proper illumination of the reflector with a prescribed amplitude distribution and minimum spillover and correct polarization with minimum cross polarization; the feed must also be capable of handling the required peak and average power levels without breakdown under all operational environments. These are the basic factors involved in the choice or design of a feed for a reflector antenna. Other considerations include operating bandwidth and whether the antenna is a single-beam, multibeam, or monopulse antenna.

Rectangular (pyramidal) waveguide horns propagating the dominant TE_{01} mode are widely used because they meet the high power and other requirements, although in some cases circular waveguide feeds with conical flares propagating the TE_{11} mode have been used. These single-mode, simply flared horns suffice for pencil-beam antennas with just one linear polarization.

When more demanding antenna performance is required, such as polarization diversity, multiple beams, high beam efficiency, or ultralow sidelobes, the feeds become correspondingly more complex. For such antennas segmented, finned, multimode, and/or corrugated horns are used. Figure 6.22 illustrates a number of feed types, many of which are described in more detail in antenna references.[3,36,37]
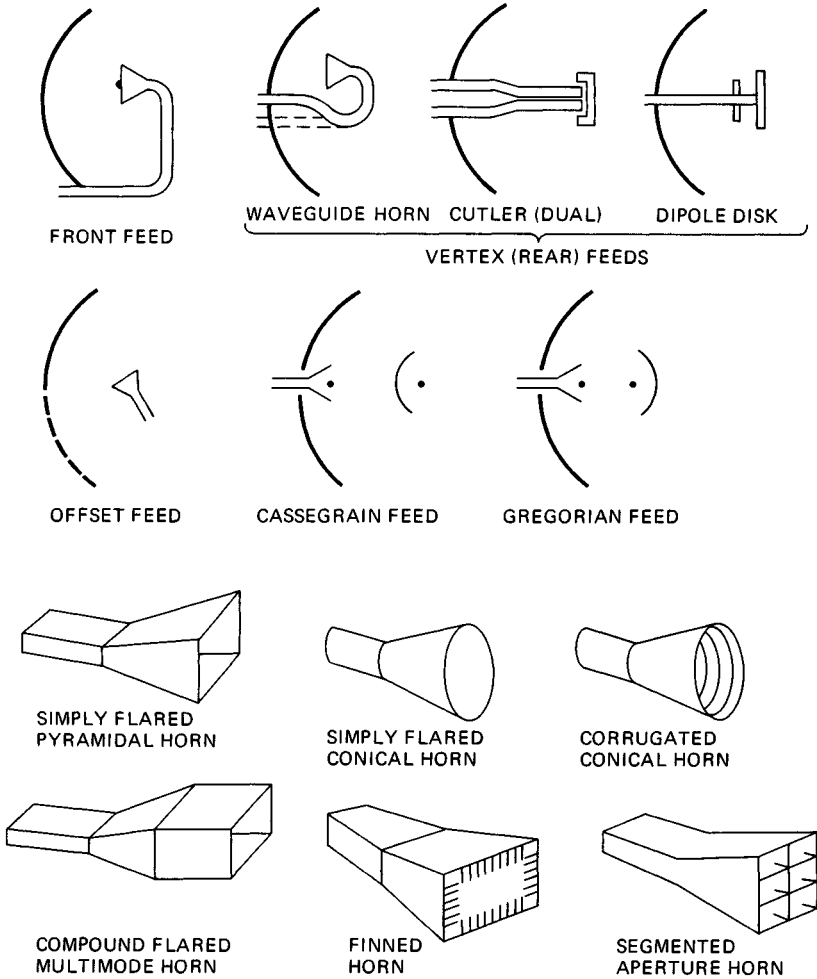
FIG. 6.22 Various types of feeds for reflector antennas.

6.5 REFLECTOR ANTENNA ANALYSIS

In calculating the antenna radiation pattern, it is assumed that the reflector is a distance from the feed such that the incident field on it has a spherical wavefront. There are two methods[2,38] for obtaining the radiation field produced by a reflector antenna. The first method, known as the current-distribution method, calculates the field from the currents induced on the reflector because of the primary

field of the feed. The second method, known as the aperture-field method, obtains the far field from the field distribution in the aperture plane. Both the current and aperture-field distributions are obtained from geometrical optics considerations. The two methods predict the same results in the limit $\lambda/D \rightarrow$?. However, in contrast to the aperture-field method, the current-distribution method can explain the effect of antenna surface curvature on the sidelobe levels and on the polarization. While the aperture-field method is handy for approximations and estimates, another problem is that it assumes that the reflection from the surface forms a planar wavefront. This is true for a paraboloidal reflector fed at its focal point, but otherwise it is not true. For this reason, the analysis that follows is devoted to the more general current-distribution method, or induced-current approach.

Although most antennas are reciprocal devices (have the same patterns in receive and transmit), analysis typically follows the transmit situation in which the signal begins at the feed element and its progress is tracked to the far field. Also at the feed, the polarization is in its purest form, so the vector properties are best known at this point and are described in many textbooks. In the analyses that follow, the constants are usually stripped away from the textbook versions since the antenna designer's primary goal in analysis is normally to determine the antenna's gain and pattern for main and cross polarizations. Therefore, the designer will normally integrate the power radiated from a feed into a sphere to determine the normalization factors needed. The magnetic field $\bar{H}$ of the feed is chosen because it leads to the reflector surface current $\bar{J}$ via the normal to the surface $\hat{n}$, by $\bar{J} = \hat{n} \times \bar{H}$.
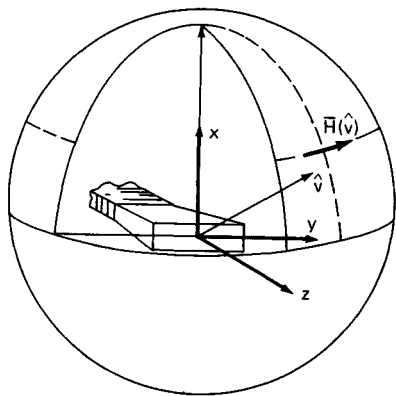
FIG. 6.23   Feed and normalization geometry.

The primary feed is assumed to radiate with the H field, $\bar{H}(\hat{v})$, perpendicular to the direction unit vector $\hat{v}$ (Fig. 6.23). It is dependent on feed type, say, horn or dipole. This H field is normalized such that the total power into a surrounding sphere (the magnitude of the tangent field squared) is equal to 1 W. This may be done by numerical integration using as much symmetry as possible to reduce computation time.

Radar reflectors are normally used to shape and distribute energy, which is more complicated than the case of the symmetrical paraboloidal reflector. Thus, where the focused paraboloid reflects into a common plane over the entire reflector, the shaped reflector focuses into many planes and the most general analysis is to treat the problem as an incremental summation of E fields. Another advantage of this analysis method is that the reflector outline can also be most general.

The surface of the reflector is divided into rectangular grid regions of area dA (Fig. 6.24), which intercept the feed-radiated field. The surface current then is the cross product of the H field with the normal to the surface $\hat{n}$ modified by differential area and a phase term,

$$\bar{J} = \hat{n} \times \bar{H}(\hat{v}) \, (\hat{v} \cdot \hat{n}) \, e^{-ikr} \, dA/4\pi r \qquad (6.24)$$
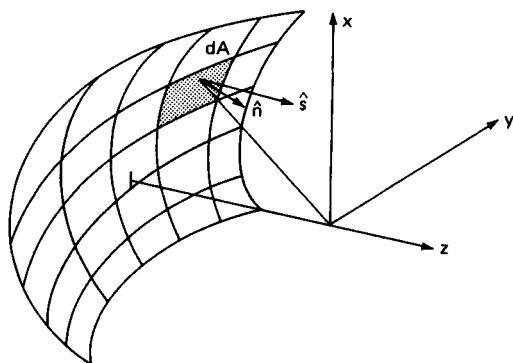
FIG. 6.24 Reflector geometry.

where r = the distance from the feed to the reflecting surface and $k = 2\pi/\lambda$ = the wavenumber.

Each reflector grid region represents the reflection of a small uniformly illuminated section. It has a gain factor and also a direction of reflection, which follows Snell's law. The direction of reflection $\hat{s}$ can be written

$$\hat{s} = \hat{v} - 2 \, (\hat{n} \cdot \hat{v}) \, \hat{n} \tag{6.25}$$

and the differential surface reflection at each grid region is modified by a pattern factor represented by a uniformly illuminated reflection steered in the direction of unit vector $\hat{s}$ and determined in the pattern direction unit vector $\hat{p}$ by

$$\text{Pattern factor} = \frac{4\pi dA}{\lambda^2 |\hat{n} \cdot \hat{s}|} \frac{\sin \pi \Delta x(s_x - p_x)}{\pi \Delta x(s_x - p_x)} \frac{\sin \pi \Delta y(s_y - p_y)}{\pi \Delta y(s_y - p_y)} \tag{6.26}$$

This factor modifies the surface current $\bar{J}$ as seen in the far field, projected in the direction of interest. At a distant spherical surface, vector $\hat{p}$ is normal to the surface. Two vectors are determined at the surface for the polarizations of interest, both perpendicular to p and perpendicular to each other, considered main and cross-polarized directions. The dot product of each of these unit vectors with $\bar{J}$ gives the field in the main and cross-polarized directions.

This pattern solution is found by a compromise between the number of grid regions and the time to compute all the parameters for each grid point. When the pattern is desired far off the pattern peak, one must consider the artificial grating lobes created by the computation method itself, in which case the grid density must be increased. With grid size Δx, the artificially induced grating lobe will appear at the angle found from $\sin \theta = \lambda/\Delta x$. Frequently, the user of such computational tools will trade off grid density in the orthogonal plane to enhance computation accuracy in the plane of interest.

Typically, the computer time-consuming operations in this type of pattern computation are trigonometric functions, sines and cosines, and square roots used in length and thus phase calculations. Extensive techniques are usually de-

vised to minimize repetitive calculations through the use of arrays containing the unit vectors in the pattern directions of interest, polarization vectors for those directions, and symmetry.

The literature contains many articles showing how *geometric theory of diffraction* (GTD) techniques can be used to compute reflector patterns.[7,39] The problem one encounters with GTD is in generalizing the situation, such as an irregular antenna outline. The primary use one finds for GTD is in defining the antenna's operation in the back hemisphere, but for many antennas the irregularity of the edges requires an agonizingly complex description of the antenna. This is often found to be impractical to implement into a GTD analysis. Sometimes simple analyses are performed at special angles of interest.

6.6 SHAPED-BEAM ANTENNAS

Rotating search radars typically require antenna patterns which have a narrow azimuth beamwidth for angular resolution and a shaped elevation pattern designed to meet multiple requirements. When circular polarization is also one of the system requirements, a shaped reflector is almost always the practical choice, since circularly polarized arrays are quite expensive.

A typical range coverage requirement might look like that shown in Fig. 6.25. At low elevation angles, the maximum range is the critical requirement. Above the height–range limit intersection, altitude becomes the governing requirement, resulting in a cosecant-squared pattern shape. At still higher elevation angles,
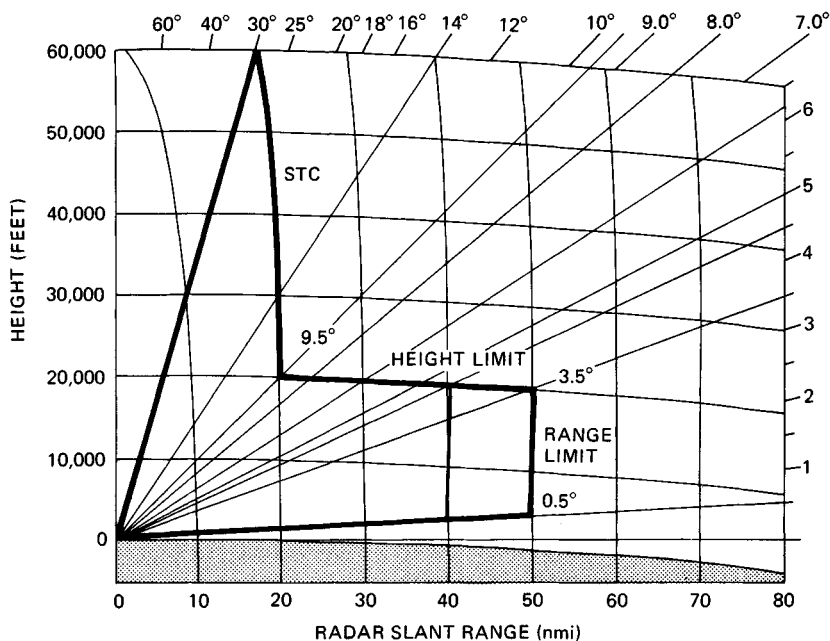
FIG. 6.25 Typical two-way coverage requirement example.

CHAPTER 7
PHASED ARRAY RADAR ANTENNAS

Theodore C. Cheston
Naval Research Laboratory

Joe Frank
Technology Service Corporation

7.1 INTRODUCTION

Phased Array Radars

Multifunction Radar. Early radar systems used antenna arrays formed by the combination of individual radiators. Such antennas date back to the turn of the twentieth century.[1,2,3] Antenna characteristics are determined by the geometric position of the radiators and the amplitude and phase of their excitation. As radars progressed to shorter wavelengths, arrays were displaced by simpler antennas such as parabolic reflectors. For modern radar applications the advent of electronically controlled phase shifters and switches has once more directed attention to array antennas. The aperture excitation may now be modulated by controlling the phase of the individual elements to give beams that are scanned electronically. This chapter will be devoted to arrays of this type.

The capability of rapidly and accurately switching beams permits multiple radar functions to be performed, interlaced in time or even simultaneously. An electronically steered array radar may track a great multiplicity of targets, illuminate a number of targets with RF energy and guide missiles toward them, perform complete hemispherical search with automatic target selection, and hand over to tracking. It may even act as a communication system, directing high-gain beams toward distant receivers and transmitters. Complete flexibility is possible; search and track rates may be adjusted to best meet particular situations, all within the limitations set by the total use of time. The antenna beamwidth may be changed to search some areas more rapidly with less gain. Frequency agility is possible with the frequency of transmission changing at will from pulse to pulse or, with coding, within a pulse. Very high powers may be generated from a multiplicity of amplifiers distributed across the aperture. Electronically controlled array antennas can give radars the flexibility needed to perform all the various functions in a way best suited for the specific task at hand. The functions may be programmed adaptively to the limit of one's capability to exercise effective automatic management and control.

Phased array theory was studied intensively in the 1960s, bringing understanding. Technology advanced and led to a series of operational systems in the 1980s; many publications became available.[4-15] In terms of performance improvement, ultralow sidelobes (less than −40 dB) were demonstrated first in the 1970s by Westinghouse Electric Corporation's AWACS (Airborne Warning and Control System) and brought about tight tolerances in construction and phase settings. The advent of more and better computer modeling and sophisticated test equipment such as network analyzers has led to improved methods of designing well-matched apertures. Better components such as radiating elements, phase shifters, and power dividers are now available. More economical solid-state devices and memory chips have led to precision aperture phase control with corrections for frequency and temperature variations. Solid-state microwave devices hold great promise for future systems where a solid-state module is associated with each radiating element; improvements in terms of aperture control, reliability, and efficiency continue. Phased arrays can be controlled adaptively, particularly for sidelobe cancellation. This is an area where theory and understanding have advanced much. Also great progress has been made with indoor near-field antenna ranges,[16] where computer-controlled precision two-dimensional radiation patterns are derived at multiple frequencies, and with scanning.

Phased arrays are very expensive. As technology advances, costs are reduced, particularly in the areas of phase shifters and drivers. At the same time, the quest for better performance with lower sidelobes and wider bandwidth keeps the costs high. The greatest potential for cost reductions is believed to lie in the application of solid-state systems with a transmit/receive module at each element.

Phased Array Antennas.   The phased array antenna has an aperture that is assembled from a great many similar radiating elements, such as slots or dipoles, each element being individually controlled in phase and amplitude. Accurately predictable radiation patterns and beam-pointing directions can be achieved.

The general planar array characteristics are readily obtained from a few simple equations, given here but discussed later in greater detail. With the elements spaced by $\lambda/2$ (λ = wavelength) to avoid the generation of multiple beams (grating lobes), the number of radiating elements N for a pencil beam is related to the beamwidth by

$$N \approx \frac{10,000}{(\theta_B)^2}$$

$$\theta_B \approx \frac{100}{\sqrt{N}}$$

where θ_B is the 3 dB beamwidth in degrees. The corresponding antenna gain, when the beam points broadside to the aperture, is

$$G_0 \approx \pi N \eta \approx \pi N \eta_L \eta_a$$

where η accounts for antenna losses (η_L) and reduction in gain due to unequal weighting of the elements with a nonuniform amplitude distribution (η_a). When scanning to an angle θ_0, the gain of a planar array is reduced to that of the projected aperture:

$$G(\theta_0) \approx \pi N \eta \cos \theta_0$$

Similarly, the scanned beamwidth is increased from the broadside beamwidth (except in the vicinity of endfire, $\theta_0 = 90°$):

$$\theta_B \text{ (scanned)} \approx \frac{\theta_B(\text{broadside})}{\cos \theta_0}$$

The total number of beams M (with broadside beamwidth and square stacking) that fit into a sphere is approximately equal to the gain and with $\eta \approx 1$ is thus simply related to N by

$$M \approx \pi N$$

In a planar array where the beamwidth changes with the scan angle, the number of beams that can actually be generated and fitted into a sphere is

$$M' \approx \frac{\pi}{2} N$$

An array where the elements are fed in parallel (Sec. 7.8) and which is scanned by phase shift, modulo 2π, has limited bandwidth since for wideband operation constant path lengths rather than constant phases are required. The limit is given by

$$\text{Bandwidth } (\%) \approx \text{beamwidth (deg)}$$

This is equivalent to limitations given by

$$\text{Pulse length} = 2 \times \text{aperture size}$$

With these criteria, the scanned radiation pattern at 60° is steered by ± one-fourth of the local scanned beamwidth as the frequency is changed over the band. If all the frequencies in the band are used with equal weighting, then twice the bandwidth (half the pulse length) becomes acceptable. At a scan angle θ_0 the beam steers with frequency through an angle $\Delta\theta$ so that

$$\delta\theta \approx \frac{\delta f}{f} \tan \theta_0 \qquad \text{rad}$$

For wider bandwidths, time-delay networks have to be introduced to supplement the phase shifters.

Conformal Arrays.[17,18]   Phased arrays may conform to curved surfaces as required, for example, for flush-mounting on aircraft or missiles. If the surface has a large radius of curvature so that all the radiating elements point to substantially the same direction, then the characteristics are similar to those of a planar array even though the exact 3D position of the element has to be taken into account to calculate the required phase. A small radius of curvature is found with cylindrical (or spherical) arrays used for 360° coverage. Elements are switched to avoid sections of the antenna where they point away from the desired beam direction. Difficulties may be encountered in matching the radiating elements and in maintaining polarization purity. This geometry has not yet found use in radar systems.

The discussions in this chapter will concentrate on planar phased arrays.

3D Volumetric Search.   3D volumetric radar search is possible with electronic scanning in both azimuth and elevation; important regions (e.g., the horizon) may be emphasized at will and searched more frequently. The radar may operate with a higher than normal false-alarm rate since targets can easily be confirmed by repeated interrogation. Phase control allows beams to be widened, for example, to reduce search time for the more elevated regions, where reduced ranges need less antenna gain. A separate rotating surveillance radar system may be added for extra coverage (at a second frequency) and to allow more emphasis on tracking.

Monopulse Track.   Phased array radars are well suited for monopulse tracking. The radiating elements of the array can be combined in three different ways to give the sum pattern and the azimuth and elevation difference patterns. Contradictory requirements in optimum amplitude distribution for sum and difference patterns exist,[19] but, as with other antenna systems, they may be independently satisfied. The sum and difference patterns are scanned simultaneously.

The difference-pattern null in a phased array system gives good beam-pointing accuracy. Absolute beam-pointing accuracies to within less than one-fiftieth of a (scanned) beamwidth have been measured with scans up to 60°.[20] The accuracy is limited by phase and amplitude errors. Since phase shift rather than time delay is used, as the frequency is changed, the direction of the null of the scanned beam is also changed, and the beam moves toward broadside with an increase in frequency. The amplitude at boresight of the difference-pattern output then increases linearly with a change in frequency. With a scan angle of 60° this change is from a null at the design frequency to a value of about −9 dB relative to the sum pattern, at the edge of the band, where the band is defined by bandwidth (percent) = beamwidth (deg). This is discussed more fully in Sec. 7.7.

Shaped Beams.   The radiation pattern of an array may be shaped by modifying the aperture distribution. Good pattern approximations can be obtained by using phase only. In particular, the beam may be broadened by applying a spherical phase distribution to the aperture or by approximating it with a gable (triangular) phase distribution. Beams of this type are of particular interest since they are easily generated. They may be used for transmission in a system where the receiving antenna has a cluster of simultaneous beams, or, as previously discussed, they may be used in a search system to reduce the number of angular cells in regions of shorter range.

Monitoring.   Electronically scanned arrays are composed of very many parts and include electronic circuitry to drive the phase shifters or switches that steer the beam. The overall reliability of such arrays can be great; graceful degradation has been claimed, since the failure of as much as 10 percent of the components leads to a loss in gain of only 1 dB. There is, however, a degradation of (low) sidelobes. Nevertheless, the functioning of the antenna is complex, and there is need for providing test or monitoring circuitry. The decision to point a beam in a certain direction is made somewhere in the radar control system and is normally defined by two direction cosines. A test or monitoring circuit should establish the correct functioning of all components, including all beam-pointing computations, electronic drivers and phase shifters or switches, and all their interconnections. Frequent indications that the antenna system is functioning or is capable of functioning should be available. In one possible method the phase shifters are programmed to focus on a nearby monitor probe and scan past it.[21] This will yield a close approximation of the complete radiation pattern, where gain and sidelobes can be measured and compared with previous results. The contribution of indi-

vidual elements and their phase shifters (and drivers) can also be checked with this configuration. The phase at each element is sequentially rotated at some low frequency; the amplitude and phase of this modulation as received by the probe relate directly to both the relative amplitude excitation of the element and its relative phase setting.[22] Other methods have been proposed[23] where measurements are compared with previously recorded ones.

Deployment of Apertures. With planar arrays, scanning is limited by the loss in gain and the increase in beamwidth corresponding to the reduction of the aperture to its projected area. Practical extreme values of scanning are therefore in the region of 60 to 70°. A minimum of three planar array apertures is then necessary for hemispherical coverage. For shipborne use, a minimum of four apertures appears desirable since, with pitch and roll, more than hemispherical coverage is necessary. The antennas may be positioned as shown in Fig. 7.1, permitting a view that is unimpeded by the central superstructure. The apertures would normally be tilted back from the vertical to balance the scan angles.

Radiating Elements. The most commonly used radiators for phased arrays are dipoles, slots, open-ended waveguides (or small horns), and printed-circuit "patches" (originally called *Collings radiator* after their inventor[24]). The element has to be small enough to fit in the array geometry, thereby limiting the element to an area of a little more than $\lambda^2/4$. In addition, many radiators are required, and the radiating element should be inexpensive and reliable and have identical characteristics from unit to unit.

Since the impedance and pattern of a radiator in an array are determined predominantly by the array geometry (Sec. 7.4), the radiating element may be chosen to suit the feed system and the physical requirements of the antenna. For example, if the radiator is fed from a stripline phase shifter, a stripline dipole

FIG. 7.1 Guided missile cruiser showing two out of four phased array antennas. (*Courtesy of Ingalls Shipbuilding Division of Litton.*)

would be a logical choice. If a waveguide phase shifter is used, an open-ended waveguide or a slot might be convenient. At the lower frequencies, where coaxial components are prevalent, dipoles have been favored. A ground plane is usually placed about $\lambda/4$ behind an array of parallel dipoles so that the antenna forms a beam in only one hemisphere. At the higher frequencies open-ended waveguides and slots are frequently used. Considerable bandwidth (perhaps 50 percent) can be obtained, even with patch radiators, provided they are fed similarly to a dipole.[25]

For limited scanning (say, less than 10°), it is possible to use directive radiators having dimensions of height and width of several wavelengths. With such separation, the mutual coupling effects can be small, and the pattern and impedance of an element in the array approach those of the isolated element.

The element must be chosen to give the desired polarization, usually vertical or horizontal. The special case of circular polarization is discussed below.

If polarization diversity is required or if an array is required to transmit one polarization and receive the orthogonal or both polarizations, either crossed dipoles or circular or square radiators seem suitable. With appropriate feed systems, both are capable of providing vertical and horizontal polarization independently and may be combined to provide any desired polarization, including circular. Such polarization diversity adds considerable complexity, requiring two feed systems or switches at the radiating element level.

Circular Polarization. From the point of view of the antenna designer, circular polarization is possible, though difficulties may be encountered in matching for large scan angles. On scanning, a component of the undesired orthogonal polarization will be generated,[26] and some provision should be made to absorb that energy.[27] With a conventional circularly polarized antenna, such as a parabolic dish with a circularly polarized feed, good circularity may be obtained over part of the main beam, with rapid deterioration over the rest of the pattern. With a planar array the relevant beamwidth is the beamwidth of the element in the array rather than the array beamwidth. The element beamwidth is broad, and good circularity may be expected over wide angles, including the main beam and sidelobes.

With circular polarization, the signal returned from a single-bounce target will require an antenna matched to the opposite sense of circular polarization from that transmitted. If the same antenna is used, then single-bounce targets are rejected. Such a system can therefore give a measure of suppression of rain echoes,[28] ideally amounting to

$$20 \log (e^2 + 1)/(e^2 - 1) \quad \text{dB}$$

where e is the voltage-ellipticity ratio. An early model of a Raytheon reflectarray gave an ellipticity ratio of less than 1.5 dB with scans up to 30°, corresponding to a theoretical rain rejection of at least 15 dB. At the same time, an aircraft target would typically lose approximately 3 dB, leaving a relative net improvement of 12 dB of rain rejection.

Phased Arrays with Very Wide Bandwidth. A radar system that has the capability of changing frequency over a very wide band can, with advantage, adapt its transmission to take into account frequency-dependent multipath characteristics, target response, environmental conditions, interference, and jamming. Further, wideband processing can give fine range resolution.

Phased arrays have the potential of operating over very wide bandwidths. Some ferrite phase shifters operate over two octaves,[29] and digital diode phase shifters that switch line lengths may function over even wider bands. The high

end of the frequency band is limited by the physical size of the elements, which must be spaced close enough in the array to avoid the generation of grating lobes. For wide instantaneous bandwidth (rather than tunable bandwidth), time delays have to be added to prevent the beam from being scanned as the frequency is changed.

The impedance of the radiating element at the aperture (with closely spaced elements) is approximately independent of frequency, but the element must be matched over the wide band. This is difficult to achieve without exciting harmful surface waves when scanning. Nevertheless, matching with octave bandwidth for scanning to $\pm 60°$ appears possible.

Limited Scanning.[30] If scanning is limited to a small angular volume, considerable simplifications become possible. The total number of active phase-shifter controls can be reduced to about equal the total number of beams. Subarrays may be formed, each with only one phase control and of a size such that its beamwidth includes all the scan angles. Alternatively, a small phased array could be placed in the focal region of a large reflector to scan the narrow beamwidth of the reflector over a limited scan angle.

Scanning of Arrays

Phase Scanning.   The beam of an antenna points in a direction that is normal to the phase front. In phased arrays, this phase front is adjusted to steer the beam by individual control of the phase of excitation of each radiating element. This is indicated in Fig. 7.2a. The phase shifters are electronically actuated to permit rapid scanning and are adjusted in phase to a value between 0 and 2π rad. With an interelement spacing s, the incremental phase shift ψ between adjacent elements for a scan angle θ_0 is $\psi = (2\pi/\lambda)s \sin \theta_0$. If the phase ψ is constant with frequency, the scan angle θ_0 is frequency-dependent.

Time-Delay Scanning.   Phase scanning was seen to be frequency-sensitive. Time-delay scanning is independent of frequency. Delay lines are used instead of
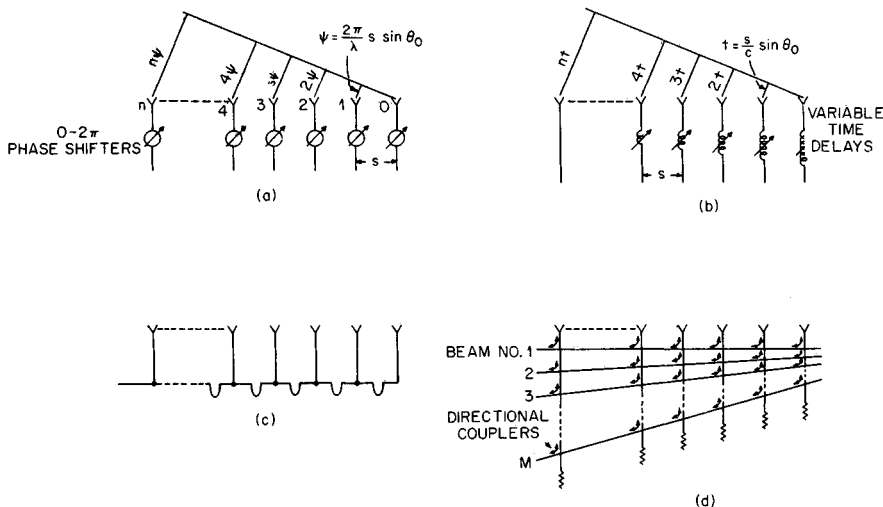

FIG. 7.2   Generation of scanned beams. (*a*) Phased array. (*b*) Time-delay array. (*c*) Frequency-scanned array. (*d*) Blass-type array.

phase shifters, as shown in Fig. 7.2b, providing an incremental delay from element to element of $t = (s/c) \sin \theta_0$, where c = velocity of propagation. Individual time-delay circuits (Sec. 7.7) are normally too cumbersome to be added to each radiating element. A reasonable compromise may be reached by adding one time-delay network to a group of elements (subarray) where each element has its own phase shifter.

Frequency Scanning.[31]   Frequency rather than phase may be used as the active parameter to exploit the frequency-sensitive characteristics of phase scanning. Figure 7.2c shows the arrangement. At one particular frequency all radiators are in phase. As the frequency is changed, the phase across the aperture tilts linearly, and the beam is scanned. Frequency-scanning systems are relatively simple and inexpensive to implement. They have been developed and deployed in the past to provide elevation-angle scanning in combination with mechanical horizontal rotation for 3D radars. A chapter in the first edition of this handbook was devoted to this approach, which since then has received much less attention; frequency is usually considered too important a parameter to give up for scanning.

IF Scanning.   For receiving, the output from each radiating element may be heterodyned (mixed) to an intermediate frequency (IF). All the various methods of scanning are then possible, including the beam-switching system described below, and can be carried out at IF, where amplification is readily available and lumped constant circuits may be used.

Digital Beamforming.[32-34]   For receiving, the output from each radiating element may be amplified and digitized. The signal is then transferred to a computer for processing, which can include the formation of multiple simultaneous beams (formed with appropriate aperture illumination weighting) and adaptively derived nulls in the beam patterns to avoid spatial interference or jamming. Limitations are due to the availability and cost of analog-to-digital (A/D) converters and to their frequency and dynamic-range characteristics. Partial implementation is possible by digitizing at subarray levels only.

Beam Switching.   With properly designed lenses or reflectors, a number of independent beams may be formed by feeds at the focal surface. Each beam has substantially the gain and beamwidth of the whole antenna. Allen[35] has shown that there are efficient equivalent transmission networks that use directional couplers and have the same collimating property. A typical form, after Blass,[36] is shown in Fig. 7.2d. The geometry can be adjusted to provide equal path lengths, thus providing frequency-independent time-delay scanning. Another possible configuration providing multiple broadband beams uses parallel plates containing a wide-angle microwave lens[37,38] (Gent, Rotman). Each port corresponds to a separate beam. The lens provides appropriate time delays to the aperture, giving frequency-invariant scanning. The beams may be selected through a switching matrix requiring $M-1$ single-pole–double-throw (SPDT) switches to select one out of M beams. The beams are stationary in space and overlap at about the 4 dB points. This is in contrast to the previously discussed methods of scanning, where the beam can be steered accurately to any position. The beams all lie in one plane. The system may be combined with mechanical rotation of the antenna, giving vertical switched scanning for 3D coverage. Much greater complexity is required for a system switching beams in both planes.

Multiple Simultaneous Beams.   Instead of switching the beams, as described in the preceding paragraph, all the beams may be connected to separate receivers, giving multiple simultaneous receive beams. The transmitter radiation pattern would need to be wide to cover all the receive beams. Such multibeam sys-

tems have found application in combination with mechanical rotation for 3D coverage.

   Multiple Independently Steered Beams.   Independent multiple beams may be generated with a single beamformer by modifying both amplitude and phase at the aperture. This can be seen from Fig. 7.3, where, for example, two independent beams are generated. Both beams have the same amplitude (voltage) distribution $F(x)$ but differently inclined linear phased fronts. The total aperture excitation with both beams is

$$F(x,\psi) = F(x)e^{j2\psi_1(x/a)} + F(x)e^{j2\psi_2(x/a)} = 2F(x)\left[\cos(\psi_1 - \psi_2)\frac{x}{a}\right]e^{j(\psi_1 + \psi_2)(x/a)}$$

That is, the aperture amplitude distribution required for two separate beams varies cosinusoidally, and the phase distribution is linear and has the average inclination.
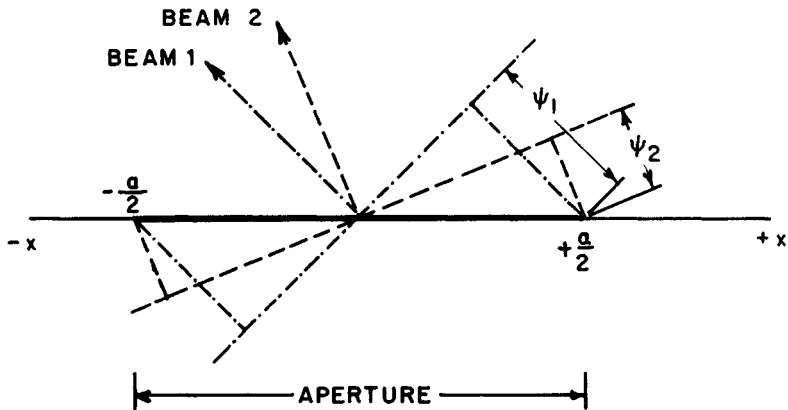
FIG. 7.3   Aperture distribution giving two beams.

   In most phased array systems only the phase can be controlled. Ignoring the required amplitude variations still leads to good approximations for forming multiple beams, by superimposing the various required phase-shifter settings (modulo 2π). In the case of two beams, the aperture phase slope has the average inclination and varies periodically from 0 to π.

   It should be noted that when a multilobed radiation pattern is received or transmitted in one channel, the gain is shared between the lobes. When the beams are contained in separate (beamforming) channels, however, each channel has the full gain of the aperture.

   Vertical Scan Only.   A greatly simplified phased array system becomes possible if there is no need for multifunction capabilities, including fire control, where a beam may have to be pointed in any given direction at any time. The array is scanned in the vertical plane only and mechanically rotated to give azimuth coverage. The number of phase control points is then reduced to the number of horizontal rows. In the case of a ship's surveillance radar, the antenna should be positioned as high as possible to avoid shadowing by the superstruc-

ture, but the pedestal need not be stabilized since stabilization can be achieved by electronic beam steering. Scanning can be in the form of phase scanning or beam switching, or multiple simultaneous beams may be used on receive with a wide antenna pattern on transmit. Many systems of this type have been developed for both naval and land-based use (Sec. 7.11).

7.2 ARRAY THEORY

Array with Two Elements.   Figure 7.4 shows two isotropic radiators which are spaced by a distance s and excited with equal amplitude and phase. With unity input power, the vector sum of their contributions, added at a great distance as a function of θ, is the radiation pattern

$$E_a(\theta) = \frac{1}{\sqrt{2}} [e^{j(2\pi/\lambda)(s/2) \sin \theta} + e^{-j(2\pi/\lambda)(s/2) \sin \theta}]$$

where θ is measured from the broadside direction. Normalizing, to give unity amplitude when $\theta = 0$, and simplifying give

$$E_a(\theta) = \cos \left[\pi \frac{s}{\lambda} \sin \theta \right] \tag{7.1}$$

The absolute value of $E_a(\theta)$ is plotted in Fig. 7.4 as a function of $\pi(s/\lambda) \sin \theta$. If the plot had been in terms of the angle θ, the lobes would have been found to increase in width as $|\theta|$ increased. The main lobe occurs when $\sin \theta = 0$. The other lobes have the same amplitude as the main lobe and are referred to as *grating lobes*. They occur at angles given by $\sin \theta = \pm [m/(s/\lambda)]$, where m is an integer. For the half space given by $-90° < \theta < +90°$, there are $2m'$ grating lobes, where m' is the largest integer smaller than s/λ. If $s < l$, grating-lobe maxima do not occur, and the value at $\pm 90°$ is $\cos (\pi s/\lambda)$. This value is for isotropic radiators and is reduced if the radiators have directivity.

Linear Array.[39]   With a linear array of N isotropic radiators, excited with equal amplitudes and phase and separated by distances s, as shown in Fig. 7.5, the condition for the occurrence of grating lobes is unchanged from the simpler case just considered. They occur for the same values of $\pi(s/\lambda) \sin \theta$, but the width of the lobes is reduced, and they are separated by minor lobes. Summing the vector contributions from all elements, with element 0 as phase reference, gives

$$E_a(\theta) = \frac{1}{\sqrt{N}} \sum_{n=0}^{n=N-1} e^{j(2\pi/\lambda) ns \sin \theta}$$

The factor $1/\sqrt{N}$ shows that each element is energized with $1/N$ of the (unity) input power. Normalizing the gain to unity at broadside, $\theta = 0$, gives the pattern

$$E_a(\theta) = \frac{\sin [N\pi(s/\lambda) \sin \theta]}{N \sin [\pi(s/\lambda) \sin \theta]} \tag{7.2}$$

$E_a(\theta)$ gives the radiation pattern for isotropic radiators and is known as the *array factor*. It is shown in Fig. 7.6 for $N = 10$. The pattern is repetitive, and the lo-
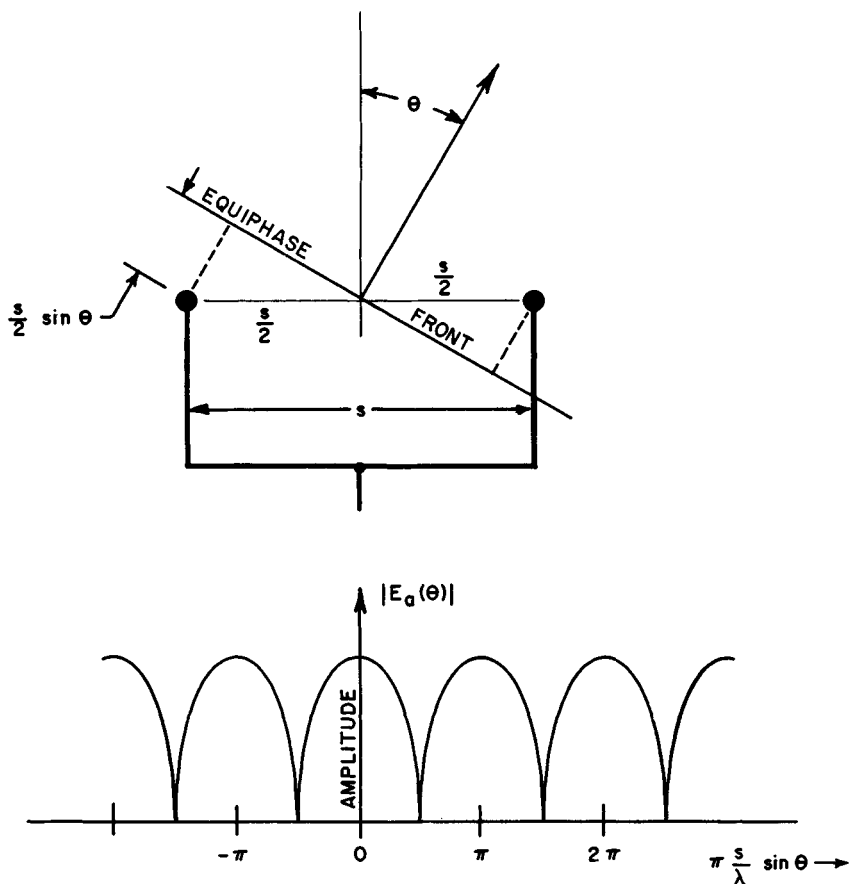
FIG. 7.4   Radiation pattern of two isotropic radiators.

cations of the adjacent grating lobes at angles θ_1 and θ_2 are separated by $\pi(s/\lambda)$ $(\sin \theta_1 - \sin \theta_2) = \pi$.

The radiating elements are not isotropic but have a radiation pattern $E_e(\theta)$, known as the *element factor* or *element pattern*; then the complete radiation pattern $E(\theta)$ is the product of the array factor and the element pattern:

$$E(\theta) = E_e(\theta)E_a(\theta) = E_e(\theta)\frac{\sin \left[N\pi(s/\lambda) \sin \theta\right]}{N \sin \left[\pi(s/\lambda) \sin \theta\right]} \tag{7.3}$$

An approximation to the pattern of Eq. (7.2) is in the form

$$E(\theta) = \frac{\sin \left[\pi(a/\lambda) \sin \theta\right]}{\pi(a/\lambda) \sin \theta} \tag{7.4}$$
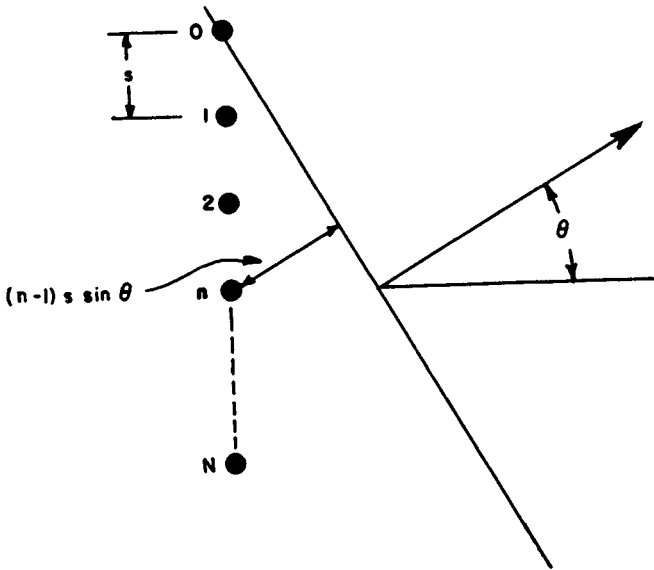
FIG. 7.5 Linear array with N radiators uniformly spaced by a distance s.
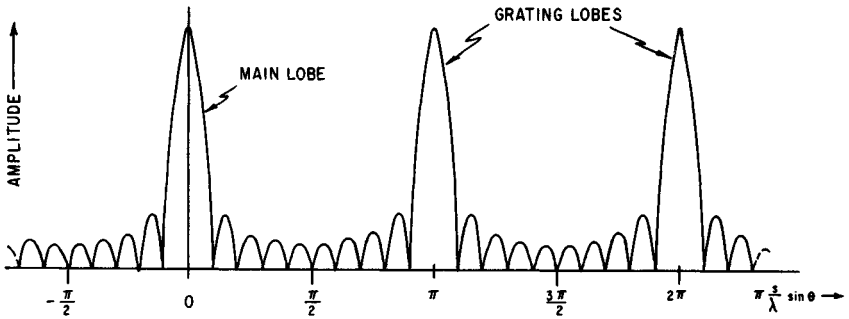
FIG. 7.6 Array factor with $N = 10$ elements.

where the effective aperture is $a = Ns$, which extends by $s/2$ beyond the centers of the end elements. In contrast to the array factor, this pattern has only one maximum and is nonrepetitive. It is the well-known Fourier transform of a continuous constant-amplitude distribution and is a reasonable approximation for small values of θ when the aperture is greater than several wavelengths. The *half-power beamwidth* is obtained from Eq. (7.4):

$$\theta_B = \frac{0.886}{a/\lambda} \text{ (rad)} = \frac{50.8}{a/\lambda} \text{ (deg)} \tag{7.5}$$

The first sidelobe is 13.3 dB down from the peak of the main beam.

For larger values of θ the pattern of a continuous aperture is modified from Eq. (7.4) by the obliquity factor[40,41] $\frac{1}{2}(1 + \cos \theta)$, which arises from the definition of a Huygens source. This gives

$$E(\theta) = \frac{1}{2}(1 + \cos \theta)\frac{\sin [\pi(a/\lambda) \sin \theta]}{\pi(a/\lambda) \sin \theta} \tag{7.6}$$

For closely spaced elements the obliquity factor is very similar to the amplitude pattern of a well-designed (matched) radiating element, $\sqrt{\cos \theta}$ for values up to some 60 or 70°. At greater angles the element pattern has values that are greater than those given by $\sqrt{\cos \theta}$ and that are a function of the total number of elements.[42]

Scanned Linear Arrays. The pattern of the array may be steered to an angle θ_0 by applying linearly progressive phase increments from element to element so that the phase between adjacent elements differs by $2\pi(s/\lambda) \sin \theta_0$. Equation (7.2) is then modified, giving the normalized array factor of a uniformly illuminated array as

$$E_a(\theta) = \frac{\sin N\pi(s/\lambda)(\sin \theta - \sin \theta_0)}{N \sin \pi(s/\lambda)(\sin \theta - \sin \theta_0)} \tag{7.7}$$

and the pattern is

$$E(\theta) = E_e(\theta) \frac{\sin N\pi(s/\lambda)(\sin \theta - \sin \theta_0)}{N \sin [\pi(s/\lambda \sin \theta]} \tag{7.8}$$

Equation (7.8) describes the fundamental response of a scanned array system. The array factor will have only one single major lobe, and grating-lobe maxima will not occur for $-90° < \theta < +90°$ as long as

$$\pi \frac{s}{\lambda} |\sin \theta - \sin \theta_0| \setminus \pi$$

or

$$\frac{s}{\lambda} < \frac{1}{1 + |\sin \theta_0|} \tag{7.9}$$

which is always true if $s/\lambda < \frac{1}{2}$. When scanning is limited, the value of s/λ may be increased, for example, to $s/\lambda < 0.53$ for scanning to a maximum of 60° or to $s/\lambda < 0.59$ for scanning to a maximum of ±45°.

For larger values of s/λ, grating lobes occur at angles θ_1, given by

$$\sin \theta_1 = \sin \theta_0 \pm \frac{n}{s/\lambda} \tag{7.10}$$

where n is an integer.

In the limit, the inequality (7.9) does allow a grating-lobe peak to occur at 90° when scanning to θ_0. Even though the grating lobe is reduced when multiplied by

the element pattern, it may be prudent to space the elements such that the first null of the grating lobe, rather than the peak, occurs at 90°. With N elements this more restrictive condition is given by

$$\frac{s}{\lambda} \frac{N-1}{N} \times \frac{1}{1 + |\sin \theta_0|} \tag{7.11}$$

Equation (7.8) may again be approximated by the Fourier transform of the illumination across the continuous aperture:

$$E(\theta) = \frac{1}{2}(1 + \cos \theta)\frac{\sin \pi(a/\lambda)(\sin \theta - \sin \theta_0)}{\pi(a/\lambda)(\sin \theta - \sin \theta_0)} \tag{7.12}$$

The Fourier-transform solutions for continuous apertures[19,43] may be used to approximate patterns for practical amplitude and phase distributions as long as the element-to-element spacing is small enough to suppress grating lobes.[44] *Monopulse difference patterns* may be approximated in the same way from the Fourier transforms of the corresponding continuous odd aperture distributions. For example, with a constant amplitude distribution, the difference-pattern array factor calculated by the exact vector addition of all radiating elements is

$$E_a(\theta) = \frac{1 - \cos N\pi(s/\lambda)(\sin \theta - \sin \theta_0)}{N \sin \pi(s/\lambda)(\sin \theta - \sin \theta_0)}$$

The Fourier transform gives the same expression with the *sine* in the denominator replaced by its argument, giving (in the denominator) $\pi(a/\lambda)(\sin \theta - \sin \theta_0)$, where $a = Ns$.

For small scan angles θ_0 and small values of θ the expression $\sin \theta - \sin \theta_0$ may be approximated by $\theta - \theta_0$. For larger values of θ_0, the expression $\sin \theta - \sin \theta_0$ may be expanded to give the response in the general direction of the (narrow) scanned beam in terms of the *small angle* $(\theta - \theta_0)$:

$$\sin \theta - \sin \theta_0 \approx a\theta - \theta_0) \cos \theta_0 \tag{7.13}$$

This gives, with Eq. (7.12),

$$E(\theta) \not\approx \frac{1}{2}(1 + \cos \theta)\frac{\sin [(\pi a \cos \theta_0)/\lambda](\theta - \theta_0)}{[(\pi a \cos \theta_0)/\lambda](\theta - \theta_0)} \tag{7.14}$$

Equation (7.14) measures the angle $\theta - \theta_0$ from the scanned direction. It shows that the effect of scanning is to reduce the aperture to the size of its projected area in the direction of scan. Correspondingly, the beamwidth is increased to

$$\theta_B \text{ (scanned)} \not\approx \frac{\theta_B \text{ (broadside)}}{\cos \theta_0} = \frac{0.886}{(a/\lambda) \cos \theta_0} \text{ (rad)} = \frac{50.8}{(a/\lambda) \cos \theta_0} \text{ (deg)} \tag{7.15}$$

When the beam is scanned from broadside by an angle $\theta_0 < 60°$ and the aperture

$a/\lambda \hbar 5$, Eq. (7.15) gives a beamwidth that is too narrow, the error being less than 7 percent.

When the beam is scanned to very large scan angles, toward endfire, more exact calculations become necessary.[42,45] Equation (7.8) still applies and gives, for endfire with isotropic radiators,

$$\theta_B \text{ (endfire)} = 2\sqrt{\frac{0.886}{a/\lambda}} \text{ rad} \qquad (7.16)$$

Element Factor and Gain of Planar Arrays.   The gain of a uniformly illuminated and lossless aperture of area A, with a broadside beam, is $G_{(o)} = 4\pi A/\lambda^2$. With a nonuniform aperture distribution and with losses present, the gain is reduced by the efficiency term η to

$$G_{(0)} = 4\pi \frac{A}{\lambda^2} \eta \qquad (7.17)$$

If the aperture is considered as a matched receiver, then the amount of energy arriving from a direction θ_0 is proportional to its projected area. The gain with scanning therefore is

$$G(\theta_0) = 4\pi \frac{A \cos \theta_0}{\lambda^2} \eta \qquad (7.18)$$

The variation of gain with the cosine of the scan angle agrees with the equivalent variation in beamwidth given by Eq. (7.15). The gain may be expressed in terms of the actual beamwidth, giving, from Eqs. (7.15) and (7.18),

$$G(\theta_0) \not\approx \frac{32,000}{\theta_B \, \phi_B} \eta \qquad (7.19)$$

where θ_B and ϕ_B are the beamwidths in degrees in the two principal planes with the beam scanned to θ_0.

If the aperture is made up of N equal radiating elements and is matched to accept the incident power, then the contribution to the overall gain is the same from all elements, whence

$$G(\theta) = NG_e(\theta)\eta \qquad (7.20)$$

where G_e is the gain per element. It follows from Eq. (7.18) that the matched-element power pattern is

$$G_e(\theta) = 4\pi \frac{A}{N\lambda^2} \cos \theta \qquad (7.21)$$

and the normalized radiation amplitude pattern of the (matched) element or (matched) *element pattern* is

$$E_e(\theta) = \sqrt{\cos \theta} \qquad (7.22)$$

It has already been noted that the matched-element pattern is very similar to the obliquity factor $\frac{1}{2}(1 + \cos \theta)$ and differs markedly only near endfire, where the number of elements begins to matter.[42]

For a given element spacing s the total number of radiators N in the area A is $N = A/s^2$, and Eq. (7.21) gives

$$G_e(\theta) = 4\pi \left[\frac{s}{\lambda}\right]^2 \cos \theta$$

When the element spacing is $s = \lambda/2$, then the power pattern of an element that is perfectly matched at all scan angles is

$$G_e(\theta) = \pi \cos \theta \tag{7.23}$$

and the peak antenna gain in the direction of scan, θ_0, is

$$G(\theta_0) = \pi N \eta \cos \theta_0 \tag{7.24}$$

where the efficiency term η accounts for losses and for a nonuniform aperture distribution. For a broadside beam $\theta_0 = 0$ and

$$G_0 = \pi N \eta \tag{7.25}$$

and the element gain is $G_e = \pi$.

The effects of the element pattern are most marked with wide beams. Figure 7.7 shows the array and element factors and the resulting pattern for a 10-element array, with element spacing $s = \lambda/2$, scanned to 60°. The pattern maximum is noted to occur at less than 60° because the gain of the element pattern increases toward broadside. The pattern value at 60° is $\cos 60° = 0.5$ in power or 0.707 in amplitude, relative to the maximum at broadside, as expected. The sidelobes in the general region of broadside are not reduced since in that region the element
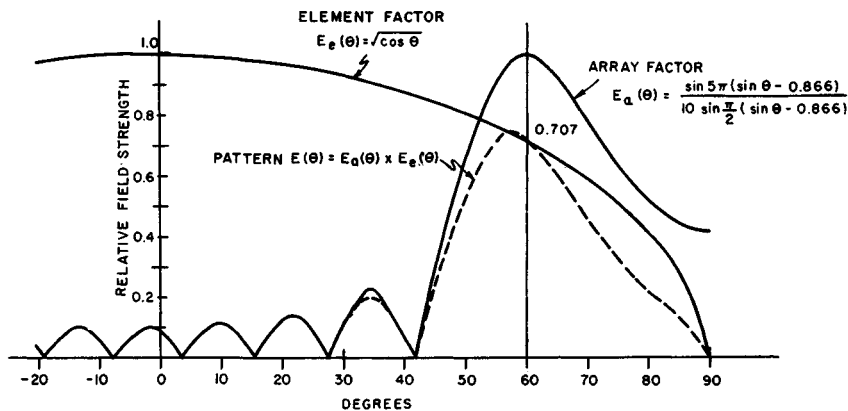
FIG. 7.7 Ten-element linear array scanned to 60°. Element spacing $s = \lambda/2$.

pattern is approximately unity. Relative to the beam maximum, therefore, the sidelobes near broadside are increased by approximately 3 dB.

7.3 PLANAR ARRAYS AND BEAM STEERING

Planar Arrays. A planar array is capable of steering the beam in two dimensions. In a spherical-coordinate system the two coordinates θ and ϕ define points on the surface of a unit hemisphere. As shown in Fig. 7.8, θ is the angle of scan measured from broadside and ϕ is the plane of scan measured from the x axis. Von Aulock[46] has presented a simplified method for visualizing the patterns and the effect of scanning. He considers the projection of the points on a hemisphere onto a plane (Fig. 7.9); the axes of the plane are the direction cosines $\cos \alpha_x$, $\cos \alpha_y$. For any direction on the hemisphere the direction cosines are

$$\cos \alpha_x = \sin \theta \cos \phi$$
$$\cos \alpha_y = \sin \theta \sin \phi$$

The direction of scan is indicated by the direction cosines $\cos \alpha_{xs}$, $\cos \alpha_{ys}$. Here the plane of scan is defined by the angle ϕ measured counterclockwise from the $\cos \alpha_x$ axis and is given by

PHASE AT mnth ELEMENT IS $mT_{xs} + nT_{ys}$

$$T_{xs} = \frac{2\pi d_x}{\lambda} \sin \theta \cos \phi$$

$$T_{ys} = \frac{2\pi d_y}{\lambda} \sin \theta \sin \phi$$
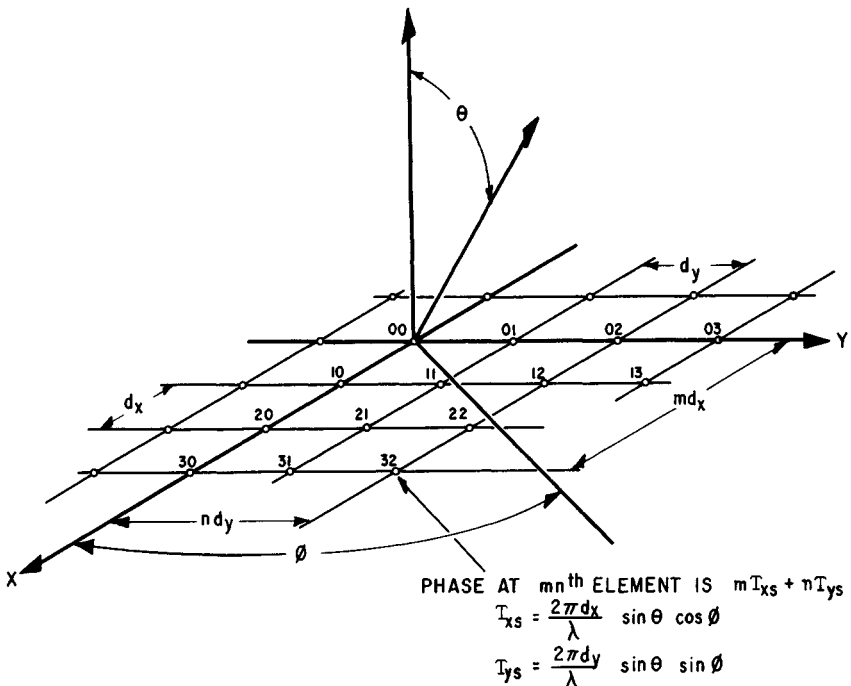
FIG. 7.8 Planar-array-element geometry and phasing.
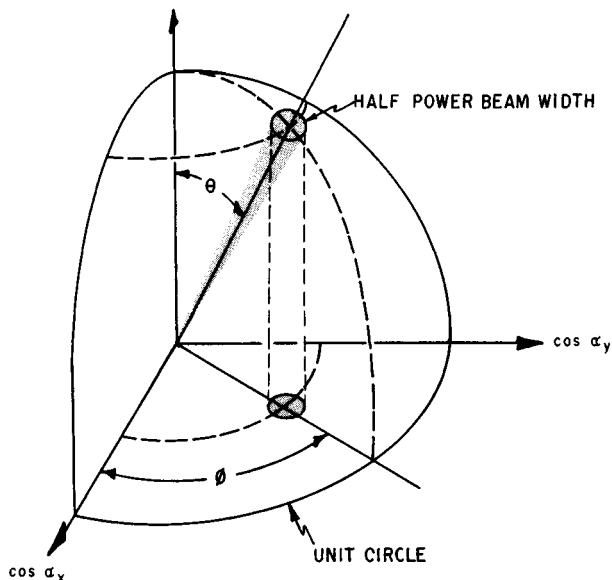
FIG. 7.9 Projection of points on a hemisphere onto the plane of the array.

$$\phi = \tan^{-1} \frac{\cos \alpha_{ys}}{\cos \alpha_{xs}}$$

The angle of scan θ is determined by the distance of the point ($\cos \alpha_{xs}$, $\cos \alpha_{ys}$) from the origin. This distance is equal to $\sin \theta$. For this reason a representation of this sort is called sin θ space. A feature of sin θ space is that the antenna pattern shape is invariant to the direction of scan. As the beam is scanned, every point on the plot is translated in the same direction and by the same distance as is the beam maximum.

The region inside the unit circle where

$$\cos^2 \alpha_x + \cos^2 \alpha_y \leq 1$$

is defined as *real space*, the hemisphere into which energy is radiated. The infinite region outside the unit circle is referred to as *imaginary space*. Although no power is radiated into imaginary space, the concept is useful for observing the motion of grating lobes as the array is scanned. In addition, the pattern in imaginary space represents stored energy and contributes to the element impedance in the array.

The most common element lattices have either a rectangular or a triangular grid. As shown in Fig. 7.8, the mnth element is located at (md_x, nd_y). The triangular grid may be thought of as a rectangular grid where every other element has been omitted. The element locations can be defined by requiring that $m + n$ be even.

Calculations for the element-steering phases are greatly simplified by the

adoption of the direction cosine coordinate system. In this system the linear-phase tapers defined by the beam-steering direction ($\cos \alpha_{xs}$, $\cos \alpha_{ys}$) may be summed at each element so that the phasing at the mnth element is given by

$$\psi_{mn} = mT_{xs} + nT_{ys}$$

where $T_{xs} = (2\pi/\lambda)d_x \cos \alpha_{xs}$ = element-to-element phase shift in the x direction
$T_{ys} = (2\pi/\lambda)d_y \cos \alpha_{ys}$ = element-to-element phase shift in the y direction

The array factor of a two-dimensional array may be calculated by summing the vector contribution of each element in the array at each point in space. For an array scanned to a direction given by the direction cosines $\cos \alpha_{xs}$ and $\cos \alpha_{ys}$, the array factor of an $M \times N$ rectangular array of radiators may be written

$$E_a(\cos \alpha_{xs}, \cos \alpha_{ys}) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |A_{mn}|e^{j[m(T_x - T_{xs}) + n(T_y - T_{ys})]}$$

where $T_x = (2\pi/\lambda) \, d_x \cos \alpha_x$
$T_y = (2\pi/\lambda) \, d_y \cos \alpha_y$
A_{mn} = amplitude of mnth element

An array may be visualized as having an infinite number of grating lobes only one of which (namely, the main beam) is desired in real space. It is convenient to plot the position of the grating lobes when the beam is phased for broadside and observe the motion of these lobes as the beam is scanned. Figure 7.10 shows the grating-lobe locations for both rectangular and triangular spacings. For a rectangular array the grating lobes are located at

$$\cos \alpha_{xs} - \cos \alpha_x = \pm \frac{\lambda}{d_x} p$$
$$\cos \alpha_{ys} - \cos \alpha_y = \pm \frac{\lambda}{d_y} q$$

$$p, q = 0, 1, 2, \ldots$$

The lobe at $p = q = 0$ is the main beam. A triangular grid is more efficient for suppressing grating lobes than a rectangular grid,[47] so that for a given aperture size fewer elements are required. If the triangular lattice contains elements at (md_x, nd_y), where $m + n$ is even, the grating lobes are located at

$$\cos \alpha_{xs} - \cos \alpha_x = \pm \frac{\lambda}{2d_x} p$$
$$\cos \alpha_{ys} - \cos \alpha_y = \pm \frac{\lambda}{2d_y} q$$

where $p + q$ is even.

Since only one main lobe is normally desired in real space, an appropriate design will place all but one maximum in imaginary space for all angles of scan. With scanning, lobes that were originally in imaginary space may move into real space if the element spacing is greater than $\lambda/2$. As the array is scanned away from broadside, each grating lobe (in $\sin \theta$ space) will move a distance equal to the sine of the angle of scan and in a direction determined by the plane of scan. To ensure that no grating lobes enter real space, the element spacing must be
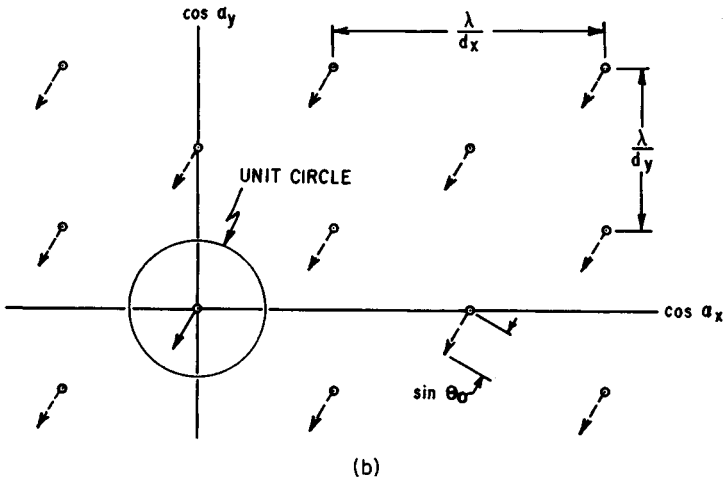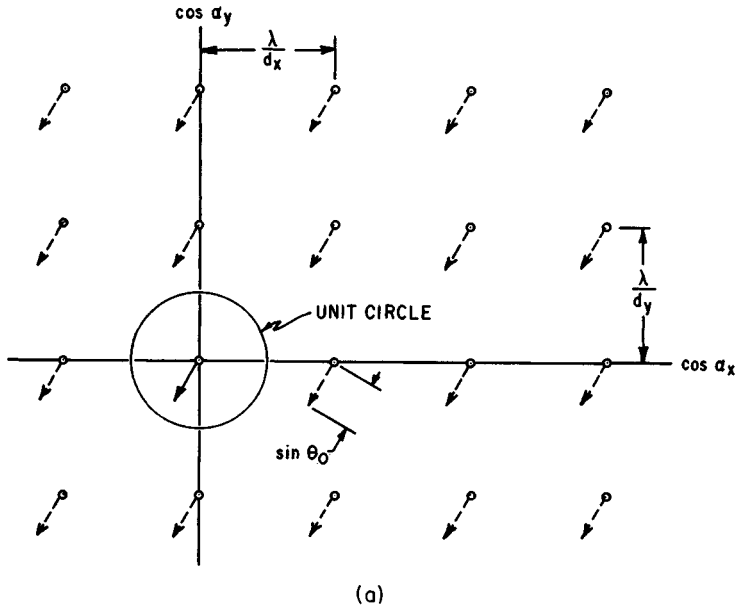
(a)

(b)

FIG. 7.10 Grating-lobe positions for (*a*) rectangular and (*b*) triangular grids, showing the motion of the lobes as the beam is scanned an angle θ_o.

chosen so that for the maximum scan angle θ_m the movement of a grating lobe by $\sin\theta_m$ does not bring the grating lobe into real space. If a scan angle of 60° from broadside is required for every plane of scan, no grating lobes may exist within a circle of radius $1 + \sin\theta_m = 1.866$. The square grid that meets this requirement has

$$\frac{\lambda}{d_x} = \frac{\lambda}{d_y} = 1.866 \qquad \text{or} \qquad d_x = d_y = 0.536\lambda$$

Here, the area per element is

$$d_x d_y = (0.536\lambda)^2 = 0.287\lambda^2$$

For an equilateral-triangular array, the requirement is satisfied by

$$\frac{\lambda}{d_y} = \frac{\lambda}{\sqrt{3}\,d_x} = 1.866 \qquad \text{or} \qquad d_y = 0.536\lambda \qquad d_x = 0.309\lambda$$

Since elements are located only at every other value of mn, the area per element is

$$2d_x d_y = 2(0.536\lambda)(0.309\lambda) = 0.332\lambda^2$$

For the same amount of grating-lobe suppression, the square geometry requires approximately 16 percent more elements.

Element-Phasing Calculations. A computer is usually required to perform the steering computations for a phased array antenna. It can compensate for many of the known phase errors caused by the microwave components, the operating environment, and the physical placement of the elements. For example, if the insertion and differential phase variations (which may occur from phase shifter to phase shifter) are known, they may be taken into account in the computations. Known temperature variations across the array that would induce phase errors may be compensated for. Finally, many feeds (e.g., optical and series feeds) do not provide equal phase excitation at the input to each phase shifter. The relative phase excitation caused by these feeds is a known function of frequency. In these cases, the computer must provide a correction based on the location of the element in the array and on the frequency of operation.

For a large array with thousands of elements, many calculations are required to determine the phasing of the elements. These calculations must be performed in a short period of time. The use of the orthogonal phase commands mT_{xs}, nT_{ys} helps to minimize these calculations. Once the element-to-element phase increments T_{xs}, T_{ys} have been computed for a given beam-pointing direction, the integral multiples of T_{ys} may be used to steer the columns (Fig. 7.8). If an adder is located at each element, the row-and-column values mT_{xs} and nT_{ys} may be summed at the element. It is also possible to put two phase shifters in series so that the summation can be done at microwave frequencies. This may be implemented with the use of a series feed, as shown in Fig. 7.11. Here the row steering commands apply equally to all rows. An amplifier between a row phase shifter and a series feed is desirable so that the generated power does not have to take the loss of two phase shifters in series. In addition, the row phase shifters must be capable of accurate phasing. Since relatively few row phase shifters are required, it is reasonable to make them considerably more accurate than the phase shifters in the array. Any corrections for a phase taper
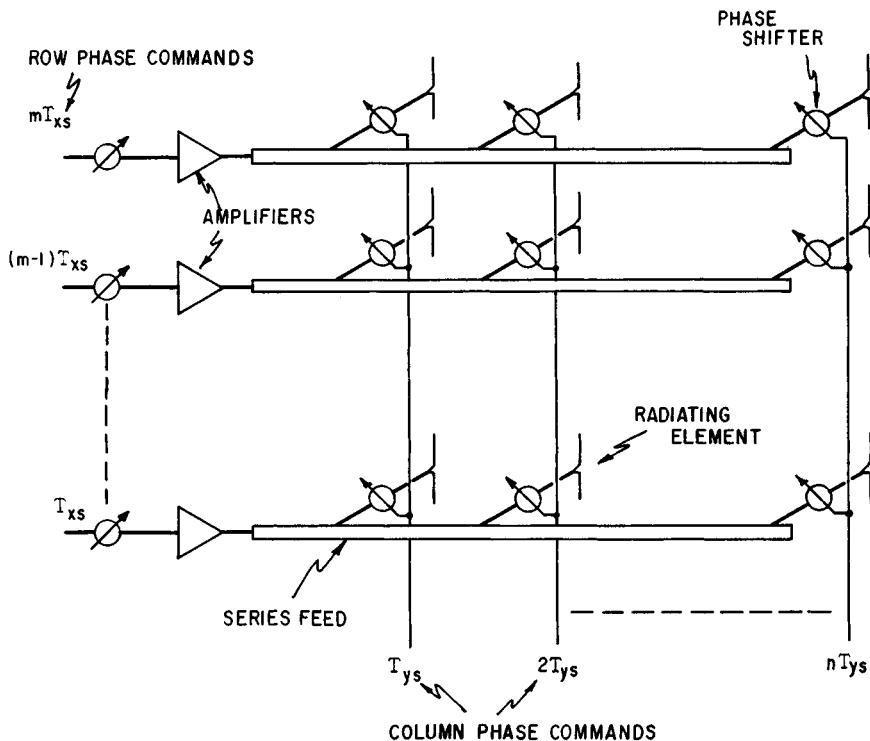
FIG. 7.11 Microwave addition of orthogonal phasing commands by means of a series feed.

across the series feed may be applied for an entire column if all the feeds have the same phase characteristics.

A large array requires many electronic phase-shifter drivers and very complex wiring to provide control signals and energy. The problem is complicated by the relatively close spacing of elements in the array. Further, many phase shifters are of the digital type and require a driver and control signal for each bit. The problems are eased somewhat in the system described above with two RF phase shifters in series, since many elements may use the same steering commands and the same drivers. In other systems it may be necessary to provide each element in the array with an independent phase command. The command may include a phase correction for the correction of errors due to component tolerances. An adder at each element provides rapid steering through the use of row-and-column steering commands. If high-speed phase shifting is not required, the computer may compute sequentially and store the phases for each of the elements. All the phase commands can then be delivered simultaneously.

7.4 APERTURE MATCHING AND MUTUAL COUPLING [48]

Significance of Aperture Matching. An antenna is a device that acts as a transformer to provide a good match between a source of power and free

space. If the antenna is not matched to free space, power will be reflected back toward the generator, resulting in a loss in radiated power. In addition, a mismatch produces standing waves on the feed line to the antenna. The voltage at the peaks of these standing waves is $(1 + |\Gamma|)$ times greater than the voltage of a matched line, where Γ is the voltage reflection coefficient. This corresponds to an increased power level that is $(1 + |\Gamma|)^2$ times as great as the actual incident power. Therefore, while the antenna is radiating less power, individual components must be designed to handle more peak power. With antennas that do not scan, the mismatch may often be tuned out by conventional techniques, preferably at a point as close to the source of the mismatch as possible.

In a scanning array the impedance of a radiating element varies as the array is scanned, and the matching problem is considerably more complicated. Unlike a conventional antenna, where mismatch affects only the level of the power radiated and not the shape of the pattern, spurious lobes in the scanning array may appear as a consequence of the mismatch. Further, there are conditions where an antenna that is well matched at broadside may have some angle of scan at which most of the power is reflected.

The variation in element impedance and element pattern is a manifestation of the mutual coupling between radiating elements that are in close proximity to one another. For a practical design, two empirical techniques are of great value:

1. Waveguide simulators provide a means for determining the element impedance in an infinite array with the use of only a few elements. The effectiveness of a matching structure based on these measurements may also be determined in the simulator.

2. A small array is the best technique for determining the active element pattern. The active element pattern, obtained by exciting one element and terminating its neighbors, is the best overall measure of array performance other than the full array itself. If a large reflection occurs at some angle of scan, it can be recognized by a null in the element pattern. The small array can also provide data on the coupling between elements. This data can be used to calculate the variation in impedance as the array is scanned.

Both these techniques will be discussed later in this section.

Effects of Mutual Coupling.   When two antennas (or elements) are widely separated, the energy coupled between them is small and the influence of one antenna on the current excitation and pattern of the other antenna is negligible. As the antennas are brought closer together, the coupling between them increases. In general, the magnitude of the coupling is influenced by the distance between the elements, the pattern of the elements, and the structure in the vicinity of the elements. For example, the radiation pattern of a dipole has a null in the $\theta = \pm 90°$ direction and is omnidirectional in the $\theta = 0°$ plane. Therefore it can be expected that dipoles in line will be loosely coupled and parallel dipoles will be tightly coupled. When an element is placed in an array of many elements, the effects of coupling are sufficiently strong that the pattern and impedance of the element in the array are drastically altered.

The terms *active element pattern* and *element impedance* refer to an element in its operating environment (i.e., in an array with its neighboring elements excited). In the array, each excited element couples to every other element. The coupling from several elements to a typical central element, element 00, is shown in Fig. 7.12. The $C_{mn,pq}$ are mutual-coupling coefficients relating the voltage (amplitude and phase) induced in the *mn*th element to the voltage excitation at the
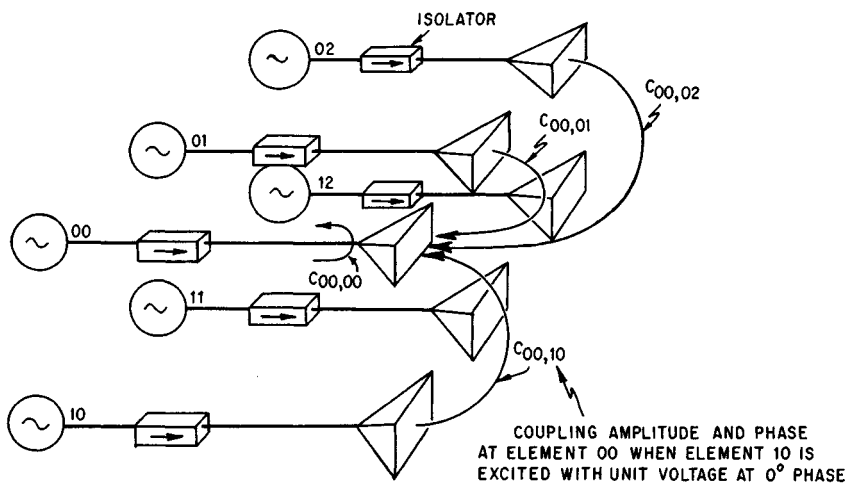
FIG. 7.12   Coupled signals to a central element from neighboring elements.

*pq*th element. The coupled signals add vectorially to produce a wave traveling toward the generator of element 00 that appears to be a reflection from the radiator of element 00. As the phases of the neighboring elements are varied to scan the beam, the vector sum of the coupled signals changes and causes an apparent change in the impedance of element 00. For some scan angles the coupled voltages tend to add in phase, causing a large reflection and possibly the loss of the main beam. Large reflections often occur at scan angles just prior to the emergence of a grating lobe into real space, but in some instances such reflections may occur at smaller scan angles.

The description of the impedance variation given above made no reference to the feed network or the phase shifters and assumed that the only coupling between elements is via the radiating aperture. The coupling coefficients would be measured, and by superposition the phased-voltage contributions from every element in the array (or at least those in the immediate vicinity) would be added vectorially to produce the voltage reflected back toward the generator. In a practical array the impedance variation depends upon the feed system and the phase shifter. If these are taken into account, the impedance variation may be different from what the above model might predict. In most analyses only the coupling at the aperture is considered. This description provides insight into the intrinsic impedance variation of the aperture when it is isolated from other effects, as in the case where each element has an independent feed (e.g., its own generator and isolator). In this case it is a simple matter to measure the voltage–standing-wave ratio (VSWR) in any line and determine exactly the extent of the impedance and mismatch variation. For many feed systems this is not possible, and a measurement of the reflected energy will provide erroneous information and a false sense of security. Unless all the reflections are collimated back at some central point (or independent feeds are used), some of the reflected energy will generally be re-reflected and contribute to large undesirable sidelobes.

For large arrays the impedance of an element located near the center of the array is often taken as typical of the impedance of every element in the array. As might be expected, this element is most strongly influenced by elements in its immediate vicinity. When the array is scanned, the influence of elements several

wavelengths distant is also significant. For dipoles above a ground plane the magnitude of the coupling between elements decays rapidly with distance. For a reasonable indication of array performance, an element in the center of a 5 by 5 array may be taken as typical of an element in a large array. For dipoles with no ground plane (or the dual, slots in a ground plane) the coupling between elements does not decay so rapidly, and a 9 by 9 array appears reasonable. For an array of open-ended waveguides, a 7 by 7 array should suffice (see Fig. 7.22, below). If accurate prediction of the array performance is required, many more elements are needed than are indicated above.[49,50]

It is often convenient to assume that the array is infinite in extent and has a uniform amplitude distribution and a linear-phase taper from element to element. In this manner every element in the array sees exactly the same environment, and the calculations for any element apply equally to all. These assumptions provide a significant simplification in the calculation of the element impedance variations. In addition, impedance measurements made in simulators correspond to the element impedance in an infinite array. In spite of the assumptions, the infinite-array model has predicted with good accuracy the array impedance and the impedance variations. Even arrays of modest proportions (less than 100 elements) have been in reasonable agreement with the results predicted for an infinite array.[51]

Element Pattern. From energy considerations the directional gain of a perfectly matched array with constant amplitude distribution ($\eta = 1$) will vary as the projected aperture area, from Eq. (7.18)

$$G(\theta_0) = \frac{4\pi A}{\lambda^2} \cos \theta_0$$

If it is assumed that each of the N elements in the array shares the gain equally, the gain of a single element is [Eq. (7.21)]

$$G_e(\theta) = \frac{4\pi A}{N\lambda^2} \cos \theta_0$$

If the element is mismatched, having a reflection coefficient $\Gamma(\theta, \phi)$ that varies as a function of scan angle, the element gain pattern is reduced to

$$G_e(\theta) = \frac{4\pi A}{N\lambda^2} (\cos \theta)[1 - |\Gamma(\theta, \phi)|^2]$$

The element pattern is seen to contain information pertaining to the element impedance.[52-55] The difference between the total power radiated in the element pattern and the power delivered to the antenna terminals must equal the reflected power. In terms of the radiation patterns of the scanning array, this means that since the scanned antenna patterns trace out the element pattern, it follows that the average power lost from the scanned patterns is equal to the power lost from the element pattern because of reflections. It is not enough to match one element in the presence of all its terminated neighbors. The element will deliver power to its neighbors, and this loss in power corresponds to the average power lost when scanning. An ideal although not necessarily realizable element pattern would place all the radiated power into the scan region, giving a pattern like a cosine on a pedestal and thereby providing maximum antenna gain for the number of elements used.

Thinned Arrays. The number of radiating elements in an array may be reduced to a fraction of those needed completely to fill the aperture without suffering serious degradation in the shape of the main beam. However, average sidelobes are degraded in proportion to the number of elements removed. The element density may be thinned so as to effectively taper the amplitude distribution, and the spacing is such that no coherent addition can occur to form grating lobes. A thinned aperture, where elements have been removed randomly from a regular grid,[56] is shown in Fig. 7.13. The gain is that due to the actual number of elements $NG_e(\theta)$, but the beamwidth is that of the full aperture. For example, if the array has been thinned so that only 10 percent of
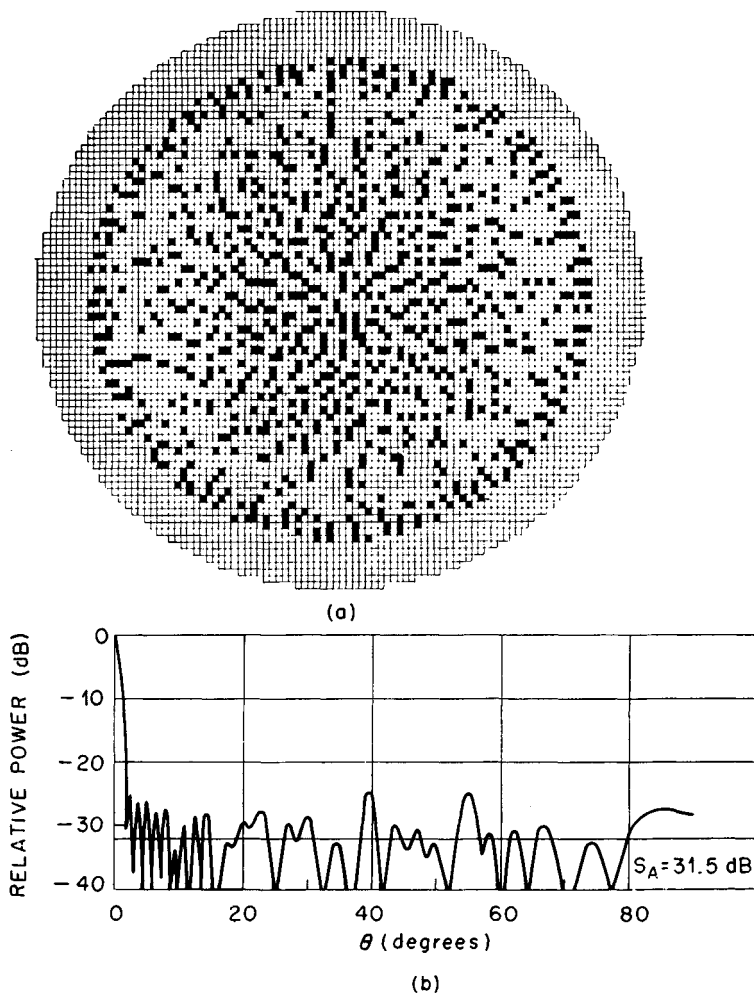

(a)

(b)

FIG. 7.13 (a) Thinned array with a 4000-element grid containing 900 elements. (b) Typical pattern for a thinned array. S_A is the average sidelobe level. (*From Willey,*[56] *courtesy of Bendix Radio.*)

the elements are used, the gain of the array will drop by 10 dB. However, since the main beam is virtually unchanged, about 90 percent of the power is delivered to the sidelobe region.

If the removed elements (in a regular thinned array) are replaced with elements with matched loads, the element pattern is identical to that of one in the regular array with all elements excited. The element pattern is independent of the array excitation, and the same fractional amount of power will be lost (because of mismatch) whether the array is thinned, tapered, or uniformly illuminated. It should be noted that the concept of an element pattern that applies equally to every element is valid only when isolating feeds are used and edge effects are ignored.

A thinned array may also be implemented with an irregular element spacing, although this is not common. In this case the element gain (and impedance) will vary from element to element, depending upon the environment of a given element. To obtain the gain of the array, it is necessary to sum all the different element gains $G_{en}(\theta)$. Thus

$$G(\theta) = \sum_n G_{en}(\theta)$$

Impedance Variation of Free Space. It is of interest to examine the case of a large continuous aperture which may be considered to be the limiting case of an array of many very small elements.[57] The free-space impedance E/H varies as $\cos \theta$ for scanning in the E plane and as $\sec \theta$ for scanning in the H plane. The impedance of a medium is thus dependent upon the direction of propagation, and the impedance variation of a scanning aperture is a natural consequence of this dependence. The continuous aperture appears to represent a lower limit to the impedance variation with scanning. This is indicated by Allen's results,[58] where the impedance variation with scanning was calculated for dipoles above a ground plane. In spite of increased mutual coupling, or perhaps because of it, the more closely the dipoles were spaced, the smaller the impedance variation with scanning (Fig. 7.14). Although the impedance variation decreased, the absolute impedance of the dipoles also decreased, making them more difficult to match at broadside. It is expected that to obtain an impedance variation smaller than that of free space some impedance compensation must be employed.

Element Impedance. The simplest and most straightforward method for computing the variation in reflection coefficient (and impedance) is by means of the scattering matrix of mutual-coupling coefficients. The mutual-coupling coefficients may be easily measured for elements of all types by exciting one element and terminating each of the other elements in a matched load. The ratio of the induced voltage at element mn to the excitation voltage at element pq gives the amplitude and phase of the coupling coefficient $C_{mn,pq}$. Once these coefficients are determined, it is a simple matter to compute the mismatch for any set of phasing conditions.

Consider the two-element array shown in Fig. 7.15 where each element is provided with an isolating feed. The incident wave in each element is represented by V_1, V_2, and the total reflected wave in each element is represented by V_1', V_2'. It should be apparent that the total reflected wave in any element
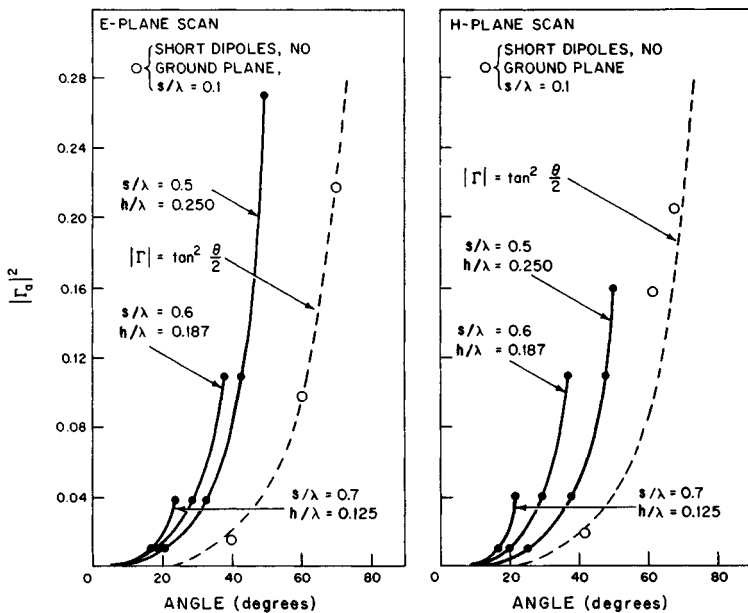
FIG. 7.14 Scanned mismatch variation for different element spacings (h/λ is the dipole spacing above a ground plane). (*After Allen.*[58])
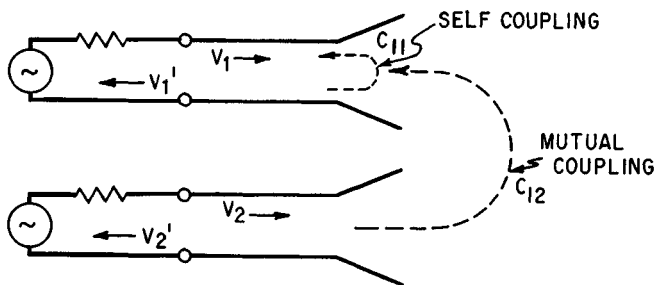
FIG. 7.15 Scattering-matrix model for a two-element array.

is the vector sum of the couplings from all elements, including its own reflection as a self-coupling:

$$V_1' = C_{11} V_1 + C_{12} V_2$$

$$V_2' = C_{21} V_1 + C_{22} V_2$$

The reflection coefficient in each element is obtained by dividing the reflected voltage by the incident voltage in the channel:

$$\Gamma_1 = \frac{V_1'}{V_1} = C_{11} \frac{V_1}{V_1} + C_{12} \frac{V_2}{V_1}$$

$$\Gamma_2 = \frac{V_2{}'}{V_2} = C_{21}\frac{V_1}{V_2} + C_{22}\frac{V_2}{V_2}$$

Note that all the quantities must contain both phase and amplitude and that as the phases of V_1, V_2 are varied to scan the beam, the reflection coefficients (Γ_1, Γ_2) will vary. Although only two elements have been used in this example, the technique is completely general. For a large array the reflection coefficient of the mnth element is given by

$$\Gamma_{mn} = \sum_{\text{all } pq} C_{mn,pq}\frac{V_{pq}}{V_{mn}}$$

The general case is treated in more detail by Oliner and Malech.[59] No restrictions need be placed on either the amplitude or the phase of the excitation at each element. There are also no restrictions on the spacing between the elements as long as the coupling coefficients are measured for the spacing and environment to be used. A considerable simplification is obtained by assuming that each of the elements sees the same environment and has the same voltage excitation. Then $|V_{pq}|/|V_{mn}|$ will always be unity, and the reflection coefficient at element mn is merely the sum of the mutual-coupling coefficients with the excitation phase at each element taken into account:

$$\Gamma_{mn} = \sum_{\text{all } pq} C_{mn,pq}e^{j(m-p)T_{xs}}e^{j(n-q)T_{ys}}$$

where $e^{j(m-p)T_{xs}}$ and $e^{j(n-q)T_{ys}}$ give the relative phase excitations in the x and y directions of the pqth element with respect to the mnth element. The impedance variation relative to a matched impedance at broadside may be obtained immediately from

$$\frac{Z_{mn}(\theta,\phi)}{Z_{mn}(0,0)} = \frac{1 + \Gamma_{mn}(\theta,\phi)}{1 - \Gamma_{mn}(\theta,\phi)}$$

Analytical Techniques. Stark[60] presents a thorough description of analytical techniques and insight to the problem of mutual coupling. He derives necessary and sufficient conditions for array blindness (i.e., nulls in the active element pattern). Evidence is provided to demonstrate that more closely spaced elements reduce the variation of impedance due to scanning in spite of the increase in mutual coupling.

Elliott and coworkers[61,62] have developed design procedures for slot arrays which are fed by either air-filled or dielectric-filled waveguide. These procedures take into account the differences in mutual coupling for central elements in an array as well as for the edge elements. This is particularly of interest when designing a small array.

Munk and colleagues[63,64] have developed a procedure for reducing the variations in impedance with scanning by matching elements with dielectric slabs. Their analytical results show a VSWR of less than 1.5 for scan angles of $\pm 80°$ in

each of the principal planes. The penalty they incur for placing a thick (0.4λ) and potentially heavy dielectric slab in front of the aperture is mitigated by the low dielectric constant of the slab ($\epsilon_r = 1.3$). This suggests that a lightweight loaded foam material could be used. Dielectric slabs on the surface of an array can often create surface waves and array blindness. In this case, however, the dielectric constant is quite low. The surface wave can be avoided by spacing the elements more closely. Here, as in some other cases, it is seen that closely spaced elements are beneficial for improving impedance matching for scanning arrays. Obviously, reducing the element spacing increases the number of elements and hence the cost without any increase in gain or reduction in beamwidth.

Another very useful tool for the computation of the impedance variation with scanning is the grating-lobe series,[65,66] which describes the impedance variation of an infinite array of regularly spaced elements.

Nonisolating Feeds. When nonisolating feeds are used, the mutual-coupling effects become dependent on whether the phase-shifting element is reciprocal or nonreciprocal. Figure 7.16 shows a space-fed array where it is assumed that the initial excitations at the input aperture are of equal amplitude and phase. If the phase shifters are nonreciprocal, the round-trip phase of a signal reflected from the radiating aperture is independent of phase-shifter setting. Therefore, when reflections occur at the radiating aperture, the reflected signal is phased so that the input aperture appears as a mirror as seen from the feed side, with the magnitude of the reflection being determined by the radiating aperture. Since the reflected beam does not scan, it should be possible to provide a good match at the input aperture. Matching the input aperture, in this case, is equivalent to providing the radiating aperture with independent feeds. Any secondary reflection from the input aperture will radiate in the original direction of scan.

If reciprocal phase shifters are used, the energy reflected from the radiating aperture will pick up equal additional phase shift on reflection from the radiating aperture, resulting in a beam at the input aperture that is phased to scan to twice the original scan angle (in sin θ space). Some energy will undergo secondary re-
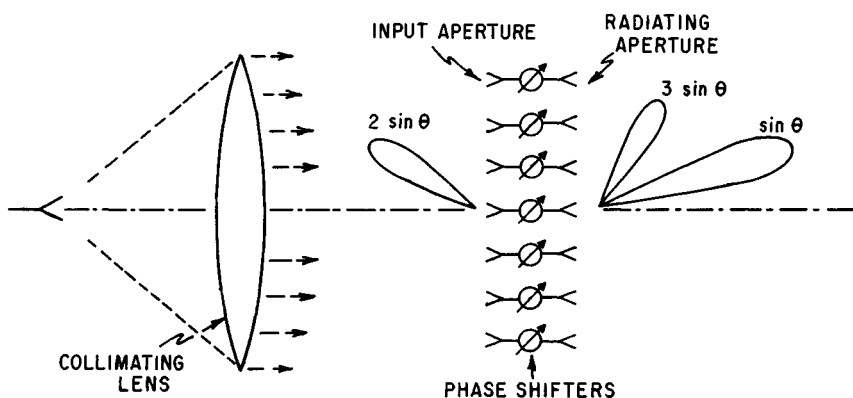

FIG. 7.16   Nonisolating feed showing spurious lobes when reciprocal phase shifters are used.

flection from the input aperture (which now has a mismatch corresponding to a scan of 2 sin θ) and will be phase-shifted once again to produce a beam at the radiating aperture in the direction 3 sin θ. Additional reflections acquire twice the original phase shift for each round trip, resulting in radiated beams at positions of 5 sin θ, 7 sin θ, etc. The magnitude of these beams is approximately equal to the product of the voltage reflection coefficients if the number of bounces is taken into account. For example, let $\Gamma_r(\sin θ)$, $\Gamma_i(2 \sin θ)$ denote the reflection coefficients corresponding to scan angles of sin θ at the radiating aperture and 2 sin θ at the input aperture. If $\Gamma_r(\sin θ) = 0.2$ and $\Gamma_i(2 \sin θ) = 0.5$, the magnitude of the radiated lobe directed at 3 sin θ would be $\Gamma_r(\sin θ) \Gamma_i(2 \sin θ) = 0.1$, or 20 dB down from the main lobe. Similar results can be expected from series feeds[67] and from reactive parallel feeds.[48]

Mutual Coupling and Surface Waves. The mutual coupling between two small isolated dipoles[68] should decrease as $1/r$ in the H plane and $1/r^2$ in the E plane (E and H planes are interchanged for slots). Coupling measurements[69] have shown that in the array environment the rate of decay is slightly greater than predicted above, indicating that some of the energy is delivered to other elements in the array and may be dissipated and reradiated from these elements. The same measurements have shown that the phase difference of the energy coupled to elements is directly proportional to their distance from the excited elements, indicative of a surface wave traveling along the array, leaking energy to each of the elements. For best performance the velocity of the surface wave should be very close to that of free space. If the array contains waveguides or horns loaded with dielectric, the velocity will decrease slightly. Further, if the dielectric protrudes from the radiators or if a dielectric sheet is used in front of the array, the velocity of the surface wave may decrease dramatically. This surface wave is important since it can cause a large reflection (and an accompanying loss of the beam) for some angles of scan. This can best be seen by examining the condition of phasing for which the couplings from many elements will add in phase to cause a large reflection in a typical element.

Consider an array in which the velocity of the surface wave is that of free space. The difference in the phase of the voltages coupled from an adjacent pair of elements to element 00 (Fig. 7.17) is related to the scan angle by
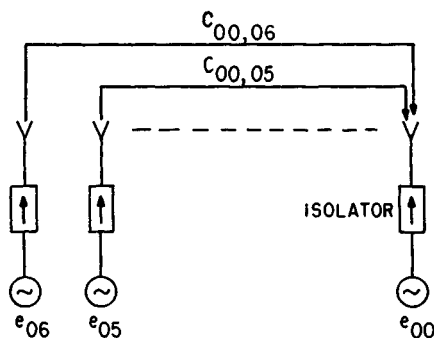

FIG. 7.17 Two adjacent elements coupling to another element in the same row.

$$\Delta\chi = \frac{2\pi s}{\lambda} + \frac{2\pi s}{\lambda} \sin \theta_0 = \frac{2\pi s}{\lambda}(1 + \sin \theta_0)$$

The couplings will be in phase when $\Delta\psi = 2\pi$ or when

$$\frac{s}{\lambda} = \frac{1}{1 + \sin \theta_0}$$

This is seen to be exactly the same condition as previously determined for the emergence of a grating lobe into real space. Therefore, it may be expected that when a grating lobe is about to emerge into real space, the coupled voltages tend to add in phase and cause a large mismatch. If the dielectric protrudes from the aperture or if a dielectric sheet covers the aperture (this is one technique for scan compensation discussed below), a large reflection may occur well before the grating lobe reaches real space.[70] If a surface-wave velocity of v_s is assumed, the couplings will add in phase when

$$\frac{s}{\lambda} = \frac{1}{c/v_s + \sin \theta_0}$$

For the purposes of mutual coupling, a slow wave across the aperture may be envisioned as being equivalent to spacing the elements farther apart in free space.

A phenomenon that produces similar effects can come about without dielectric in front of the aperture, e.g., by using a periodic structure of baffles. Under certain conditions an array of open-ended waveguides will perform as though a slow wave were propagating across the aperture. This effect has been analyzed and studied experimentally,[71,72] and it has been shown that as the array is scanned, higher-order modes are excited in the waveguides. Even though these modes are cut off, they contribute to the active impedance. At certain angles, almost all the energy is reflected, causing a null in the element pattern. To guard against these reflections it is best to design the radiators so that higher-order modes are well into the cutoff region. Figure 7.18 shows the results obtained by Diamond,[72] using a waveguide array. When only the dominant TE_{10} mode was taken into account, the null could not be explained. When the TE_{20} mode (which was only slightly cut off) was taken into account, the null showed up clearly. Since a null in the element pattern indicates in-phase addition of the mutually coupled signals, it appears that the array of open-ended waveguides causes the couplings between elements to have a phase variation corresponding to a velocity slower than that of free space.

Regardless of the cause of the null, it will show up in the element pattern. If only a few elements surround the central element, the null will normally be shallow and broad. If many elements are used, the null will be deep and sharp. The null will also show up if the mutual-coupling coefficients are measured and used to calculate the reflection coefficient.

Array Simulators. A good deal of effort has gone into matching a radiator in the presence of an array of radiators. The use of waveguide simulators as developed by Wheeler Laboratories has made it possible to determine the matching structure experimentally without the need of building an array. A waveguide, operating in a TE_{10} mode, may be considered to contain two inclined-plane waves propagating down the guide. The angle that each of the plane waves makes with the longitudinal direction (Fig. 7.19) is determined by
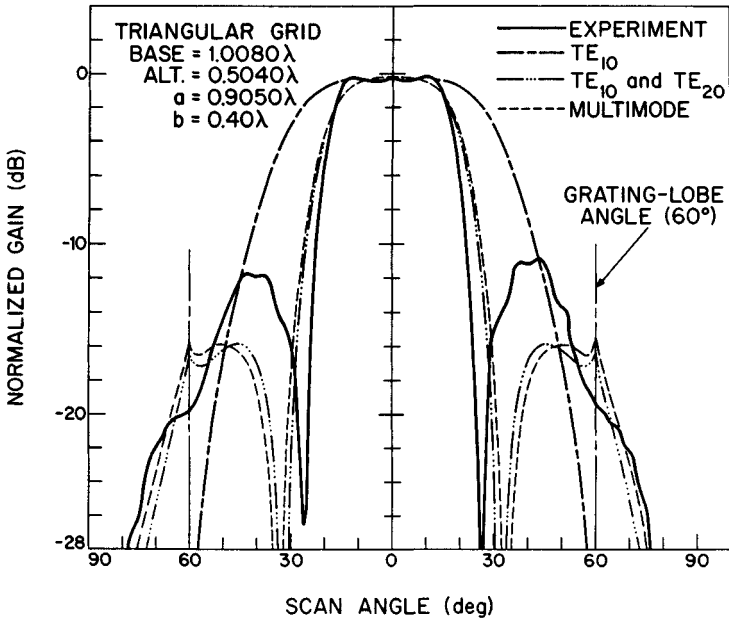
FIG. 7.18 Comparison of the theoretical and an experimental H-plane element pattern: triangular array of waveguides with a 2:1 ratio of width to height. (*From Diamond.*[72])
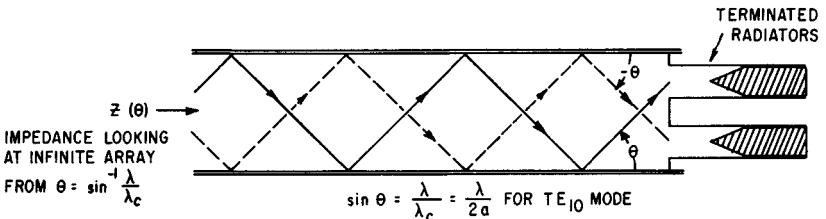
FIG. 7.19 Array simulator terminated with two dummy elements.

the H dimension of the waveguide and simulates the angle of scan of an infinite array:

$$\sin \theta = \frac{\lambda}{\lambda_c} \qquad (7.26)$$

where θ = scan angle
      λ = free-space wavelength
      λ_c = cutoff wavelength of guide

Additional scan angles may be simulated by exciting other modes. The waveguide dimensions are chosen so that a radiating element or elements placed

in the waveguide sees mirror images in the walls of the waveguide that appear to be at the same spacing as the array to be simulated. Both rectangular and triangular arrays may be simulated, as shown in Fig. 7.20. The impedance measurements are made by looking into a waveguide simulator that is terminated with dummy elements. This is equivalent to looking at an infinite array from free space at a scan angle given by Eq. (7.26). A matching structure, designed from the simulator impedance data, may be placed into the simulator to measure its effectiveness. Several simulator designs, results, and a complete discussion of the topic have been presented by Hannan and Balfour.[73] The technique is limited in that only discrete scan angles can be simulated. Several scan angles in both planes of scan give a general idea of the array impedance but may miss a large reflection of the type described in this section under "Mutual Coupling and Surface Waves." Nevertheless, the array simulator is the best method available for empirically determining the array impedance without building an array.
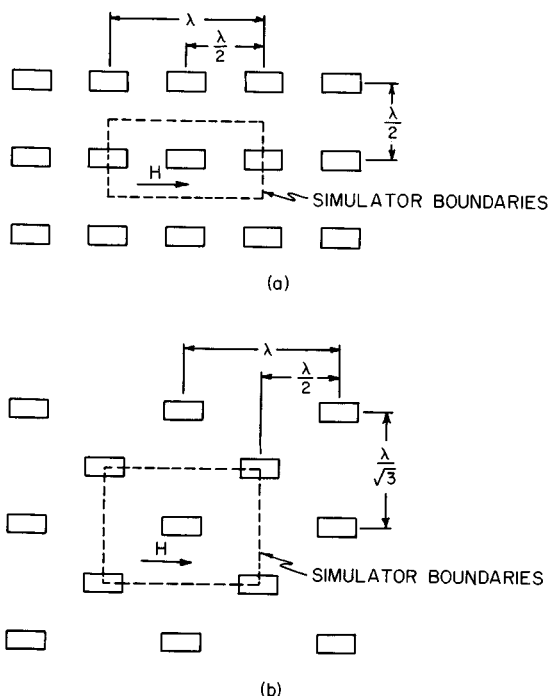
FIG. 7.20 Rectangular- and triangular-array geometries with simulator boundaries superimposed. (*a*) Square array with simulator superimposed. (*b*) Triangular array with simulator superimposed.

Under certain conditions, the matched element impedance may be determined analytically by envisioning a simulator.[74] For example, if the waveguide simulator is terminated by a single dummy element, the matched impedance of the element is exactly the impedance of the waveguide. Since waveguide impedances are well known, the matched impedance at the simulator scan angle is also known.

Compensation for Scanned Impedance Variation. The impedance of an element in an array has been discussed and has been shown to vary as the array is scanned. An array that is matched at broadside can be expected to have at least a 2:1 VSWR at a 60° angle of scan. To compensate for the impedance variation, it is necessary to have a compensation network that is also dependent on scanning.

One method uses a thin sheet of high-dielectric-constant material (e.g., alumina) spaced a fraction of a wavelength from the array, as shown in Fig. 7.21.This is in contrast to the method used by Munk[63,64] and described under "Analytical Techniques." The properties of a thin dielectric sheet (less than a quarter wave in the medium) are such that to an incident plane wave it appears as a susceptance that varies with both the plane of scan and the angle of scan. Magill and Wheeler[75] describe the technique in greater detail and present the results of a particular design using simulators. An alumina matching sheet is attractive because it simultaneously provides a natural radome. It should be cautioned that a dielectric sheet in front of the aperture may produce a slow surface wave and a possible null in the element pattern. However, a thin dielectric matching sheet has been used for a 400-element array,[51] and some compensation has been achieved without any noticeable slow-wave phenomenon.
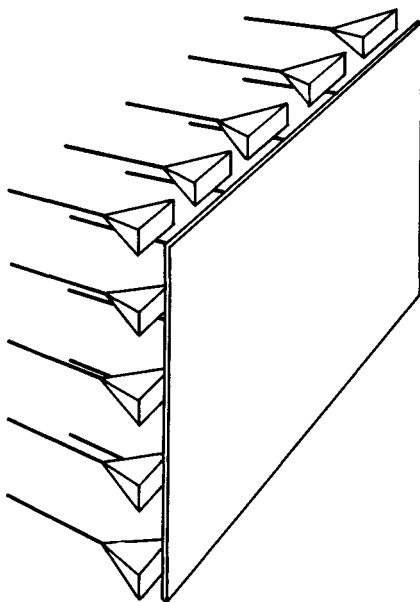
FIG. 7.21   Planar array with a thin dielectric sheet spaced a fraction of a wavelength from the radiators.

Small Arrays. The element pattern is the best single indicator of impedance matching in a scanning array. One way of determining the element pattern is to build a small array. A central element is excited, and all other elements are terminated. The pattern of this central element is the active element pattern.

Diamond[76] has examined the number of elements required in a small array to provide a reasonable approximation to an element in an infinite array. He concludes that 25 to 37 elements are required to provide a good indication. Figure 7.22 shows the change in the measured active element pattern as the number of elements is increased. For a 41-element array the null is very pronounced. Even for the 23-element array it is clear that the gain variation with scanning is dramatically greater than cos θ.

The small array can also be used to measure coupling coefficients as demonstrated by Fig. 7.12. These coupling coefficients can be used to calculate the impedance variation as the array is scanned. Grove, Martin, and Pepe[77] have noted that for the element to be matched in its operating environment the self-coupling C_{11} must exactly cancel the coupling from all other elements. They have used this technique to provide a good match on an ultralow-sidelobe, wideband phased array. After adjusting C_{11} to cancel the sum of the other mutuals, they further adjusted the amplitude of C_{11} to compensate for the variation in mismatch loss due to the resultant change in C_{11}. Results were obtained for an array of greater than two-octave bandwidth. In this design grating lobes were permitted to enter real space. At this point a high mismatch was observed. The mismatch improved as the grating lobe moved further into real space.

The combination of waveguide simulators and small arrays provides powerful empirical tools to supplement the analytical techniques. Experience has demonstrated that a large antenna should not be built until the element pattern has been verified with a small array.
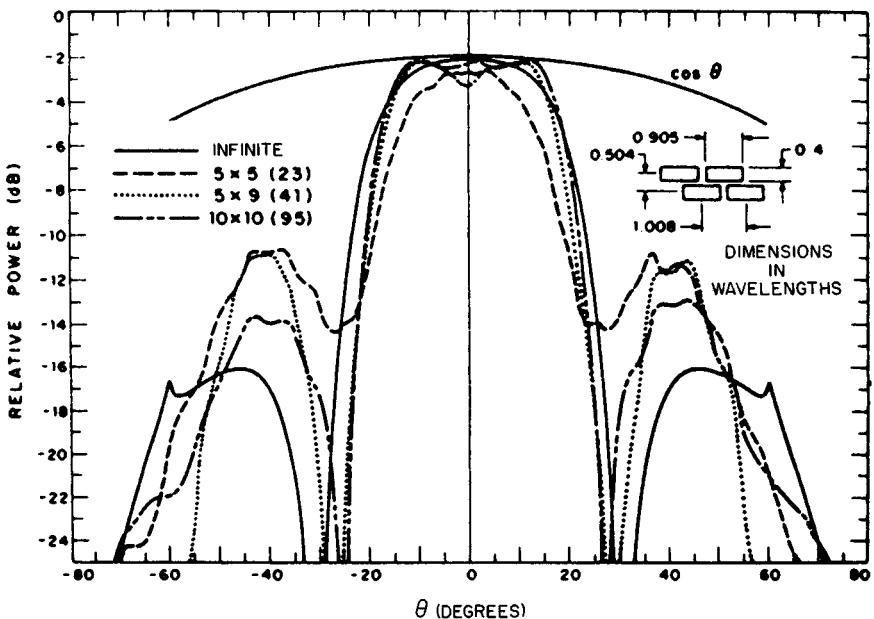
FIG. 7.22 Experimental *H*-plane patterns of the center elements of waveguide arrays. (*After Diamond.*[76])

CHAPTER 8
AUTOMATIC DETECTION, TRACKING, AND SENSOR INTEGRATION

G. V. Trunk
Naval Research Laboratory

8.1 INTRODUCTION

Since the invention of radar, radar operators have detected and tracked targets by using visual inputs from a variety of displays. Although operators can perform these tasks very accurately, they are easily saturated and quickly become fatigued. Various studies have shown that operators can manually track only a few targets. To correct this situation, automatic detection and tracking (ADT) systems were attached to many radars. As digital processing increases in speed and hardware decreases in cost and size, ADT systems will become associated with almost all but the simplest radars.

In this chapter, automatic detection, automatic tracking, and sensor integration systems for air surveillance radar will be discussed. Included in this discussion are various noncoherent integrators that provide target enhancement, thresholding techniques for reducing false alarms and target suppression, and algorithms for estimating target position and resolving targets. Then, an overview of the entire tracking system is given, followed by a discussion of its various components such as tracking filter, maneuver-following logic, track initiation, and correlation logic. Next, multiscan approaches to automatic tracking such as maximum likelihood are discussed. Finally, the chapter concludes with a discussion of sensor integration and radar netting, including both colocated and multisite systems.

8.2 AUTOMATIC DETECTION

The statistical framework necessary for the development of automatic detection was applied to radar in the 1940s by Marcum,[1] and later Swerling[2] extended the work to fluctuating targets. They investigated many of the statistical problems

associated with the noncoherent detection of targets in Rayleigh noise. (NOTE: If the quadrature components are gaussian-distributed, the envelope is Rayleigh-distributed and the power is exponentially distributed.) Marcum's most important result was the generation of curves of probability of detection (P_D) versus signal-to-noise ratio (S/N) for a detector which sums N envelope-detected samples (either linear or square-law) under the assumption of equal signal amplitudes. However, in a search radar, as the beam sweeps over the target, the returned signal amplitude is modulated by the antenna pattern. Many authors investigated various detectors (weightings), comparing detection performance and angular estimation results with optimal values; and many of these results are presented later in this section.

In the original work on detectors, the environment was assumed known and homogeneous, so that fixed thresholds could be used. However, a realistic environment (e.g., containing land, sea, and rain) will cause an exorbitant number of false alarms for a fixed-threshold system that does not utilize excellent coherent processing. Three main approaches, adaptive thresholding, nonparametric detectors, and clutter maps, have been used to solve the false-alarm problem. Both adaptive thresholding and nonparametric detectors are based on the assumption that homogeneity exists in a small region about the range cell that is being tested. The adaptive thresholding method assumes that the noise density is known except for a few unknown parameters (e.g., the mean and the variance). The surrounding reference cells are then used to estimate the unknown parameters, and a threshold based on the estimated density is obtained. Nonparametric detectors obtain a constant false-alarm rate (CFAR) by ranking the test samples (ordering the samples from smallest to largest), usually with the reference cells. Under the hypothesis that all the samples (test and reference) are independent samples from an unknown density function, the test sample has a uniform density function, and, consequently, a threshold which yields CFAR can be set. Clutter maps store an average background level for each range-azimuth cell. A target is then declared in a range-azimuth cell if the new value exceeds the average background level by a specified amount.

Optimal Detector. The radar detection problem is a binary hypothesis-testing problem in which H_0 denotes the hypothesis that no target is present and H_1 is the hypothesis that the target is present. While several criteria (i.e., definitions of optimality) can be used to solve this problem, the most appropriate for radar is the Neyman-Pearson.[3] This criterion maximizes the probability of detection P_D for a given probability of false alarm P_{fa} by comparing the likelihood ratio L [defined by Eq. (8.1)] to an appropriate threshold T which determines the P_{fa}. A target is declared present if

$$L(x_1, \ldots, x_n) \frac{p(x_1, \ldots, x_n | H_1)}{p(x_1, \ldots, x_n | H_0)} \geq T \qquad (8.1)$$

where $p(x_1, \ldots, x_n | H_1)$ and $p(x_1, \ldots, x_n | H_0)$ are the joint probability density functions of the n samples x_i under the conditions of target presence and target absence, respectively. For a linear envelope detector the samples have a Rayleigh density under H_0 and a ricean density under H_1, and the likelihood ratio detector reduces to

$$\prod_{i=1}^{n} I_0 \left(\frac{A_i x_i}{\sigma^2} \right) \geq T \qquad (8.2)$$

where I_0 is the Bessel function of zero order, σ^2 is the noise power, and A_i is the target amplitude of the ith pulse and is proportional to the antenna power pattern. For small signals ($A_i < \not\in s$), the detector reduces to the square-law detector

$$\sum_{i=1}^{n} A_i^2 x_i^2 \geq T \qquad (8.3)$$

and for large signals ($A_i > \$ s$), it reduces to the linear detector

$$\sum_{i=1}^{n} A_i x_i > T \qquad (8.4)$$

For constant signal amplitude (i.e., $A_i = A$) these detectors were first studied by Marcum[1] and were studied in succeeding years by numerous other people. Detection curves for both linear and square-law detectors are given in Chap. 2. The most important facts concerning these detectors are the following:

- The detection performances of the linear and square-law detectors are similar, differing only by less than 0.2 dB over wide ranges of P_D, P_{fa}, and n.

- Since the signal return of a scanning radar is modulated by the antenna pattern, to maximize the S/N when integrating a large number of pulses with no weighting (i.e., $A_i = 1$) only 0.84 of the pulses between the half-power points should be integrated, and the antenna beam-shape factor (ABSF) is 1.6 dB.[4] The ABSF is the number by which the midbeam S/N must be reduced so that the detection curves generated for equal signal amplitudes can be used for the scanning radar.

- The collapsing loss for the linear detector can be several decibels greater than the loss for a square-law detector[5] (see Fig. 8.1). The collapsing loss is the additional signal required to maintain the same P_D and P_{fa} when unwanted noise samples along with the desired signal-plus-noise samples are integrated. The number of signal samples integrated is N, the number of extraneous noise samples integrated is M, and the collapsing ratio $\rho = (N + M)/N$.

- Most automatic detectors are required not only to detect targets but to make angular estimates of the azimuth position of the target. Swerling[6] calculated the standard deviation of the optimal estimate by using the Cramer-Rao lower bound. The results are shown in Fig. 8.2, where a normalized standard deviation is plotted against the midbeam S/N. This result holds for a moderate or large number of pulses integrated, and the optimal estimate involves finding the location where the correlation of the returned signal and the derivative of the antenna pattern is zero. Although this estimate is rarely implemented, its performance is approached by simple estimates.

Practical Detectors.   Many different detectors (often called *integrators*) are used to accumulate the radar returns as a radar sweeps by a target. A few of the most common detectors[7] are shown in Fig. 8.3. Though they are shown in the figure as being constructed with shift registers, they would normally be implemented with random-access memory. The input to these detectors can be linear, square-law, or log video. Since linear is probably the most commonly used, the advantages and disadvantages of the various detectors will be stated for this video.
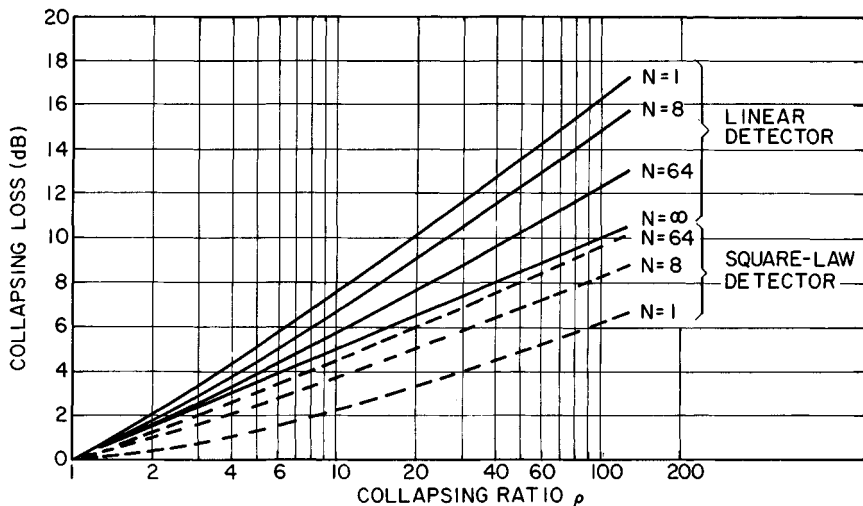
FIG. 8.1 Collapsing loss versus collapsing ratio for a probability of false alarm of 10^{-6} and a probability of detection of 0.5. (*Copyright 1972, IEEE; from Ref. 5.*)
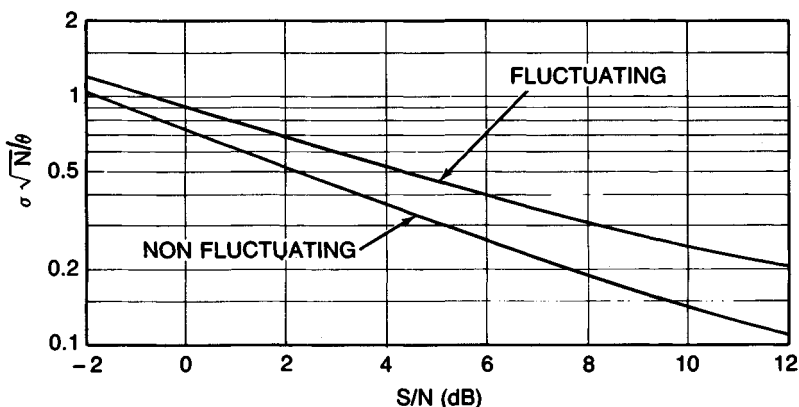
FIG. 8.2 Comparison of angular estimates with the Cramer-Rao lower bound. σ is the standard deviation of the estimation error, and N is the number of pulses within the 3-dB beamwidth, which is υ. The S/N is the value at the center of the beam. (*Copyright 1956, IEEE; after Ref. 6.*)

Moving Window. The moving window in Fig. 8.3*a* performs a running sum of n pulses in each range cell;

$$S_i = S_{i-1} + x_i - x_{i-n} \qquad (8.5)$$

where S_i is the sum at the ith pulse of the last n pulses and x_i is the ith pulse. The performance[8] of this detector for $n \approx 10$ is only 0.5 dB worse than the optimal detector given by Eq. (8.3). The detection performance can be obtained by using

INTEGRATORS
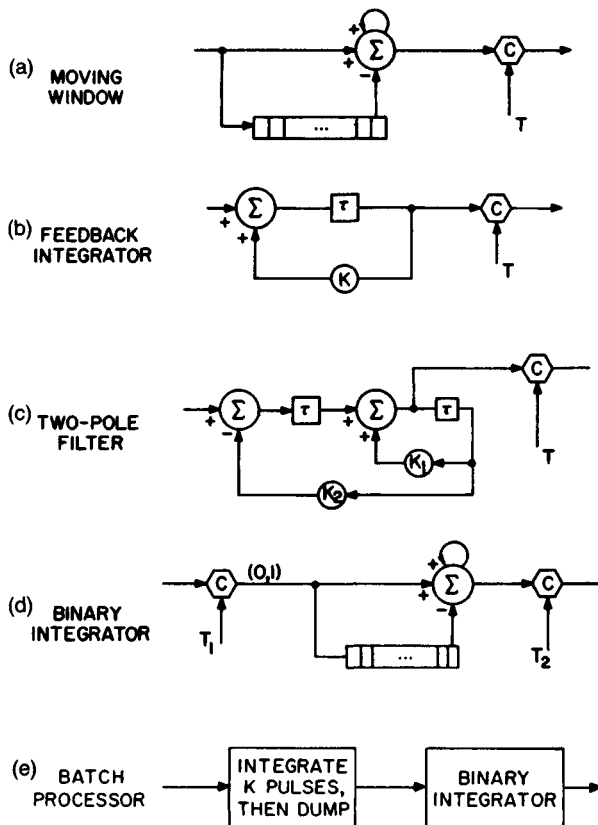
FIG. 8.3   Block diagrams of various detectors. The letter *C* indicates a comparison, τ is a delay, and loops indicate feedback. (*From Ref. 7.*)

an ABSF of 1.6 dB and the detection curves in Chap. 2. The angular estimate that is obtained by either taking the maximum value of the running sum or taking the midpoint between the first and last crossings of the detection threshold has a bias of $n/2$ pulses, which is easily corrected. The standard deviation of the estimation error of both estimators is about 20 percent higher than the optimal estimate specified by Cramer-Rao bound. A disadvantage of this detector is that it is susceptible to interference; that is, one large sample from interference can cause a detection. This problem can be minimized by using limiting. A minor disadvantage is that the last *n* pulses for each range cell must be saved, resulting in a large storage requirement when a large number of pulses are integrated. However, because of the availability of large memories of reduced size and cost, this is a minor problem.

The detection performance discussed previously is based on the assumption that the target is centered in the moving window. In the real situation the radar scans over the target, and decisions which are highly correlated are made at every pulse. Hansen[9] analyzed this situation for $N = 2, 4, 8$, and 16 pulses and
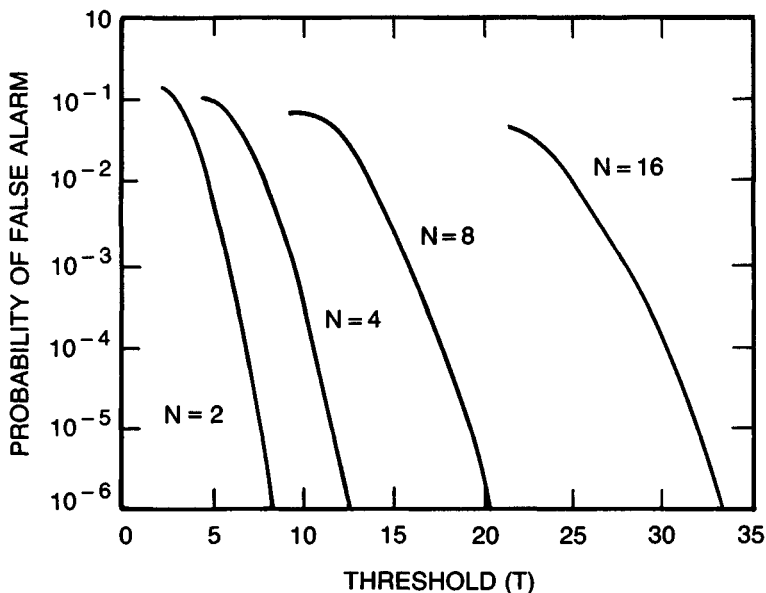
FIG. 8.4 Single-sweep false-alarm probability P_{fa} versus threshold for moving window. The noise is Rayleigh-distributed with $\sigma = 1$. (*Copyright 1970, IEEE; after Ref. 9.*)

calculated the detection thresholds shown in Fig. 8.4, the detection performance shown in Fig. 8.5, and the angular accuracy shown in Fig. 8.6. Comparing Hansen's scanning calculation with the single-point calculation, one concludes that 1 dB of improvement is obtained by making a decision at every pulse. The angular error of the beam-splitting procedure is about 20 percent greater than the optimal estimate. For large signal-to-noise ratios, the accuracy (rms error) of the beam-splitting and maximum-return procedures will be limited by the pulse
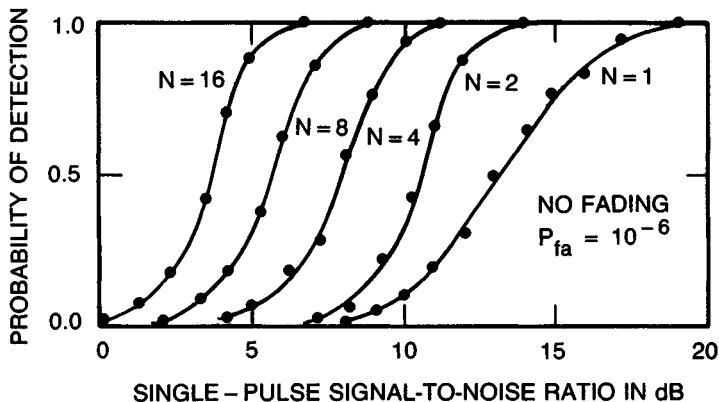
FIG. 8.5 Detection performance of the analog moving-window detector for the no-fading case. (*Copyright 1970, IEEE; after Ref. 9.*)
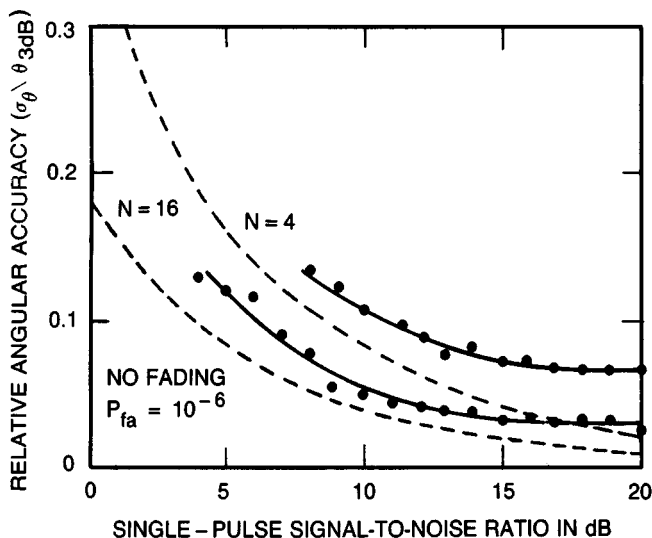
FIG. 8.6 Angular accuracy obtained with beam-splitting estimation procedure for the no-fading case. Broken-line curves are lower bounds derived by Swerling,[6] and points shown are simulation results. (*Copyright 1970, IEEE; after Ref. 9.*)

spacing[10] and will approach

$$\sigma(\hat{\theta}) = \Delta\theta/\sqrt{12} \tag{8.6}$$

where $\Delta\theta$ is the angular rotation between transmitted pulses. Consequently, if the number of pulses per beamwidth is small, the angular accuracy will be poor. For instance, if pulses are separated by 0.5 beamwidth, $\sigma(\hat{\theta})$ is bounded by 0.14 beamwidth. However, improved accuracy can be obtained by using the amplitudes of the radar returns. An accurate estimate of the target angle is given by

$$\hat{\theta} = \theta_1 + \frac{\Delta\theta}{2} + \frac{1}{2a\Delta\theta}\ln(A_2/A_1) \tag{8.7}$$

where

$$a = 1.386/(\text{beamwidth})^2 \tag{8.8}$$

and A_1 and A_2 are the two largest amplitudes of the returned samples and occur at angles θ_1 and $\theta_2 = \theta_1 + \Delta\theta$ respectively. Since the estimate should lie between θ_1 and θ_2 and Eq. (8.7) will not always yield such an estimate, $\hat{\theta}$ should be set equal to θ_1 if $\hat{\theta} < u_1$ and $\hat{\theta}$ should be set equal to θ_2 if $\hat{\theta} > u_2$. The accuracy of this estimator is given in Fig. 8.7 for the case of $n = 2$ pulses per beamwidth. This estimation procedure can also be used to estimate the elevation angle of a target in multibeam systems where θ_1 and θ_2 are the elevation-pointing angles of adjacent beams.
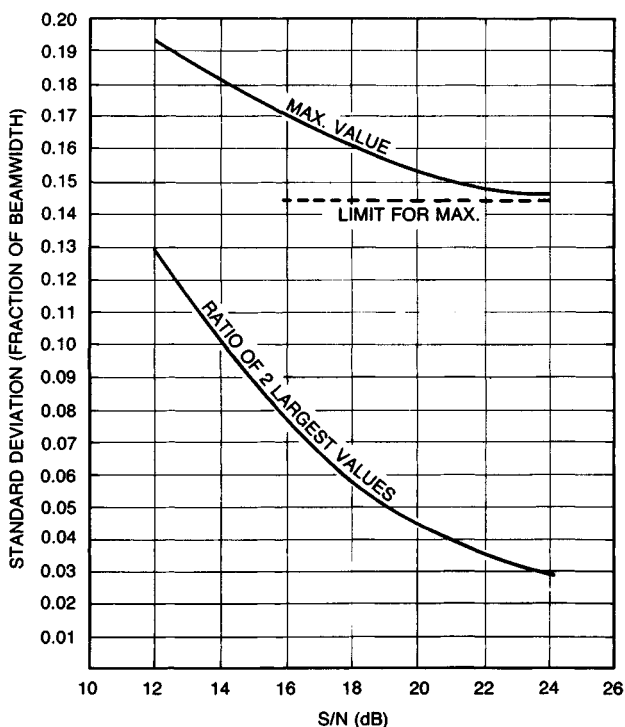
FIG. 8.7 Angular accuracy using two-pulse estimates.

Feedback Integrator. The amount of storage required can be reduced significantly by using a feedback integrator shown in Fig. 8.3b:

$$S_i = KS_{k-1} + x_i \tag{8.9}$$

For a feedback value of K, the effective number of pulses integrated M is $M = 1/(1 - K)$, and for optimal (maximum P_D) performance $M = 0.63 \, N$, where N is the number of pulses between the 3-dB antenna beamwidth.[11] The detection performance is given by the detection curves for M pulses with ABSF = 1.6 dB. Although the feedback integrator applies an exponential weighting into the past, its detection performance is only 1 dB less than that of the optimal integrator.[8] Unfortunately, difficulties are encountered when using the feedback integrator to estimate the azimuth position.[11] The threshold-crossing procedure yields estimates only 20 percent greater than the lower bound, but the bias is a function of S/N and must be estimated. On the other hand, the maximum value, though it has a constant bias, has estimates that are 100 percent greater than the lower bound. Furthermore, the exponential weighting function essentially destroys the radar antenna sidelobes. Because of these problems, the feedback integrator has limited utility.

Two-Pole Filter. The two-pole filter in Fig. 8.3c requires the storage of an intermediate calculation in addition to the integrated output and is described mathematically by

$$y_i = x_i - k_2 z_{i-1} \tag{8.10}$$

and
$$z_i = y_{i-1} + k_1 z_{i-1} \tag{8.11}$$

where x_i is the input, y_i is the intermediate calculation, z_i is the output, and k_1 and k_2 are the two feedback values. The values[12,13] which maximize P_D are given by

$$k_1 = 2 \exp\left(-\xi\omega_d\tau/\sqrt{1 - \xi^2}\right) \cos\left(\omega_d\tau\right) \tag{8.12}$$

and
$$k_2 = \exp\left(-2\xi\omega_d\tau/\sqrt{1 - \xi^2}\right) \tag{8.13}$$

where $\xi = 0.63$, $N\omega_d\tau = 2.2$, and N is the number of pulses between the 3-dB points of the antenna. With this rather simple device a weighting pattern similar to the antenna pattern can be obtained. The detection performance is within 0.15 dB of the optimal detector, and its angular estimates are about 20 percent greater than the Cramer-Rao lower bound. If the desired number of pulses integrated is changed (e.g., because of a change in the antenna rotation rate of the radar), it is only necessary to change the feedback values k_1 and k_2. The problems with this detector are that (1) it has rather high detector sidelobes, 15 to 20 dB, and (2) it is extremely sensitive to interference (i.e., the filter has a high gain resulting in a large output for a single sample that has a high value).

Binary Integrator.   The binary integrator is also known as the dual-threshold detector, M-out-of-N detector, or rank detector (see "Nonparametric Detectors" later in this section), and numerous individuals have studied it.[14–18] As shown in Fig. 8.3d, the input samples are quantized to 0 or 1, depending on whether or not they are less than a threshold T_1. The last N zeros and ones are summed and compared with a second threshold $T_2 = M$. For large N, the detection performance of this detector is approximately 2 dB less than the moving-window integrator because of the hard limiting of the data, and the angular estimation error is about 25 percent greater than the Cramer-Rao lower bound. Schwartz[16] showed that within 0.2 dB the optimal value of M for maximum P_D is given by

$$M = 1.5\sqrt{N} \tag{8.14}$$

when $10^{-10} < P_{fa} < 10^{-5}$ and $0.5 < P_D < 0.9$. The optimal value of P_n, the probability of exceeding T_1 when only noise is present, was calculated by Dillard[18] and is shown in Fig. 8.8. The corresponding threshold T_1 is

$$T_1 = \sigma(-2 \ln P_N)^{1/2} \tag{8.15}$$

A comparison of the optimal (best value of M) binary integrator with various other procedures is given in Figs. 8.9 and 8.10 for $P_D = 0.5$ and 0.9, respectively.

The binary integrator is used in many radars because (1) it is easily implemented, (2) it ignores interference spikes which cause trouble with integrators that directly use signal amplitude, and (3) it works extremely well when the noise has a non-Rayleigh density.[19] For $N = 3$, comparison of the optimal binary integrator (3 out of 3), another binary integration (2 out of 3), and the moving-window detector in log-normal interference (an example of a non-Rayleigh density) is shown in Fig. 8.11. The optimal binary integrator is much better than the moving-
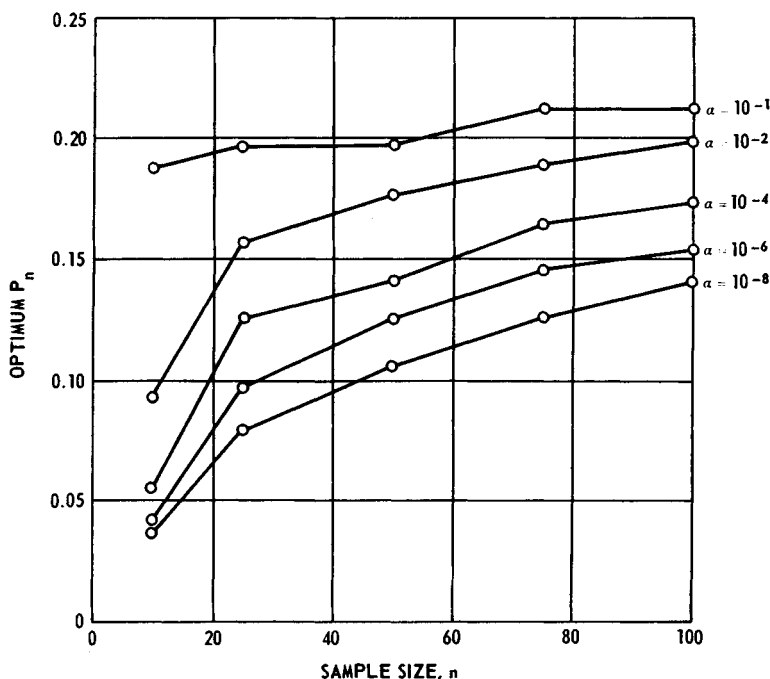
FIG. 8.8 Optimum values of P_N as a function of the sample size n and the probability of false alarm α; Ricean distribution with $S/N = 0$ dB per pulse. (*Copyright 1967, IEEE; from Ref. 18.*)

window integrator. The optimal values for log-normal interference were calculated by Schleher[19] and are $M = 3, 8$, and 25 and $N = 3, 10$, and 30, respectively.

A modified version of binary integration is sometimes used when there is a large number of pulses. It also has flexibility to integrate a different number of pulses. The modified binary moving window (MBMW) differs from the ordinary binary moving window (OBMW) by the introduction of a third threshold. When the second threshold is reached, one counts the number of consecutive pulses for which the second threshold is exceeded. When this number equals the third threshold, a target is declared. The performance of the MBMW and a comparison with the OBMW were given in Ref. 20. The major conclusion to be drawn is that the larger the value of N, the larger the difference in performance between the MBMW and OBMW detectors. For instance, with respect to the OBMW, the MBMW incurs losses of 0.15, 0.53, 1.80, and 2.45 dB for $N = 8, 16, 24$, and 32 pulses, respectively.

Batch Processor.   The batch processor (Fig. 8.3*e*) is very useful when a large number of pulses are in the 3-dB beamwidth. If KN pulses are in the 3-dB beamwidth, K pulses are summed (batched) and either a 0 or a 1 is declared, depending on whether or not the batch is less than a threshold T_1. The last N zeros and ones are summed and compared with a second threshold M. An alternative version of this detector is to put the batches through a moving-window detector.

The batch processor, like the binary integrator, is easily implemented, ignores interference spikes, and works extremely well when the noise has a non-Rayleigh
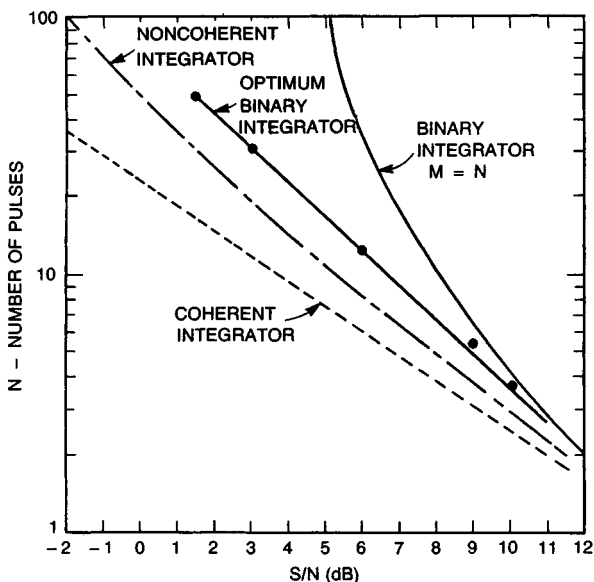
FIG. 8.9 Comparison of binary integrator (M out of N) with other integration methods ($P_{fa} = 10^{-10}$; $P_D = 0.5$). (*Copyright 1956, IEEE; after Ref. 16.*)
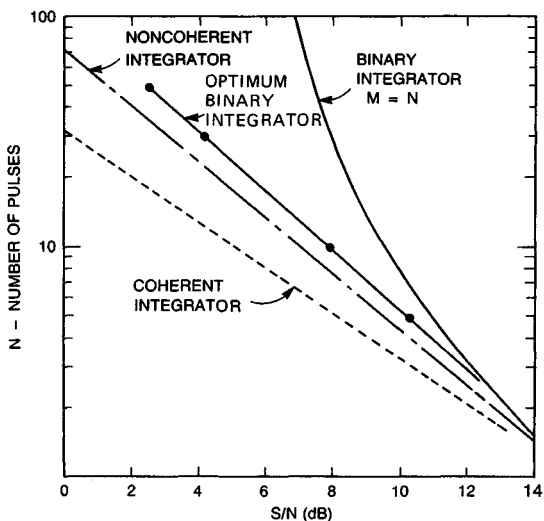
FIG. 8.10 Comparison of binary integrator (M out of N) with other integration methods ($P_{fa} = 10^{-10}$; $P_D = 0.90$). (*Copyright 1956, IEEE; after Ref. 16.*)
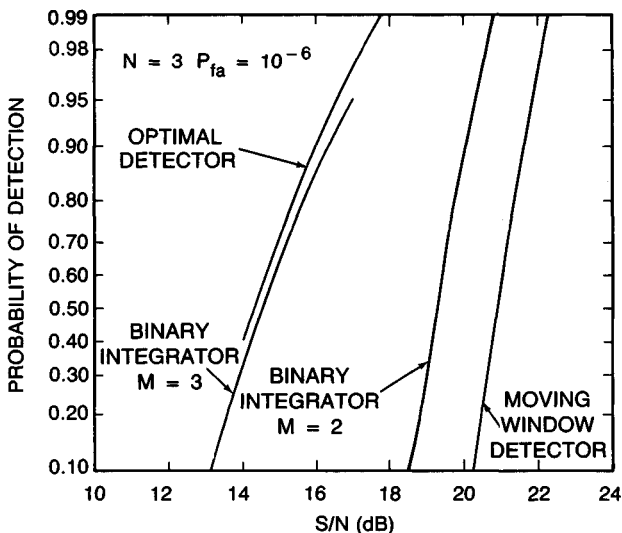
FIG. 8.11 Comparison of various detectors in log-normal ($\sigma = 6$ dB) interference ($N = 3$; $P_{fa} = 10^{-6}$). (*Copyright 1975, IEEE; after Ref. 19.*)

density. Furthermore, the batch processor requires less storage, detects better, and estimates angles more accurately than the binary integrator. For instance, if there were 80 pulses on target, one could batch 16 pulses, quantize this result to a 0 or a 1, and declare a target with a 3-out-of-5 (or 2-out-of-5) binary integrator. With an 8-bit analog-to-digital converter, the storage requirement per range cell is 17 bits (12 bits for the batch and 5 for the binary integrator) for the batch processor as opposed to 80 bits for the binary integrator and 640 bits for the moving window. The detection performance of the batch processor for a large number of pulses integrated is approximately 0.5 dB worse than the moving window. The batch processor has been successfully implemented by the Applied Physics Laboratory[21] of Johns Hopkins University. To obtain an accurate azimuth estimate $\hat{\theta}$, approximately 20 percent greater than the lower bound,

$$\hat{\theta} = \frac{\Sigma B_i \theta_i}{\Sigma B_i} \tag{8.16}$$

is used, where B_i is the batch amplitude and θ_i is the azimuth angle corresponding to the center of the batch.

False-Alarm Control. In the presence of clutter, if fixed thresholds are used with the previously discussed integrators, an enormous number of detections will occur and will saturate and disrupt the tracking computer associated with the radar system. Four important facts should be noted:

• A tracking system should be associated with the automatic detection system (the only exception is when one displays multiple scans of detections).

• The P_{fa} of the detector should be as high as possible without saturating the tracking computer.

- Random false alarms and unwanted targets (e.g., stationary targets) are not a problem if they are removed by the tracking computer.
- Scan-to-scan processing can be used to remove stationary point clutter or moving-target indication (MTI) clutter residues.

One can limit the number of false alarms with a fixed-threshold system by setting a very high threshold. Unfortunately, this would reduce target sensitivity in regions of low noise (clutter) return. Three main approaches—adaptive threshold, nonparametric detectors, and clutter maps—have been used to reduce the false-alarm problem. Adaptive thresholding and nonparametric detectors assume that the samples in the range cells surrounding the test cell (called *reference cells*) are independent and identically distributed. Furthermore, it is usually assumed that the time samples are independent. Both kinds of detectors test whether the test cell has a return sufficiently larger than the reference cells. Clutter maps allow variation in space, but the clutter must be stationary over several (typically 5 to 10) scans. Clutter maps store an average background level for each range-azimuth cell. A target is then declared in a range-azimuth cell if the new value exceeds the average background level by a specified amount.

Adaptive Thresholding. The basic assumption of the adaptive thresholding technique is that the probability density of the noise is known except for a few unknown parameters. The surrounding reference cells are then used to estimate the unknown parameters, and a threshold based on the estimated parameters is obtained. The simplest adaptive detector, shown in Fig. 8.12, is the cell-averaging CFAR (constant false-alarm rate) investigated by Finn and Johnson.[22] If the noise has a Rayleigh density, $p(x) = x \exp(-x^2/2\sigma^2)/\sigma^2$, only the parameter σ (σ^2 is the noise power) needs to be estimated, and the threshold is of the form $T = K\Sigma x_i = Kn\sqrt{\pi/2}\hat{\sigma}$, where $\hat{\sigma}$ is the estimate of σ. However, since T is set by an estimate $\hat{\sigma}$, it has some error and must be slightly larger than the threshold that one would use if σ were known exactly a priori. The raised threshold causes a loss in target sensitivity and is referred to as a CFAR loss. This loss has been
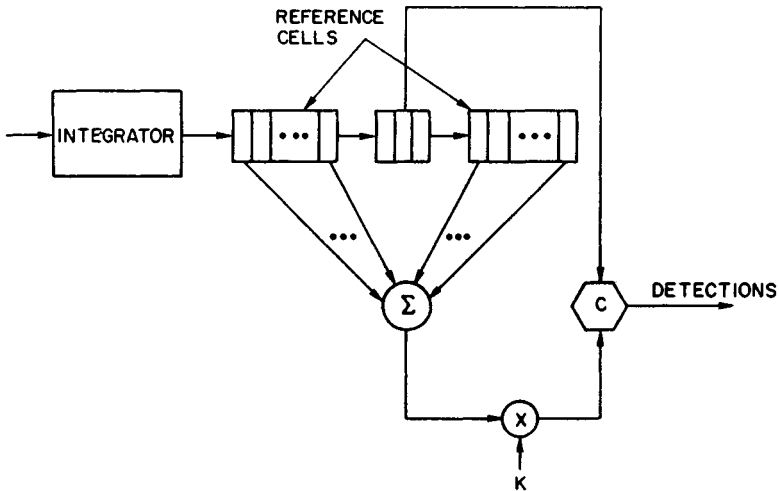
FIG. 8.12 Cell-averaging CFAR. The letter C indicates a comparison. (*From Ref. 7.*)

TABLE 8.1 CFAR Loss for $P_{fa} = 10^{-6}$ and $P_D = 0.9$*

Number of pulses integrated	Loss for various numbers of reference cells, dB					
	1	2	3	5	10	∞
1	. . .	15.3	7.7	3.5	0	
3	. . .	7.8	5.1	3.1	1.4	0
10	6.3	3.3	2.2	1.3	0.7	0
30	3.6	2.0	1.4	1.0	0.5	0
100	2.4	1.4	1.0	0.6	0.3	0

*After Ref. 23.

calculated[23] and is summarized in Table 8.1. As can be seen, for a small number of reference cells the loss is large because of the poor estimate of σ. Consequently, one would prefer to use a large number of reference cells. However, if one does this, the homogeneity assumption (i.e., all the reference cells are statistically similar) might be violated. A good rule of thumb is to use enough reference cells so that the CFAR loss is below 1 dB and at the same time not let the reference cells extend beyond 1 nmi on either side of the test cell. For a particular radar this might not be feasible.

If there is uncertainty about whether or not the noise is Rayleigh-distributed, it is better to threshold individual pulses and use a binary integrator as shown in Fig. 8.13. This detector is tolerant of variations in the noise density because by setting K to yield a 1 with probability 0.1, a $P_{fa} \approx 10^{-6}$ can be obtained by using a 7-out-of-9 detector. While noise may be non-Rayleigh, it will probably be very Rayleigh-like out to the tenth percentile. Furthermore, one can use feedback based on several
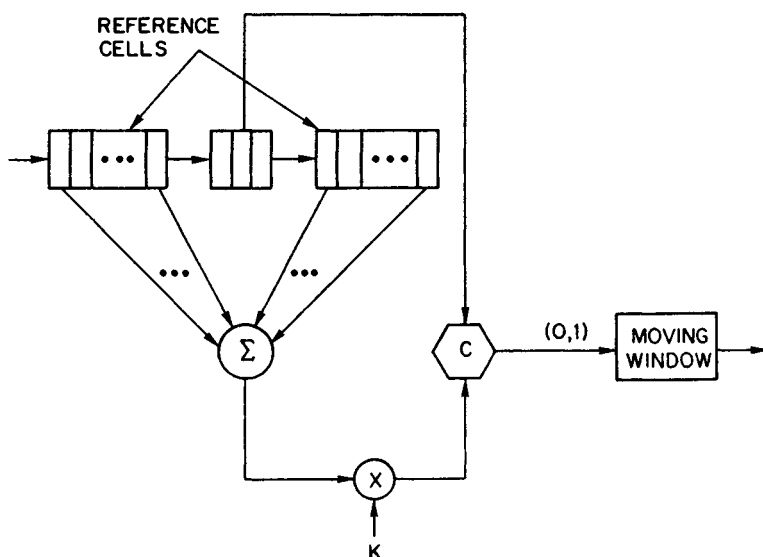
FIG. 8.13 Implementation of a binary integrator. The letter C indicates a comparison. (*From Ref. 7.*)

scans of data to control K in order to maintain a desired P_{fa} either on a scan or a sector basis. This demonstrates a general rule: to maintain a low P_{fa} in various environments, adaptive thresholding should be placed in front of the integrator.

If the noise power varies from pulse to pulse (as it would in jamming when frequency agility is employed), one must CFAR each pulse and then integrate. While the binary integrator performs this type of CFAR action, analysis[24,25] has shown that the ratio detector in Fig. 8.14 is a better detector. The ratio detector sums signal-to-noise ratios and is specified by

$$\sum_{i=1}^{n} \frac{x_i^2\,(j)}{\dfrac{1}{2m}\displaystyle\sum_{k=1}^{m}[x_i^2(j+1+k) + x_i^2(j-1-k)]} \tag{8.17}$$

where $x_i(j)$ is the ith envelope-detected pulse in the jth range cell and $2m$ is the number of reference cells. The denominator is the maximum-likelihood estimate of σ_i^2, the noise power per pulse. It will detect targets even though only a few returned pulses have a high signal-to-noise ratio. Unfortunately, this will also cause the ratio detector to declare false alarms in the presence of narrow-pulse interference. To reduce the number of false alarms when narrow-pulse interference is present, the individual power ratios can be soft-limited[25] to a small enough value so that interference will cause only a few false alarms. A comparison of the ratio detector with other commonly used detectors is shown in Figs. 8.15 and 8.16 for nonfluctuating and fluctuating targets. A typical performance in sidelobe jamming when the jamming level varies by 20 dB per pulse is shown in Fig. 8.17. By employing a second test to identify the presence of narrow-pulse interference, a detection performance approximately halfway between the limiting and nonlimiting ratio detectors can be obtained.
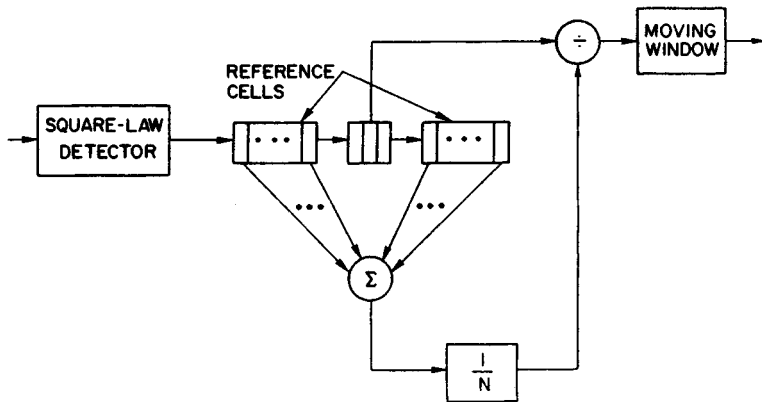
FIG. 8.14 Ratio detector. (*From Ref. 7.*)

If the noise samples are dependent in time or have a non-Rayleigh density such as the chi-square density or log-normal density, it is necessary to estimate two parameters and the adaptive detector is more complicated. Usually several pulses are integrated so that one can assume the integrated output has a gaussian probability density. Then the two parameters that must be estimated are the
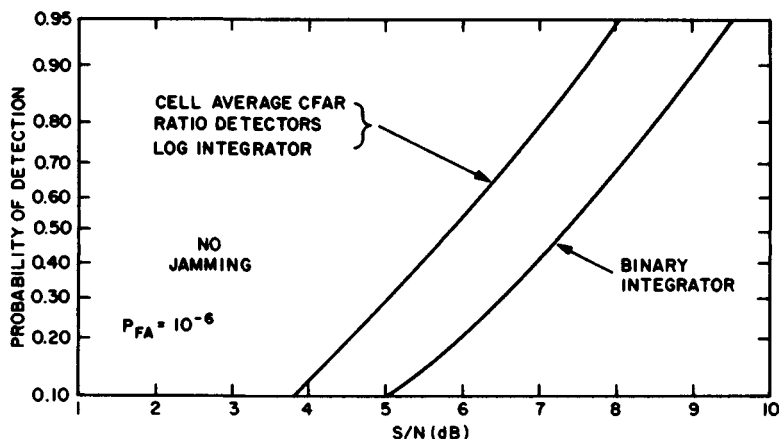
FIG. 8.15 Curves of probability of detection versus signal-to-noise ratio per pulse for the cell-averaging CFAR, ratio detectors, log integrator, and binary integrator: nonfluctuating target, $N = 6$, and probability of false alarm = 10^{-6}. *(From Ref. 25.)*
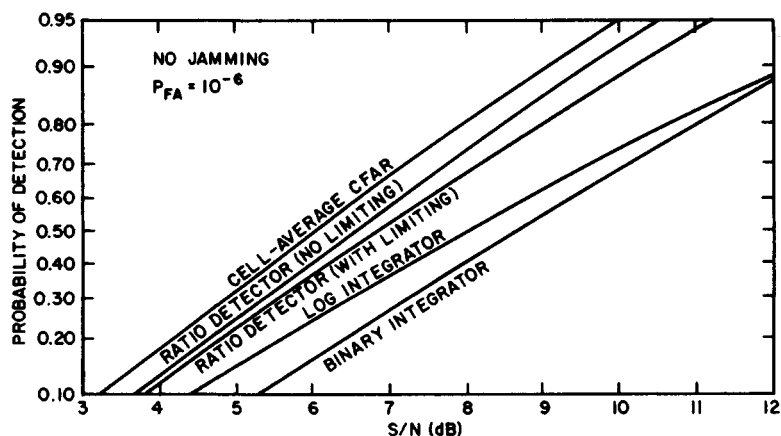
FIG. 8.16 Curves of probability of detection versus signal-to-noise ratio for the cell-averaging CFAR, ratio detectors, log integrator, and binary integrator: Rayleigh, pulse-to-pulse fluctuating target, $N = 6$, and probability of false alarm = 10^{-6}. *(From Ref. 25.)*

mean and the variance, and a threshold of the form $T = \hat{\mu} + K\hat{\sigma}$ is used. Though the mean is easily obtained in hardware, the usual estimate of the standard deviation

$$\hat{\sigma} = \left[\frac{1}{N}\Sigma(x_i - \bar{x})^2\right]^{1/2} \tag{8.18}$$

where

$$x = \frac{1}{N}\Sigma x_i \tag{8.19}$$
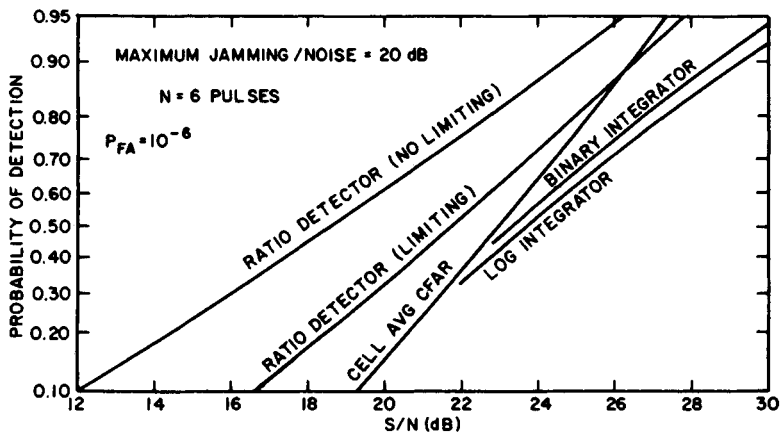
FIG. 8.17 Curves of probability of detection versus signal-to-noise ratio for the cell-averaging CFAR, ratio detectors, log integrator, and binary integrator: Rayleigh, pulse-to-pulse fluctuations, probability of false alarm = 10^{-6}, and maximum jamming-to-noise ratio = 20 dB. (*From Ref. 25.*)

is more difficult to implement. Consequently, the mean deviate defined by

$$\sigma = A\Sigma|x_i - \bar{x}| \tag{8.20}$$

is sometimes used because of its ease of implementation. Nothing can be done to the binary integrator to yield a low P_{fa} if the noise samples are correlated. Thus, it should not be used in this situation. However, if the correlation time is less than a batching interval, the batch processor will yield a low P_{fa} without modifications.

Target Suppression. Target suppression is the loss in detectability caused by other targets or clutter residues in the reference cells. Basically, there are two approaches to solving this problem: (1) remove large returns from the calculation of the threshold,[26–28] or (2) diminish the effects of large returns by either limiting or using log video. The technique that should be used is a function of the particular radar system and its environment.

Rickard and Dillard[27] proposed a class of detectors D_K, where the K largest samples are censored (removed) from the reference cells. A comparison of D_0 (no censoring) with D_1 and D_2 for a Swerling 2 target and a single square-law detected pulse is shown in Fig. 8.18, where N is the number of reference cells, β is the ratio of the power of the interfering target to the target in the test cell, and the bracketed pair (m, n) indicates the Swerling models of the target and the interfering target, respectively. As shown in Fig. 8.18, when one has an interfering target, the P_D does not approach 1 as S/N increases. Another approach[26] which censors samples in the reference cell if they exceed a threshold is briefly discussed in the subsection "Nonparametric Detectors."

Finn[28] investigated the problem of the reference cells spanning two continuous different "noise" fields (e.g., thermal noise, sea clutter, land clutter, etc.). On the basis of the samples, he estimated the statistical parameters of the two noise fields and the separation point between them. Then, only those reference cells which are in the noise field containing the test cell are used to calculate the adaptive threshold.
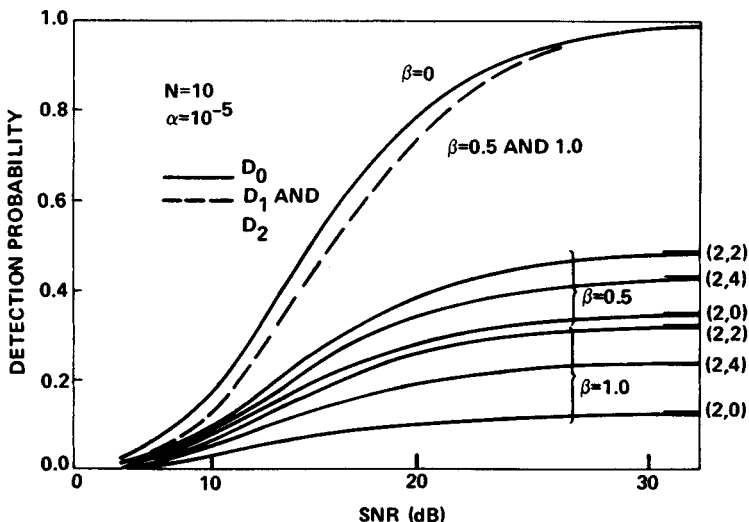
FIG. 8.18   Detection probability versus *SNR* for Swerling Case 2 primary target. (*Copyright 1977, IEEE; from Ref. 27.*)

An alternative approach for interfering targets is to use log video. By taking the log, large samples in the reference cells will have little effect on the threshold. The loss associated with using log video is 0.5 dB for 10 pulses integrated and 1.0 dB for 100 pulses integrated.[29] An implementation of the log CFAR[30] is shown in Fig. 8.19. In many systems the antilog shown in Fig. 8.19 is not taken. To maintain the same CFAR loss as for linear video, the number of reference cell $M_{\log}$ for the log CFAR should equal

$$M_{\log} = 1.65\, M_{\lin} - 0.65 \tag{8.21}$$

where $M_{\lin}$ is the number of reference cells for linear video. The effect of target suppression with log video is discussed later in this section (Table 8.2).
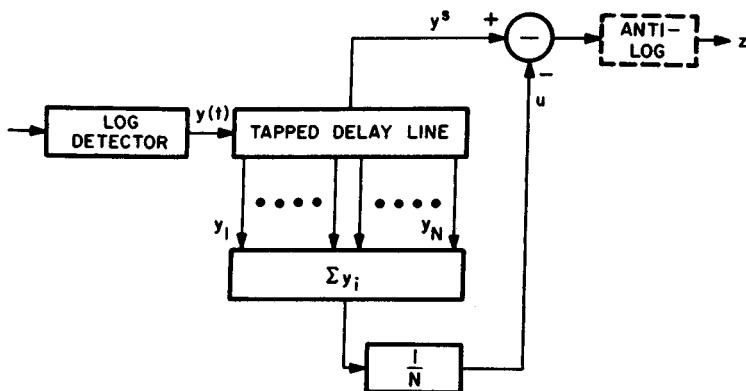
FIG. 8.19   Block diagram of cell-averaging log-CFAR receiver. (*Copyright 1972, IEEE; from Ref. 30.*)

Nonparametric Detectors.   Usually nonparametric detectors obtain CFAR by ranking the test sample with the reference cells.[31,32] Ranking means that one orders the samples from the smallest to the largest and replaces the smallest with rank 0, the next smallest with rank 1,..., and the largest with rank $n-1$. Under the hypothesis that all the samples are independent samples from an unknown density function, the test sample has equal probability of taking on any of the n values. For instance, referring to the ranker in Fig. 8.20, the test cell is compared with 15 of its neighbors. Since in the set of 16 samples, the test sample has equal probability of being the smallest sample (or equivalently any other rank), the probability that the test sample takes on values 0, 1,..., 15 is 1:16. A simple rank detector is constructed by comparing the rank with a threshold K and generating a 1 if the rank is larger, a 0 otherwise. The 0s and 1s are summed in a moving window. This detector incurs a CFAR loss of about 2 dB but achieves a fixed P_{fa} for any unknown noise density as long as the time samples are independent. This detector was incorporated into the ARTS-3A postprocessor used in conjunction with the Federal Aviation Administration airport surveillance radar (ASR). The major shortcoming of this detector is that it is fairly susceptible to target suppression (e.g., if a large target is in the reference cells, the test cell cannot receive the highest ranks).

If the time samples are correlated, the rank detector will not yield CFAR. A modified rank detector, called the modified generalized sign test (MGST),[26] maintains a low P_{fa} and is shown in Fig. 8.21. This detector can be divided into three parts: a ranker, an integrator (in this case a two-pole filter), and a threshold (decision process). A target is declared when the integrated output exceeds two thresholds. The first threshold is fixed (equals $\mu + T_1/K$ in Fig. 8.21) and yields $P_{fa}=10^{-6}$ when the reference cells are independent and identically distributed. The second threshold is
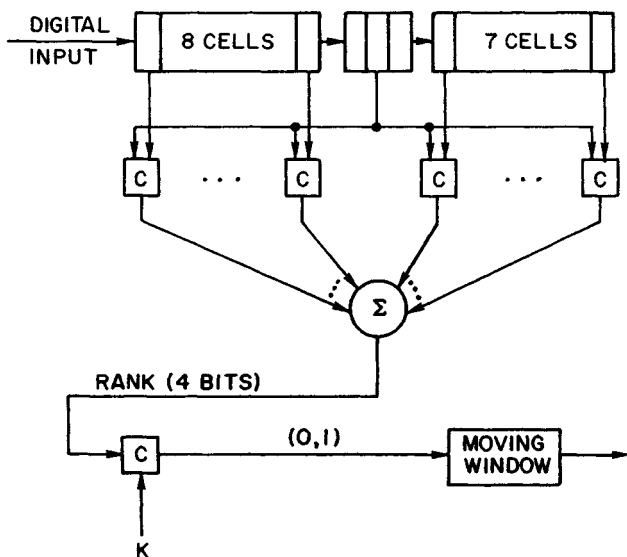
FIG. 8.20   Rank detector: output of a comparator C is either a zero or a one. (*From Ref. 7.*)

adaptive and maintains a low P_{fa} when the reference samples are correlated. The device estimates the standard deviation of the correlated samples with the mean deviator estimator, where extraneous targets in the reference cells have been excluded from the estimate by use of a preliminary threshold T_2.

The rank and MGST detectors are basically two-sample detectors. They decide that a target is present if the ranks of the test cell are significantly greater than the ranks of the reference cells. Target suppression occurs at all interfaces (e.g., land, sea) where the homogeneity assumption is violated. However, some tests exist, such as the Spearman Rho and Kendall Tau tests,[33] that depend only on the test cell. These tests use the fact that as the antenna beam sweeps by a point target, the signal return increases and then decreases. Thus, for the test cell the ranks should follow a pattern, first increasing and then decreasing. Although these detectors do not require reference cells and hence have the useful property of not requiring homogeneity, they are not generally used because of the large CFAR loss that occurs for moderate sample sizes. For instance, the CFAR losses are approximately 10 dB for 16 pulses on target and 6 dB for 32 pulses on target.[33]

The basic disadvantages of all nonparametric detectors are that (1) they have relatively large CFAR losses, (2) they have problems with correlated samples, and (3) one loses amplitude information, which can be a very important discriminant between target and clutter.[34] For example, a large return ($\sigma \hbar 1000 \text{ m}^2$) in a clutter area is probably just clutter breakthrough. See "Contact Entry Logic" in Sec. 8.3.

Clutter Mapping. A clutter map uses adaptive thresholding where the threshold is calculated from the return in the test cell on previous scans rather than from the surrounding reference cells on the same scan. This technique has
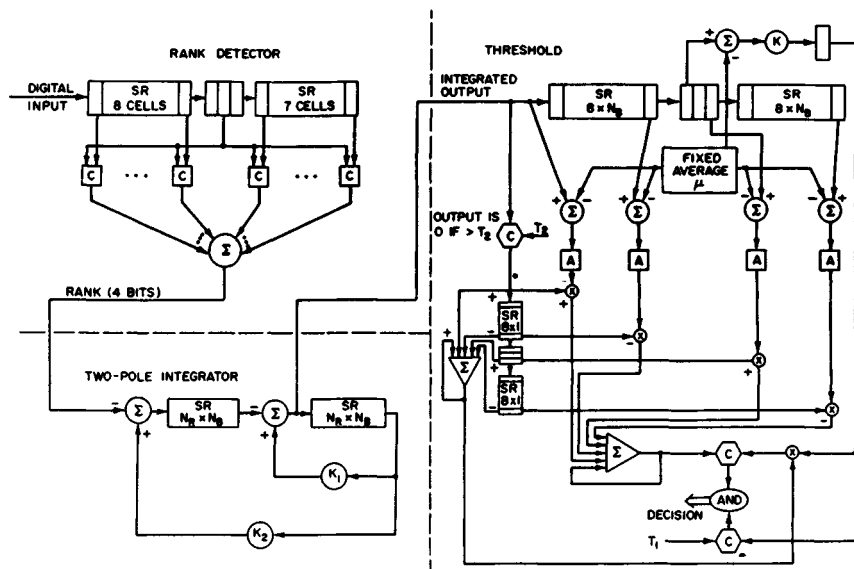
FIG. 8.21 Modified generalized sign test processor. (*Copyright 1974, IEEE; from Ref. 26.*)

the advantage that for essentially stationary environments (e.g., land-based radar against ground clutter), the radar has interclutter visibility—it can see between large clutter returns. Lincoln Laboratory[35] in its moving-target detector (MTD) used a clutter map for the zero-doppler filter very effectively. The decision threshold T for the ith cell is

$$T = A \, S_{i-1} \tag{8.22}$$

where $$S_i = K \, S_{i-1} + X_i \tag{8.23}$$

S_i is the average background level, X_i is the return in the ith cell, K is the feedback value which determines the map time constant, and A is the constant which determines the false-alarm rate. In the MTD used for ASR application K is 7:8, which effectively averages the last eight scans. The main utility of clutter maps is with fixed-frequency land-based radars. While clutter maps can be used with frequency-agile radars and on moving platforms, they are not nearly as effective in these environments.

**Target Resolution.** In automatic detection systems, a single large target will probably be detected many times, e.g., in adjacent range cells, azimuth beams, and elevation beams. Therefore, automatic detection systems have algorithms for merging the individual detections into a single centroided detection. Most algorithms have been designed so that they will rarely split a single target into two targets. This procedure results in poor range resolution capability. A merging algorithm[36] often used is the adjacent-detection merging algorithm, which decides whether a new detection is adjacent to any of the previously determined sets of adjacent detections. If the new detection is adjacent to any detection in the set of adjacent detections, it is added to the set. Two detections are adjacent if two of their three parameters (range, azimuth, and elevation) are the same and the other parameter differs by the resolution element: range cell ΔR, azimuth beamwidth θ, or elevation beamwidth γ.

A simulation[36] was run to compare the resolving capability of three common detection procedures: linear detector with $T = \hat{\mu} + A\hat{\sigma}$, linear detector with $T = B\hat{\mu}$, and log detector with $T = C + \hat{\mu}$. The constants A, B, and C are used to obtain the same P_{fa} for all detectors. The estimates $\hat{\mu}$ and $\hat{\sigma}$ of μ and σ were obtained from either (1) all the reference cells or (2) the leading or lagging half of the reference cells, choosing the half with the lower mean value. The simulation involved two targets separated by 1.5, 2.0, 2.5, or 3.0 range cells and a third target 7.0 range cells from the first target. When the two closely spaced targets were well separated, either 2.5 or 3.0 range cells apart, the probability of detecting both targets (P_{D2}) was < 0.05 for the linear detector with $T = \hat{\mu} + A\hat{\sigma}$; $0.15 < P_{D2} < 0.75$ for the linear detector with $T = B\hat{\mu}$; and $P_{D2} > 0.9$ for the log detector. A second simulation, involving only two targets, investigated the effect of target suppression on log video, and the results are summarized in Table 8.2. One notes an improved performance for small S/N (10 to 13 dB) when one calculates the threshold using only the half of the reference cells with the lower mean value. The resolution capability of the log detector which uses only the half of the reference cells with the lower mean is shown in Fig. 8.22. The probability of resolving two equal-amplitude targets does not rise above 0.9 until they are separated in range by 2.5 pulse widths.

By assuming that the target is small with respect to the pulse width and that the pulse shape is known, the resolution capability can be improved by fitting

TABLE 8.2 Probability of Detecting with Log Video Two Targets Separated by 1.5, 2.0, 2.5, or 3.0 Range Cells*

Thresholding technique	Target separation	S/N of target no. 2				
		10	13	20	30	40
All reference cells	1.5	0.0	0.04	0.0	0.00	0.00
	2.0	0.0	0.22	0.54	0.14	0.10
	2.5	0.04	0.24	0.94	0.62	0.32
	3.0	0.0	0.24	0.88	0.92	0.76
Reference cells with minimum mean value	1.5	0.0	0.0	0.00	0.0	0.02
	2.0	0.10	0.32	0.44	0.12	0.04
	2.5	0.18	0.58	0.98	0.46	0.28
	3.0	0.22	0.66	0.98	0.82	0.74

*S/N of target 1 is 20 dB. S/N of target 2 is 10, 13, 20, 30, or 40 dB. After Ref. 36.

the known pulse shape to the received data and comparing the residue square error with a threshold.[37] If only one target is present, the residue should be only noise and hence should be small. If two or more targets are present, the residue will contain signal from the remaining targets and should be large. The results of resolving two targets with $S/N = 20$ dB are shown in Fig. 8.23. These targets can be resolved at a resolution probability of 0.9 at separations
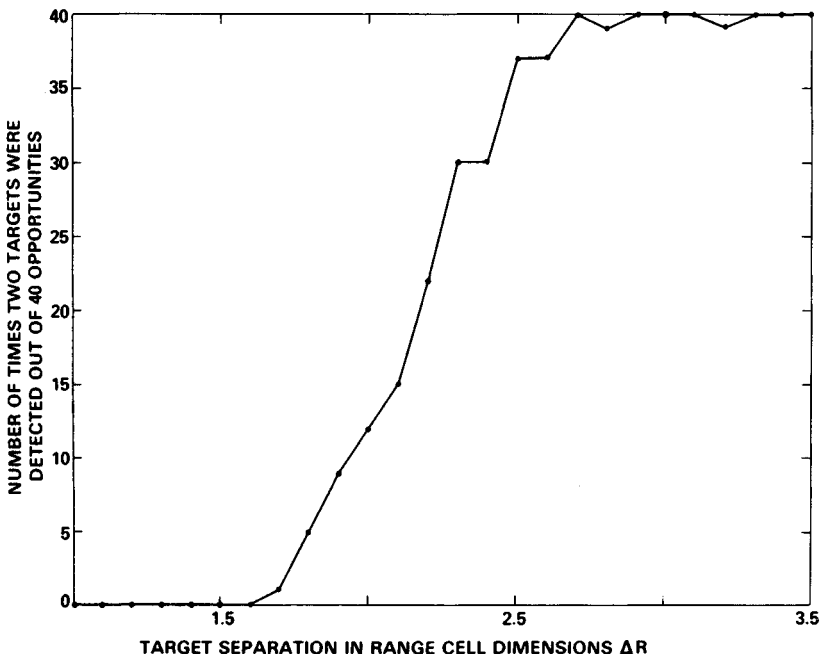
FIG. 8.22 Resolution capability of a log detector which uses the half of the reference cells with the lower mean. (*Copyright 1978, IEEE; from Ref. 36.*)

varying between one-fourth and three-fourths of a pulse width, depending on the relative phase difference between the two targets. Furthermore, this result can be improved further by processing multiple pulses.
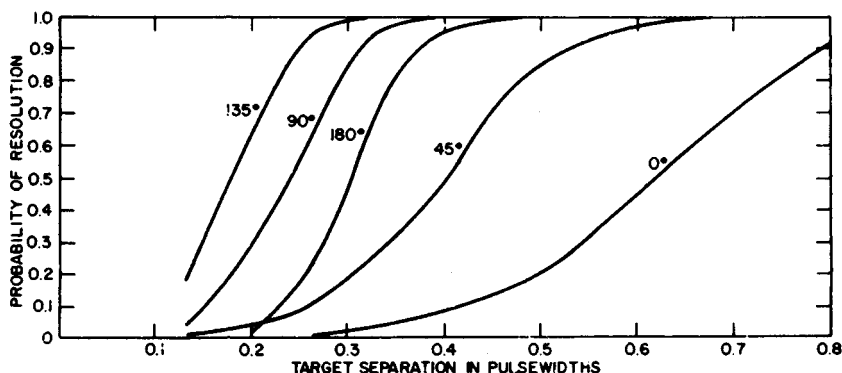
FIG. 8.23 Probability of resolution as a function of range separation: sampling rate $\Delta R = 1.5$ samples per pulse width; target strengths—nonfluctuating, $A_1 = A_2 = 20$ dB; phase differences = 0°, 45°, 90°, 135°, and 180°. (*Copyright 1984, IEEE; from Ref. 37.*)

Detection Summary.   When only 2 to 4 samples (pulses) are available, a binary integrator should be used to avoid false alarms due to interference. When a moderate number of pulses (5 to 16) are available, a binary integrator, a rank detector, or a moving-window integrator should be used. If the number of pulses is large (greater than 20), a batch processor or a two-pole filter should be used. If the samples are independent, a one-parameter (mean) threshold can be used. If the samples are dependent, one can either use a two-parameter (mean and variance) threshold or adapt a one-parameter threshold on a sector basis. These rules should serve only as a general guideline. It is *highly recommended* that before a detector is chosen the radar video from the environment of interest be collected and analyzed and that various detection processes be simulated on a computer and tested against the recorded data.

8.3 AUTOMATIC TRACKING

Track-while-scan (TWS) systems are tracking systems for surveillance radars whose nominal scan time (revisit time) is from 4 to 12 s for aircraft targets. If the probability of detection (P_D) per scan is high, if accurate target location measurements are made, if the target density is low, and if there are only a few false alarms, the design of the correlation logic (i.e., associating detections with tracks) and tracking filter (i.e., filter for smoothing and predicting track positions) is straightforward. However, in a realistic radar environment these assumptions are seldom valid, and the design of the automatic tracking system is complicated. In actual situations one encounters target fades (changes in signal strength due to multipath propagation, blind speeds, and atmospheric conditions), false alarms (due to noise, clutter, interference, and jamming), and poor radar parameter estimates (due to noise, unstabilized antennas, unresolved targets, target splits,

CHAPTER 9
ELECTRONIC COUNTER-COUNTERMEASURES

A. Farina
Radar Department
Selenia S.p.A.

9.1 INTRODUCTION

Since World War II both radar and electronic warfare (EW) have achieved a very high state of performance.[1,2] Modern military forces depend heavily on electromagnetic systems for surveillance, weapon control, communication, and navigation. Electronic countermeasures (ECM) are likely to be taken by hostile forces to degrade the effectiveness of electromagnetic systems. As a direct consequence, electromagnetic systems are more and more frequently equipped with so-called electronic counter-countermeasures (ECCM) to ensure effective use of the electromagnetic spectrum despite an enemy's use of EW actions.

This chapter is devoted to the description of the ECCM techniques and design principles to be used in radar systems when they are subject to an ECM threat. Section 9.2 starts with a recall of the definitions pertaining to EW and ECCM. The topic of radar signals interception by EW devices is introduced in Sec. 9.3; the first strategy to be adopted by radar designers is to try to avoid interception by the opponent electronic devices. Section 9.4 is dedicated entirely to the analysis of the major ECM techniques and strategies. It is important to understand the ECM threat to a radar system in order to be able to efficiently react to it. To facilitate the description of the crowded family of ECCM techniques (Secs. 9.6 through 9.10), a classification is attempted in Sec. 9.5. Then, the techniques are introduced according to their use in the various sections of radar, namely, antenna, transmitter, receiver, and signal processing. A key role is also played by those ECCM techniques which cannot be classified as electronic, such as human factors, methods of radar operation, and radar deployment tactics (Sec. 9.10).

The ensuing Sec. 9.11 shows the application of the aforementioned techniques to the two most common radar families, namely, surveillance and tracking radars. The main design principles (e.g., selection of transmitter power, frequency, and antenna gain) as dictated by the ECM threat are also discussed in some detail.

The chapter ends with an approach to the problem of evaluating the efficacy of ECCM and ECM techniques (Sec. 9.12). There is a lack of theory to properly

quantify the endless battle between ECCM and ECM techniques. Nevertheless, a commonly adopted approach to determine the ECM effect on a radar system is based on evaluation of the radar range under jamming conditions. The advantage of using specific ECCM techniques can also be taken into account by calculating the radar range recovery.

9.2 TERMINOLOGY

Electronic warfare is defined as a military action involving the use of electromagnetic energy to determine, exploit, reduce, or prevent radar use of the electromagnetic spectrum.[3-6] EW is organized into two major categories: electronic warfare support measures (ESM) and electronic countermeasures (ECM). Basically, the EW community takes as its job the degradation of radar capability. The radar community takes as its job the successful application of radar in spite of what the EW community does; the goal is pursued by means of ECCM techniques. The definitions of ESM, ECM, and ECCM are listed below.[3,6,7]

ESM is that division of electronic warfare involving actions taken to search for, intercept, locate, record, and analyze radiated electromagnetic energy for the purpose of exploiting such radiations in the support of military operations. Thus, electronic warfare support measures provide a source of electronic warfare information required to conduct electronic countermeasures, threat detection, warning, and avoidance. ECM is that division of electronic warfare involving actions taken to prevent or reduce a radar's effective use of the electromagnetic spectrum. ECCM comprises those radar actions taken to ensure effective use of the electromagnetic spectrum despite the enemy's use of electronic warfare.

The topic of EW is extremely rich in terms, some of which are also in general use in other electronic fields. A complete glossary of terms in use in the ECM and ECCM fields is found in the literature.[3,6,8]

9.3 ELECTRONIC WARFARE SUPPORT MEASURES

ESM is based on the use of intercept or warning receivers and relies heavily on a previously compiled directory of both tactical and strategic electronic intelligence (ELINT).[4,9,10] ESM is entirely passive, being confined to identification and location of radiated signals. Radar interception, which is of particular interest in this section, is based on the information gleaned from analysis of the signals transmitted by radar systems. The scenario in which ESM should operate is generally crowded with pulsed radar signals: figures of 500,000 to 1 million pulses per second (pps) are frequently quoted in the literature.[4] The train of interleaved pulses is processed in the ESM receiver to identify for each pulse the center frequency, amplitude, pulse width, time of arrival (TOA), and bearing. This information is then input to a pulse-sort processor which deinterleaves the pulses into the pulse repetition interval (PRI) appropriate to each emitter. Further comparison against a store of known radar types permits the generation of an emitter list classified with its tactical

value. The ESM receiver is used to control the deployment and operation of ECM; the link between ESM and ECM is often automatic.

A single received radar pulse is characterized by a number of measurable parameters. The availability, resolution, and accuracy of these measurements must all be taken into account when designing the deinterleaving system because the approach used depends on the parameter data set available. Obviously, the better the resolution and accuracy of any parameter measurement, the more efficiently the pulse-sort processor can carry out its task. However, there are limitations on the measurement process from outside the ESM system (e.g., multipath), from inside the system (e.g., timing constraints, dead time during reception), and from cost-effectiveness considerations. Angle of arrival is probably the most important sorting parameter available to the deinterleaving process since the target bearing does not vary from pulse to pulse. A rotating directional antenna could be used for direction finding (DF); however, an interferometric system with more than one antenna is preferred because the probability of interception is higher than with the system having only one antenna.

The carrier frequency is the next most important pulse parameter for deinterleaving. A common method of frequency measurement is to use a scanning superheterodyne receiver that has the advantage of high sensitivity and good frequency resolution.[4] Unfortunately, this type of receiver has a poor probability of intercept for the same reasons as the rotating bearing measurement system. The situation is much worse if the emitter is also frequency-agile (random variation) or frequency-hopping (systematic variation). One method of overcoming this problem would be to use banks of contiguous receiver channels. This approach is today feasible owing to the availability of accurate surface acoustic-wave (SAW) filters and the integrated optic spectrum analyzer which utilizes the Bragg refraction of an optical guided beam by an SAW to perform spectral analysis.[4] Pulse width is an unreliable sorting parameter because of the high degree of corruption resulting from multipath transmission. Multipath effects can severely distort the pulse envelope, for example, by creating a long tail to the pulse and even displacing the position of the peak.

The TOA of the pulse can be taken as the instant that a threshold is crossed, but in the presence of noise and distortion this becomes a very variable measurement. Nevertheless, the TOA is used for deriving the PRI of the radar. The amplitude of the pulse is taken as the peak value. Dynamic-range considerations must take into account at least some three orders of magnitude for range variation and three orders of magnitude for scan pattern variations. In practice, 60-dB instantaneous dynamic range sounds like a minimum value; in many applications it should be larger. The amplitude measurement is used (along with TOA) for deriving the scan pattern of the emitter.[4]

Radar intercept receivers are implemented at varying levels of complexity. The simplest is the radar warning receiver (RWR), which in an airborne installation advises of the presence of threats such as a missile radar, supplying the relative bearing on a cockpit-based display. It is an unsophisticated low-sensitivity equipment which is preset to cover the bandwidth of expected threats, and it exploits the range advantage to indicate the threat before it comes into firing range. Receivers then increase in complexity through tactical ESM to the full ELINT (intelligence-gathering) capability. The specification of an ideal ELINT receiver for today's applications demands an instantaneous frequency coverage of 0.01 to 40 GHz, a sensitivity of better than −60 dBm, an instantaneous dynamic range greater than 60 dB, and a frequency resolution of 1 to 5 MHz. A diversity of signals, such as pulsed, CW, frequency-agile, PRI-agile, and intrapulse-modulated

(chirp, multiphase-shift-keyed, etc.), must all be accommodated with a high probability of intercept (POI) and a low false-alarm rate (FAR).[4]

The range at which a radar emission is detected by an RWR depends primarily on the sensitivity of the receiver and the radiated power of the victim radar. The calculation of the warning range can be obtained by the basic *one-way beacon equation*, which provides the signal-to-noise ratio at the RWR:

$$\left(\frac{S}{N}\right)_{\text{at RWR}} = \left(\frac{P}{4\pi R^2}\right)G_t\left(\frac{G_r\lambda^2}{4\pi}\right)\left(\frac{1}{kT_SB}\right)\frac{1}{L} \qquad (9.1)$$

where P is the radar radiated power, R is the range from the RWR to the radar, G_t is the transmitting-antenna gain of the radar, G_r is the receiving-antenna gain of the RWR, λ is the radar wavelength, the quantity kT_SB is the total system noise power of the RWR, and L is the losses.

Equation (9.1) is the basis of performance calculation for an RWR. It is noted that the RWR detection performance is inversely proportional to R^2 rather than to R^4 of the radar target detection equation. For this reason, the RWR can detect a radiating radar at distances far beyond those of radar's own target detection capability. The radar-versus-interceptor problem is a battle in which the radar's advantage lies in the use of matched filtering, which cannot be duplicated by the interceptor (it does not know the exact radar waveform), while the interceptor's advantage lies in the fundamental R^2 advantage of one-way versus two-way radar propagation.[10]

9.4 ELECTRONIC COUNTERMEASURES

The objectives of an ECM system are to deny information (detection, position, track initiation, track update, and classification of one or more targets) that the radar seeks or to surround desired radar echoes with so many false targets that the true information cannot be extracted.

ECM tactics and techniques may be classified in a number of ways, i.e., by main purpose, whether active or passive, by deployment-employment, by platform, by victim radar, or by a combination of them.[8,11] An encyclopedia of ECM tactics and techniques can be found in the literature.[8,12] Here it is intended to limit description to the most common types of ECM.

ECM includes both jamming and deception. *Jamming* is the intentional and deliberate transmission or retransmission of amplitude, frequency, phase, or otherwise modulated intermittent, CW, or noiselike signals for the purpose of interfering with, disturbing, exploiting, deceiving, masking, or otherwise degrading the reception of other signals that are used by radar systems.[8] A jammer is any ECM device that transmits a signal of any duty cycle for the sole or partial purpose of jamming a radar system.[8]

Radio signals by special transmitters intended for interfering with or precluding the normal operation of a victim radar system are called *active jamming*. They produce at the input of a victim system a background which impedes the detection and recognition of useful signals and determination of their parameters. The most common forms of active noise jamming are spot and barrage noises. Spot noise is used when the center frequency and bandwidth of the victim system

to be jammed are known and confined to a narrow band. However, many radars are frequency-agile over a wide band as an ECCM against spot jamming. If the rate of frequency agility is slow enough, the jammer can follow the frequency changes and maintain the effect of spot jamming. Barrage or broadband jamming is simultaneously radiated across the entire band of the radar spectrum of interest. This method is used against frequency-agile systems whose rates are too fast to follow or when the victim's frequency parameters are imprecisely known.

Jammer *size* is characterized by the *effective radiated power*; ERP = $G_j P_j$, where G_j is the transmit antenna gain of the jammer and P_j is the jammer power.

Passive ECM is synonymous with chaff, decoys, and other reflectors which require no prime power. The chaff is made of elemental passive reflectors or absorbers which can be floated or otherwise suspended in the atmosphere or exoatmosphere for the purpose of confusing, screening, or otherwise adversely affecting the victim electronic system. Examples are metal foils, metal-coated dielectrics (aluminum, silver, or zinc over fiberglass or nylon being the most common), string balls, rope, and semiconductors.[8] The basic properties of chaff are effective scatter area, the character and time of development of a chaff cloud, the spectra of the signals reflected by the cloud, and the width of the band that conceals the target.[4,12–14] Chaff consists of dipoles cut to approximately a half wavelength of the radar frequency. It is usually packaged in cartridges which contain a broad range of dipole lengths designed to be effective over a wide frequency band. From a radar viewpoint, the properties of chaff are very similar to those of weather clutter, except that its broadband frequency capability extends down to VHF. The mean doppler frequency of the chaff spectrum is determined by the mean wind velocity, while the spectrum spread is determined by wind turbulence and a shearing effect due to different wind velocities as a function of altitude.[12]

Decoys, which are another type of passive ECM, are a class of physically small radar targets whose radar cross sections are generally enhanced by using reflectors or a Luneburg lens to simulate fighter or bomber aircraft. The objective of decoys is to cause a dilution of the assets of the defensive system, thereby increasing the survivability of the penetrating aircraft.

The other major type of active jammer is deceptive ECM (DECM). *Deception* is the intentional and deliberate transmission or retransmission of amplitude, frequency, phase, or otherwise modulated intermittent or continuous-wave (CW) signals for the purpose of misleading in the interpretation or use of information by electronic systems.[8] The categories of deception are manipulative and imitative. *Manipulative* implies the alteration of friendly electromagnetic signals to accomplish deception, while *imitative* consists of introducing radiation into radar channels which imitates a hostile emission. DECM is also divided into *transponders* and *repeaters*.[12] Transponders generate noncoherent signals which emulate the temporal characteristics of the actual radar return. Repeaters generate coherent returns which attempt to emulate the amplitude, frequency, and temporal characteristics of the actual radar return. Repeaters usually require some form of memory for microwave signals to allow anticipatory returns to be generated; this is usually implemented by using a microwave acoustic memory or a digital RF memory (DRFM).[12]

The most common type of deception jammer is the range-gate stealer, whose function is to pull the radar tracking gate from the target position through the introduction of a false target into the radar's range-tracking circuits. A repeater jammer sends back an amplified version of the signal received from the radar. The deception jammer signal, being stronger than the radar's return signal, cap-

tures the range-tracking circuits. The deception signal is then progressively delayed in the jammer by using an RF memory, thereby "walking" the range gate off the actual target (range-gate pull-off, or RGPO, technique). When the range gate is sufficiently removed from the actual target, the deception jammer is turned off, forcing the tracking radar into a target reacquisition mode.[12]

Another DECM technique is called *inverse-gain jamming*; it is used to capture the angle-tracking circuits of a conical-scan tracking radar.[8] This technique repeats a replica of the received signal with an induced amplitude modulation which is the inverse of the victim radar's combined transmitting and receiving antenna scan patterns. Against a conically scanning tracking radar, an inverse-gain repeater jammer has the effect of causing positive feedback, which pushes the tracking-radar antenna away from the target rather than toward the target. Inverse-gain jamming and RGPO are combined in many cases to counter conical-scan tracking radars.[12]

A different form of DECM used against the main beam of surveillance radar attempts to cover the target's skin return with a wide pulse in order to confuse the radar's signal-processing circuitry into suppressing the actual target return.

In the deployment-employment of ECM, five classes can be singled out.[12] In the standoff jammer (SOJ) case, the jamming platform remains close to but outside the lethal range of enemy weapon systems and jams these systems to protect the attacking vehicles. Standoff ECM systems employ high-power noise jamming which must penetrate through the radar antenna receiving sidelobes at long ranges. *Escort jamming* is another ECM tactic in which the jamming platform accompanies the strike vehicles and jams radars to protect the strike vehicles.

Mutual-support, or *cooperative*, ECM involves the coordinated conduct of ECM by combat elements against acquisition and weapon control radars. One advantage of mutual-support jamming is the greater ERP available from a collection of platforms in contrast with a single platform. However, the real value of mutual-support jamming is in the coordinated tactics which can be employed. A favorite tactic employed against tracking radars, for example, is to switch between jammers located on separate aircraft within the radar's beamwidth. This blinking has the effect of introducing artificial glint into the radar tracking circuits, which, if introduced at the proper rate (typically 0.1 to 10 Hz), can cause the radar to break angle track. In addition, blinking has the desirable effect of confusing radiation homing missiles which might be directed against the jammer radiations.[12]

A *self-screening jammer* (SSJ) is used to protect the carrying vehicle. This situation stresses the capability of an ECM system relating to its power, signal-processing, and ESM capabilities.

Stand-forward jamming is an ECM tactic in which the jamming platform is located between the weapon systems and the strike vehicles and jams the radars to protect the strike vehicles. The stand-forward jammer is usually within the lethal range of defensive weapon systems for a considerable time. Therefore, only the use of relatively low-cost remotely piloted vehicles (RPVs) might be practical. RPVs can assist strike aircraft or missiles in penetrating radar-defended areas by jamming, ejecting chaff, dropping expendable jammers or decoys, acting as decoys themselves, and performing other related ECM tasks.

According to the platform, the jammer can be classified as space-borne, airborne, missile-borne, based on the ground, or based on the sea surface.

A special class of missile-borne threat is the antiradiation missile (ARM), having the objective of homing on and destroying the victim radar. The sorting and acquisition of radar signals is preliminarily made by an ESM system; afterwards it cues the ARM, which continues homing on the victim radar by means of its own

antenna, receiver, and signal processor. Acquisition depends on the direction of arrival, operating band, carrier frequency, pulse width, PRI, scan rate, and other parameters of the victim radar. An ARM homes on the continuous radiation from the radar sidelobes or on the flash of energy from the main beam. ARM benefits from the one-way-only radar signal attenuation. However, ARM receiver sensitivity is affected by mismatching losses, and accuracy in locating the victim radar is affected by the limited dimension of the ARM antenna.

9.5 OBJECTIVES AND TAXONOMY OF ECCM TECHNIQUES

The primary objective of ECCM techniques when applied to a radar system is to allow the accomplishment of the radar intended mission while countering the effects of the enemy's ECM. In greater detail, the benefits of using ECCM techniques may be summarized as follows: (1) prevention of radar saturation, (2) enhancement of the signal-to-jamming ratio, (3) discrimination of directional interference, (4) rejection of false targets, (5) maintenance of target tracks, (6) counteraction of ESM, and (7) radar system survivability.[3]

There are two broad classes of ECCM: (1) electronic techniques (Secs. 9.6–9.9) and (2) operational doctrines (Sec. 9.10). Specific electronic techniques take place in the main radar subsystems, namely, the antenna, transmitter, receiver, and signal processor. Suitable blending of these ECCM techniques can be implemented in the surveillance and tracking radars, as discussed in Sec. 9.11.

The ensuing description is limited to the major ECCM techniques; the reader should be aware that an alphabetically listed collection of 150 ECCM techniques and an encyclopedia of ECCM tactics and techniques can be found in the literature.[3,15] Many other references describe the ECCM problem, among which Refs. 13, 16, and 17 are worth noting.

9.6 ANTENNA-RELATED ECCM

Since the antenna represents the transducer between the radar and the environment, it is the first line of defense against jamming. The directivity of the antenna in the transmission and reception phases allows space discrimination to be used as an ECCM strategy. Techniques for space discrimination include antenna coverage and scan control, reduction of main-beam width, low sidelobes, sidelobe blanking, sidelobe cancelers, and adaptive array systems. Some of these techniques are useful during transmission, while others operate in the reception phase. Additionally, some are active against main-beam jammers, and others provide benefits against sidelobe jammers.

Blanking or turning off the receiver while the radar is scanning across the azimuth sector containing the jammer or reducing the scan sector covered are means to prevent the radar from looking at the jammer. Certain deception jammers depend on anticipation of the beam scan or on knowledge or measurement of the antenna scan rate. Random electronic scanning effectively prevents these deception jammers from synchronizing to the antenna scan rate, thus defeating this type of jammer. A high-gain antenna can be employed to spotlight a target and burn through the jammers. An antenna having multiple beams can also be

used to allow deletion of the beam containing the jammer and still maintain detection capabilities with the remaining beams. Increased angular resolution of jammers in the main beam can be reached by resorting to spectral analysis algorithms, commonly referred to as *superresolution* techniques. Although they add complexity, cost, and possibly weight to the antenna, reduction of main-beam width and control of coverage and scan are valuable and worthwhile ECCM features of all radars.

If an air defense radar operates in a severe ECM environment, the detection range can be degraded because of jamming entering the sidelobes. On transmit, the energy radiated into spatial regions outside of the main beam is subject to being received by enemy RWRs or ARMs. For these reasons, low sidelobes are desirable on both receive and transmit.[18] Sometimes the increase in main-beam width that results from low sidelobes worsens the problem of main-beam jamming; this consequence should be carefully considered in specifying the antenna radiation pattern.

Usually, specification of the sidelobes as a single number (e.g., -30 dB) means that the peak of the highest sidelobe is 30 dB below the peak of the main beam. The average, or root-mean-square (rms), sidelobe level is often more important. For example, if 10 percent of the radiated power is in the sidelobes, the average sidelobe level is -10 dB, where dB refers to the number of decibels by which the average sidelobe level is below the gain of an isotropic (ideal) radiator. In theory, extremely low sidelobes can be achieved with aperture illumination functions that are appropriately tapered. This leads to the well-known tradeoffs among gain, beamwidth, and sidelobe level.[19] In order to keep the beamwidth small with low sidelobes, a larger and more costly antenna is needed. Other design principles involved in low antenna sidelobes are the use of radar-absorbent material (RAM) about the antenna structure, the use of a fence on ground installations, and the use of polarization screens and reflectors. This means that very low sidelobe antennas are costly in terms of size and complexity when compared with conventional antennas of similar gain and beamwidth characteristics. Second, as the design sidelobes are pushed lower and lower, a point is reached where minor error contributions to scattered energy (random errors) or misdirected radiation (systematic errors) become significant. In practice, peak sidelobe levels as low as -30 to -35 dB (average level, -5 to -20 dB) can be readily realized with phased array antennas which electronically scan. To obtain sidelobes at levels -45 dB down from the main beam (average level, below -20 dB), the total phase-error budget is required to be in the order of $5°$ rms or less. This is extremely difficult in arrays which electronically scan: the errors induced by phase shifters, active components, and feed elements must be included in this budget. Arrays have been realized in practice which have peak sidelobes in the vicinity of the -45 dB level; however, these are generally mechanically scanned, and the low error budgets are achieved by using all-passive feed components. Future antenna development will yield -45 dB sidelobe antennas which do scan electronically.[12]

Two additional techniques to prevent jamming from entering through the radar's sidelobes are the so-called sidelobe blanking (SLB) and sidelobe cancelers (SLC). An example of the practical effectiveness of the SLB and SLC devices is presented in the literature, where the plan position indicator (PPI) display is shown for a radar, subject to an ECM, equipped with and without the SLB and SLC systems.[17]

Other discrimination means are based on polarization. The polarization characteristics of a radar can be exploited as ECCM techniques in two ways. First,

the cross-polarized pattern (i.e., the orthogonal polarization to the main plane of polarization) of a radar antenna should be kept as low as possible consistent with radar system cost. Ratios of copolarized main-beam peak gain to cross-polarized gain anywhere in the antenna pattern should be greater than 25 dB to provide protection against common cross-polarized jamming. This is thought of as an ECCM technique, but it is really no more than good antenna design. The cross-polarized jamming in this case attacks a design deficiency in the radar. The requirement for good cross-polarization design practice in a radar antenna system extends to any auxiliary ECCM antennas as well. If their cross-polarized gains are high, ECCM techniques such as SLC and SLB may not be effective against cross-polarized noise or repeater jammers.[15]

In the second use of polarization the radar antenna system purposely receives the cross-polarization component of the radar wave in addition to the copolarized component. The two orthogonally polarized components can be used to discriminate the useful target from chaff and jammer on the basis of their different polarizations.[20] However, limited benefits (few decibels of cancellation ratio) can be obtained at the expense of a more complex antenna system (consider, for example, a phased array with radiating elements able to separately receive and possibly transmit the two orthogonal components of a radar wave) and of a duplication of the receiver and signal processing.

Sidelobe-Blanking (SLB) System.   The purpose of an SLB system is to prevent the detection of strong targets and interference pulses entering the radar receiver via the antenna sidelobes.[21-24] A method of achieving this is to employ an auxiliary antenna coupled to a parallel receiving channel so that two signals from a single source are available for comparison. By suitable choice of the antenna gains, one may distinguish signals entering the sidelobes from those entering the main beam, and the former may be suppressed. Figure 9.1a illustrates the radiation pattern of the main antenna together with a low-gain auxiliary antenna. An implementation of the SLB processor is shown in Fig. 9.1b, where the square-law-detected outputs of the two channels, ideally identical except for the antenna patterns, are compared. The comparison is made at each range bin for each pulse received and processed by the two parallel channels. Thus, the SLB decides whether or not to blank the main channel on a single-sweep basis and for each range bin. A target A in the main beam will result in a large signal in the main receiving channel and a small signal in the auxiliary receiving channel. A proper blanking logic allows this signal to pass. Targets and/or jammers J situated in the sidelobes give small main but large auxiliary signals so that these targets are suppressed by the blanking logic. It is assumed that the gain G_A of the auxiliary antenna is higher than the maximum gain G_{sl} of the sidelobes of the radar antenna.

The performance of the SLB may be analyzed by looking at the different outcomes obtained as a consequence of the pair (u, v) of the processed signals (see Fig. 9.1b). Three hypotheses have to be tested: (1) the null hypothesis H_0 corresponding to the presence of noise in the two channels, (2) the H_1 hypothesis pertaining to the target in the main beam, and (3) the H_2 hypothesis corresponding to target or interference signal in the sidelobe region. The null and H_1 hypotheses correspond to the usual decisions of "no detection" and "target detection," respectively. The blanking command is delivered when H_2 is detected.

SLB performance can be expressed in terms of the following probabilities: (1) The probability P_B of blanking a jammer in the radar sidelobes, which is the probability of associating the received signals (u, v) with H_2 when the same hypoth-
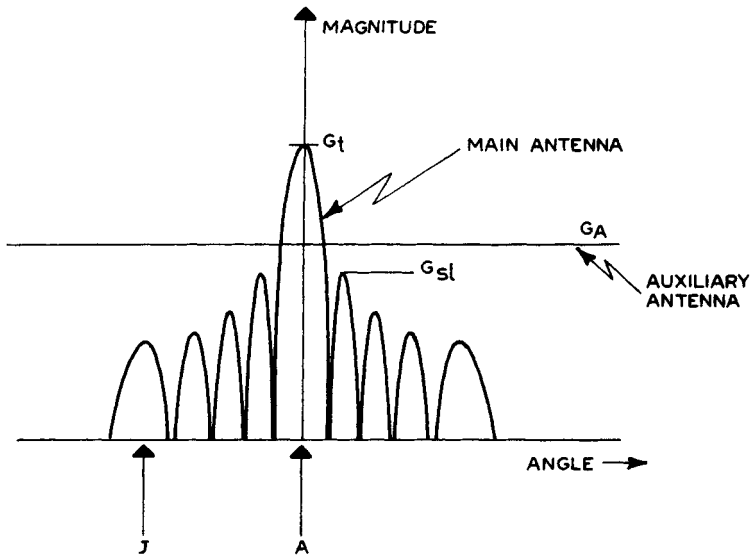
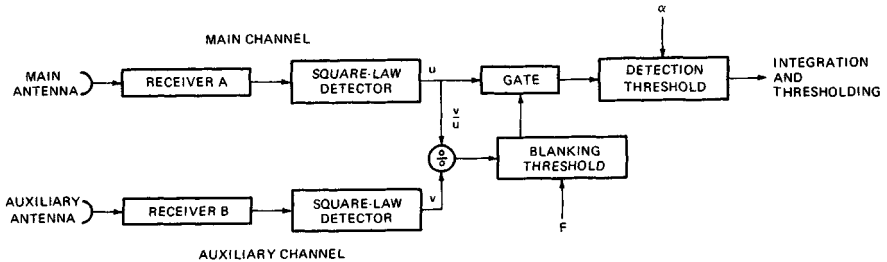**FIG. 9.1a**   Main and auxiliary antenna patterns for the SLB. (*From Ref. 21.*)

FIG. 9.1b   Scheme of sidelobe-blanking system. (*From Ref. 21.*)

esis is true; P_B is a function of the jammer-to-noise ratio (*JNR*) value, the blanking threshold F, and the gain margin $\beta = G_A/G_{sl}$ of the auxiliary antenna with respect to the radar antenna sidelobes. (2) The probability P_{FA} of false alarm, which is the probability of associating the received signals (u, v) with the hypothesis H_1 when the true hypothesis is H_0; P_{FA} is a function of the detection threshold α normalized to the noise power level and of the blanking threshold F. (3) The probability P_D of detecting a target in the main beam, which is the probability of associating the received signal (u, v) with H_1 when the same hypothesis is true; P_D depends, among other things, on the signal-to-noise power ratio *SNR*, P_{FA}, and the blanking threshold F. (4) The probability P_{FT} of detecting a false target produced by a jammer entering through the radar sidelobes. P_{FT} is the probability of associating (u, v) with H_1 when H_2 is true; it is a function of *JNR*, the thresholds α and F, and the gain margin β. (5) The probability P_{TB} of blanking a target received in the main beam. This is the probability of associating (u, v) with H_2 when H_1 is the true hypothesis. P_{TB} is related to *SNR*, F, and the auxiliary gain $w = G_A/G_t$ normalized to the gain G_t of the main beam. To complete the list of

parameters needed to describe the SLB performance, the last figure to consider is the detection loss L on the main-beam target. This can be found by comparing the SNR values required to achieve a specified P_D value for the radar system with and without the SLB. L is a function of many parameters such as P_D, P_{FA}, F, G_A, JNR, and β. A numerical evaluation of some of these performance parameters can be found in the literature.[21,24]

The SLB design requires the selection of suitable values for the following parameters: (1) the gain margin β and then the gain w of the auxiliary antenna, (2) the blanking threshold F, and the normalized detection threshold α. The a priori known parameters are the radar sidelobe level G_{sl} and the values of SNR and JNR. The design parameters can be selected by trying to maximize the detection probability P_D while keeping at prescribed values the probabilities P_B and P_{FA} and trying to minimize P_{FT}, P_{TB}, and L.

Sidelobe Canceler (SLC) System. The objective of the SLC is to suppress high duty cycle and noiselike interferences (e.g., SOJ) received through the sidelobes of the radar. This is accomplished by equipping the radar with an array of auxiliary antennas used to adaptively estimate the direction of arrival and the power of the jammers and, subsequently, to modify the receiving pattern of the radar antenna to place nulls in the jammers' directions. The SLC was invented by P. Howells and S. Applebaum.[25,26]

The conceptual scheme of an SLC system is shown in Fig. 9.2. The auxiliary antennas provide replicas of the jamming signals in the radar antenna sidelobes. To this end the auxiliary patterns approximate the average sidelobe level of the radar receiving pattern. In addition, the auxiliaries are placed sufficiently close to the phase center of the radar antenna to ensure that the samples of the interference which they obtain are statistically correlated with the radar jamming signal. It is also noted that as many auxiliary antennas are needed as there are jamming signals to be suppressed. In fact, at least N auxiliary patterns properly controlled in amplitude and phase are needed to force to zero the main-antenna receiving pattern in N given directions. The auxiliaries may be individual antennas or groups of receiving elements of a phased array antenna.

The amplitude and phase of the signals delivered by the N auxiliaries are controlled by a set of suitable weights: denote the set with the N-dimensional vector $\mathbf{W} = (W_1, W_2, \ldots, W_N)$. Jamming is canceled by a linear combination of the signals from the auxiliaries and the main antenna. The problem is to find a suitable means of controlling the weights $\mathbf{W}$ of the linear combination so that the maximum possible cancellation is achieved. Owing to the stochastic nature of the jamming signals in the radar and in the auxiliary channels and the hypothesized linear combination of signals, it is advisable to resort to the techniques of linear prediction theory for stochastic processes.[27] Denote with V_M the radar signal at a certain range bin and with $\mathbf{V} = (V_1, V_2, \ldots, V_N)$ the N-dimensional vector containing the set of signals, at the same range bin, from the N auxiliary antennas. It is assumed that all the signals have bandpass frequency spectra; therefore, the signals can be represented by their complex envelopes, which modulate a common carrier frequency that does not appear explicitly. The jamming signals in the channels may be regarded as samples of a stochastic process having zero mean value and a certain time autocorrelation function. For linear prediction problems, the set of samples $\mathbf{V}$ is completely described by its N-dimensional covariance matrix $\mathbf{M} = \mathrm{E}(\mathbf{V}^*\mathbf{V}^T)$, where $E(.)$ denotes the statistical expectation, the asterisk $*$ indicates the complex conjugate, and $\mathbf{V}^T$ is the transpose vector of $\mathbf{V}$. The statistical relationship between V_M and $\mathbf{V}$ is mathematically represented by the N-
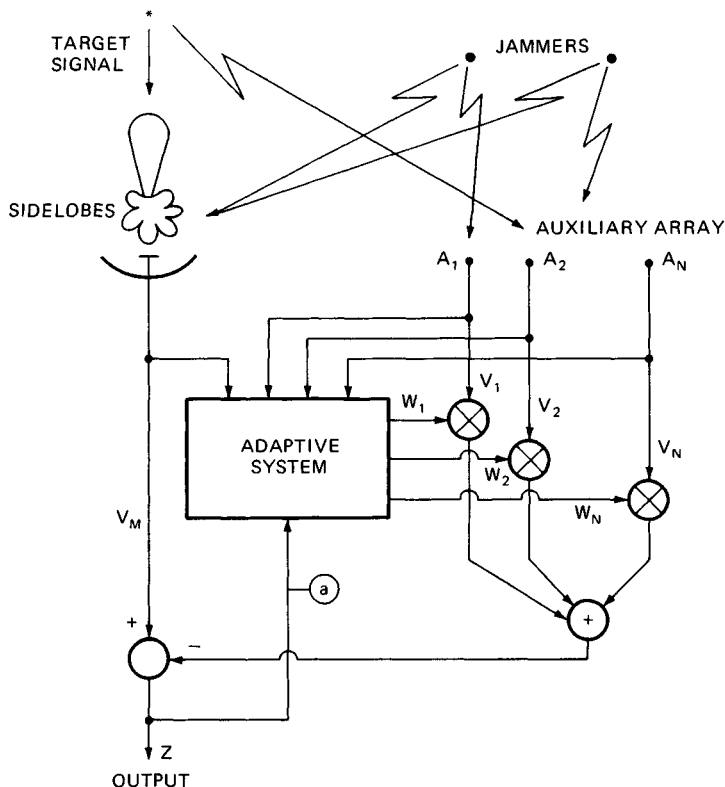
FIG. 9.2 Principle of SLC operation (connection *a* only in the closed-loop implementation techniques).

dimensional covariance vector $\mathbf{R} = E(V_M \mathbf{V}^*)$. The optimum weight vector $\hat{\mathbf{W}}$ is determined by minimizing the mean square prediction error which equals the output residual power:

$$P_Z = E\{\,|Z|^2\} = E\{|V_M - \hat{\mathbf{W}}^T \mathbf{V}|^2\} \tag{9.2}$$

where Z is the system output. It is found that the following fundamental equation applies:[27]

$$\hat{\mathbf{W}} = \mu \mathbf{M}^{-1} \mathbf{R} \tag{9.3}$$

where μ is an arbitrary constant value.

The benefit of using the SLC can be measured by introducing the jammer cancellation ratio (*JCR*), defined as the ratio of the output noise power without and with the SLC:

$$JCR = \frac{E\{|V_M|^2\}}{E\{|V_M - \hat{\mathbf{W}}^T \mathbf{V}|^2\}} = \frac{E\{|V_M|^2\}}{E\{|V_M|^2\} - \mathbf{R}^T \mathbf{M}^{-1} \mathbf{R}^*} \tag{9.4}$$

By applying Eqs. (9.3) and (9.4) to the simple case of one auxiliary antenna and one jammer, the following results are easily found:

$$\hat{W} = \frac{E\{V_M V_A^*\}}{E\{|V_A|^2\}} \triangleq \rho \qquad JCR = \frac{1}{1 - |\rho|^2} \qquad (9.5)$$

It is noted that the optimum weight is related to the correlation coefficient ρ between the main signal V_M and the auxiliary signal V_A; high values of the correlation coefficient provide high values of JCR.

The problem of implementing the optimum-weight set [Eq. (9.3)] is essentially related to the real-time estimation of $\mathbf{M}$ and $\mathbf{R}$ and to the inversion of $\mathbf{M}$. Several processing schemes have been conceived which may be classified in two main categories: (1) closed-loop techniques, in which the output residue (connection *a* of Fig. 9.2) is fed back into the adaptive system; and (2) direct-solution methods, often referred to as *open-loop*, which operate just on the incoming signals V_M and V. Broadly speaking, closed-loop methods are cheaper and simpler to implement than direct-solution methods.[27,28] By virtue of their self-correcting nature, they do not require components which have a wide dynamic range or a high degree of linearity, and so they are well suited to analog implementation. However closed-loop methods suffer from the fundamental limitation that their speed of response must be restricted in order to achieve a stable and not noisy steady state. Direct-solution methods, on the other hand, do not suffer from problems of slow convergence but, in general, require components of such high accuracy and wide dynamic range that they can only be realized by digital means.[27,29] Of course, closed-loop methods can also be implemented by using digital circuitry, in which case the constraints on numerical accuracy are greatly relaxed and the total number of arithmetic operations is much reduced by comparison with direct-solution methods.

Practical considerations often limit the SLC nulling capabilities to JCR of about 20 to 30 dB, but their theoretical performance is potentially much higher.[30,31] Examples of possible limitations are listed below:[27]

  1. Mismatch between the main and auxiliary signals including the propagation paths, the patterns of the main and auxiliary antennas, the paths internal to the system up to the cancellation point, and the crosstalk between the channels.[32,33]

  2. The limited number of auxiliary channels adopted in a practical system as compared with the number of jamming signals.[32]

  3. The limited bandwidth of the majority of the schemes implementing Eq. (9.3) as compared with the wide band of a barrage jammer which can be regarded as a cluster, spread in angle, of narrowband jammers.[28,30,34]

  4. The pulse width which limits the reaction time of the adaptive system, in order to avoid the cancellation of target signal.[33]

  5. The target signal in the auxiliary array which may result in nonnegligible steering of the auxiliaries toward the main-beam direction.[33]

  6. The presence of clutter which, if not properly removed, may capture the adaptive system, giving rise to nulls along directions different from those of the jammers.

  7. The tradeoff which has to be sought between the accuracy of weights estimation and the reaction time of the adaptive system.

8. The quantization and processing accuracy in the digital implementation.

Adaptive Arrays.   An adaptive array (Fig. 9.3) is a collection of N antennas, feeding a weighting and summing network, with automatic signal-dependent weight adjustment to reduce the effect of unwanted signals and/or emphasize the desired signal or signals in the summing network output. Output signal z is envelope-detected and compared with a suitable threshold α to detect the presence of a useful target.[28,34–40] The adaptive array is a generalization of the SLC system concept described in the preceding subsection. The basic theory of jammer cancellation and target enhancement is considered first, and attention is then focused on the use of adaptive arrays to obtain superresolution capabilities which can be of help for ECCM. The implementation of the adaptive array concept is more and more related to digital beamforming technology.[41–43]

Jammer Cancellation and Target Signal Enhancement.   Adaptive array principles have found a thorough mathematical treatment since the early 1970s.[40] The basic result is given by the expression of the optimum set of weights:

$$W = \mu M^{-1}S*$$ (9.6)

where $M = E(V^*V^T)$ is the N-dimensional covariance matrix of the overall disturbance (noise and jammer) V received by the array and S is the N-dimensional vector containing the expected signal samples in the array from a target along a certain direction of arrival. The similarity of Eq. (9.6) to Eq. (9.3) governing the SLC is immediately recognized.

With respect to SLC, adaptive array techniques offer the capability of enhancing the target signal while canceling the disturbance. The adaptive system allocates in an optimum fashion its degrees of freedom (i.e., the set of received
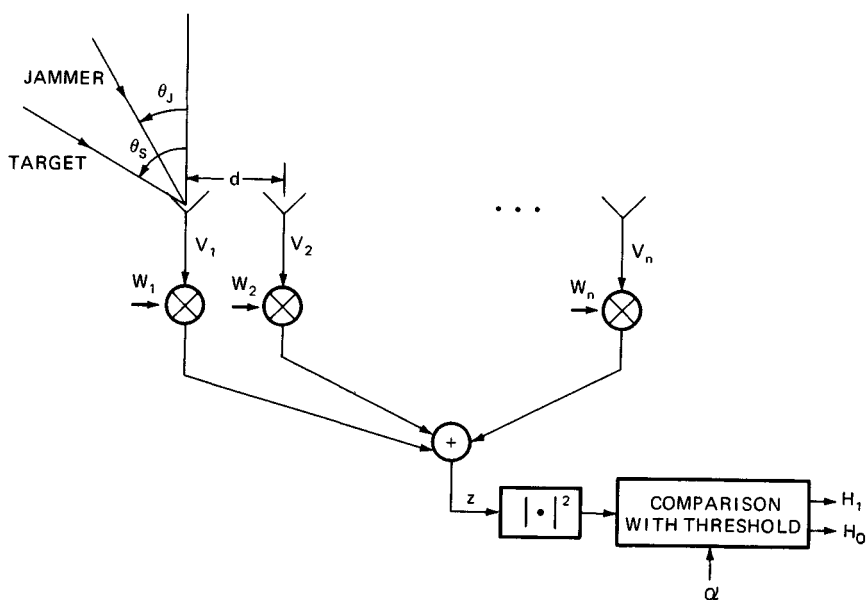

FIG. 9.3 The adaptive array scheme.

pulses from each antenna of the array) to the enhancement of the target signal and to the cancellation of clutter, chaff, and jammer.

Several generalizations of the basic theory have been considered, including: (1) the target model S is not known a priori, as it is assumed in deriving Eq. (9.6); (2) in addition to spatial filtering, doppler filtering is performed to cancel clutter and chaff; and (3) the radar platform is moving as in airborne and shipborne applications.

The detection probability P_D for the optimum filter of Eq. (9.6) is:[40]

$$P_D = Q \left(\sqrt{S^T M^{-1} S^*}, \sqrt{2 \ln 1/P_{FA}} \right) \tag{9.7}$$

where $Q (. , .)$ is the Marcum Q function and P_{FA} is the prescribed probability of false alarm. It is also shown that the set of weights of Eq. (9.6) provides the maximum value of the improvement factor I_f, which is defined as follows:

$$I_f = \frac{\text{signal-to-interference power ratio at the output}}{\text{signal-to-interference power ratio at the input}} \tag{9.8}$$

The signal-to-interference power ratio $(SNR)_I$ at the input is measured at the input of an antenna of the array and refers to one echo pulse. The I_f value corresponding to the optimum set of weights of Eq. (9.6) is[40]

$$I_f = \frac{S^T M^{-1} S^*}{(SNR)_I} \tag{9.9}$$

The I_f is better suited than the cancellation ratio, adopted in the SLC, to represent the performance of the adaptive array. In fact, in the latter case the useful signal is integrated while the interference is canceled.

The implementation of an adaptive array has been limited to experimental systems with a small number of antennas (say, 10), so that the matrix inversion can be handled by practical computing systems.[44,45] Arrays with a large number of receiving elements need some form of processing reduction. One method of partial adaptivity is to arrange the array elements in subgroups which form the inputs of the adaptive processor. Careful selection of the subgroup elements is necessary to avoid grating lobes.[46,47]

Other simplifications of the fully adaptive array are deterministic spatial filtering and phase-only nulling techniques. In the first case, a fixed reduction of the sidelobes is operated in those directions or solid angles from which the interferences are expected to come. As an example, a probable region with interferences is the horizon or part of it because jammers are mostly ground-based or at long range. The weights are computed offline, by assuming an a priori known covariance matrix M, and stored in a memory where a "menu" of weights is available to an operator or an automatic decision system.[48] The idea of phase-only nulling in phased array antennas is appealing because the phase shifters are already available as part of the beam-steering system. Hence, if the same phase shifters can be employed for the dual purposes of beam steering and adaptive nulling of unwanted interferences, costly retrofitting could be unnecessary. However, phase-only null synthesis presents analytic and computational difficulties not present when both the amplitude and the phase of the element weights can be freely perturbed.[49,50] Nevertheless, experimental systems have been tested with success.[51-53]

Superresolution. The resolution of a conventional antenna is limited by the well-known Rayleigh criterion, which states that two equal-amplitude noise sources can be resolved if they are separated in angle by 0.8 λ/L, in radians, where λ is the wavelength and L is the aperture length. When the incident wave is received with a high signal-to-thermal noise ratio, an adaptive array antenna may achieve a narrower *adaptive beamwidth*, giving a sharper bearing estimation of the incident wave. This is important for ECCM purposes: very accurate strobes of the jammers can be obtained. It is also possible to measure the source strength and to obtain a spatial spectrum pattern without sidelobes. The estimated angles of the jammers can be used to form beams in the jammer directions, which are used as auxiliary channels for adaptive interference suppression.[54] The interference directions can also be used for deterministic nulling, which is of special interest for main-beam nulling.[55] In addition to the interference source directions and source strengths, this technique can provide other information as to the number of sources and any cross correlations (coherence) between the sources. Such information can be used to track and catalog the interference sources in order to properly react to them.

The superresolution concept was mainly developed and analyzed by W. F. Gabriel.[56] Different methods for bearing estimation were described by Gabriel and, subsequently, by other authors.[39,57-59] One is the maximum-entropy method (MEM). It works well with a Howells-Applebaum adaptive beamformer, which has an omnidirectional receiving pattern except where signals are present. The presence of signals is indicated by nulls in the receiving pattern. Since nulls are always sharper than antenna lobes, signal bearings can be obtained more accurately from the adaptive beam pattern, and superresolution is the result. The desired spatial spectrum pattern is obtained as simply the inverse of the adapted pattern. As Gabriel points out, there is not a true antenna pattern because there is no linear combination of the signals from an array that could produce such a peaked spatial pattern. It is simply a function computed from the reciprocal of a true adapted antenna pattern. Superresolution and adaptive antennas are identical mathematically, use the same algorithms, and have identical hardware. Roughly speaking, the difference is that one produces a pattern with the nulls down (adaptive antenna for jammer cancellation) and the other with the nulls up (superresolution of jammers).

The achievable degree of superresolution depends heavily on the way in which the algorithms are implemented. The required accuracy of signal quantization and the matching of channels are comparable with those of adaptive nulling. The heavy computational task required by the algorithms can be handled by resorting to systolic array processors.[60] Experiments indicate that the resolution limit is determined rather more by implementation factors like channel mismatching errors than by the pure *SNR*. Two incoherent sources separated by a quarter beamwidth seem to be the lower limit for superresolution with the current technology for achieving equality between the channels, offset compensation, equality of I and Q channel amplification, etc. Resolution is worse for more than two sources.[58]

9.7 TRANSMITTER-RELATED ECCM

The different types of ECCM are related to the proper use and control of the power, frequency, and waveform of the radiated signal. One brute-force ap-

proach to defeating noise jamming is to increase the radar's transmitter power. This technique, when coupled with "spotlighting" the radar antenna on the target, results in an increase of the radar's detection range. Spotlighting or burnthrough modes might be effective, but a price must be paid. As the radar dwells in a particular direction, it is not looking elsewhere, where it is supposed to look. In addition, the burnthrough mode is not effective against chaff, decoys, repeaters, spoofers, and so on.

More effective is the use of complex, variable, and dissimilar transmitted signals which place a maximum burden on ESM and ECM. Different ways of operation refer to the change of the transmitted frequency in frequency-agility or frequency-diversity modes or to the use of wide instantaneous bandwidth.[61-64] *Frequency agility* usually refers to the radar's ability to change the transmitter frequency on a pulse-to-pulse or batch-to-batch basis. The batch-to-batch approach allows doppler processing, which is not compatible with frequency agility on a pulse-to-pulse basis. In a waveform with pulse-to-pulse frequency agility, the center frequency of each transmitted pulse is moved, in either a random or a programmed schedule, between a large number of center frequencies on a pulse-to-pulse basis. The frequency of the next pulse cannot generally be predicted from the frequency of the current pulse.[65] *Frequency diversity* refers to the use of several complementary radar transmissions at different frequencies, either from a single radar (e.g., a radar having stacked beams in elevation by employing different frequencies on each elevation beam) or from several radars. The objective of frequency agility and diversity is to force the jammer to spread its energy over the entire agile bandwidth of the radar; this corresponds to a reduction of the jammer density and resulting ECM effectiveness.[15] Signals with wide instantaneous bandwidth exhibit considerable variation of the frequency within each transmitted pulse. A spread of about 10 percent of the transmitter center frequency can be proper. Three of the more common coded-pulse waveforms are recalled: (1) the linear frequency-modulated signal, where the carrier frequency is varied linearly within the pulse; (2) the frequency-shift-coded signal, where the carrier frequency is changed in a stepwise fashion within the pulse; and (3) the phase-coded signal, in which the phase of the RF carrier is shifted at a rate equal to the bandwidth of the waveform.

Frequency agility, diversity, and instantaneous wideband techniques represent a form of ECCM in which the information-carrying signal is spread over as wide a frequency (or space, or time) region as possible to reduce detectability by ESM and/or ARM and make jamming more difficult. This ECCM technique pertains to the realm of waveform coding.[12,29] Waveform coding includes pulse-repetition-frequency (PRF) jitter, PRF stagger, and, perhaps, shaping of the transmitted radar pulse. All these techniques make deception jamming or spoofing of the radar difficult, since the enemy should not know or anticipate the fine structure of the transmitted waveform; as a consequence, they give assurance of maximum range performance against jamming. Intrapulse coding to achieve pulse compression may be particularly effective in improving target detection capability by radiation of enough average radar power without exceeding peak power limitations within the radar and by improving range resolution (larger bandwidth), which in turn reduces chaff returns and resolves targets to a higher degree.

Some advantage can be gained by including the capability to examine the jammer signals, find holes in their transmitted spectra, and select the radar frequency with the lowest level of jamming. This approach is particularly useful against pulsed ECM, spot noise, and nonuniform barrage noise; its effectiveness depends

primarily on the extent of the radar agile bandwidth and the acquisition speed and frequency tracking of an "intelligent" jammer. A technique suited to this purpose is referred to as automatic frequency selection (AFS).[64,66]

Another method to reduce the effect of main-beam noise jamming is to increase the transmitter frequency (as an alternative means to the use of a larger antenna) in order to narrow the antenna's beamwidth. This restricts the sector which is blanked by main-beam jamming and also provides a strobe in the direction of the jammer. Strobes from two or three spatially separated radars allow the jammer to be located.

9.8 RECEIVER-RELATED ECCM

Jamming signals that survive the antenna ECCM expedients can, if large enough, saturate the radar processing chain. Saturation results in the virtual elimination of information about targets. Wide dynamic range (i.e., log and lin-log) receivers are normally used to avoid saturation.

Other special processing circuits can be used in the radar to avoid saturation, i.e., fast-time-constant (FTC) devices, automatic gain control (AGC), and constant-false-alarm rate (CFAR).[3,15,17] However, they cannot be said to be ECCM techniques. For example, FTC allows the detection of signals that are greater than clutter by preventing the clutter from saturating the display. FTC does not provide subclutter visibility. AGC keeps the radar receiver operating within its dynamic range, preventing system overload and providing proper normalization so as to furnish signals of standardized amplitude to radar range, velocity, and angle processing-tracking circuits. CFAR is a technique made necessary because of the limitations of the computer in automatic systems. It prevents the computer from being overloaded by lowering the capability of the radar to detect desired targets. In conclusion, these devices have a place in the radar but not as means for fighting the ECM battle.

A log (logarithmic) receiver is a device whose video output is proportional to the logarithm of the envelope of the RF input signal over a specified range. It is useful in preventing receiver saturation in the presence of variable intensities of jamming noise, rain, clutter, and chaff. Log receivers have the ECCM advantage of permitting the radar receiver to detect target returns that are larger than jamming noise, chaff, or clutter levels. By comparison with a linear receiver of low dynamic range, moderate jamming noise levels will normally cause the display to saturate so that the target signal will not be detected. However, the disadvantage lies in the fact that low-level jamming signals will be amplified more than higher-level target signals, thereby reducing the signal-to-jamming ratio and allowing a low-level noise jammer to be more effective. Another disadvantage is that a log characteristic causes spectral spreading of the received echoes. It would not be possible to maintain clutter rejection in an MTI (moving-target indicator) or pulse doppler radar if the spectrum of clutter echoes were to spread into the spectral region in which target returns were expected.[13,15]

In a lin-log (linear-logarithmic) receiver the output signal amplitude is closely proportional to the logarithm of the envelope of the RF input signal amplitude for high input signal amplitudes, while the output signal amplitude is directly proportional to the envelope of the RF input signal amplitude for low input signal am-

CHAPTER 10
PULSE COMPRESSION RADAR

Edward C. Farnett
George H. Stevens
RCA Electronic Systems Department
GE Aerospace

10.1 INTRODUCTION

Pulse compression involves the transmission of a long coded pulse and the processing of the received echo to obtain a relatively narrow pulse. The increased detection capability of a long-pulse radar system is achieved while retaining the range resolution capability of a narrow-pulse system. Several advantages are obtained. Transmission of long pulses permits a more efficient use of the average power capability of the radar. Generation of high peak power signals is avoided. The average power of the radar may be increased without increasing the pulse repetition frequency (PRF) and, hence, decreasing the radar's unambiguous range. An increased system resolving capability in doppler is also obtained as a result of the use of the long pulse. In addition, the radar is less vulnerable to interfering signals that differ from the coded transmitted signal.

A long pulse may be generated from a narrow pulse. A narrow pulse contains a large number of frequency components with a precise phase relationship between them. If the relative phases are changed by a phase-distorting filter, the frequency components combine to produce a stretched, or expanded, pulse. This expanded pulse is the pulse that is transmitted. The received echo is processed in the receiver by a compression filter. The compression filter readjusts the relative phases of the frequency components so that a narrow or compressed pulse is again produced. The pulse compression ratio is the ratio of the width of the expanded pulse to that of the compressed pulse. The pulse compression ratio is also equal to the product of the time duration and the spectral bandwidth (time-bandwidth product) of the transmitted signal.

A pulse compression radar is a practical implementation of a matched-filter system. The coded signal may be represented either as a frequency response $H(\omega)$ or as an impulse time response $h(t)$ of a coding filter. In Fig. 10.1a, the coded signal is obtained by exciting the coding filter $H(\omega)$ with a unit impulse. The received signal is fed to the matched filter, whose frequency response is the complex conjugate $H^*(\omega)$ of the coding filter. The output of the matched-filter section is the compressed pulse, which is given by the inverse Fourier transform of the product of the signal spectrum $H(\omega)$ and the matched-filter response $H^*(\omega)$:
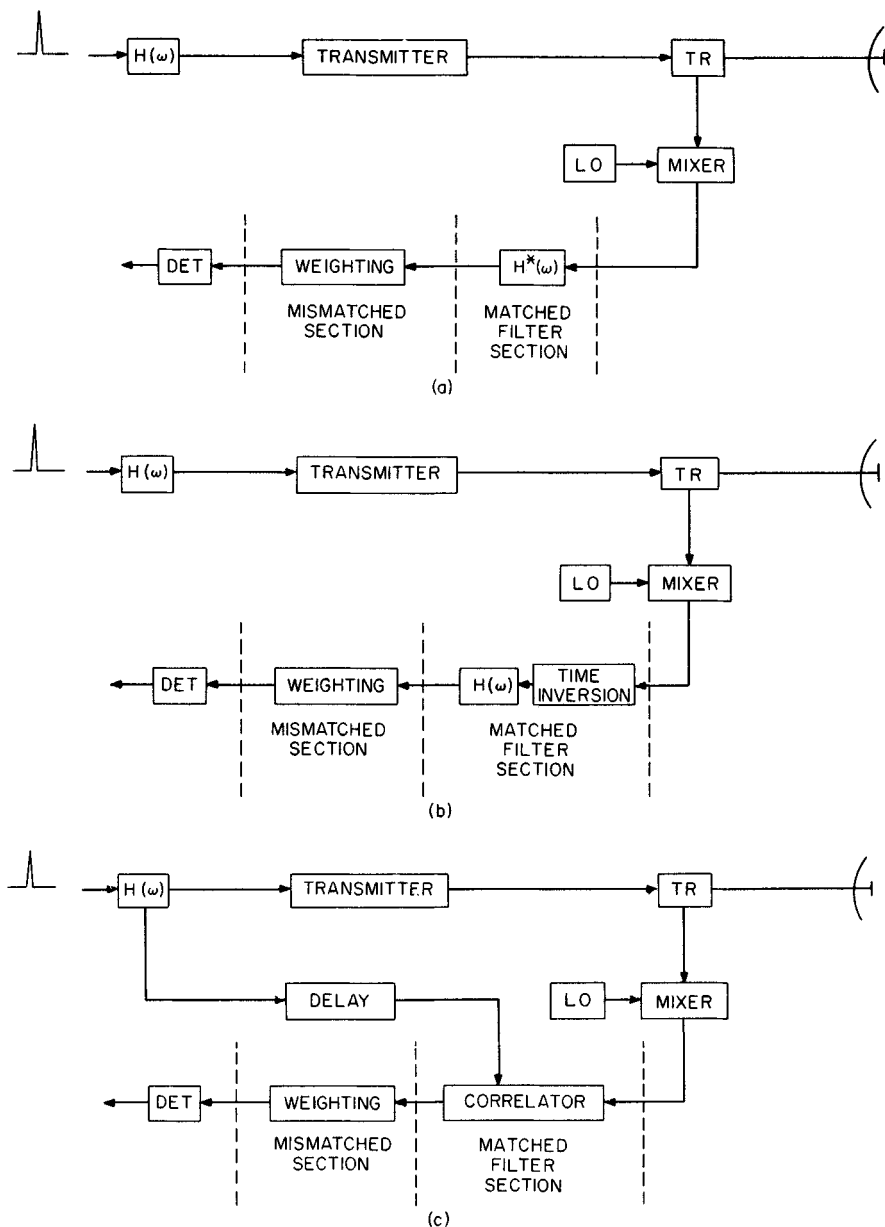
FIG. 10.1 Pulse compression radar using (*a*) conjugate filters, (*b*) time inversion, and (*c*) correlation.

$$y(t) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |H(\omega)|^2 e^{j\omega t}d\omega$$

The implementation of Fig. 10.1a uses filters which are conjugates of each other for the expansion and compression filters.

A filter is also matched to a signal if the signal is the complex conjugate of the time inverse of the filter's response to a unit impulse. This is achieved by applying the time inverse of the received signal to the compression filter, as shown in Fig. 10.1b. Identical filters may be used for both expansion and compression, or the same filter may be used for both expansion and compression with appropriate switching between the transmitting and receiving functions. The output of this matched filter is given by the convolution of the signal $h(t)$ with the conjugate impulse response $h^*(-t)$ of the matched filter:

$$y(t) = \int\limits_{-\infty}^{\infty} h(\tau)h^*(t - \tau)d\tau$$

The matched filter results in a correlation of the received signal with the transmitted signal. Hence, correlation processing as shown in Fig. 10.1c is equivalent to matched filtering. In practice, multiple delays and correlators are used to cover the total range interval of interest.

The output of the matched filter consists of the compressed pulse accompanied by responses at other ranges, called time or range sidelobes. Frequency weighting of the output signals is usually employed to reduce these sidelobes. This results in a mismatched condition and leads to a degradation of the signal-to-noise output of the matched filter. In the presence of a doppler frequency shift, a bank of matched filters is required, with each filter matched to a different frequency so as to cover the band of expected doppler frequencies.

10.2 FACTORS AFFECTING CHOICE OF PULSE COMPRESSION SYSTEM

The choice of a pulse compression system is dependent upon the type of waveform selected and the method of generation and processing. The primary factors influencing the selection of a particular waveform are usually the radar requirements of range coverage, doppler coverage, range and doppler sidelobe levels, waveform flexibility, interference rejection, and signal-to-noise ratio (SNR). The methods of implementation are divided into two general classes, active and passive, depending upon whether active or passive techniques are used for generation and processing.

Active generation involves generating the waveform by phase or frequency modulation of a carrier without the occurrence of an actual time expansion. An example is digital phase control of a carrier. Passive generation involves exciting a device or network with a short pulse to produce a time-expanded coded waveform. An example is an expansion network composed of a surface-acoustic-wave (SAW) delay structure. Active processing involves mixing delayed replicas of the transmitted signal with the received signal and is a correlation-processing

approach. Passive processing involves the use of a compression network that is the conjugate of the expansion network and is a matched-filtering approach. Although a combination of active and passive techniques may be used in the same radar system, most systems employ the same type for generation and processing; e.g., a passive system uses both passive generation and passive processing.

The performance of common types of pulse compression systems is summarized in Table 10.1. The systems are compared on the assumption that information is extracted by processing a single waveform as opposed to multiple-pulse processing. The symbols B and T are used to denote, respectively, the bandwidth and the time duration of the transmitted waveform. Ripple loss refers to the SNR loss incurred in active systems because of the fluctuation or ripple in the SNR that occurs as a target moves from range cell to range cell. Clutter rejection performance of a single waveform is evaluated on the basis of doppler response rather than range resolution; pulse compression provides a means for realizing increased range resolution and, hence, greater clutter rejection. In applications where an insufficient doppler frequency shift occurs, range resolution is the chief means for seeing a target in clutter.

10.3   LINEAR FM

The linear-FM, or chirp, waveform is the easiest to generate. The compressed-pulse shape and SNR are fairly insensitive to doppler shifts. Because of its great popularity, more approaches for generating and processing linear FM have been developed than for any other coded waveform.[1] The major disadvantages are that (1) it has excessive range-doppler cross coupling which introduces errors unless either range or doppler is known or can be determined (i.e., a shift in doppler causes an apparent change in range and vice versa); and (2) weighting is usually required to reduce the time sidelobes of the compressed pulse to an acceptable level. Time and frequency weighting are nearly equivalent for linear FM and cause a 1 to 2 dB loss in SNR. Passive linear-FM generation and processing may be used as in Fig. 10.1a or b, where conjugate networks or a single network is employed. Active linear-FM generation and processing may be used as in Fig. 10.1c.

10.4   NONLINEAR FM

The nonlinear-FM waveform has attained little acceptance although it has several distinct advantages. The nonlinear-FM waveform requires no time or frequency weighting for range sidelobe suppression since the FM modulation of the waveform is designed to provide the desired amplitude spectrum. Matched-filter reception and low sidelobes become compatible in this design. Thus, the loss in signal-to-noise ratio associated with weighting by the usual mismatching techniques is eliminated. If a symmetrical FM modulation is used with time weighting to reduce the frequency sidelobes, the nonlinear-FM waveform will have a near-ideal ambiguity function. A symmetrical waveform typically has a frequency that increases (or decreases) with time during the first half of the pulse and decreases (or increases) during the last half of the pulse. A nonsymmetrical waveform is

TABLE 10.1 Summary of Performance of Various Pulse Compression Implementations

|	Linear FM		Nonlinear FM		Phase-coded	
	Active	Passive	Active	Passive	Active	Passive
Range coverage	Limited range coverage per active correlation processor.	Provides full range coverage.	Limited range coverage per active correlation processor.	Provides full range coverage.	Limited range coverage per active correlation processor.	Provides full range coverage.
Doppler coverage	Covers any doppler up to $\pm B/10$, but a range error is introduced. *SNR* and time-sidelobe performance poor for larger doppler.		Multiple doppler channels required, spaced by $(1/T)$ Hz.			
Range sidelobe level	Requires weighting to reduce the range sidelobes below $(\sin x)/x$ falloff.		Good range sidelobes possible with no weighting. Sidelobes determined by waveform design.		Good range sidelobes. $N^{-1/2}$ for an N-element code.	
Waveform flexibility	Bandwidth and pulse width can be varied.	Limited to one bandwidth and pulse width per compression network.	Bandwidth and pulse width can be varied.	Limited to one bandwidth and pulse width per compression network.	Bandwidth, pulse width, and code can be varied.	
Interference rejection	Poor clutter rejection.		Fair clutter rejection.		Fair clutter rejection.	
SNR	Reduced by weighting and by ripple loss versus range.	Reduced by weighting.	Reduced by ripple loss versus range.	No *SNR* loss.	Reduced by ripple loss versus range.	No *SNR* loss.
Comments	1. Very popular with the advent of high-speed digital devices. 2. Extremely wide bandwidths achievable.	1. Widely used in past. 2. Well-developed technology.	1. Limited use. 2. Waveform generation by digital means most popular.	1. Limited use. 2. Extremely limited development.	1. Widely used. 2. Waveform very easy to generate.	1. Limited use. 2. Waveform moderately difficult to generate.
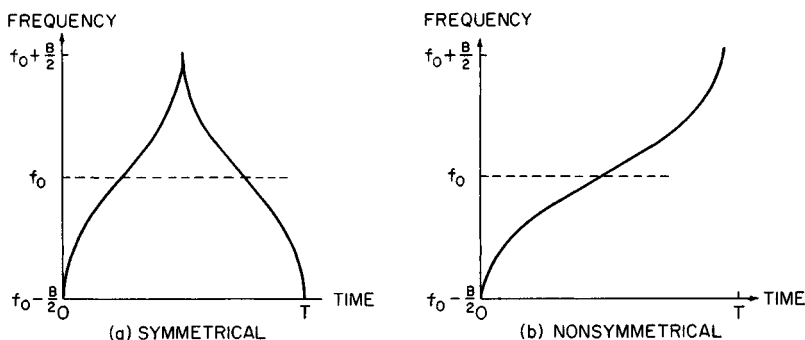
FIG. 10.2   Nonlinear-FM waveforms with 40 dB Taylor weighting.

obtained by using one-half of a symmetrical waveform (Fig. 10.2). However, the nonsymmetrical waveform retains some of the range-doppler cross coupling of the linear-FM waveform.

The disadvantages of the nonlinear-FM waveform are (1) greater system complexity, (2) limited development of nonlinear-FM generation devices, and (3) the necessity for a separate FM modulation design for each amplitude spectrum to achieve the required sidelobe level. Because of the sharpness of the ambiguity function, the nonlinear waveform is most useful in a tracking system where range and doppler are approximately known.

To achieve a 40 dB Taylor time-sidelobe pattern, the frequency-versus-time function of a nonsymmetrical transmitted pulse of bandwidth W is[2]

$$f(t) = W\left(\frac{t}{T} + \sum_{n=1}^{7} K_n \left| \sin \left| \frac{2\pi n t}{T} \right| \right.\right)$$

where K_1 = $-$ 0.1145
K_2 = $+$ 0.0396
K_3 = $-$ 0.0202
K_4 = $+$ 0.0118
K_5 = $-$ 0.0082
K_6 = $+$ 0.0055
K_7 = $-$ 0.0040

For a symmetrical frequency-versus-time function based on the above waveform, the first half ($t \leq T/2$) of the frequency-versus-time function will be the $f(t)$ given above, with T replaced with $T/2$. The last half ($t \geq T/2$) of the frequency-versus-time function will be the $f(t)$ above, with T replaced with $T/2$ and t replaced with $T/2 - t$.

10.5   PULSE COMPRESSION DEVICES

Major advances are continually being made in the devices used in pulse compression radars. Significant advances are evident in the digital and SAW techniques.

These two techniques allow the implementation of more exotic signal waveforms such as nonlinear FM. The digital approach has blossomed because of the manyfold increase in the computational speed and also because of the size reduction and the speed increase of the memory units. SAW technology has expanded because of the invention of the interdigital transducer,[3] which provides efficient transformation of an electrical signal into acoustic energy and vice versa. In spite of these advanced technologies, the most commonly used pulse compression waveforms are still the linear-FM and the phase-coded signals. Improved techniques have enhanced the processing of these "old standby" waveforms.

Digital Pulse Compression. Digital pulse compression techniques are routinely used for both the generation and the matched filtering of radar waveforms. The digital generator uses a predefined phase-versus-time profile to control the signal. This predefined profile may be stored in memory or be digitally generated by using appropriate constants. The matched filter may be implemented by using a digital correlator for any waveform or else a "stretch" approach for a linear-FM waveform.

Digital pulse compression has distinct features that determine its acceptability for a particular radar application. The major shortcoming of a digital approach is that its technology is restricted in bandwidths under 100 MHz. Frequency multiplication combined with stretch processing would increase this bandwidth limitation. Digital matched filtering usually requires multiple overlapped processing units for extended range coverage. The advantages of the digital approach are that long-duration waveforms present no problem, the results are extremely stable under a wide variety of operating conditions, and the same implementation could be used to handle multiple-waveform types.

Figure 10.3 shows the digital approach[4] for generating the radar waveform. This technique is normally used only for FM-type waveforms or polyphase-coded waveforms. Biphase coding can be achieved in a simpler manner, as shown in Sec. 10.6. The phase control element supplies digital samples of the in-phase component I and the quadrature component Q, which are converted to their analog equivalents. These phase samples may define the baseband components of the desired waveform, or they may define the waveform components on a low-frequency carrier. If the waveform is on a carrier, the balanced modulator is not required and the filtered components would be added directly. The sample-and-hold circuit is to remove the transients due to the nonzero transition time of the digital-to-analog (D/A) converter. The low-pass filter smooths (or interpolates) the analog signal components between waveform samples to provide the equivalent of a much higher waveform-sampling rate. The $I(t)$ component modulates a 0° carrier signal, and the $Q(t)$ component modulates a 90° phase-shifted carrier signal. The desired waveform is the sum of the 0°-modulated carrier and the 90°-modulated carrier. As mentioned earlier, when the digital phase samples include the carrier component, the I and Q components are centered on this carrier fre-
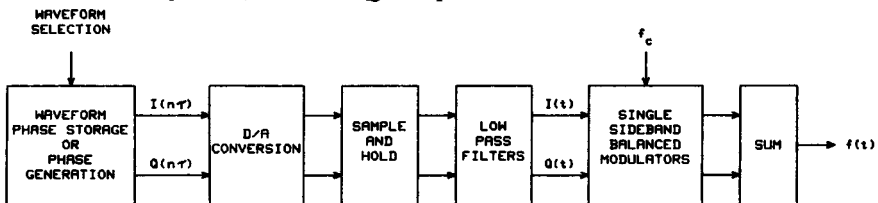

FIG. 10.3 Digital waveform generation.

quency and the low-pass filter can be replaced with a bandpass filter centered on the carrier.

Digital waveform generators are very stable devices with a well-defined distortion. As a result, the generated waveform may be frequency-multiplied to achieve a much wider waveform bandwidth. With multiplication, the distortion components are increased in magnitude by the multiplication factor, and tighter control of the distortion is required.

When a linear-FM waveform is desired, the phase samples follow a quadratic pattern and can be generated by two cascaded digital integrators. The input digital command to the first integrator defines this quadratic phase function. The digital command to the second integrator is the output of the first integrator plus the desired carrier frequency. This carrier may be defined by the initial value of the first integrator. The desired initial phase of the waveform is the initial value of the second integrator or else may be added to the second-integrator output.

Figure 10.4 illustrates two digital approaches to providing the matched filter for a pulse compression waveform. These approaches provide only limited range coverage, and overlapped processors are needed for all-range performance. Figure 10.4a shows a digital implementation of a correlation processor that will provide matched-filter performance for any radar waveform. Figure 10.4b shows a
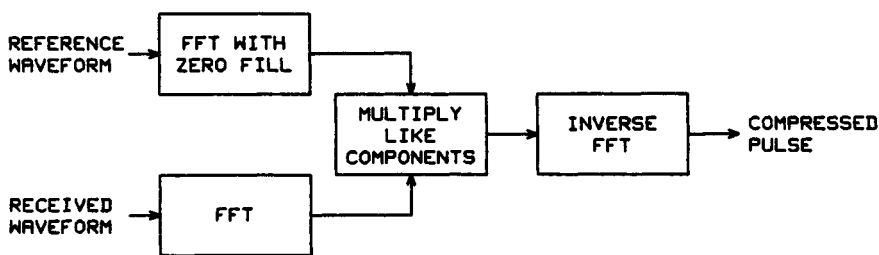
(a)
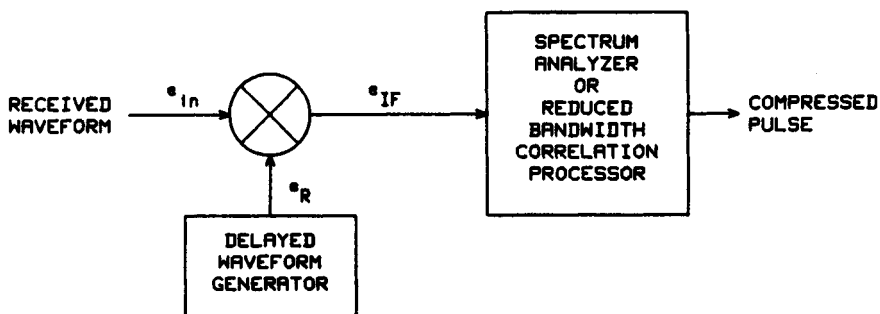
(b)

FIG. 10.4 Digital matched filter. (a) Correlation processor. (b) Stretch processor.

stretch processor for a linear-FM waveform. The delayed waveform has a bandwidth that is equal to or somewhat less than the transmitted waveform and a length that exceeds the duration of the transmitted waveform. This excess length equals the range window coverage.

The digital correlation processor[5] operates on the principle that the spectrum of the time convolution of two waveforms is equal to the product of the spectrum of these two signals. If M range samples are to be provided by one correlation processor, the number of samples in the fast Fourier transform (FFT) must equal M plus the number of samples in the reference waveform. These added M samples are filled with zeros in the reference waveform FFT. For extended range coverage, repeated correlation processor operations are required with range delays of M samples between adjacent operations. This correlation processor can be used with any waveform, and the reference waveform can be offset in doppler to achieve a matched filter at this doppler.

A stretch processor[6] can expand or contract the time scale of the compressed-pulse waveform within any defined time window. This general technique can be applied to any waveform, but it is much easier to use with a linear-FM waveform. For any waveform other than linear FM, an all-range pulse expansion approach is required in the received waveform path ahead of the mixer of Fig. 10.4b. Time contraction has not been applied to radar situations, as it requires an increased bandwidth for the compressed pulse. The stretch processing consideration will be restricted to time expansion of a linear-FM waveform.

Figure 10.4b shows the basic configuration of a time-expansion stretch processor for a linear-FM waveform. Let the received waveform be given by

$$e_{in} = A \, \text{rect}\left(t - \frac{\tau_{in}}{T_{in}}\right) \sin \left[2\pi(f_0 + f_d)(t - \tau_{in}) + \pi\alpha_{in}(t - \tau_{in})^2 + \phi\right]$$

where rect (X/T) is a unit amplitude pulse of duration T for $|X| \leq T/2$; τ_{in}, T_{in}, and α_{in} are the target time delay, the time pulse length, and the input frequency slope, respectively. The delayed waveform generator output will be

$$e_R = 2 \, \text{rect}\left(t - \frac{\tau_r}{T_R}\right) \sin \left[2\pi f_R(t - \tau_R) + \pi\alpha_R(t - \tau_R)^2 + \phi\right]$$

where the constants are the reference waveform equivalent of the received waveform constants. The intermediate-frequency (IF) input to the pulse compressor can easily be shown to be

$$e_{IF} = A \, \text{rect}\left(t - \frac{\tau_{in}}{T}\right)\text{rect}\left(t - \frac{\tau_R}{T_R}\right)$$
$$\cos \left[2\pi(f_o + f_d - f_R)(t - \tau_{in}) + \pi(\alpha_{in} - \alpha_R)\right.$$
$$\left.(t - \tau_{in})^2 + 2\pi\alpha_R(\tau_R - \tau_{in})(t - \tau_{in}) + \psi\right]$$

The resultant waveform is a reduced-frequency-slope linear-FM waveform with a target-range-dependent frequency offset riding on the doppler-shifted IF carrier frequency. Note that the frequency slope of the received waveform will be modified by the target's velocity.

For the special case where the two frequency slopes are equal, the IF

waveform is a constant-frequency pulse with an offset of $f_d + \alpha_R (\tau_R - \tau_{in})$. A spectrum analysis of this IF signal will yield the relative target range $(\tau_R - \tau_{in})$ information. This frequency offset (exclusive of the target doppler) can be rewritten as $B (\Delta T/T)$, where B is the transmitted waveform bandwidth and ΔT is the time separation between the two waveforms. If the waveform bandwidth is 1 GHz and the analyzer can process only a 10-MHz bandwidth, the range coverage is restricted to under 1 percent of the transmitted waveform length. To increase the range coverage, a wider processing bandwidth is required. This stretch approach allows the full range resolution of a wide-bandwidth waveform to be realized with a restricted bandwidth processor. Note that the duration of the reference waveform should exceed the duration of the received waveform by the range processed interval, or else an S/N loss will occur.

A stretch processor with unequal-frequency-slope waveforms requires pulse compression of the residual linear FM. A linear FM with a frequency slope of $\alpha_{in} - \alpha_R$ occurs at the target's range. This linear FM will be offset in frequency by $\alpha_R \Delta_T$. With the range-doppler coupling of the linear-FM waveform, the apparent range of this target will be

$$\tau_{app} = -\alpha_R \Delta T/(\alpha_{in} - \alpha_R)$$

This results in a time-expansion factor of $\alpha_R/(\alpha_{in} - \alpha_R)$ for the compressed pulse. Again the range coverage capability of the system depends on the processing bandwidth that can be implemented.

Surface-Wave Pulse Compression. A SAW pulse compression unit consists of an input transducer and an output transducer mounted on a piezoelectric substrate. These transducers are usually implemented as interdigital devices which consist of a metal film deposited on the surface of the acoustic medium. This metal film is made of fingers (see Fig. 10.5) that dictate the frequency characteristic of the unit. The input transducer converts an electrical signal into a sound wave with over 95 percent of the energy traveling along the surface of the medium. The output transducer taps a portion of this surface sound wave and converts it back into an electric signal.

The SAW device[7-9] has unique features that dictate its usefulness for a given radar application. The major shortcomings of the SAW approach are that the waveform length is restricted to under 200 μs by the physical size of available crystals and that each waveform requires another design. The advantages of the SAW device are its compact size, the wide bandwidths that can be attained, the ability to tailor the transducers to a particular waveform, the all-range coverage of the device, and the low cost of reproducing a given design.

SAW pulse compression devices depend on the interdigital transducer finger locations or else the surface-etched grating to determine its bandpass characteristic. Figure 10.5 shows three types of filter determination approaches. Figure 10.5a has a wideband input transducer and a frequency-selective (dispersive) output transducer. When an impulse is applied to the input, the output signal is initially a low frequency that increases (based on the output transducer finger spacings) at later portions of the pulse. This results in an up-chirp waveform which would be a matched filter for a down-chirp transmitted waveform. In Figure 10.5b, both the input transducer and the output transducer are dispersive. This would result in the same impulse response as that of Fig. 10.5a. For a given crystal length and material, the waveform duration for approaches in Fig. 10.5a and b would be the same and is limited to the time that it takes an acoustic wave to
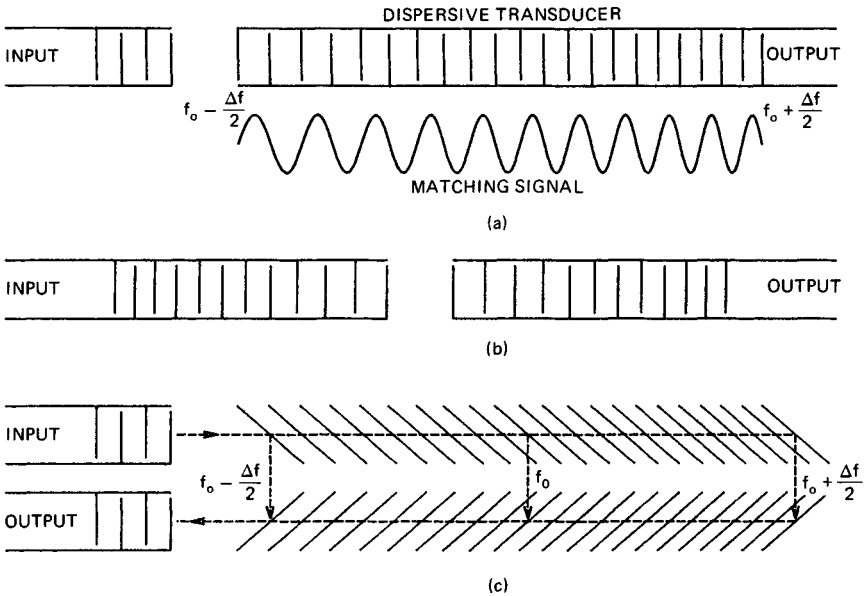
FIG. 10.5 SAW transducer types. (*a*) Dispersive output. (*b*) Both input and output dispersive. (*c*) Dispersive reflections.

traverse the crystal length. Figure 10.5*c* shows a reflection-array-compression (RAC) approach[10] which essentially doubles the achievable pulse length for the same crystal length. In an RAC, the input and output transducers have a broad bandwidth. A frequency-sensitive grating is etched on the crystal surface to reflect a portion of the surface-wave signal to the output transducer. This grating coupling does not have a significant impact on the surface-wave energy. Except for a 2:1 increase in the waveform duration, the impulse response of the RAC is the same as for approaches in Fig. 10.5*a* and *b*. Thus, these three approaches yield a similar impulse response.

Figure 10.6 shows a sketch of a SAW pulse compression device with dispersive input and output transducers. As the energy in a SAW device is concentrated in its surface wave, the SAW approach is much more efficient than bulk-wave devices, where the wave travels through the crystal. The propagation velocity of the surface wave is in the range of 1500 to 4000 m/s, depending on the crystal material, and allows a large delay in a compact device. Acoustic absorber material is required at the crystal edges to reduce the reflections and, hence, the spurious responses. Figure 10.7 shows the limit that can be expected from an SAW device and shows that bandwidths up to 1 GHz and delays up to 200 μs are achievable. The upper frequency limit depends on the accuracy that can be achieved in the fabrication of the interdigital transducer. The SAW device must provide a response that is centered on a carrier, as the lowest frequency of operation is about 20 MHz and is limited by the crystal. A matched-filter SAW pulse compression device can use variable finger lengths to achieve frequency weighting, and this internal weighting can correct for the Fresnel wiggles[11] in the FM spectrum. With this correction, 43 dB time-sidelobe levels can be achieved
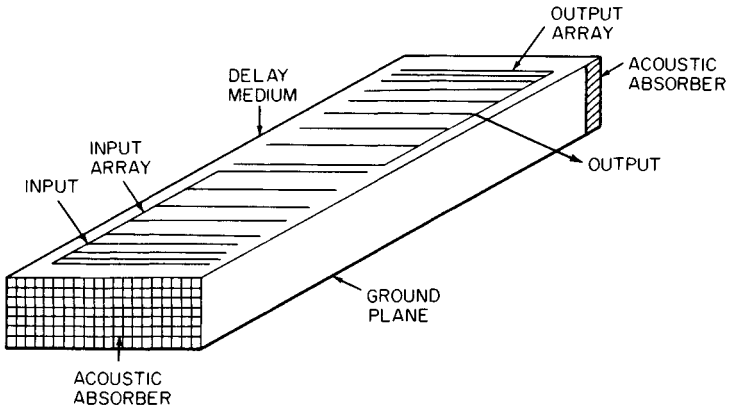
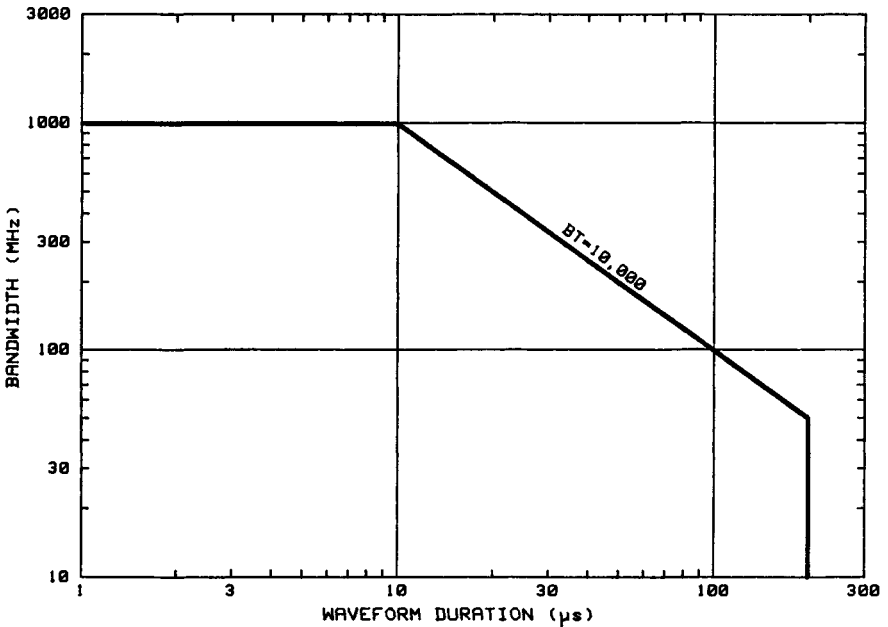**FIG. 10.6**   Surface-wave delay line.

FIG. 10.7   Waveform limits for a SAW device.

for a linear-FM waveform with a *BT* as low as 15. The dynamic range is limited to under 80 dB by nonlinearities in the crystal material. The most common SAW materials are quartz and lithium niobate.

Other Passive Linear-FM Devices.   Table 10.2 summarizes the general characteristics of several other passive devices that are used for linear-FM

TABLE 10.2 Characteristics of Passive Linear-FM Devices

	B, MHz	T, μs	BT	f_0, MHz	Typical loss, dB	Typical spurious, dB
Aluminum strip delay line	1	500	200	5	15	−60
Steel strip delay line	20	350	500	45	70	−55
All-pass network	40	1000	300	25	25	−40
Perpendicular diffraction delay line	40	75	1000	100	30	−45
Surface-wave delay line	40	50	1000	100	70	−50
Wedge-type delay line	250	65	1000	500	50	−50
Folded-tape meander line	1000	1.5	1000	2000	25	−40
Waveguide operated near cutoff	1000	3	1000	5000	60	−25
YIG crystal	1000	10	2000	2000	70	−20

pulse compression. These passive devices fall into two broad classes: (1) bulk ultrasonic devices in which an electrical signal is converted into a sonic wave and propagates through the medium and (2) electrical devices that use the dispersive characteristic of an electrical network. The main objectives in designing and selecting a device are (1) a flat-amplitude characteristic over the bandwidth B, (2) a linear delay slope with a differential delay T across the bandwidth B, (3) minimum spurious responses and minimum distortion to achieve low sidelobes, and (4) a low insertion loss.

In a bulk ultrasonic device the input electrical signal is transformed into an acoustic wave, propagates through a medium at sonic speeds, and is then converted back to an electrical signal at the output. Since the wave propagates at sonic speeds, longer delays are achieved than with an electrical device of comparable size. A major disadvantage of ultrasonic devices is that the transducers required for coupling electrically to the acoustic medium are inefficient energy converters and hence cause high insertion losses. The most common types of bulk ultrasonic dispersive devices are (1) strip delay lines, (2) perpendicular diffraction delay lines, (3) wedge delay lines, and (4) yttrium iron garnet (YIG) crystals. The strip delay line and the YIG crystal depend on the dispersive nature of the medium for their operation. The other two types use a nondispersive medium and depend upon the diffraction characteristics of the input and output transducers for their operation; hence they are called grating-type delay lines.

A strip delay line[12-15] is made of a long, thin strip of material with transducers at opposite ends. Since the strips must be extremely thin (of the order of a few milli-inches), metal is selected because of its ruggedness. Aluminum and steel are the only metals that have found wide application. The dispersive strip delay line uses the phenomenon that if acoustic energy is propagated through a medium as a longitudinal wave, the medium exhibits a nearly linear delay-versus-frequency characteristic over an appreciable frequency range. The strip width is not critical as long as it is greater than 10 acoustic wavelengths. The thickness, however, is very critical and must be about one-half of an acoustic wavelength at a frequency equal to the center of the linear delay-versus-frequency characteristic. The length of the strip is a linear function of the differential delay required, but the bandwidth is independent of length. The differential delay corresponds to the time

separation between the initial frequency and the final frequency of the waveform and is usually equal to the expanded pulse width T.

Because the thickness is very critical and cannot be controlled adequately, the stripline is placed in an oven whose temperature is adjusted to control the final operating frequency. One side of the strip is treated with an absorbing material to prevent reflections which could excite a wave that is not longitudinal and could thus introduce spurious signals.

Aluminum strip delay lines have the lowest losses, but their center frequency and bandwidth must be kept low. It is necessary to operate these lines below about 5 MHz if differential delays of over 50 μs are required. Aluminum lines have a midband delay of 7 to 10 μs/in.

Steel strip delay lines have high losses but operate at higher center frequencies, permitting wider bandwidths. Steel lines have typical losses of 70 to 80 dB and operating frequencies between 5 and 45 MHz. Steel lines have midband delays of 9 to 12 μs/in.

The perpendicular diffraction delay line[13,14,16] uses a nondispersive delay medium, such as quartz, with nonuniform input and output array transducers arranged on adjacent, perpendicular faces of the medium to produce the dispersion. The array element spacings decrease with increasing distance from the vertex of the right angle between the arrays. Thus only a positive slope of delay versus frequency can be produced. The bandwidth of the device is dictated by the array designs, and the delay is controlled by the size of the device. Errors in the array spacings produce phase errors which generate amplitude ripples and delay nonlinearities. Since many paths exist at a given frequency, these delay and amplitude errors tend to average out. Because of the averaging of the phase errors, the best delay linearity is achieved when the maximum number of grating lines is used. The center-frequency delay is limited to less than 75 μs for normal lines and 225 μs for polygonal lines because of limitations on the size of the quartz. In polygonal lines, the acoustic wave reflects off several reflecting faces in traveling from the input to the output array.

The wedge-type dispersive delay line[14] uses a wedge of quartz crystal and a frequency-selective receiver array to produce a linear delay-versus-frequency characteristic. The input transducer has a wide bandwidth, and the receiving-array elements are spaced in a quadratic manner. Reversal of the spacing of the output-array elements will change the output from an up-chirp waveform to a down-chirp waveform. The delay slope is dependent on the output-array configuration and the wedge angle. This device is fairly sensitive to grating phase errors since there is only one delay path per frequency.

YIG crystals[15,17] provide a dispersive microwave delay. YIG devices do not have a linear delay-versus-frequency characteristic, but their delay characteristic is very repeatable. The crystals require an external magnetic field, and the bandwidth and center frequency increase with the field strength. The delay of a YIG is determined by the crystal length. The maximum crystal length is limited to about 1.5 cm, corresponding to a delay of about 10 μs.

In the electrical-network class of linear-FM waveform generators, a signal is passed through an electrical delay network designed to have a linear delay-versus-frequency characteristic. The most common electrical networks that are used to generate linear-FM waveforms are (1) all-pass networks, (2) folded-tape meander lines, and (3) waveguide operated near its cutoff frequency. The all-pass network is a low-frequency device that uses lumped constant elements. The other two networks operate at very high frequencies and depend upon distributed parameters for delay.

An all-pass time-delay network[18,19] is ideally a four-terminal lattice network with constant gain at all frequencies and a phase shift that varies with the square of the frequency to yield a constant delay slope. The networks have equal input and output impedances so that several networks can be cascaded to increase the differential delay.

The folded-tape meander line[20] is the UHF or microwave analog of the low-frequency, all-pass network. A meander line consists of a thin conducting tape extending back and forth midway between two ground planes. The space between tape meanders and between the tape and the ground plane is filled with dielectric material. The center frequency of a meander loop is the frequency at which the tape length is λ/4. The time delay per meander loop is a function of the dimensions of the loop and the distance from the ground plane. To achieve a linear delay-versus-frequency curve, several loops with staggered delay characteristics are used in series. The number of meander loops required is greater than $B\Delta T$.

Other microwave dispersive networks include a waveguide operated near its cutoff frequency and stripline all-pass networks. If a section of rectangular waveguide is operated above its cutoff frequency, the time delay through the waveguide decreases with frequency. Over a limited frequency band, delay is a linear function of frequency. The usable frequency band and the delay linearity are significantly improved by employing a tapered-waveguide structure. Since stripline all-pass networks are microwave counterparts of the low-frequency all-pass networks, the synthesis of these networks is usually based on the low-frequency approach.

Voltage-Controlled Oscillator. A voltage-controlled oscillator (VCO) is a frequency generation device in which the frequency varies with an applied voltage. Ideally, the frequency is a linear function of the applied voltage, but most devices have a linearity error of over 1 percent. If a linear voltage ramp is applied to an ideal VCO, a linear-FM waveform is generated. A linear voltage ramp can be generated by applying a voltage step to an analog integrator. The integrator must be reset at the end of the generated pulse. If the VCO has a defined nonlinearity characteristic, the voltage into the integrator can be varied during the pulse so that the voltage ramp compensates for the VCO nonlinearity. Precompensation of this type is often employed. The characteristics of several common VCO devices are given in Table 10.3. The frequency-versus-voltage characteristic of the backward-wave oscillator is exponential; all the others have a linear characteristic. If coherent operation of the VCO is required, the output signal must be phased-locked to a coherent reference signal.

10.6 PHASE-CODED WAVEFORMS

Phase-coded waveforms differ from FM waveforms in that the pulse is subdivided into a number of subpulses. The subpulses are of equal duration, and each has a particular phase. The phase of each subpulse is selected in accordance with a given code sequence. The most widely used phase-coded waveform employs two phases and is called binary, or biphase, coding. The binary code consists of a sequence of either 0s and 1s or +1s and −1s. The phase of the transmitted signal alternates between 0° and 180° in accordance with the sequence of elements,

TABLE 10.3 Characteristics of VCO Devices

VCO device	Center-frequency range	Maximum frequency deviation as percent of center frequency, %	Maximum linearity as percent of deviation, %	Maximum center-frequency stability	Comments
LC oscillator	Up to 50 MHz	± 15	± 0.5	± 10 to ±100 ppm	
Crystal oscillator	100 kHz to 300 MHz	± 0.25	± 1	± 1 to ±10 ppm	
Three-terminal gallium arsenide oscillator	60 to 2500 MHz	± 2	± 2	±1%	
Voltage-tunable magnetron	100 to 10,000 MHz	± 50	± 1	±0.2%	Requires anode-voltage-control range of 750 to 3000 V.
Backward-wave oscillator	2 to 18 GHz	± 20	± 0.3*	±0.2%	Requires helix-voltage-control range of 400 to 1500 V.

*Deviation from an exponential frequency-versus-voltage curve.

0s and 1s or $+1$s and -1s, in the phase code, as shown in Fig. 10.8. Since the transmitted frequency is not usually a multiple of the reciprocal of the subpulse width, the coded signal is generally discontinuous at the phase-reversal points.
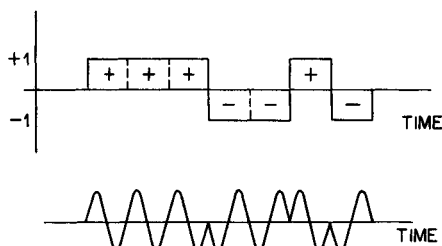
FIG. 10.8 Binary phase-coded signal.

Upon reception, the compressed pulse is obtained by either matched filtering or correlation processing. The width of the compressed pulse at the half-amplitude point is nominally equal to the subpulse width. The range resolution is hence proportional to the time duration of one element of the code. The compression ratio is equal to the number of subpulses in the waveform, i.e., the number of elements in the code.

Optimal Binary Sequences. Optimal binary sequences are binary sequences whose peak sidelobe of the aperiodic autocorrelation function (see Fig. 10.10b below) is the minimum possible for a given code length. Codes whose autocorrelation function, or zero-doppler responses, exhibit low sidelobes are desirable for pulse compression radars. Responses due to moving targets will differ from the zero-doppler response. However, with proper waveform design the doppler/bandwidth ratio can usually be minimized so that good doppler response is obtained over the target velocities of interest. The range-doppler response, or ambiguity diagram, over this velocity region then approximates the autocorrelation function.

Barker Codes. A special class of binary codes is the Barker[21] codes. The peak of the autocorrelation function is N, and the magnitude of the minimum peak sidelobe is 1, where N is the number of subpulses or length of the code. Only a small number of these codes exist. All the known Barker codes are listed in Table 10.4 and are the codes which have a minimum peak sidelobe of 1. These codes would be ideal for pulse compression radars if longer lengths were available. However, no Barker codes greater than 13 have been found to exist.[22–24] A pulse compression radar using these Barker codes would be limited to a maximum compression ratio of 13.

Allomorphic Forms. A binary code may be represented in any one of four allomorphic forms, all of which have the same correlation characteristics. These forms are the code itself, the inverted code (the code written in reverse order), the complemented code (1s changed to 0s and 0s to 1s), and the inverted complemented code. The number of codes listed in Table 10.4 is the number of codes, not including the allomorphic forms, which have the same minimum peak sidelobe. For example, the following 7-bit Barker codes all have the same

TABLE 10.4 Optimal Binary Codes

Length of code N	Magnitude of minimum peak sidelobe	No. of codes	Code (octal notation* for $N > 13$)
2	1	2	11,10
3	1	1	110
4	1	2	1101,1110
5	1	1	11101
6	2	8	110100
7	1	1	1110010
8	2	16	10110001
9	2	20	110101100
10	2	10	1110011010
11	1	1	11100010010
12	2	32	110100100011
13	1	1	1111100110101
14	2	18	36324
15	2	26	74665
16	2	20	141335
17	2	8	265014
18	2	4	467412
19	2	2	1610445
20	2	6	3731261
21	2	6	5204154
22	3	756	11273014
23	3	1021	32511437
24	3	1716	44650367
25	2	2	163402511
26	3	484	262704136
27	3	774	624213647
28	2	4	1111240347
29	3	561	3061240333
30	3	172	6162500266
31	3	502	16665201630
32	3	844	37233244307
33	3	278	55524037163
34	3	102	144771604524
35	3	222	223352204341
36	3	322	526311337707
37	3	110	1232767305704
38	3	34	2251232160063
39	3	60	4516642774561
40	3	114	14727057244044

*Each octal digit represents three binary digits:

0	000	4	100
1	001	5	101
2	010	6	110
3	011	7	111

autocorrelation peak value and the same minimum peak sidelobe magnitude: 1110010, 0100111, 0001101, 1011000. For symmetrical codes, the code and its inverse are identical.

Other Optimal Codes.   Table 10.4 lists the total number of optimal binary codes for all N up through 40 and gives one of the codes for each N. As an example, the minimum peak sidelobe for a 19-bit code is 2. There are two codes having this minimum peak sidelobe, one of which is 1610445 = 1 110 001 000 100 100 101. Computer searches are generally used to find optimal codes.[25] However, the search time becomes excessively long as N increases, and recourse is often made to using other sequences which may not be optimal but possess desirable correlation characteristics.

Maximal-Length Sequences.   The maximal-length sequences are of particular interest. They are the maximum-length sequences that can be obtained from linear-feedback shift-register generators. They have a structure similar to random sequences and therefore possess desirable autocorrelation functions. They are often called pseudorandom (PR) or pseudonoise (PN) sequences. A typical shift-register generator is shown in Fig. 10.9. The n stages of the shift register are initially set to all 1s or to combinations of 0s and 1s. The special case of all 0s is not allowed, since this results in an all-zero sequence. The outputs from specific individual stages of the shift register are summed by modulo-2 addition to form the input. Modulo-2 addition depends only on the number of 1s being added. If the number of 1s is odd, the sum is 1; otherwise, the sum is 0. The shift register is pulsed at the clock-frequency, or shift-frequency, rate. The output of any stage is then a binary sequence. When the feedback connections are properly chosen, the output
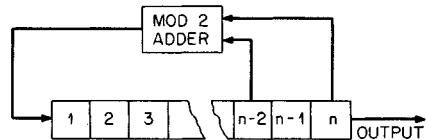
FIG. 10.9   Shift-register generator.

is a sequence of maximal length. This is the maximum length of a sequence of 1s and 0s that can be formed before the sequence is repeated.

The length of the maximal sequence is $N = 2^n - 1$, where n is the number of stages in the shift-register generator. The total number M of maximum-length sequences that may be obtained from an n-stage generator is

$$M = \frac{N}{n}\Pi\left(1 - \frac{1}{p_i}\right)$$

where p_i are the prime factors of N. The fact that a number of different sequences exist for a given value of n is important for applications where different sequences of the same length are required.

The feedback connections that provide the maximal-length sequences may be determined from a study of primitive and irreducible polynomials. An extensive list of these polynomials is given by Peterson and Weldon.[26]

Table 10.5 lists the length and number of maximal-length sequences obtainable from shift-register generators consisting of various numbers of stages. A feedback connection for generating one of the maximal-length sequences is also given for each. For a seven-stage generator, the modulo-2 sum of stages 6 and 7 is fed back to the input. For an eight-stage generator, the modulo-2 sum of stages 4, 5, 6, and 8 is fed back to the input. The length N of the maximal-length sequence is

TABLE 10.5 Maximal-Length Sequences

Number of stages, n	Length of maximal sequence, N	Number of maximal sequences, M	Feedback-stage connections
2	3	1	2,1
3	7	2	3,2
4	15	2	4,3
5	31	6	5,3
6	63	6	6,5
7	127	18	7,6
8	255	16	8,6,5,4
9	511	48	9,5
10	1,023	60	10,7
11	2,047	176	11,9
12	4,095	144	12,11,8,6
13	8,191	630	13,12,10,9
14	16,383	756	14,13,8,4
15	32,767	1,800	15,14
16	65,535	2,048	16,15,13,4
17	131,071	7,710	17,14
18	262,143	7,776	18,11
19	524,287	27,594	19,18,17,14
20	1,048,575	24,000	20,17

equal to the number of subpulses in the sequence and is also equal to the time-bandwidth product of the radar system. Large time-bandwidth products can be obtained from registers having a small number of stages. The bandwidth of the system is determined by the clock rate. Changing both the clock rate and the feedback connections permits the generation of waveforms of various pulse lengths, bandwidths, and time-bandwidth products. The number of zero crossings, i.e., transitions from 1 to 0 or from 0 to 1, in a maximal-length sequence is 2^{n-1}.

Periodic waveforms are obtained when the shift-register generator is left in continuous operation. They are sometimes used in CW radars. Aperiodic waveforms are obtained when the generator output is truncated after one complete sequence. They are often used in pulsed radars. The autocorrelation functions for these two cases differ with respect to the sidelobe structure. Figure 10.10 gives the autocorrelation functions for the periodic and aperiodic cases for a typical 15-element maximal-length code obtained from a four-stage shift-register generator. The sidelobe level for the periodic case is constant at a value of -1. The periodic autocorrelation function is repetitive with a period of $N\tau$ and
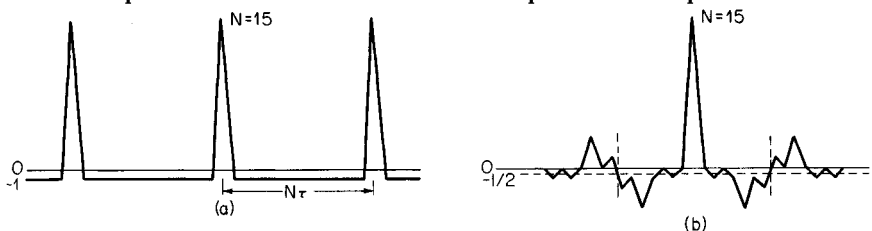
FIG. 10.10 Autocorrelation functions for (a) the periodic case and (b) the aperiodic case.

a peak value of N, where N is the number of subpulses in the sequence and τ is the time duration of each subpulse. Hence the peak-sidelobe–voltage ratio is N^{-1}.

For the aperiodic case, the average sidelobe level along the time axis is $-\frac{1}{2}$. The sidelobe structure of each half of the autocorrelation function has odd symmetry about this value. The periodic autocorrelation function may be viewed as being constructed by the superposition of successive aperiodic autocorrelation functions, each displaced in time by $N\tau$ units. The odd symmetry exhibited by the aperiodic function causes the sidelobe structure for the periodic function to have a constant value of -1. When the periodic waveform is truncated to one complete sequence, this constant sidelobe property is destroyed. For large N the peak-sidelobe–voltage ratio is approximately $N^{-1/2}$ for the aperiodic case.

Maximal-length sequences have characteristics which approach the three randomness characteristics ascribed to truly random sequences,[27] namely, that (1) the number of 1s is approximately equal to the number of 0s; (2) runs of consecutive 1s and 0s occur with about half of the runs having a length of 1, a quarter of length 2, an eighth of length 3, etc.; and (3) the autocorrelation function is thumbtack in nature, i.e., peaked at the center and approaching zero elsewhere. Maximal-length sequences are of odd length. In many radar systems it is desirable to use sequence lengths of some power of 2. A common procedure is to insert an extra 0 in a maximal-length sequence. This degrades the autocorrelation function sidelobes somewhat. An examination of sequences with an inserted 0 will yield the sequence with the best autocorrelation characteristics.

Quadratic Residue Sequences. Quadratic residue (p. 254 of Ref. 26), or Legendre, sequences offer a greater selection of code lengths than are available from maximal-length sequences. Quadratic residue sequences satisfy two of the randomness characteristics: the periodic autocorrelation function is as shown in Fig. 10.10a having a peak of N and a uniform sidelobe level of -1, and the number of 1s is approximately the same as the number of 0s.

A quadratic residue sequence of length N exists if $N = 4t - 1$, with N a prime and t any integer. The code elements a_i for $i = 0, 1, 2, \ldots, N - 1$ are 1 if i is a quadratic residue modulo N and -1 otherwise. Quadratic residues are the remainders where x^2 is reduced modulo N for $x = 1, 2, \ldots, (N - 1)/2$. As an example, the quadratic residues for $N = 11$ are 1, 3, 4, 5, 9. Hence the code elements a_i for $i = 1, 3, 4, 5, 9$ are 1, and the sequence is $-1, 1, -1, 1, 1, 1, -1, -1, -1, 1, -1$, or 10100011101. The periodic autocorrelation function of this sequence has a peak of 11 and a uniform sidelobe level of -1. Also, the numbers of 1s and 0s are approximately equal; the number of 1s is one more than the number of 0s.

Complementary Sequences. Complementary sequences consist of two sequences of the same length N whose aperiodic autocorrelation functions have sidelobes equal in magnitude but opposite in sign. The sum of the two autocorrelation functions has a peak of $2N$ and a sidelobe level of zero. Figure 10.11 shows the individual autocorrelation functions of the complementary sequences for length 26 and also the sum of the two autocorrelation functions. Golay[28,29] and Hollis[30] discuss general methods for forming complementary codes. In general, N must be an even number and the sum of two squares. In a practical application, the two sequences must be separated in time, frequency, or polarization, which results in decorrelation of radar returns so that complete sidelobe cancellation may not occur. Hence they have not been widely used in pulse compression radars.
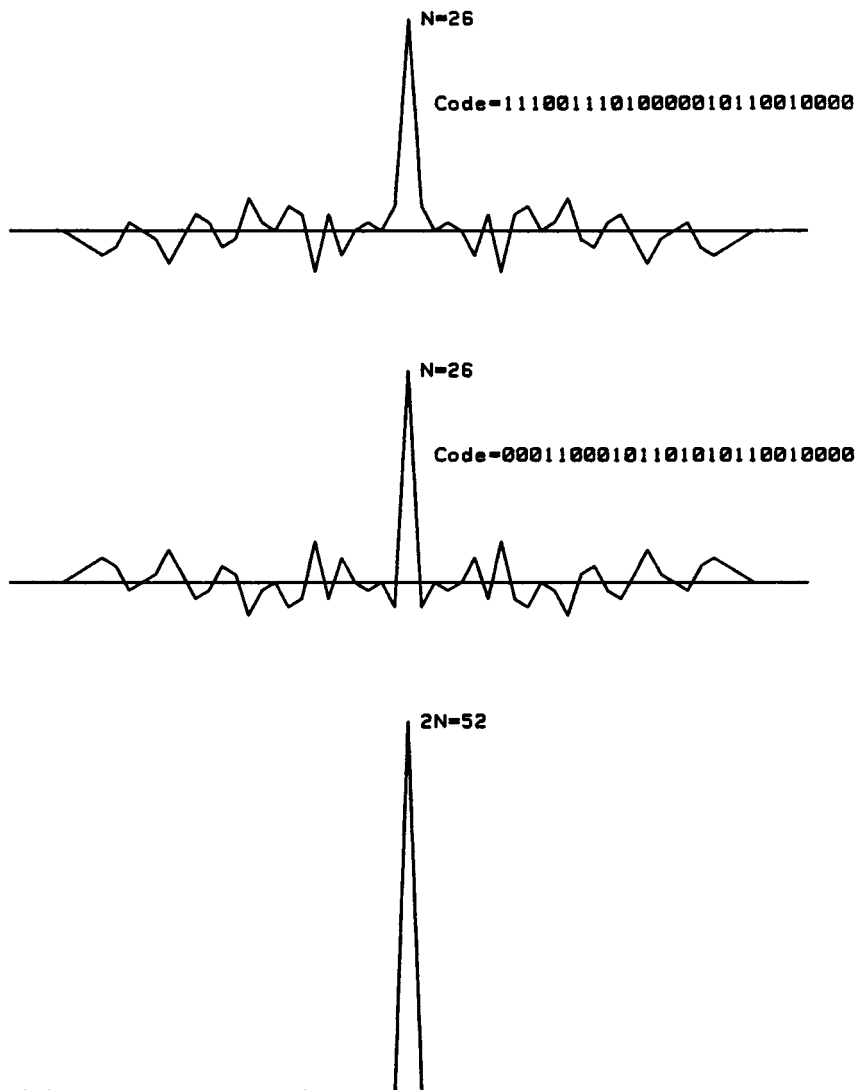
FIG. 10.11   Complementary-code aperiodic autocorrelation function.

Implementation of Biphase-Coded Systems. Digital implementation is generally used to perform the pulse compression operation in biphase-coded systems. A block diagram of a digital pulse compression system is given in Fig. 10.12. The code generator generates the binary sequence, which is sent to the RF modulator and transmitter and to the correlators. Received IF signals are passed through a bandpass filter matched to the subpulse width and are demodulated by I and Q phase detectors. The I and Q detectors compare the phase of the received IF signal with the phase of a local-oscillator (LO) signal at the same IF frequency. The LO signal is also used in the RF modulator to
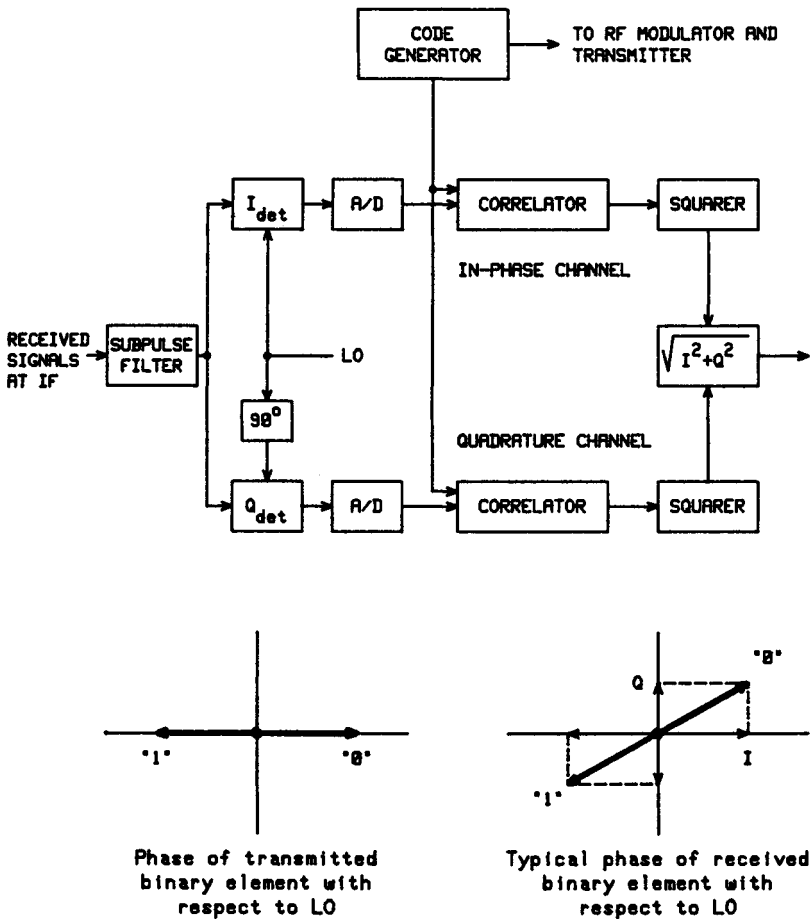
FIG. 10.12   Digital pulse compression for phase-coded signals.

generate the biphase-modulated transmitted signal. The phase of each transmitted binary element is 0° or 180° with respect to the LO signal. The phase of the received signal with respect to the LO signal, however, is shifted by an amount depending upon the target's range and velocity. Two processing channels are used, one which recovers the in-phase components of the received signal and the other which recovers the quadrature components. These signals are converted to digital form by analog-to-digital (A/D) converters, correlated with the stored binary sequence and combined, e.g., by the square root of the sum of the squares. A processing system of this type, which contains an in-phase and quadrature channel and two matched filters or correlators, is called a homodyne or zero IF system. There is an average loss in signal-to-noise ratio of 3 dB if only one channel is implemented instead of both I and Q channels. Each correlator may actually consist of several correlators, one for each quantization bit of the digitized signal.

Two methods of implementing the correlators are shown in Fig. 10.13. Fig-

ure 10.13*a* shows a fixed reference correlator; i.e., only one binary sequence is used. The received input sequence is continuously clocked into a shift register whose number of stages is equal to the number of elements in the sequence. The output of each stage is multiplied by weight a_i, which is either $+1$ or -1 in accordance with the reference sequence. The summation circuit provides the output correlation function or compressed pulse.

Figure 10.13*b* shows an implementation where the reference may be changed for each transmitted pulse. The transmitted reference sequence is fed into the reference shift register. The received input sequence is continuously clocked into the signal shift register. In each clock period the comparison counter forms the sum of the matches minus the sum of the mismatches between corresponding stages of the two shift registers, which is the output correlation function. In some systems, only the sum of the matches is counted and an offset of $-N/2$ is added to the sum.

Doppler Correction. In many applications the effect of doppler is negligible over the expanded pulse length, and no doppler correction or compensation is required. These applications transmit a short-duration phase-coded pulse, and
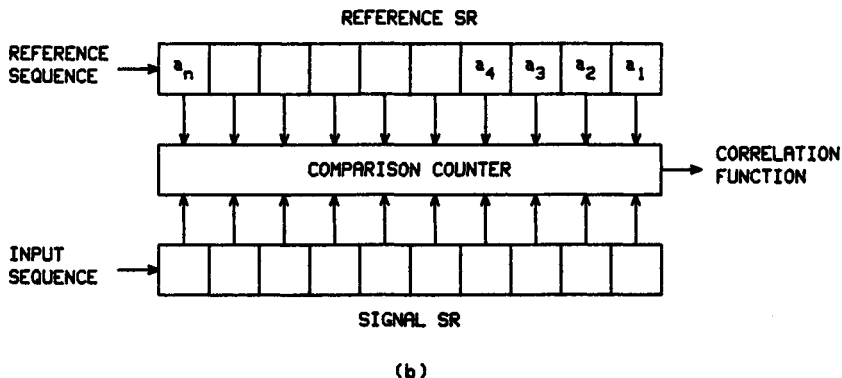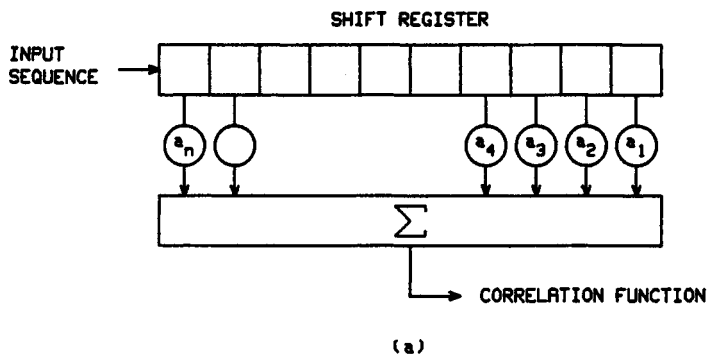
FIG. 10.13 Digital correlation with (*a*) fixed and (*b*) variable references.

the phase shift due to doppler over each expanded pulse width is negligible. Pulse compression is performed on each pulse. When the doppler shift over the expanded pulse width is not negligible, multiple doppler channels are required to minimize the loss in *SNR*. The received signals may be mixed with multiple LO signals (see Fig. 10.12), each offset in frequency by an amount corresponding to a doppler resolution element which is the reciprocal of the expanded pulse length. The processing following the subpulse filter in Fig. 10.12 is then duplicated for each doppler channel.

An alternative technique is to use a single LO signal and single-bit A/D converters in Fig. 10.12. Doppler compensation is performed on the outputs of the A/D converters prior to the correlators. This doppler compensation is in the form of inverting data bits, i.e., changing 1s to 0s and 0s to 1s, at time intervals corresponding to 180° phase shifts of the doppler frequency. As an example, the first doppler channel corresponds to a doppler frequency which results in a 360° phase shift over the pulse width. The bits are inverted after every half pulse width and remain inverted for a half pulse width. Bit inversion occurs at intervals of a quarter pulse width for the second doppler channel, an eighth pulse width for the third doppler channel, etc. Negative doppler frequency channels are handled in the same manner as for positive doppler frequency channels, but bits that were inverted in the corresponding positive channel are not inverted in the negative channel, and bits that were not inverted in the positive channel are inverted in the negative channel. No bit inversion occurs in the zero doppler channel. Each doppler channel consists of the single-bit *I* and *Q* correlators and the combiner, e.g., square root of the sum of the squares. After initial detection occurs, linear doppler processing may then be used to reduce the SNR loss. For example, the LO signal in Fig. 10.12 would then correspond to the doppler which resulted in the initial detection, and full A/D conversion is used. Some radar systems use long-duration pulses with single-bit doppler compensation to obtain initial detection and then switch to shorter-duration pulses which require no doppler compensation.

Polyphase Codes. Waveforms consisting of more than two phases may also be used.[31,32] The phases of the subpulses alternate among multiple values rather than just the 0° and 180° of binary phase codes. The Frank polyphase codes[33] derive the sequence of phases for the subpulses by using a matrix technique. The phase sequence can be written as $\phi_n = 2\pi i(n - 1)/P^2$, where P is the number of phases, $n = 0, 1, 2, \ldots, P^2 - 1$, and $i = n$ modulo P. For a three-phase code, $P = 3$, and the sequence is 0, 0, 0, 0, $2\pi/3$, $4\pi/3$, 0, $4\pi/3$, $2\pi/3$.

The autocorrelation function for the periodic sequence has time sidelobes of zero. For the aperiodic sequence, the time sidelobes are greater than zero. As P increases, the peak-sidelobe–voltage ratio approaches $(\pi P)^{-1}$. This corresponds to approximately a 10 dB improvement over pseudorandom sequences of similar length. The ambiguity response over the range-doppler plane grossly resembles the ridgelike characteristics associated with linear-FM waveforms, as contrasted with the thumbtack characteristic of pseudorandom sequences. However, for small ratios of doppler frequency to radar bandwidth, good doppler response can be obtained for reasonable target velocities.

Lewis and Kretschmer[34] have rearranged the phase sequence to reduce the degradation that may occur by receiver band limiting prior to pulse compression. The rearranged phase sequence is

$$\phi_n = \frac{n\pi}{P}\left[1 - P + \frac{2(n - i)}{P}\right] \quad \text{for } P \text{ odd}$$

$$\phi_n = \frac{\pi}{2P}(P - 1 - 2i)\left[P - 1 - \frac{2(n - i)}{P}\right] \quad \text{for } P \text{ even}$$

where P, n, and i are as defined above for the Frank code. For $P = 3$, the phase sequence is 0, $-2\pi/3$, $-4\pi/3$, 0, 0, 0, 0, $2\pi/3$, $4\pi/3$.

Generation and processing of polyphase waveforms use techniques similar to those for the FM waveforms of Sec. 10.5.

10.7 TIME-FREQUENCY-CODED WAVEFORMS

A time-frequency-coded waveform (Fig. 10.14) consists of a train of N pulses with each pulse transmitted at a different frequency. The ambiguity response for a periodic waveform of this type consists of a central spike plus multiple spikes or ridges displaced in time and frequency. The objective is to create a high-resolution, thumbtacklike central spike with a clear area around it; measurement is then performed on the high-resolution central spike. The range resolution or compressed pulse width is determined by the total bandwidth of all the pulses, and the doppler resolution is determined by the waveform duration T. For example, a typical waveform in this class has N contiguous pulses of width τ, whose spectra of width $1/\tau$ are placed side by side in frequency to eliminate gaps in the composite spectrum. Since the waveform bandwidth is now N/τ, the nominal compressed-pulse width is τ/N. Relationships are summarized in Table 10.6.
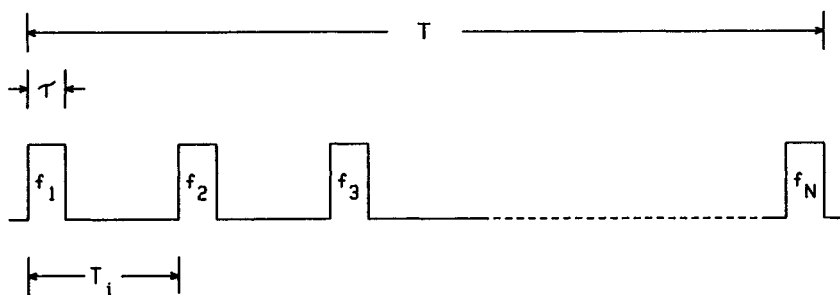
FIG. 10.14 Time-frequency-coded waveform.

TABLE 10.6 N Pulses Contiguous in Time and Frequency

Waveform duration, T	$N\tau$
Waveform bandwidth, B	N/τ
Time-bandwidth product, TB	N^2
Compressed pulse width, $1/B$	$\tau/N = T/N^2$

Shaping of the high-resolution central spike area as well as the gross structure of the ambiguity surface can be accomplished by variations of the basic waveform parameters such as amplitude weighting of the pulse train, staggering

of the pulse repetition interval, and frequency or phase coding of the individual pulses.[35]

10.8 WEIGHTING AND EQUALIZATION

The process of shaping the compressed-pulse waveform by adjustment of the amplitude of the frequency spectrum is known as *frequency weighting*. The process of shaping the doppler response by control of the waveform envelope shape is called *time weighting*. The primary objective of weighting in either domain is to reduce sidelobes in the other domain. Sidelobes can severely limit resolution when the relative magnitudes of received signals are large.

Paired Echoes and Weighting. A description of the weighting process is facilitated by the application of paired-echo theory.[36–39] The first seven entries in Table 10.7 provide a step-by-step development of Fourier transforms useful in frequency and time weighting, starting with a basic transform pair. The last entry pertains to phase-distortion echoes. The spectrum $G(f)$ of the time function $g(t)$ is assumed to have negligible energy outside the frequency interval $-B/2$ to $+B/2$, where B is the bandwidth in hertz. The transform pairs of Table 10.7 are interpreted as follows:
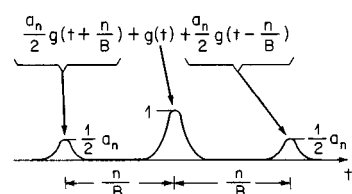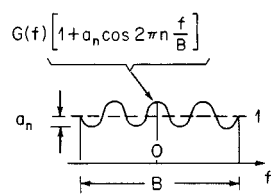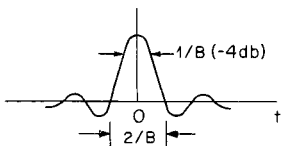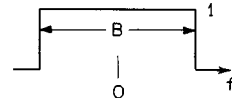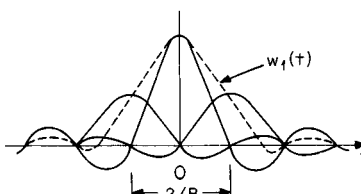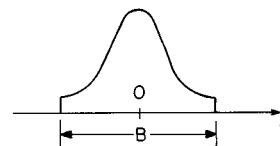
Pair 1. Cosinusoidal amplitude variation over the passband creates symmetrical paired echoes in the time domain in addition to the main signal $g(t)$, whose shape is uniquely determined by $G(f)$. The echoes are replicas of the main signal, delayed and advanced from it by n/B s and scaled in amplitude by $a_n/2$.

Pair 2. The rectangular frequency function $W_0(f)$, that is, uniform weighting over the band, leads to a $(\sin x)/x$ time function $w_0(t)$ with high-level sidelobes, which can be objectionable in some cases. A normalized logarithmic plot of the magnitude of this time function is shown by curve A in Fig. 10.15. (All functions illustrated are symmetrical about $t = 0$.) The sidelobe adjacent to the main lobe has a magnitude of -13.2 dB with respect to the main-lobe peak. The sidelobe falloff rate is very slow.

Pair 3. Taper is applied by introducing one amplitude ripple ($n = 1$) in the frequency domain to form $W_1(f)$. By pairs 1 and 2, the time function is the superposition of the three time-displaced and weighted $(\sin x)/x$ functions.[39] Low time sidelobes are attainable in the resultant function $w_1(t)$ by the proper choice of the coefficient F_1. In particular, $F_1 = 0.426$ corresponds to Hamming weighting[40–42] and to the time function whose magnitude is represented by the solid curve B in Fig. 10.15.

Pair 4. The frequency-weighting function includes a Fourier series of $\bar{n} - 1$ cosine terms, where the selection of $\bar{n}$ is determined by the required compressed pulse width and the desired sidelobe falloff. By pairs 1 and 2, the time function includes the superposition of $2(\bar{n} - 1)$ echoes that occur in $\bar{n} - 1$ symmetrical pairs. If the coefficients F_m are selected to specify the Taylor weighting function[39,42,43] $W_{Tay}(f)$, the corresponding resultant time function $w_{Tay}(t)$ exhibits good resolution characteristics by the criterion of small main-lobe width for a specified sidelobe level. Taylor coefficients chosen for a -40 dB sidelobe level, with $\bar{n}$ selected as 6, lead to the main-sidelobe structure indicated by curve C of Fig. 10.15.
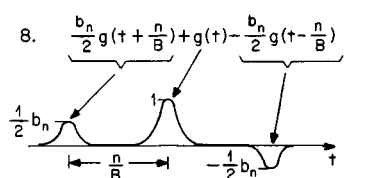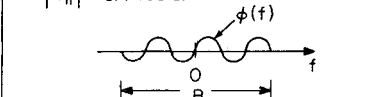
TABLE 10.7 Paired-Echo and Weighting Transforms

$g(t)=\int_{-\infty}^{\infty} G(f)\exp(j2\pi ft)df$	$G(f)=\int_{-\infty}^{\infty} g(t)\exp(-j2\pi ft)dt$
PAIRED ECHOES:	n AMPLITUDE RIPPLES:
1. $\dfrac{a_n}{2}g(t+\dfrac{n}{B})+g(t)+\dfrac{a_n}{2}g(t-\dfrac{n}{B})$	$G(f)\left[1+a_n\cos 2\pi n\dfrac{f}{B}\right]$
	(REFS. 36–39)
HIGH SIDELOBES (−13.2db):	UNIFORM WEIGHTING:
2. $w_0(t)=B\dfrac{\sin \pi Bt}{\pi Bt}$	$W_0(f)=\begin{cases}1 & \lvert f\rvert<\frac{1}{2}B \\ 0 & \lvert f\rvert>\frac{1}{2}B\end{cases}$
LOW SIDELOBES:	TAPER:
3. $w_1(t)=$ $F_1 w_0(t+\dfrac{1}{B})+w_0(t)+F_1 w_0(t-\dfrac{1}{B})$	$W_1(f)=$ $W_0(f)\left[1+2F_1\cos 2\pi\dfrac{f}{B}\right]$
	(REFS. 39–42)

Pairs 5 to 7. The duality theorem 5 permits the interchange of time and frequency functions in each of the preceding pairs. Functions may be interchanged if the sign of the parameter t is reversed. Examples are pairs 6 and 7 obtainable from pairs 2 and 4 with the substitution of T s for B Hz. Taylor time weighting is applied in pair 7 to achieve good frequency resolution when the coefficients are selected for a specified sidelobe level.

Pair 8. Similarly to the amplitude variations of pair 1, sinusoidal phase variation over the passband creates symmetrical paired echoes in the time domain in addition to the main signal $g(t)$. The echoes are replicas of the main signal, de-

TABLE 10.7 Paired-Echo and Weighting Transforms (*Continued*)

	TAYLOR WEIGHTING:
4. $w_{Tay}(t) = \sum_{m=-\infty}^{\infty} F_m w_0(t - \frac{m}{B})$ where $\qquad F_0 = 1$, $F_m = 0$ for $\|m\| \geq \bar{n}$ and $\qquad F_m = F_{-m}$	$W_{Tay}(f) =$ $W_0(f)\left[1 + 2 \sum_{m=1}^{\bar{n}-1} F_m \cos 2\pi m \frac{f}{B}\right]$ (REFS. 39, 42, 43)
DUALITY THEOREM: 5. $\qquad\qquad G(-t)$	$g(f)$
6. $W_0(t) = \begin{cases} 1 & \|t\| < \frac{T}{2} \\ 0 & \|t\| > \frac{T}{2} \end{cases}$	$w_0(f) = T \frac{\sin \pi f T}{\pi f T}$
7. $W_{Tay}(t) =$ $W_0(t)\left[1 + 2 \sum_{m=1}^{\bar{n}-1} F_m \cos 2\pi m \frac{t}{T}\right]$	$w_{Tay}(f) = \sum_{m=-\infty}^{\infty} F_m w_0(f - \frac{m}{T})$ (SEE PAIR No. 4)
PAIRED ECHOES: 8. $\underbrace{\frac{b_n}{2} g(t + \frac{n}{B})} + g(t) - \underbrace{\frac{b_n}{2} g(t - \frac{n}{B})}$ 	n PHASE RIPPLES: $G(f)e^{jb_n \sin 2\pi n \frac{f}{B}} \cong$ $\left[1 + jb_n \sin 2\pi n \frac{f}{B}\right]G(f)$ $\|b_n\| < 0.4$ radian (REFS. 36 – 39)

layed and advanced from it by n/B s, scaled in amplitude by $b_n/2$, and opposite in polarity.

Comparison of Weighting Functions. The performance achieved with various frequency-weighting functions is summarized in Table 10.8. With a change in parameter, the table also applies to time weighting (or weighting of the aperture distribution of an antenna). Pedestal height H is defined in all cases as the weighting-function amplitude at the band edge ($f = \pm B/2$) when the function has been normalized to unit amplitude at the band center ($f = 0$). The loss in the signal-to-noise ratio is based on the assumption that the transmitted amplitude spectrum is rectangular.

Item 1, uniform weighting, thus provides matched-filter operation with no
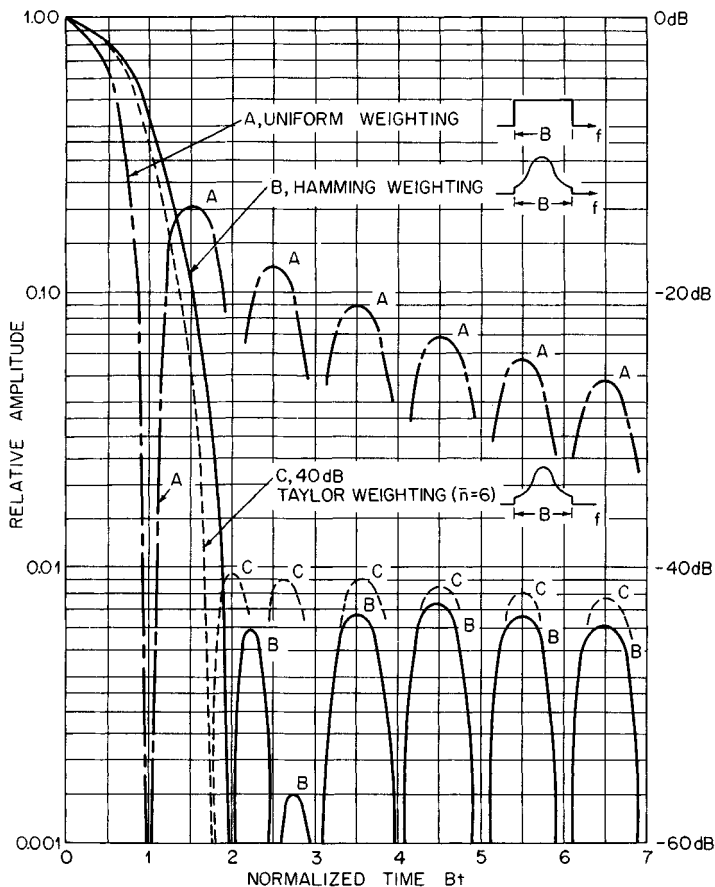
FIG. 10.15 Comparison of compressed-pulse shapes for three frequency-weighting functions.

SNR loss. Weighting in other cases is applied by a mismatch of the receiver amplitude characteristic. Item 2, Dolph-Chebyshev[44] weighting, is optimum in the sense of producing the minimum main-lobe width for a specified sidelobe level. However, the Dolph-Chebyshev function is physically unrealizable[39,41,42] for the continuous spectra under discussion. Item 3, Taylor weighting, provides a realizable approximation to Dolph-Chebyshev weighting. Time sidelobes have little decay in the region $B|t| \setminus \bar{n} - 1$ but decay at 6 dB per octave when $B|t| \hbar \bar{n}$. Item 4, cosine-squared-plus-pedestal weighting, becomes equivalent, after normalization and use of a trigonometric identity, to the weighting function $W_1(f)$ of pair 3 in Table 10.7. The normalized pedestal height H is related to the taper coefficient F_1 by $H = (1 - 2F_1)/(1 + 2F_1)$. The Hamming function produces the lowest sidelobe level attainable under category 4 of Table 10.8. Item 4b, 3:1 *taper ratio* (that is, $1/H = 3$), is analogous to a typical antenna distribution with power tapering to about 10 percent at the aperture edges.[45] Cosine-squared weighting without pedestal

TABLE 10.8 Performance for Various Frequency-Weighting Functions

	Weighting function	Pedestal height H, %	SNR loss, dB	Main-lobe width, -3 dB	Peak sidelobe level, dB	Far sidelobe falloff
1	Uniform	100	0	0.886/B	-13.2	6 dB/octave
2	Dolph-Chebyshev			1.2/B	-40	No decay
3	Taylor ($\bar{n} = 8$)	11	1.14	1.25/B	-40	6 dB*/octave
4	Cosine-squared plus pedestal:$H + (1 - H)$ $\cos^2 (\pi f/B)$					
	a. Hamming	8	1.34	1.33/B	-42.8	6 dB/octave
	b. 3:1 "taper ratio"	33.3	0.55	1.09/B	-25.7	6 dB/octave
5	$\cos^2 (\pi f/B)$	0	1.76	1.46/B	-31.7	18 dB/octave
6	$\cos^3 (\pi f/B)$	0	2.38	1.66/B	-39.1	24 dB/octave
7	$\cos^4 (\pi f/B)$	0	2.88	1.94/B	-47	30 dB/octave
8	Triangular: $1 - 2|f|/B$	0	1.25	1.27/B	-26.4	12 dB/octave

*In the region $|f| \hbar$ 8/B.

($H = 0$, $F_1 = 1/2$), listed as item 5, achieves a faster decay in far-off sidelobes and may simplify implementation. Entries 6 to 8 are of interest primarily because of the sidelobe falloff rate. The falloff rate can be shown to be related to the manner in which the frequency function and its derivatives behave at cutoff points, $f = \pm B/2$.[46,47]

Taylor versus Cosine-Squared-Plus-Pedestal Weighting. Figure 10.16a plots the taper coefficient F_1 and pedestal height H versus the peak sidelobe level for cosine-squared-plus-pedestal weighting. Table 10.9 lists Taylor coefficients F_m and main-lobe widths for various sidelobe levels and selections of $\bar{n}$.[48] The table illustrates that, for low design sidelobe levels, F_1 is much greater than $|F_m|$ when $m > 1$, indicating that Taylor weighting is closely approximated by the cosine-squared-plus-pedestal taper. A larger value of F_1 is required, however, in the latter case to yield the same sidelobe level. $F_1 = 0.426$ ($H = 0.08$), corresponding to Hamming weighting, produces the lowest level, -42.8 dB, attainable with this function. As indicated in Fig. 10.16a, larger values of $F_1(H < 0.08)$ increase the sidelobe level. For a given peak sidelobe level, Taylor weighting offers theoretical advantages in pulse width and SNR performance, as illustrated in Fig. 10.16b and c.

Taylor Weighting with Linear FM. The spectrum of a linear-FM pulse with a rectangular time envelope is not exactly rectangular in amplitude, nor is its phase exactly matched by the linear group delay of the compression filter.[2,39,42] The discrepancy is particularly severe for small time-bandwidth products. Therefore, the use of 40 dB Taylor weighting based on a simplified model which assumes a rectangular amplitude spectrum and a parabolic phase spectrum (that can be matched by the linear group delay) fails to achieve a -40 dB sidelobe level. Further degradation results when there is a doppler shift. Figure 10.17 plots the peak sidelobe level versus the target's doppler
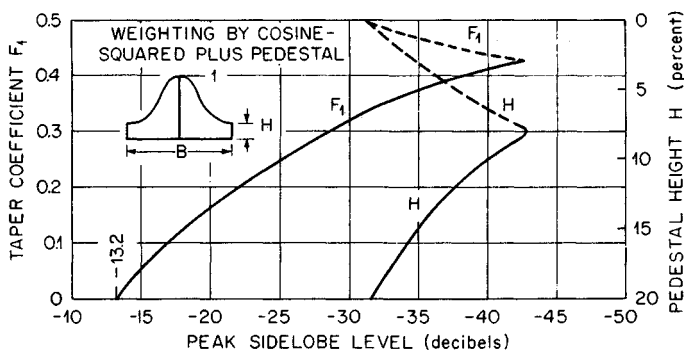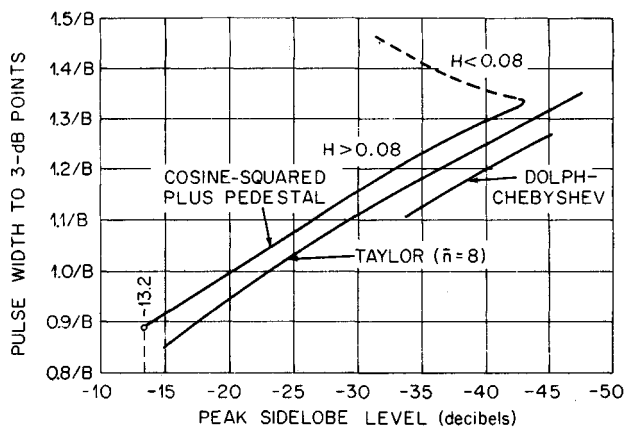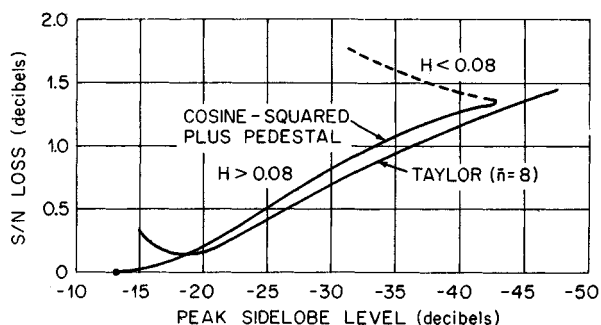
FIG. 10.16 (a) Taper coefficient and pedestal height versus peak side-lobe level. (b) Compressed-pulse width versus peak sidelobe level. (c) *SNR* loss versus peak sidelobe level.

TABLE 10.9 Taylor Coefficients F_m*

Design sidelobe ratio, dB	−30	−35	−40	−40	−45	−45	−50
$\bar{n}$	4	5	6	8	8	10	10
Main lobe width, −3 dB	1.13/B	1.19/B	1.25/B	1.25/B	1.31/B	1.31/B	1.36/B
F_1	0.292656	0.344350	0.389116	0.387560	0.428251	0.426796	0.462719
F_2	−0.157838(−1)	−0.151949(−1)	−0.945245(−2)	−0.954603(−2)	0.208399(−3)	−0.682067(−4)	0.126816(−1)
F_3	0.218104(−2)	0.427831(−2)	0.488172(−2)	0.470359(−2)	0.427022(−2)	0.420099(−2)	0.302744(−2)
F_4		−0.734551(−3)	−0.161019(−2)	−0.135350(−2)	−0.193234(−2)	−0.179997(−2)	−0.178566(−2)
F_5			0.347037(−3)	0.332979(−4)	0.740559(−3)	0.569438(−3)	0.884107(−3)
F_6				0.357716(−3)	−0.198534(−3)	0.380378(−5)	−0.382432(−3)
F_7				−0.290474(−3)	0.339759(−5)	−0.224597(−3)	0.121447(−3)
F_8						0.246265(−3)	−0.417574(−5)
F_9						−0.153486(−3)	−0.249574(−4)

*$F_0 = 1$; $F_{-m} = F_m$; floating decimal notation: $-0.945245(-2) = -0.00945245$.
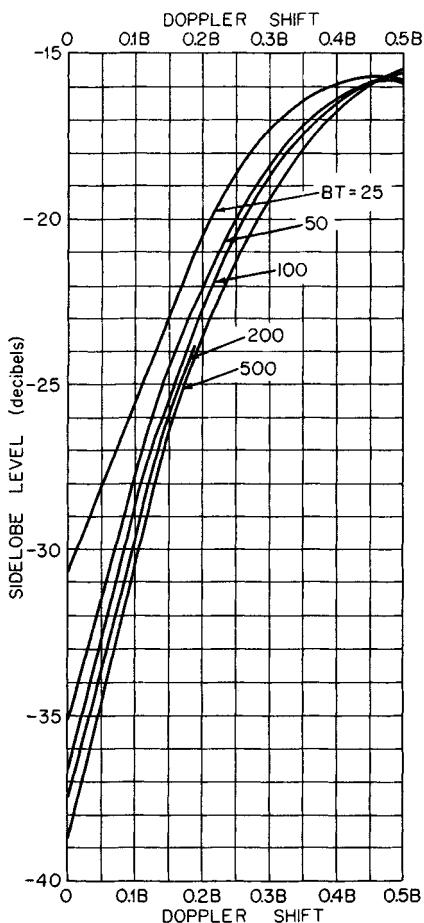
DOPPLER SHIFT

FIG. 10.17 Peak sidelobe level versus doppler shift for linear FM.
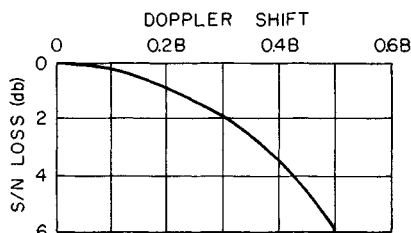
FIG. 10.18 Loss in signal-to-noise ratio versus doppler shift for linear FM.

frequency. As the time-bandwidth product is increased, the model rectangular spectrum with parabolic phase is approached, and the sidelobe level in the absence of doppler shift approaches −40 dB. Unless SAW compression networks that compensate for the nonideal spectrum are employed, equalization techniques described later in this section are needed when sidelobe levels lower than about −30 dB are required. In Fig. 10.18 the loss in signal-to-noise ratio is plotted as a function of doppler shift. To obtain the total SNR loss with respect to that achieved with matched-filter reception, it is necessary to add 1.15 dB (see Fig. 10.16c for Taylor weighting) to the loss of Fig. 10.18.

Discrete Time Weighting[2]. A stepped-amplitude function for the reduction of doppler sidelobes is shown in Fig. 10.19. It is symmetrical about the origin, with N denoting the number of steps on each side. Table 10.10 lists stepped-amplitude functions optimized to yield minimum peak sidelobes for $N = 2, 3, 4,$ and 5. $N = 1$, corresponding to the rectangular time envelope, is included for comparison. For $N = 2, 3,$ and 4, the list corresponds very closely to stepped-antenna-aperture distributions[49] optimized by the criterion of maximizing the percentage energy included between the first nulls of the antenna radiation pattern.

Amplitude and Phase Distortion. The ideal compressed pulse has an amplitude spectrum that exactly matches the frequency-weighting function chosen to meet time-sidelobe requirements. Its phase spectrum is linear, corresponding to constant group delay over the band. Amplitude and phase distortion represent a departure of the actual spectrum from this ideal. All radar components are potential sources of distortion which can

TABLE 10.10   Optimum Stepped-Amplitude Time-Weighting Functions

N	Peak sidelobe, dB	Main-lobe width, −3 dB	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5
1	− 13.2	0.886/T	1					1				
2	− 20.9	1.02/T	0.5	0.5				1	0.55			
3	− 23.7	1.08/T	0.35	0.35	0.30			1	0.625	0.350		
4	− 27.6	1.14/T	0.25	0.25	0.25	0.25		1	0.78	0.56	0.34	
5	− 29.6	1.16/T	0.300	0.225	0.235	0.170	0.070	1	0.72	0.54	0.36	0.18
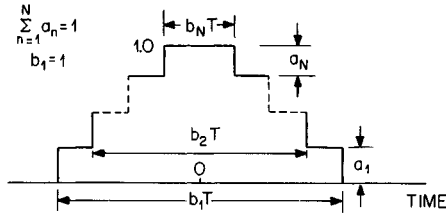
FIG. 10.19   Stepped-amplitude time weighting.

contribute to cumulative radar system distortion. Distortion degrades system performance usually by increasing the sidelobe level and, in extreme cases, by reducing the *SNR* and increasing the pulse width.

The paired-echo concept is useful in estimating distortion tolerances necessary to achieve a required time-sidelobe level.[50] Pair 1 of Table 10.7 shows
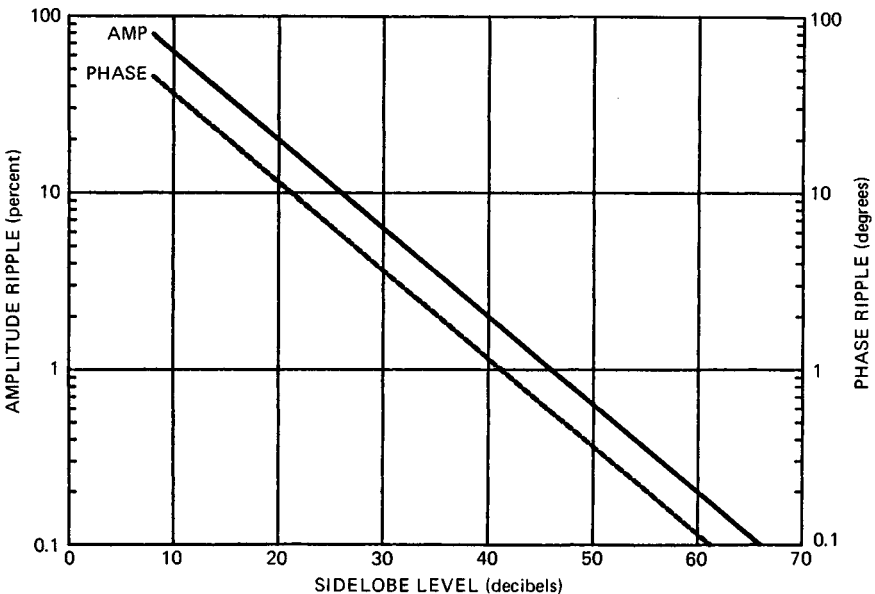
FIG. 10.20   Distortion tolerances versus time sidelobes.

DISTORTED
COMPRESSED PULSE

INPUT

IF TAPPED DELAY LINE

-3   -2   -1   0   1   N

$\frac{1}{B}$

a_{-3} a_{-2} a_{-1} a_0 a_1 a_N

AMPLITUDE
CONTROLS
$(a_0 = 1)$

b_{-3} b_{-2} b_{-1} b_0 b_1 b_N

PHASE
CONTROLS
$(b_0 = 0)$

SUMMING BUS

CORRECTED
OUTPUT

$\frac{1}{B}$
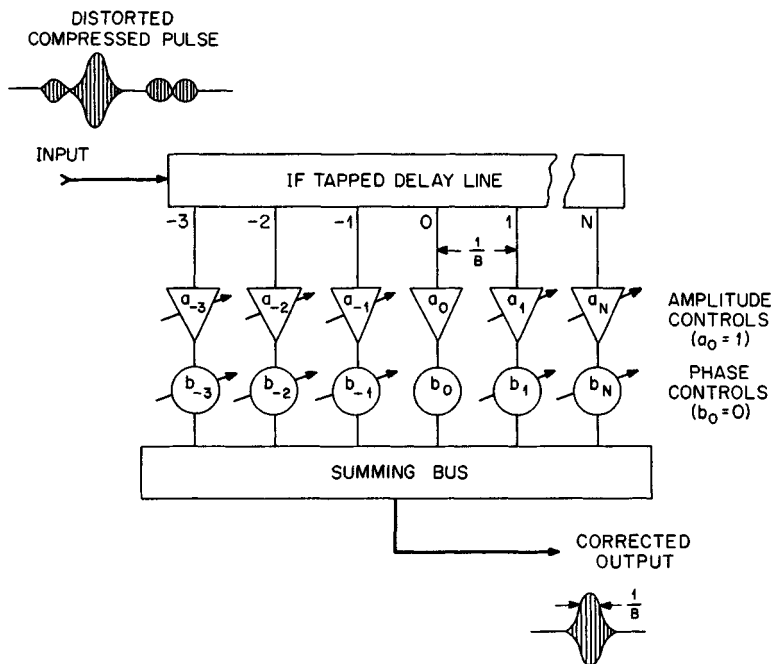
FIG. 10.21 Transversal filter.

that an amplitude ripple results in time sidelobes around the compressed pulse. Pair 8 of Table 10.7 shows that a phase ripple also results in time sidelobes around the compressed pulse. Figure 10.20 shows the amplitude and phase tolerances versus sidelobe level. To obtain time sidelobes of 40 dB below the compressed pulse, the amplitude and phase tolerances are 2 percent and 1.15°, respectively.

Equalization. The transversal filter[51,52] is widely used in the equalization of cumulative amplitude and phase distortion. One version of the transversal filter is shown in Fig. 10.21. It consists of a wideband, dispersion-free IF tapped delay line connected through each of its taps to a summing bus by amplitude and phase controls. The zeroth tap couples the distorted compressed pulse, unchanged except for delay, to the bus. The other taps make it possible to "buck out" distortion echoes of arbitrary phase and amplitude over a compensation interval equal to the total line delay. Reducing time sidelobes to an acceptable level is in effect synthesizing an equalizing filter, which makes the spectrum of the output pulse approach the ideal one described above. Because the transversal filter provides the means for reducing time sidelobes, it eliminates the need for a separate weighting filter since frequency weighting (see pairs 3 and 4 of Table 10.7) can be incorporated in the filter.

REFERENCES

1. Delay Devices for Pulse Compression Radar, *IEE (London) Conf. Publ.* 20, February 1966.

2. Murakami, T.: Optimum Waveform Study for Coherent Pulse Doppler, *RCA Final Rept.*, prepared for Office of Naval Research, Contract Nonr 4649(00)(x), Feb. 28, 1965. AD641391.

3. Morgan, D. P.: Surface Acoustic Wave Devices and Applications, *Ultrasonics*, vol. 11, pp. 121–131, 1973.

4. Eber, L. O., and H. H. Soule, Jr.: Digital Generation of Wideband LFM Waveforms, *IEEE Int. Radar Conf. Rec.*, pp. 170–175, 1975.

5. Hartt, J. K., and L. F. Sheats: Application of Pipeline FFT Technology in Radar Signal and Data Processing, *EASCON Rec.*, pp. 216–221, 1971; reprinted in David K. Barton, *Radars*, vol. 3, Books on Demand UMI, Ann Arbor, Michigan, 1975.

6. Caputi, W. J., Jr.: Stretch: A Time-Transformation Technique, *IEEE Trans.*, vol. AES-7, pp. 269–278, March 1971.

7. Gautier, H., and P. Tournois: Signal Processing Using Surface-Acoustic-Wave and Digital Components, *IEE Proc.*, vol. 127, pt. F, pp. 92–93, April 1980.

8. Slobodnik, A. J., Jr.: Surface Acoustic Waves and SAW Materials, *Proc. IEEE*, vol. 64, pp. 581–594, May 1976.

9. Bristol, T. W.: Acoustic Surface-Wave-Device Applications, *Microwave J.*, vol. 17, pp. 25–27, January 1974.

10. Williamson, R. C.: Properties and Applications of Reflective-Array Devices, *Proc. IEEE*, vol. 64, pp. 702–703, May 1976.

11. Judd, G. W.: Technique for Realizing Low Time Sidelobe Levels in Small Compression Ratio Chirp Waveforms, *Proc. IEEE Ultrasonics Symp.*, pp. 478–481, 1973.

12. Coquin, G. A., T. R. Meeker, and A. H. Meitzler: Attenuation of Longitudinal and Flexural Wave Motions in Strips, *IEEE Trans.*, vol. SU-12, pp. 65–70, June 1965.

13. May, J. E., Jr.: Ultrasonic Traveling-Wave Devices for Communications, *IEEE Spectrum*, vol. 2, pp. 73–85, October 1965.

14. Eveleth, J. H.: A Survey of Ultrasonic Delay Lines Operating Below 100 Mc/s, *Proc. IEEE*, vol. 53, pp. 1406–1428, October 1965.

15. Improved Delay Line Technique Study, *RADC Tech. Rept. RADC-TR-65-45*, May 1965. ASTIA AD617693.

16. Coquin, G. A., and R. Tsu: Theory and Performance of Perpendicular Diffraction Delay Lines, *Proc. IEEE*, vol. 53, pp. 581–591, June 1965.

17. Rodrigue, G. P.: Microwave Solid-State Delay Line, *Proc. IEEE*, vol. 53, pp. 1428–1437, October 1965.

18. O'Meara, T. R.: The Synthesis of "Band-Pass," All-Pass, Time Delay Networks with Graphical Approximation Techniques, *Hughes Aircraft Co. Res. Rept.* 114, June 1959.

19. Peebles, P. Z., Jr.: Design of a 100:1 Linear Delay Pulse Compression Filter and System, master thesis, Drexel Institute of Technology, Philadelphia, December 1962.

20. Hewett, H. S.: Highly Accurate Compression Filter Design Technique, *Stanford University, Electron. Lab. Tech. Rept.* 1965-3, November 1967. See also H. S. Hewett: A Computer Designed, 720 to 1 Microwave Compression Filter, *IEEE Trans.*, vol. MTT-15, pp. 687–694, December 1967.

21. Barker, R. H.: Group Synchronization of Binary Digital Systems, in Jackson, W. (ed.): "Communication Theory," Academic Press, New York, 1953, pp. 273–287.

22. Turyn, R., and J. Stover: On Binary Sequences, *Proc. Am. Math. Soc.*, vol. 12, pp. 394–399, June 1961.

23. Luenburger, D. G.: On Barker Codes of Even Length, *Proc. IEEE*, vol. 51, pp. 230–231, January 1963.

24. Turyn, R.: On Barker Codes of Even Length, *Proc. IEEE* (correspondence), vol. 51, p. 1256, September 1963.

25. Lindner, J.: Binary Sequences Up to Length 40 with Best Possible Autocorrelation Function, *Electron. Lett.*, vol. 11, p. 507, October 1975.

26. Peterson, W. W., and E. J. Weldon, Jr.: "Error Correcting Codes," app. C, M.I.T. Press, Cambridge, Mass., 1972.

27. Golomb, S. W.: "Shift Register Sequences," Holden-Day, Oakland, Calif., 1967, chap. 3.

28. Golay, M. J. E.: Complementary series, *IRE Trans.*, vol. IT-7, pp. 82–87, April 1961.

29. Golay, M. J. E.: Note on complementary series, *Proc. IRE*, vol. 50, p. 84, January 1962.

30. Hollis, E. E.: Another type of complementary series, *IEEE Trans.*, vol. AES-11, pp. 916–920, September 1975.

31. Golomb, S. W., and R. A. Scholtz: Generalized Barker Sequences, *IEEE Trans.*, vol. IT-11, pp. 533–537, October 1965.

32. Somaini, U., and M. H. Ackroyd: Uniform Complex Codes with Low Autocorrelation Sidelobes, *IEEE Trans.*, vol. IT-20, pp. 689–691, September 1974.

33. Frank, R. L.: Polyphase Codes with Good Nonperiodic Correlation Properties, *IEEE Trans.*, vol. IT-9, pp. 43–45, January 1963.

34. Lewis, B. L., and F. F. Kretschmer, Jr.: A New Class of Polyphase Pulse Compression Codes and Techniques, *IEEE Trans.*, vol. AES-17, pp. 364–372, May 1981. (See correction, *IEEE Trans.*, vol. AES-17, p. 726, May 1981.)

35. Rihaczek, A. W.: "Principles of High-Resolution Radar," McGraw-Hill Book Company, New York, 1969, chap. 8.

36. Wheeler, H. A.: The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, *Proc. IRE*, vol. 27, pp. 359–385, June 1939.

37. MacColl, L. A.: unpublished manuscript referred to by H. A. Wheeler (see Ref. 36, p. 359, footnote 1).

38. Burrow, C. R.: Discussion on Paired Echo Distortion Analysis, *Proc. IRE*, vol. 27, p. 384, June 1939.

39. Klauder, J. R., A. C. Price, S. Darlington, and W. J. Albersheim: The Theory and Design of Chirp Radars, *Bell Syst. Tech. J.*, vol. 39, pp. 745–808, July 1960.

40. Blackman, R. B., and J. W. Tukey: "The Measurement of Power Spectra," Dover Publications, New York, 1958.

41. Temes, C. L.: Sidelobe Suppression in a Range Channel Pulse-Compression Radar, *IRE Trans.*, vol. MIL-6, pp. 162–169, April 1962.

42. Cook, C. E., and M. Bernfield: "Radar Signals: An Introduction to Theory and Application," Academic Press, New York, 1967.

43. Taylor, T. T.: Design of Line-Source Antennas for Narrow Beamwidth and Low Sidelobes, *IRE Trans.*, vol. AP-3, pp. 16–28, January 1955.

44. Dolph, C. L.: A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Sidelobe Level, *Proc. IRE*, vol. 34, pp. 335–348, June 1946.

45. Ramsay, J. F.: Fourier Transforms in Aerial Theory, *Marconi Rev.*, vol. 9, October–December 1946.

46. Cummings, R. D., M. Perry, and D. H. Preist: Calculated Spectra of Distorted Gaussian Pulses, *Microwave J.*, pp. 70–75, April 1965.

47. Mason, S. J., and H. J. Zimmerman: "Electronic Circuits, Signals and Systems," John Wiley & Sons, New York, 1960, p. 237.

48. Spellmire, R. J.: Tables of Taylor Aperture Distributions, *Hughes Aircraft Co., Syst. Dev. Lab. Tech. Mem.* 581, October 1958.

49. Nash, R. T.: Stepped Amplitude Distributions, *IEEE Trans.*, vol. AP-12, pp. 515–516, July 1964.

50. DiFranco, J. V., and W. L. Rubin: Signal Processing Distortion in Radar Systems, *IRE Trans.*, vol. MIL-6, pp. 219–225, April 1962.

51. Kallmann, H. E.: Transversal Filters, *Proc. IRE*, vol. 28, pp. 302–310, July 1940.

52. Pratt, W. R.: Transversal Equalizers for Suppressing Distortion Echoes in Radar Systems, *Proc. Symp. Pulse Compression Techniques*, pp. 119–128, *Rome Air Dev. Center, RADC-TDR*-62-580, April 1963.

CHAPTER 11
RADAR CROSS SECTION

Eugene F. Knott
The Boeing Company

11.1 INTRODUCTION

A radar detects or tracks a target, and sometimes can identify it, only because there is an echo signal. It is therefore critical in the design and operation of radars to be able to quantify or otherwise describe the echo, especially in terms of such target characteristics as size, shape, and orientation. For that purpose the target is ascribed an effective area called the *radar cross section*. It is the projected area of a metal sphere which would return the same echo signal as the target had the sphere been substituted for the target.

Unlike the echo of the sphere, however, which is independent of the viewing angle, the echoes of all but the simplest targets vary significantly with orientation. As such, one must mentally allow the size of this fictitious sphere to vary as the aspect angle of the target changes. As will be shown, the variation can be quite rapid, especially for targets many wavelengths in size.

The echo characteristics depend in strong measure on the size and nature of the target surfaces exposed to the radar beam. The variation is small for electrically small targets (targets less than a wavelength in size) because the incident wavelength is too long to resolve target details. On the other hand, the flat, singly curved and doubly curved surfaces of electrically large targets each give rise to different echo characteristics. Reentrant structures like jet engine intakes and exhausts generally have large echoes, and even the trailing edges of airfoils can be significant echo sources. The characteristics of some common targets and target features are discussed in Sec. 11.2.

The radar cross sections of simple bodies can be computed exactly by a solution of the wave equation in a coordinate system for which a constant coordinate coincides with the surface of the body. The exact solution requires that the electric and magnetic fields just inside and just outside the surface satisfy certain conditions that depend on the electromagnetic properties of the material of which the body is made.

While these solutions constitute interesting academic exercises and can, with some study, reveal the nature of the scattering mechanisms that come into play, there are no known tactical targets that fit the solutions. Thus, exact solutions of

the wave equation are, at best, guidelines for gauging other (approximate) methods of computing scattered fields.

An alternative approach is the solution of the integral equations governing the distribution of induced fields on target surfaces. The most useful approach at solution is known as the *method of moments*, in which the integral equations are reduced to a system of linear homogeneous equations. The attraction of the method is that the surface profile of the body is unrestricted, allowing the computation of the scattering from truly tactical objects. Another is that ordinary methods of solution (matrix inversion and gaussian elimination, for example) may be employed to effect a solution. The method is limited by computer memory and execution time, however, to objects a few dozen wavelengths in size at best.

Alternatives to these exact solutions are several approximate methods that may be applied with reasonable accuracy to electrically large target features. They include the theories of geometrical and physical optics, the geometrical and physical theories of diffraction, and the method of equivalent currents. These approximations are discussed in Sec. 11.3. Other approximate methods not discussed here are explored in detail in some of the references listed at the end of this chapter.

The practical engineer cannot rely entirely on predictions and computations and must eventually measure the echo characteristics of some targets. This may be done by using full-scale test objects or scale models thereof. Small targets often may be measured indoors, but large targets usually must be measured on an outdoor test range. The characteristics of both kinds of test facilities are described in Sec. 11.4.

Control of the echo characteristics of some targets is of vital tactical importance. There are only two practical ways of doing so: shaping and radar absorbers. Shaping is the selection or design of surface profiles so that little or no energy is reflected back toward the radar. Because target contours are difficult to change once the target has become a production item, shaping is best implemented in the concept definition stage before production decisions have been made. Radar-absorbing materials actually soak up radar energy, also reducing the energy reflected back to the radar. However, the application of such materials can be expensive, whether gauged in terms of nonrecurring engineering costs, lifetime maintenance, or reduced mission capabilities. The two methods of echo control are discussed in Sec. 11.5.

Unless otherwise noted, the time convention used in this chapter is exp $(-i\omega t)$, with the time dependence suppressed in all equations. Readers who prefer the exp $(j\omega t)$ time convention may replace i by $-j$ wherever it appears.

11.2 THE CONCEPT OF ECHO POWER

Definition of RCS. An object exposed to an electromagnetic wave disperses incident energy in all directions. This spatial distribution of energy is called *scattering*, and the object itself is often called a *scatterer*. The energy scattered back to the source of the wave (called *backscattering*) constitutes the *radar echo* of the object. The intensity of the echo is described explicitly by the radar cross section of the object, for which the abbreviation RCS has been generally recognized. Early papers on the subject called it the *echo area* or the *effective area*, terms still found occasionally in contemporary technical literature.

The formal definition of radar cross section is

$$\sigma = \lim_{R \to \infty} 4\pi R^2 \frac{|E_s|^2}{|E_0|^2} \tag{11.1}$$

where E_0 is the electric-field strength of the incident wave impinging on the target and E_s is the electric-field strength of the scattered wave at the radar. The derivation of the expression assumes that a target extracts power from an incident wave and then radiates that power uniformly in all directions. Although the vast majority of targets do *not* scatter energy uniformly in all directions, the definition assumes that they do. This permits one to calculate the scattered power density on the surface of a large sphere of radius R centered on the scattering object. R is typically taken to be the range from the radar to the target.

The symbol σ has been widely accepted as the designation for the RCS of an object, although this was not so at first.[1,2] The RCS is the projected area of a metal sphere which is large compared with the wavelength and which, if substituted for the object, would scatter identically the same power back to the radar. The RCS of all but the simplest scatterers fluctuates greatly with the orientation of the object. As such, this imaginary sphere would have to expand and contract with changing target orientation to represent the amplitude fluctuations displayed by most objects.

The limiting process in Eq. (11.1) is not always an absolute requirement. In both measurement and analysis, the radar receiver and transmitter are usually taken to be in the far field of the target (discussed in Sec. 11.4), and at that distance the scattered field E_s decays inversely with the distance R. Thus, the R^2 term in the numerator of Eq. (11.1) is canceled by an identical but implicit R^2 term in the denominator. Consequently the dependence of the RCS on R, and the need to form the limit, usually disappears.

Radar cross section is therefore a comparison of the scattered power density at the receiver with the incident power density at the target. An equally valid definition of the RCS results when the electric-field strengths in Eq. (11.1) are replaced with the incident and scattered magnetic-field strengths. It is often necessary to measure or calculate the power scattered in some other direction than back to the transmitter, a *bistatic* situation. A bistatic RCS may be defined for this case as well as for backscattering, provided it is understood that the distance R is measured from the target to the receiver. *Forward scattering* is a special case of bistatic scattering in which the bistatic angle is 180°, whence the direction of interest is along the shadow zone behind the target.

The shadow itself can be regarded as the sum of two fields of nearly equal strength but 180° out of phase. One is the incident field, and the other is the scattered field. The formation of the shadow implies that the forward scattering is large, which is indeed the case. The fields behind the target are hardly ever precisely zero, however, because some energy usually reaches the shadow zone via diffraction from the sides of the target.

While there are few two-dimensional (infinite cylindrical) objects in the physical world, analyses of the scattering from two-dimensional structures are very useful. A two-dimensional object is, by definition, a cylinder formed by the pure translation of a plane curve to plus and minus infinity along an axis perpendicular to the plane of that curve. Many scattering problems become analytically tractable when there is no field variation along the cylindrical axis, such as when the infinite structure is illuminated by a plane wave propagating at right angles to the cylinder axis.

In this case, one defines a scattering *width* instead of a scattering area,

$$\sigma_{2D} = \lim_{\rho \to \infty} 2\pi\rho \frac{V_s^2}{V_0^2} \qquad (11.2)$$

where ρ is the distance from the cylindrical body to a remove receiver, measured perpendicularly to the cylindrical axis. We have appended the subscript 2D to distinguish the scattering width of Eq. (11.2), whose dimension is length, from the scattering cross section of Eq. (11.1), whose dimension is the square of length.

By virtue of the linear properties of electromagnetic fields, the solutions of two-dimensional problems may be resolved into two cases, one each for the electric field or the magnetic field parallel to the cylindrical axis. The ratio $|V_s|/|V_0|$ thus represents either the incident and scattered electric fields or the incident and scattered magnetic fields, depending on the case at hand. These two cases are often called E and H polarizations, respectively. They are also known as TM and TE polarizations.

Practical three-dimensional problems often involve truncated segments of two-dimensional structures, such as shown in Fig. 11.1. In the practical world, those segments may be viewed at angles other than incidence perpendicular to the cylindrical axis, as implied in the solution of two-dimensional problems. The three-dimensional RCS of a truncated two-dimensional structure may be found from the approximate relationship

$$\sigma = \frac{2\ell^2 \sigma_{2D}}{\lambda} \left| \frac{\sin(k\ell \sin \tau)}{k\ell \sin \tau} \right|^2 \qquad (11.3)$$

where ℓ is the length of the truncated structure, σ_{2D} is its two-dimensional scattering width (obtained for the infinite structure), and τ is the tilt angle of the segment measured from broadside incidence. This approximation assumes that the amplitudes of the fields induced on the three-dimensional body are identically those induced on the corresponding two-dimensional structure and that the tilt angle influences only the phase of the surface fields induced on the body. The expression should not be used for large tilt angles, for which the amplitudes obtained from the two-dimensional solution no longer apply to the three-dimensional problem.

Examples of RCS Characteristics

Simple Objects. Because of its pure radial symmetry, the perfectly conducting sphere is the simplest of all three-dimensional scatterers. Despite the simplicity of its geometrical surface, however, and the invariance of its echo with orientation, the RCS of the sphere varies considerably with electrical size.
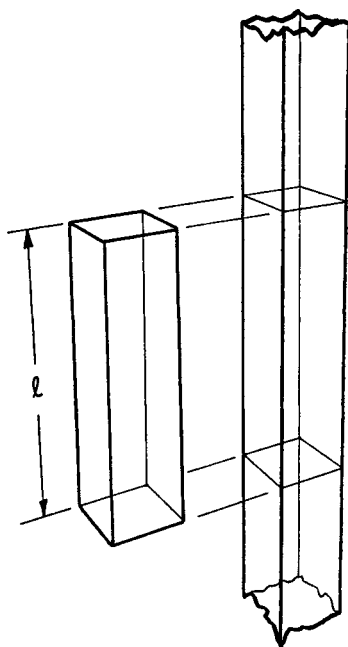
FIG. 11.1 A three-dimensional object whose profile does not vary along its length, such as the truncated rectangular cylinder on the left, is a finite chunk of an infinite (two-dimensional) structure having the same profile, such as the one on the right. Equation (11.3) relates the RCS of the two structures.

The exact solution for the scattering by a conducting sphere is known as the Mie series,[3] illustrated in Fig. 11.2.
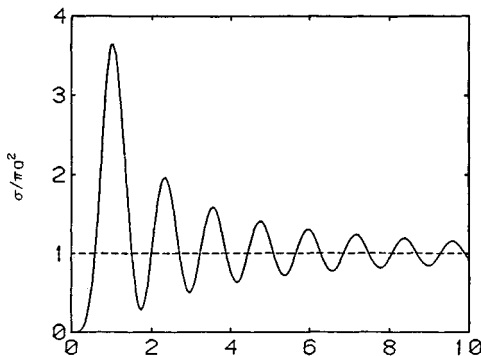
FIG. 11.2 RCS of a perfectly conducting sphere as a function of its electrical size ka.

The parameter $ka = 2\pi a/\lambda$ is the circumference of the sphere expressed in wavelengths, and the RCS is shown normalized with respect to the projected area of the sphere. The RCS rises quickly from a value of zero to a peak near $ka = 1$ and then executes a series of decaying undulations as the sphere becomes electrically larger. The undulations are due to two distinct contributions to the echo, one a *specular reflection* from the front of the sphere and the other a *creeping wave* that skirts the shadowed side. The two go in and out of phase because the difference in their electrical path lengths increases continuously with increasing ka. The undulations become weaker with increasing ka because the creeping wave loses more energy the longer the electrical path traveled around the shadowed side.

The log-log plot of Fig. 11.3 reveals the rapid rise in the RCS in the region $0 < ka < 1$, which is known as the *Rayleigh region*. Here the normalized RCS increases with the fourth power of ka, a feature shared by other electrically small or thin structures. The central region characterized by the interference between the specular and creeping-wave contributions is known as the *resonance region*. There is no clear upper boundary for this part of the curve, but a value near $ka = 10$ is generally accepted. The region $ka > 10$ is dominated by the specular return from the front of the sphere and is called the *optics region*. For spheres of these sizes the geometric optics approximation πa^2 is usually an adequate representation of the magnitude of the RCS.

The echoes of all scattering objects, and not just the perfectly conducting sphere, can be grouped according to the electrical-size characteristics of the object. The dimensions of a Rayleigh scatterer are much less than a wavelength, and the RCS is proportional to the square of the volume of the body. Resonant scatterers are generally of the order of one-half to 10 wavelengths in size, for which neither Rayleigh nor optics approximations may be very accurate. In the optics region several approximations are available for making estimates or predictions (see Sec. 11.3).
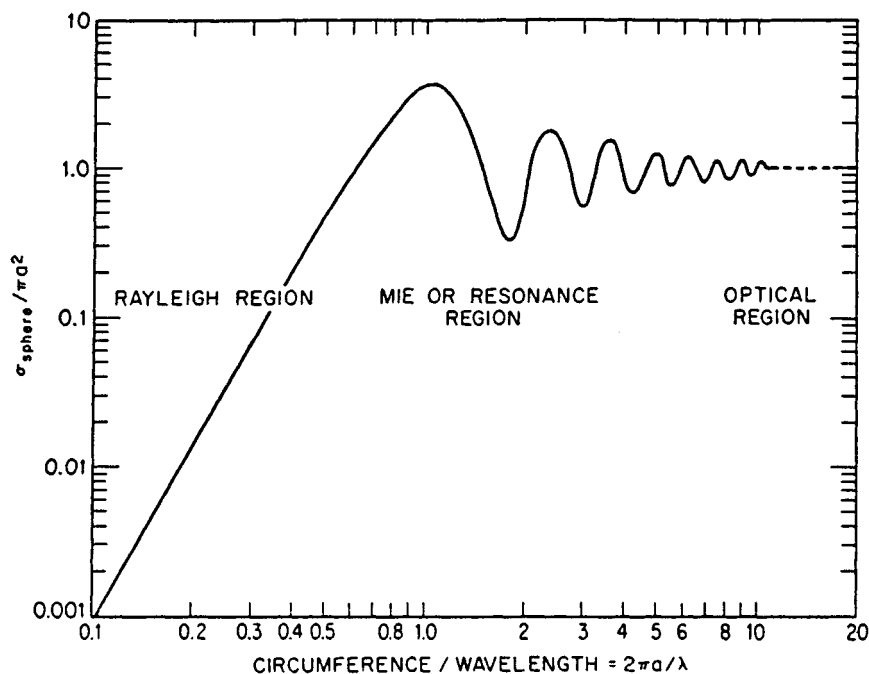
FIG. 11.3 Log-log version of the data displayed in Fig. 11.2.

The echo characteristics of permeable (dielectric) bodies can be more complicated than those of perfect conductors because energy may enter the body and suffer several internal bounces before emerging. An example is the dielectric sphere whose RCS is plotted in Fig. 11.4. Because the dielectric material is slightly lossy, as indicated by the nonzero imaginary component of the index of refraction, the RCS of the sphere decays gradually with increasing electrical size. The RCS of small dielectric bodies does not exhibit this complexity, on the other hand, because the sources of reflection are too close to each other to be resolvable by the incident wave. An example is the two-dimensional Rayleigh region RCS of a thin dielectric cylinder, plotted in Fig. 11.5. The thin dielectric cylinder has been used to model the target support lines sometimes employed in RCS measurements.[5] Note that the H-polarized echo is barely 6 dB less than that for E polarization for this particular dielectric constant.

The thin wire (a metal dipole) can have a complicated pattern, as shown in Fig. 11.6. The RCS of the wire varies with the wire length, the angle subtended by the wire and the line of sight, and on that component of the incident electric field in the plane containing the wire and the line of sight. The wire diameter has only a minor influence if it is much smaller than the wavelength. In addition to the prominent broadside lobe at the center of the pattern, there are traveling-wave lobes near the left and right sides. The traveling-wave lobes tend to disappear as the dipole becomes shorter and are closely related to those excited on traveling-wave antennas.
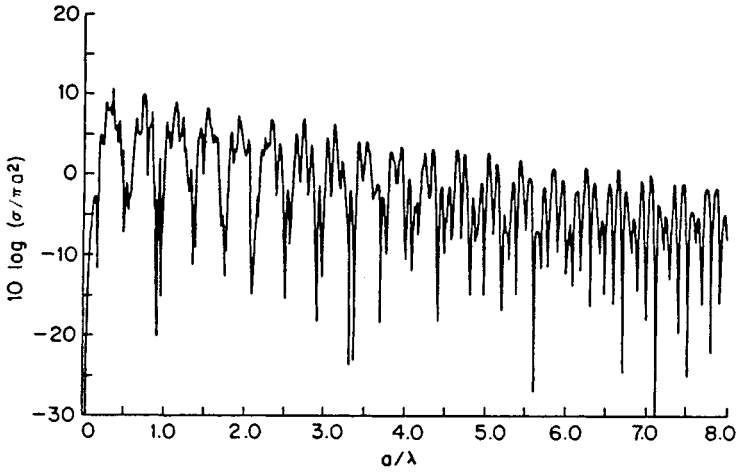
FIG. 11.4   RCS of a lossy dielectric sphere with $n = 2.5 + i0.01$. (*Copyright 1968, IEEE.*[4])
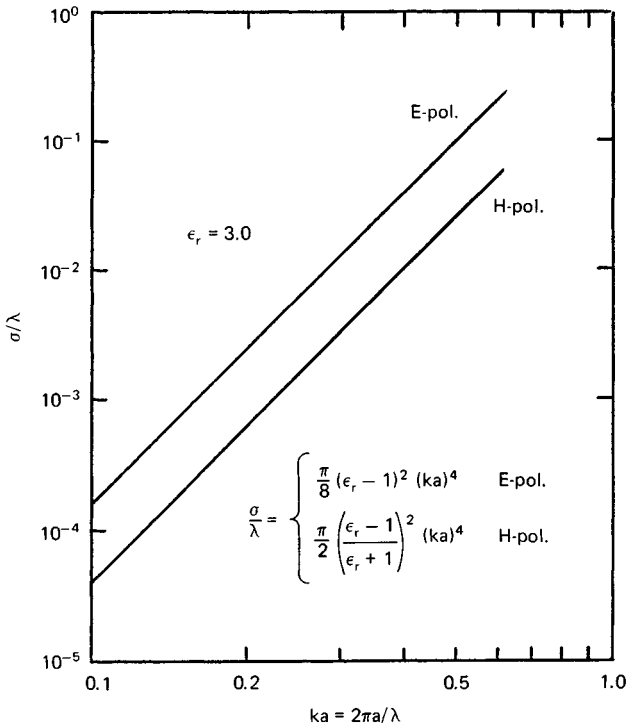
FIG. 11.5   RCS of a slender dielectric cylinder with $\epsilon_r = 3.0$.
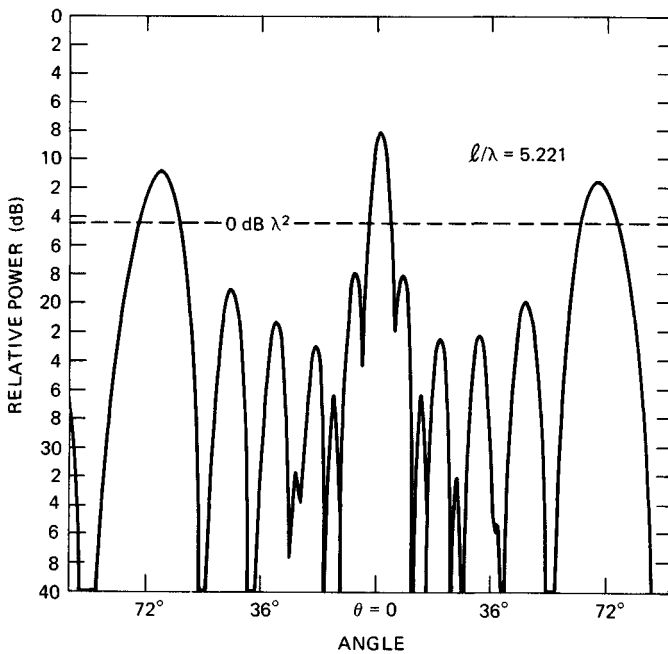
FIG. 11.6 Measured RCS pattern of a dipole 5.221λ long. (*Courtesy of University of Michigan Radiation Laboratory.*[6])

Figure 11.7 shows the broadside resonances of a wire dipole as a function of dipole length. The first resonance occurs when the dipole is just under a half wavelength long, and its magnitude is very nearly λ^2. Other resonances occur near odd multiples of a quarter wavelength, with plateaus of nearly constant re-
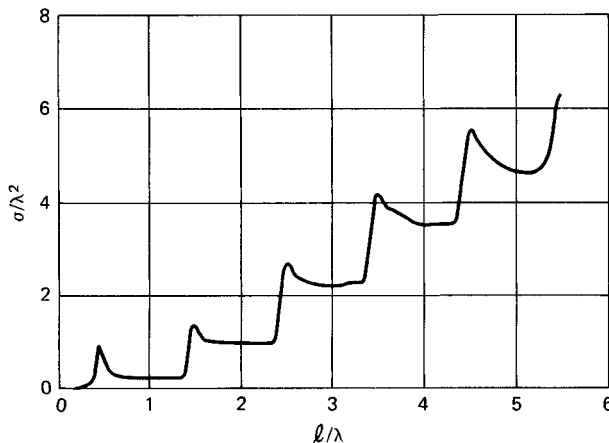
FIG. 11.7 Measured broadside returns of a thin dipole. (*Courtesy of University of Michigan Radiation Laboratory.*[6])

turn between the resonant peaks. These plateaus rise as the dipole becomes thicker, and the resonances eventually disappear.

Bodies considerably thicker than the thin wire also support surface traveling waves that radiate power in the backward direction. An example is the ogive, a spindle-shaped object formed by rotating an arc of a circle about its chord. Figure 11.8 is the RCS pattern of a 39-wavelength 15° half-angle ogive recorded for horizontal polarization (incident electric field in the plane of the ogive axis and the line of sight). The large lobe at the right side of the pattern is a specular echo in the broadside sector, and the sequence of peaks at the left side is the contribution of the surface traveling wave near end-on incidence. Note that the RCS is extremely small (not measurable in this case) at precisely end-on incidence. Theoretical predictions in the end-on region closely match the measured pattern for this particular body.
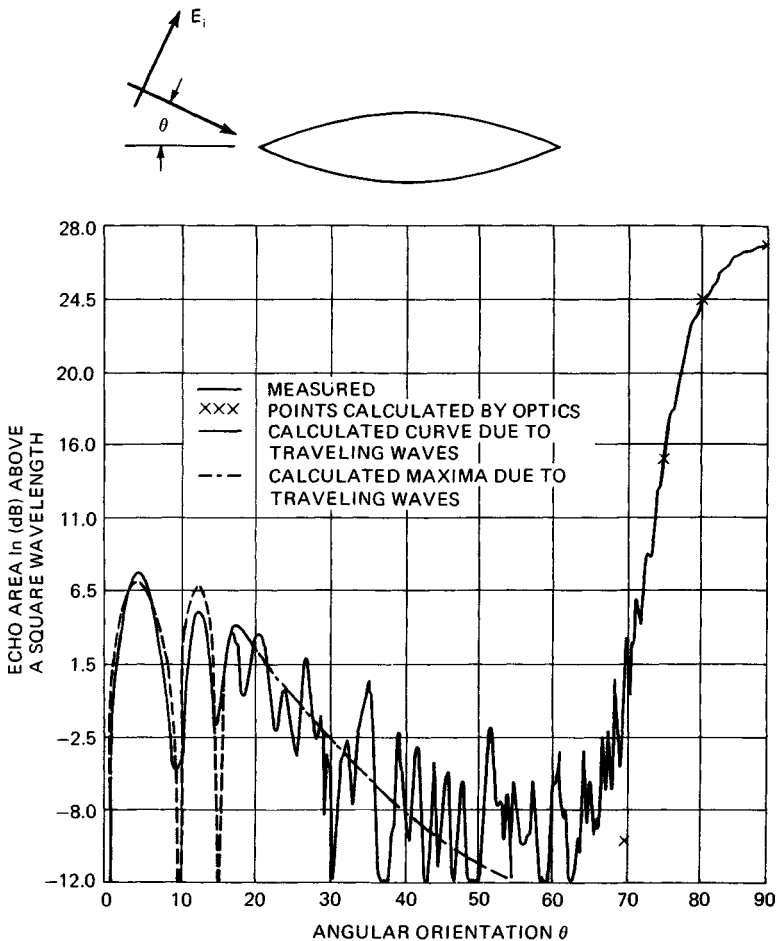
FIG. 11.8   Measured RCS pattern of a 39-wavelength 15° half-angle metal ogive. (*Copyright 1958, IEEE.*[7])

The dominant scattering mechanisms for the right circular conducting cone are the tip and the base. The return from the tip is very small in the nose-on region, and the RCS pattern is dominated by the echo from the base. Figures 11.9 and 11.10 are patterns of the RCS of a 15° (half-angle) cone with a base circumference of 12.575λ. Both patterns were measured as the cone was rotated about a vertical axis parallel to the base of the cone. The transmitted and received electric polarization was in the plane swept out by the cone axis (horizontal polarization) for Fig. 11.9 and was perpendicular to that plane (vertical polarization) for Fig. 11.10.

Nose-on incidence lies at the center of the patterns, and the sharp peaks near the sides are the specular returns from the slanted sides of the cone, also called specular *flashes*. The RCS formula for singly curved surfaces given in Table 11.1 may be used to predict the amplitudes of the specular flash within a fraction of a decibel. At precisely nose-on incidence the RCS must be independent of polarization because the cone is a body of axial symmetry. This may be verified by comparing the nose-on values in the two figures. At this angle the entire ring of the base of the cone is excited, but as the aspect angle swings away from nose-
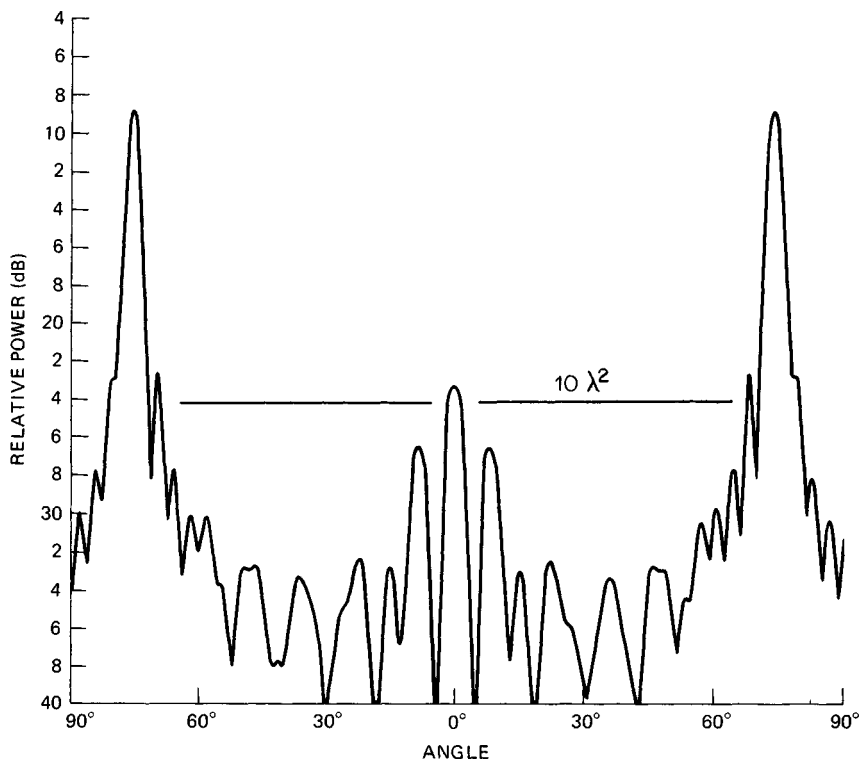
FIG. 11.9 Measured RCS of a 15° half-angle cone (horizontal polarization). The base circumference is 12.575λ. The heavy horizontal line indicates 10λ². (*Courtesy of University of Michigan Radiation Laboratory.*[8])
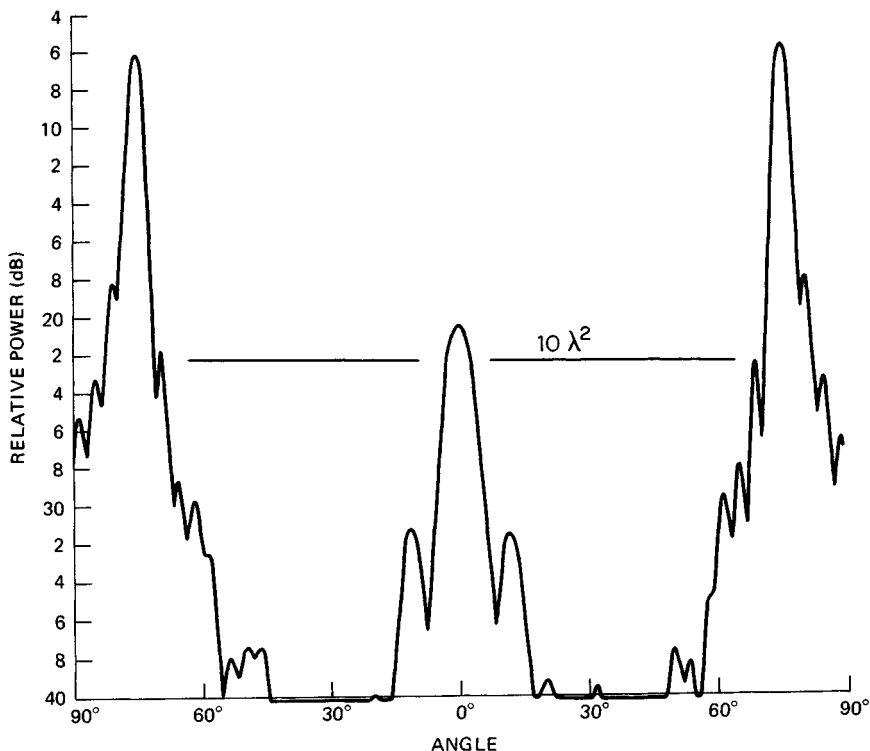
FIG. 11.10 Measured RCS of a 15° half-angle cone (vertical polarization). The base circumference is 12.575λ. The heavy horizontal line indicates 10λ². (*Courtesy of University of Michigan Radiation Laboratory.*[8])

TABLE 11.1 RCS Approximations for Simple Scattering Features

Scattering feature	Orientation (1)	Approximate RCS	
Corner reflector	Axis of symmetry along LOS	$4\pi A_{eff}^2/\lambda^2$	(2)
Flat plate	Surface perpendicular to LOS	$4\pi A^2/\lambda^2$	(3)
Singly curved surface	Surface perpendicular to LOS	$2\pi a\ell^2/\lambda$	(4)
Doubly curved surface	Surface perpendicular to LOS	$\pi a_1 a_2$	(5)
Straight edge	Edge perpendicular to LOS	ℓ^2/π	(6)
Curved edge	Edge element perpendicular to LOS	$a\lambda/2$	(7)
Cone tip	Axial incidence	$\lambda^2 \sin^4(\alpha/2)$	(8)

NOTES:
1. LOS = line of sight.
2. A_{eff} = effective area contributing to multiple internal reflections.
3. A = actual area of the plate.
4. a = mean radius of curvature; ℓ = length of slanted surface.
5. a_1, a_2 = principal radii of surface curvature in orthogonal planes.
6. ℓ = edge length.
7. a = radius of edge contour.
8. α = half angle of the cone.

on, the scattering from the base degenerates to a pair of flash points. They lie at opposite ends of a diameter across the base in the plane containing the direction of incidence and the cone axis.

The echoes from the flash points at the sides of the base weaken as the aspect angle moves away from nose-on incidence, and the sidelobes seen at $+13°$ in Fig. 11.10 are actually due to an interaction between the two flash points *across the shadowed side* of the base. (The sidelobes disappear when a pad of absorber is cemented to the base.[8]) The flash point at the far side of the base disappears when the aspect angle moves outside the backward half cone, but the near flash point remains visible, and its echo decays with increasing aspect angle. Trailing-edge contributions like these are excited by that component of the incident electric field perpendicular to the edge; therefore they are stronger for horizontal incident polarization than for vertical polarization.

A flat plate also can support multiple diffraction from one side of the plate to the other, as shown in Fig. 11.11. The axis of rotation was in the plane of the plate parallel to one edge; normal incidence to the incident wave is $0°$, at the left side of each chart, with edge-on incidence at $90°$ near the right side. The specular return from the plate is the large peak at $0°$, which is predicted with quite good accuracy by the flat-plate formula given in Sec. 11.3. The edge-on return for vertical polarization is well predicted by the straight-edge formula given in Table 11.1.
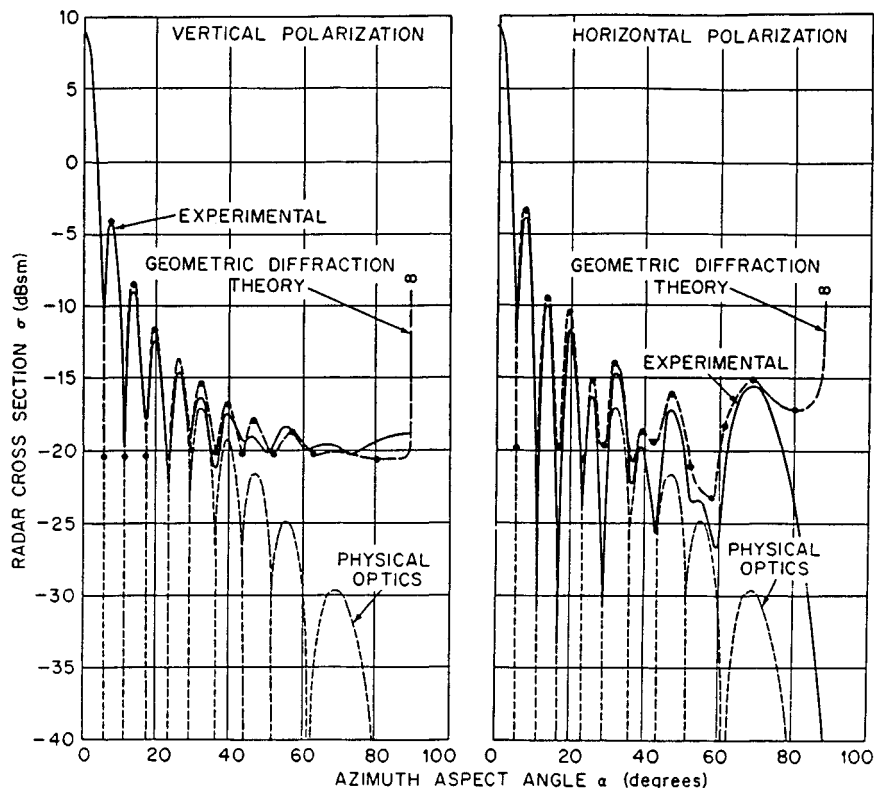
FIG. 11.11   RCS of a square flat plate 6.5 in along a side; $\lambda = 1.28$ in. (*Copyright 1966, IEEE.*[9])

These undulating patterns follow a sin x/x variation quite closely for aspect angles out to about 30°, but beyond that angle the two patterns differ by progressively wider margins. The sin x/x behavior is characteristic of a uniformly illuminated aperture, but unlike the one-way illumination function encountered in antenna work, the argument x for a flat plate includes a two-way (round-trip) illumination function. Thus, the beamwidth of the echo response of a flat plate is half the beamwidth of an antenna aperture of the same size. The prominent lobe in the horizontal pattern at 68° is a surface traveling-wave lobe closely related to the one appearing at nearly the same angle in the dipole pattern of Fig. 11.7.

In contrast to the pattern of a flat plate, the RCS pattern of a corner reflector is quite broad. This is true because the corner reflector is a reentrant structure, and no matter what its orientation (within limits, of course), internally reflected waves are directed back toward the source of the incident wave. A corner reflector is formed by two or three flat plates intersecting at right angles, and waves impinging on the first face are reflected onto the second; if there is a third face, it receives waves reflected by the first two faces. The mutual orthogonality of the faces ensures that the direction taken by waves upon final reflection is back toward the source.

The individual faces of the corner reflector may be of arbitrary shape, but the most common is an isosceles triangle for the trihedral corner; dihedral corners typically have rectangular faces. The RCS of a corner reflector seen along its axis of symmetry is identically that of a flat plate whose physical area matches the effective area of the corner reflector. The magnitude of the echo may be determined by finding the polygonal areas on each face of the corner receiving waves reflected by the other faces, and from which the final reflection is back toward the source. The effective area is determined by summing the projections of the areas of those polygons on the line of sight;[10] the RCS is then found by squaring that area, multiplying by 4π and dividing by λ^2.

Figure 11.12 is a collection of RCS patterns of a trihedral corner reflector with triangular faces. The reflector was fabricated of three triangular plywood panels, metallized to enhance their surface reflectivities. The aperture exposed to the radar was therefore an equilateral triangle, as shown in Fig. 11.13. The eight patterns in Fig. 11.12 were measured with the plane of the aperture tilted above or below the line of sight by the angle ϕ.

The broad central part of these patterns is due to a triple-bounce mechanism between the three participating faces, while the "ears" at the sides of the patterns are due to the single-bounce, flat-plate scattering from the individual faces. Along the axis of symmetry of the trihedral reflector in Fig. 11.13 ($\theta = 0°$, $\phi = 0°$), the RCS is $\pi\ell^4/3\lambda^2$, where ℓ is the length of one of the edges of the aperture. Not shown are the echo reductions obtained when the trihedral faces are angled other than at 90° from each other. The reductions resulting from changes in the angles of the corner faces depend on the size of the faces expressed in wavelengths.[11,12]

The RCS of most of the simple scattering features discussed above may be estimated by using the simple formulas listed in Table 11.1. The RCS of some complicated targets may be estimated by representing the target as a collection of features like those listed in Table 11.1, calculating the individual contributions, and then summing the contributions coherently or noncoherently. More detailed formulas are given in Sec. 11.3 that account for surface orientations not included in Table 11.1.

Complex Objects.   Objects like antennas, insects, birds, airplanes, and ships can be much more complex than those discussed above, either because of the multiplicity of scatterers on them or because of the complexity of their surface profiles and dielectric constants. Insects are examples of the latter.
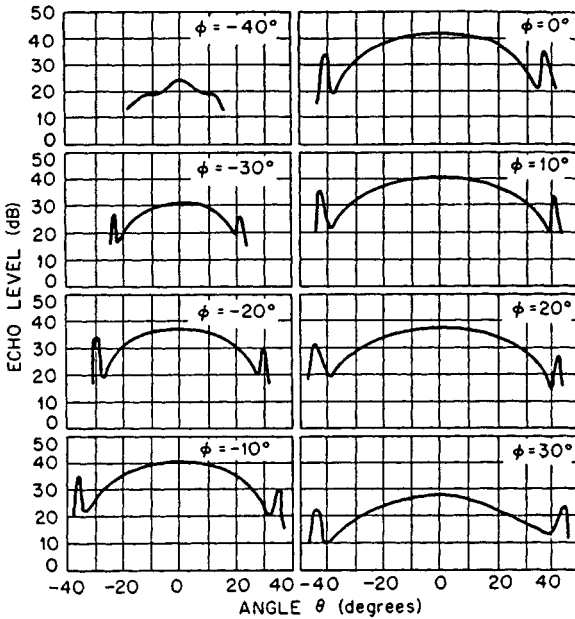
FIG. 11.12 RCS patterns of a trihedral corner reflector. Edge of aperture = 24 in; λ = 1.25 cm. (*Reprinted with permission from the* AT&T Technical Journal, *copyright 1947, AT&T.*[2])

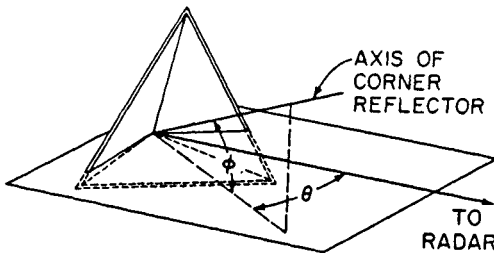FIG. 11.13 Coordinate system for the RCS patterns in Fig. 11.12. (*Reprinted with permission from the* AT&T Technical Journal, *copyright 1947, AT&T.*[2])

Measured values for a dozen species are listed in Table 11.2. (The spider is an arachnid, not an insect, of course.) The animals were live for the measurements but had been drugged to immobilize them. Figure 11.14 shows the relationship between the RCS and the mass of an insect, with the variation of a water droplet shown for comparison. Similar comparisons have been made for both birds and insects.[15] The following values have been reported for the RCS of a man:[16]

TABLE 11.2   Measured Insect RCS at 9.4 GHz[13]

Insect	Length, mm	Width, mm	Broadside RCS, dBsm	End-on RCS, dBsm
Blue-winged locust	20	4	− 30	− 40
Armyworm moth	14	4	− 39	− 49
Alfalfa caterpillar butterfly	14	1.5	− 42	− 57
Honeybee worker	13	6	− 40	− 45
California harvester ant	13	6	− 54	− 57
Range crane fly	13	1	− 45	− 57
Green bottle fly	9	3	− 46	− 50
Twelve-spotted cucumber beetle	8	4	− 49	− 53
Convergent lady beetle	5	3	− 57	− 60
Spider (unidentified)	5	3.5	− 50	− 52

NOTE: Original values reported in square centimeters have been converted here to dBsm.
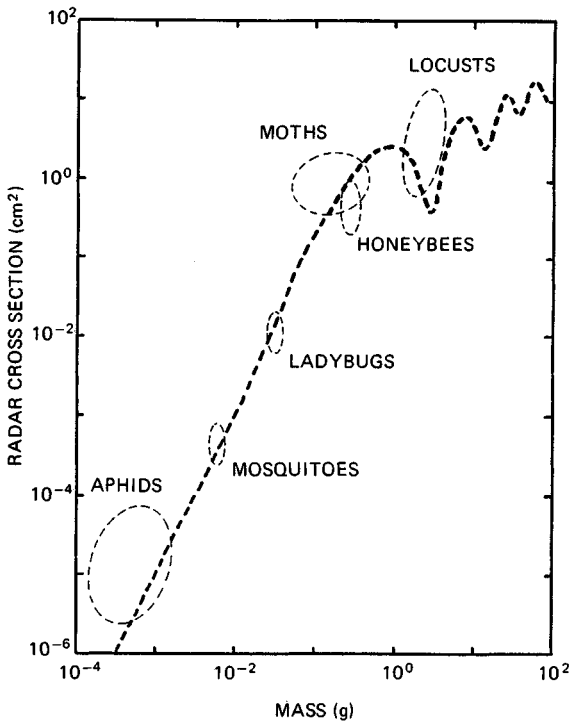
FIG. 11.14   Sample of measured RCS of insects as a function of insect mass at 9.4 GHz, based on Riley's summary. The solid trace is the calculated RCS of water droplets for comparison. (*Copyright, 1985 IEEE.*[14])

Frequency, GHz	RCS, m²
0.41	0.033–2.33
1.12	0.098–0.997
2.89	0.140–1.05
4.80	0.368–1.88
9.375	0.495–1.22

Examples of the RCS of aircraft are shown in Figs. 11.15 through 11.17. The B-26 pattern in Fig. 11.15 was measured at a wavelength of 10 cm (frequency of about 3 GHz); the polar format is useful for display purposes but is not as convenient for detailed comparisons as a rectangular format is. The RCS levels shown in the scale model Boeing 737 patterns of Fig. 11.16 are those at the measurement frequency. To obtain the corresponding full-scale values, one must add 23.5 dB (10 log 225); the full-scale frequency is one-fifteenth of the measurement frequency in this case, or 667 MHz. The patterns shown in Fig. 11.17 are medians of RCS averages taken in cells 10° square. With modern data-collecting and -recording equipment, it is feasible to plot measured results at much finer intervals than are plotted in this figure. Note that the data is relative to 1 ft²; to convert the displayed results to dBsm, one must subtract 10.3 dB (10 log 10.76 ft²/m²).
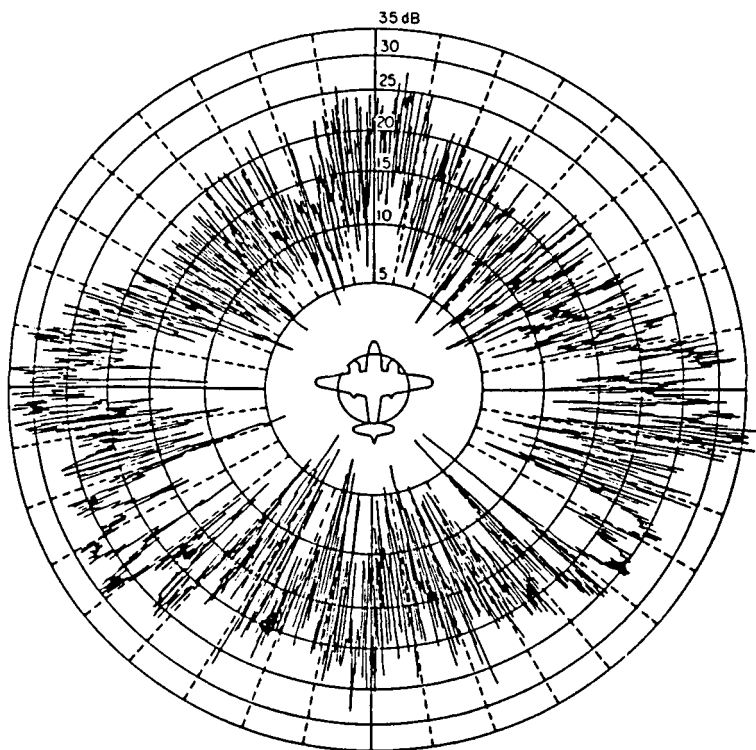
FIG. 11.15 Measured RCS pattern of a B-26 bomber at 10-cm wavelength. (*Copyright 1947, McGraw-Hill Book Company.*[17])
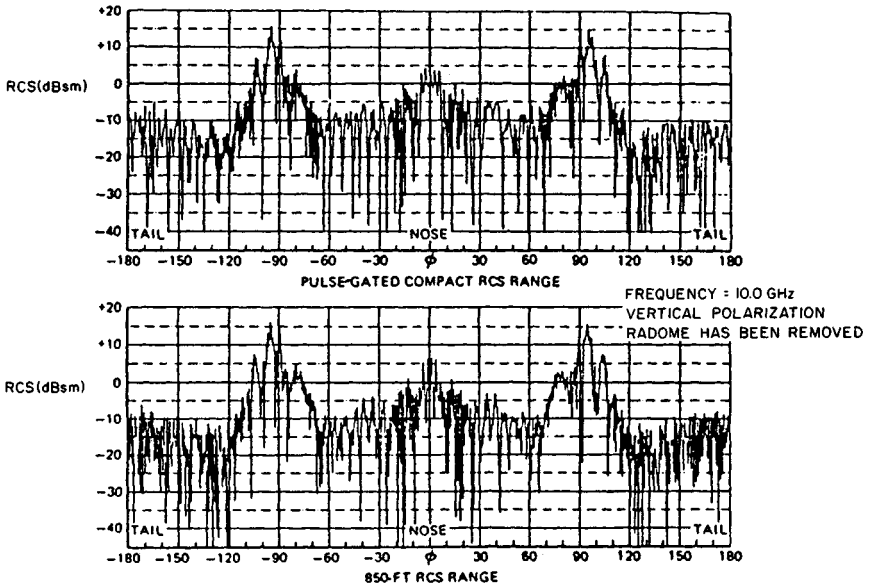
FIG. 11.16   Measured RCS of a one-fifteenth scale model Boeing 737 commercial jetliner at 10 GHz and vertical polarization. (*Copyright 1970, IEEE.*[18])

Figure 11.18 charts the RCS of a ship measured at 2.8 and 9.225 GHz at horizontal polarization. The data was collected by a shore-based radar instrumentation complex as the ship steamed in a large circle on Chesapeake Bay. The three traces in these charts are the 80, 50, and 20 percentile levels of the signals collected over aspect angle "windows" 2° wide. The patterns are not symmetrical, especially at the higher frequency. Note that the RCS can exceed 1 mi^2 (64.1 dBsm).

An empirical formula for the RCS of a naval ship is

$$\sigma = 52 f^{\frac{1}{2}} D^{\frac{3}{2}} \tag{11.4}$$

where f is the radar frequency in megahertz and D is the full-load displacement of the vessel in kilotons.[20,21] The relationship is based on measurements of several ships at low grazing angles and represents the average of the median RCS in the port and starboard bow, and quarter aspects, but excluding the broadside peaks. The statistics include data collected at nominal wavelengths of 3.25, 10.7, and 23 cm for ship displacements ranging from 2 to 17 kilotons.

Figure 11.19 summarizes the general RCS levels of the wide variety of targets discussed in this section, with the RCS of a metallic sphere shown as a function of its volume for comparison. The ordinate is the RCS in square meters, and the abscissa is the volume of the target in cubic feet. Because the chart is intended only to display the wide range in RCS that may be encountered in practice, the locations of targets on the chart are approximate at best. Within given classes of target the RCS may be expected to vary by as much as 20 or 30 dB, depending on frequency, aspect angle, and specific target characteristics. The reader requiring
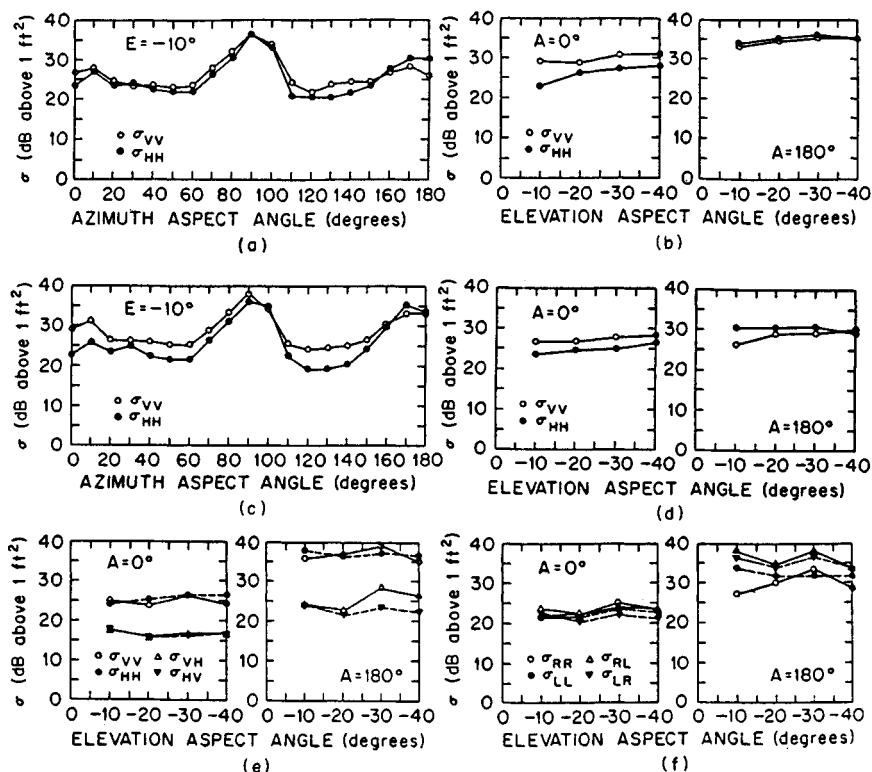
FIG. 11.17 Measured RCS of a C-54 aircraft in azimuth and elevation planes for linear and circular polarizations. Plotted values are the average RCS in a cell 10° in azimuth by 10° in elevation. Azimuth patterns *a* and *c* are for a fixed elevation angle of $-10°$. The remaining patterns are in the elevation plane for fixed nose-on or tail-on azimuths. The first and second subscripts give transmitted and received polarizations; *H* and *V* indicate horizontal and vertical polarizations, and *R* and *L* indicate right circular and left circular polarizations. (*Courtesy of I. D. Olin and F. D. Queen,*[19] *Naval Research Laboratory.*)

more explicit detail than this should consult referenced material at the end of this chapter.

11.3 RCS PREDICTION TECHNIQUES

Although the complexity and size of most scattering objects preclude the application of exact methods of radar cross-section prediction, exact solutions for simple bodies provide valuable checks for approximate methods. The exact methods are restricted to relatively simple or relatively small objects in the Rayleigh and resonant regions, while most of the approximate methods have been developed for the optics region. There are exceptions to these general limitations, of course; the exact solutions for many objects can be used for large bodies in the optics region if one uses arithmetic of sufficient precision, and many of the optics ap-
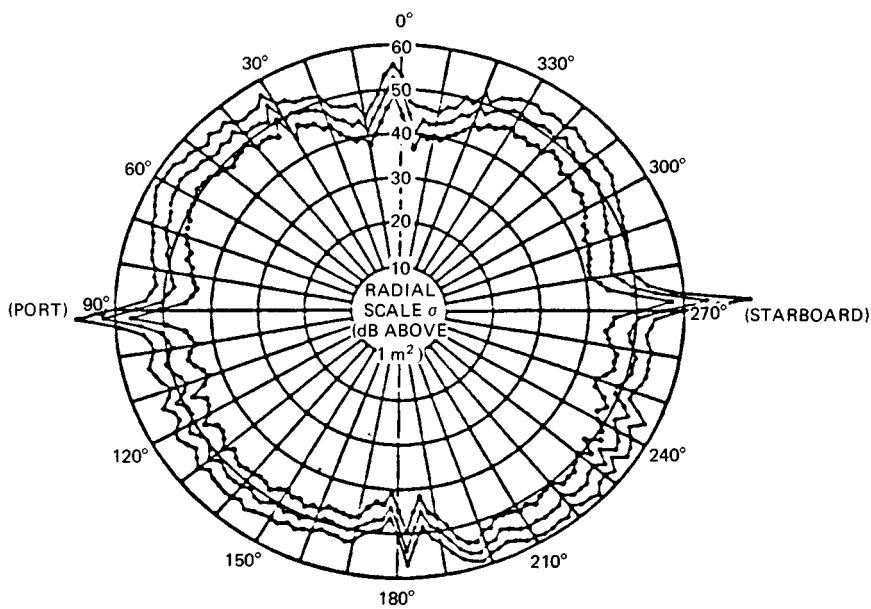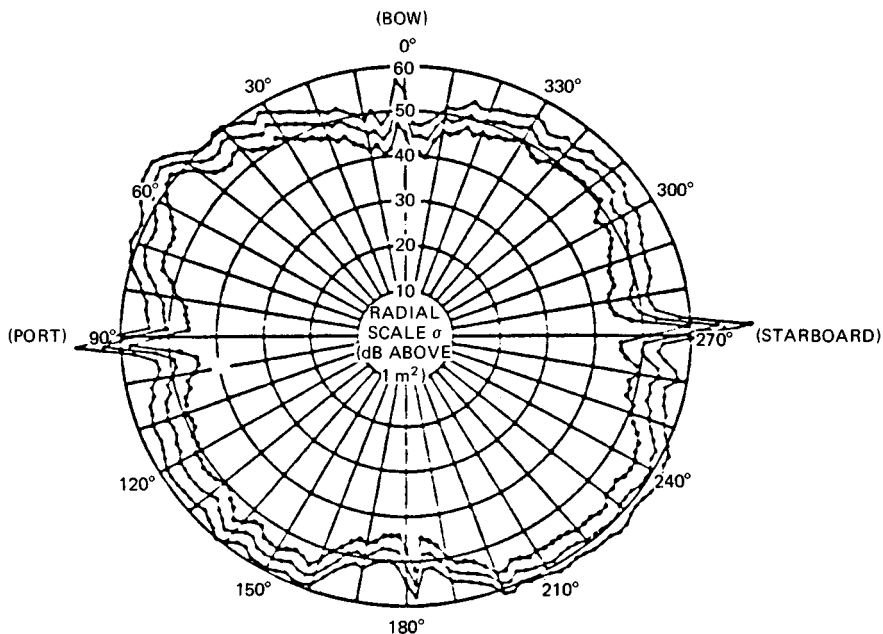
FIG. 11.18 Measured RCS of a large naval auxiliary ship for horizontal incident polarization. Upper pattern (a) is for 2.8 GHz and the lower (b) for 9.225 GHz. Shown are the 80, 50, and 20 percentile levels based on the statistics of the data over 2° aspect angle windows.
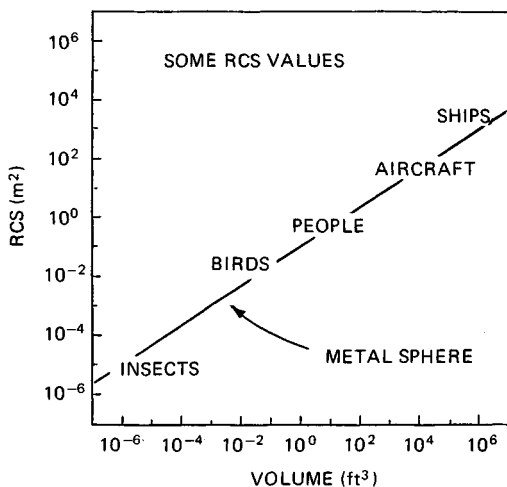
FIG. 11.19 Summary of RCS levels of targets discussed in this section. The locations of targets on the chart are general indications only.

proximations can be extended to bodies of modest electrical size in the resonance region. Low-frequency approximations developed for the Rayleigh region can extend nearly into the resonance region.

Exact Methods

Differential Equations. The exact methods are based on either the integral or differential form of Maxwell's equations. Maxwell's four differential equations constitute a succinct statement of the relationship between electric and magnetic fields produced by currents and charges and by each other.[22] The four equations may be manipulated for isotropic source-free regions to generate the wave equation

$$\nabla^2 \mathbf{F} + k^2 \mathbf{F} = 0 \tag{11.5}$$

where $\mathbf{F}$ represents either the electric field or the magnetic field. Equation (11.5) is a second-order differential equation which may be solved as a boundary-value problem when the fields on the surface of the scattering obstacle are specified. The fields are typically represented as the sum of known and unknown components (incident and scattered fields), and the boundary conditions are the known relationships that must be satisfied between the fields (both electric and magnetic) just inside and just outside the surface of the obstacle exposed to the incident wave. Those boundary conditions are particularly simple for solid conducting or dielectric objects.

The boundary conditions involve all three components of the vector fields, and the surface of the body must coincide with a coordinate of the geometrical system in which the body is described. The solution of the wave equation is most useful for those systems in which the equation is separable into ordinary differential equations in each of the variables. The scattered fields are typically expressed in terms of infinite series, the coefficients of which are to be determined in the actual solution of the problem. The solution allows the fields to be calculated at any point in space, which in RCS problems is the limit as the distance

from the obstacle becomes infinite. The product implied in Eqs. (11.1) and (11.2) is then formed from the solution of the wave equation, yielding the scattering cross section or the scattering width.

An example of a solution of the wave equation is the following infinite series for a perfectly conducting sphere:

$$\frac{\sigma}{\pi a^2} = \left| \sum_{n=1}^{\infty} \frac{(-1)^n (2n+1)}{f_n(ka)[ka\, f_{n-1}(ka) - nf_n(ka)]} \right|^2 \tag{11.6}$$

The function $f_n(x)$ is a combination of spherical Bessel functions of order n and may be formed from the two immediately lower order functions by means of the recursion relationship

$$f_n(x) = \frac{2n-1}{x} f_{n-1}(x) - f_{n-2}(x) \tag{11.7}$$

An efficient computational algorithm may be developed by using the two lowest orders as starting values,

$$f_0(x) = 1$$

$$f_1(x) = (1/x) - i$$

Equation (11.6) was used to compute the RCS characteristics plotted in Figs. 11.2 and 11.3. The infinite summation is truncated at the point where additional terms are negligible. The number of terms N required to compute the value of the bracketed term in Eq. (11.6) to six decimal places for $ka < 100$ is approximately

$$N = 8.53 + 1.21(ka) - 0.001(ka)^2 \tag{11.8}$$

The constants in Eq. (11.8) are slightly different for $ka > 100$ and are lower in value for fewer decimal places in the required accuracy.

The solution of the wave equation for the infinite, perfectly conducting circular cylinder can be resolved into two cases, one each for the incident electric or magnetic field parallel to the cylinder axis. The expressions are slightly simpler than Eq. (11.6) and involve cylindrical Bessel functions of the first and second kinds.[23] Figures 11.20 and 11.21 illustrate the backscattering behavior for the two principal polarizations as a function of the electrical circumference of the cylinder.

The response for E polarization (Fig. 11.20) is much larger than geometric optics value, πa, when the cylinder is less than a fraction of a wavelength in circumference, but it approaches the geometric optics value within a few percent for cylinders larger than about 2 wavelengths in circumference. The backscattering is markedly different for H polarization (Fig. 11.21), exhibiting the same kind of undulations noted earlier in the case of the metallic sphere. These undulations are caused by creeping waves that propagate around the rear of the cylinder just as they do around a sphere. However, the peaks and nulls of the sphere and cylinder interference patterns are not perfectly aligned with each another, suggesting that the relative phase angles between the creeping waves and specular contributions are slightly different for the two geometries.

The exact expression for the RCS of the dielectric cylinder is more complicated than for the conducting cylinder, but it accounts for the fact that energy penetrates the interior of the body. Unless the cylinder material is a perfect in-
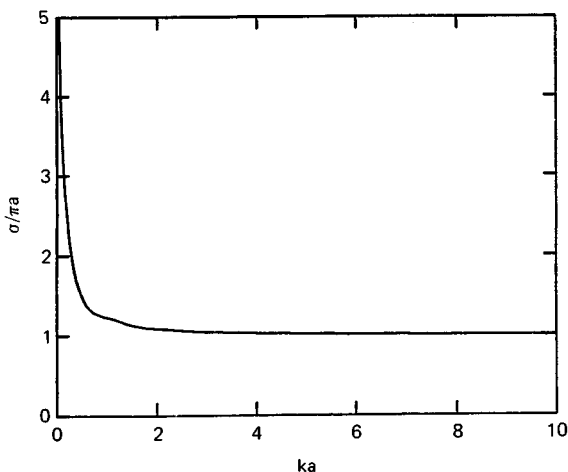
FIG. 11.20 Normalized scattering width of an infinite, perfectly conducting cylinder for E polarization (incident electric field parallel to the cylinder axis). The normalization is with respect to the geometric optics return from the cylinder.
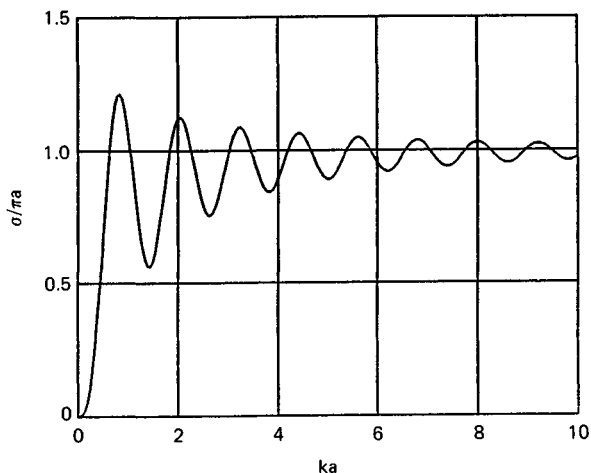

FIG. 11.21 Normalized scattering width of an infinite, perfectly conducting cylinder for H polarization (incident magnetic field parallel to the cylinder axis). The normalization is with respect to the geometric optics return from the cylinder.

sulator, its index of refraction is a complex function whose imaginary part gives rise to losses in the material. This in turn requires the computation of Bessel functions of complex argument, not an insignificant undertaking. Quite simple formulas for the scattering width may be obtained in the Rayleigh region, however, for which the cylinder diameter is much smaller than the incident wave-

length. Figure 11.5 illustrates the scattering behavior of very thin dielectric cylinders.

Integral Equations. Maxwell's equations may also be manipulated to generate a pair of integral equations (known as the Stratton-Chu equations[24]),

$$\mathbf{E}_s = \oint \{ikZ_0(\mathbf{n} \times \mathbf{H})\psi + (\mathbf{n} \times \mathbf{E}) \times \nabla\psi + (\mathbf{n} \cdot \mathbf{E})\nabla\psi\}dS \qquad (11.9)$$

$$\mathbf{H}_s = \oint \{ -ikY_0(\mathbf{n} \times \mathbf{E})\psi + (\mathbf{n} \times \mathbf{H}) \times \nabla\psi + (\mathbf{n} \cdot \mathbf{H})\nabla\psi\}dS \qquad (11.10)$$

where **n** is the unit surface normal erected at the surface patch dS and the Green's function ψ is

$$\psi = e^{ikr}/4\pi r \qquad (11.11)$$

The distance r in Eq. (11.11) is measured from the surface patch dS to the point at which the scattered fields are desired. These expressions state that if the total electric and magnetic field distributions are known over a closed surface S, the scattered fields anywhere in space may be computed by summing (integrating) those surface field distributions.

The surface field distributions may be interpreted as induced electric and magnetic currents and charges, which become unknowns to be determined in a solution. The two equations are coupled because the unknowns appear in both. Unknown quantities also appear on both sides of the equations because the induced fields include the known incident field intensity and the unknown scattered field intensity. The method of solution is known as the *method of moments* (MOM),[25] reducing the integral equations to a collection of homogeneous linear equations which may be solved by matrix techniques.

The solution of the integral equations begins with the specification of the relation between the incident and scattered fields on the surface S, as governed by the material of which the object is made. If the body is perfectly conducting or if the electric and magnetic surface fields can be related by a constant (the surface impedance boundary condition), the equations become decoupled, and only one or the other need be solved. If the body is not homogeneous, the fields must be sampled at intervals within its interior volume, complicating the solution.

Once the boundary conditions have been specified, the surface S is split into a collection of small discrete patches, as suggested in Fig. 11.22. The patches must be small enough (typically less than 0.2λ) that the unknown currents and charges on each patch are constant or at least can be described by simple functions. A weighting function may be assigned to each patch, and the problem is essentially solved when the amplitude and phase of those functions have been determined.

The point of observation is forced down to a general surface patch, whereupon the fields on the left sides of Eqs. (11.9) and (11.10) are those due to the coupling of the fields on all other patches, plus the incident fields and a "self-field." The self-field (or current or charge) is moved to the right side of the equations, leaving only the known incident field on the left side. When the process is repeated for each patch on the surface, a system of $2n$ linear homogeneous equations in $2n$ unknowns is generated. If the boundary conditions permit the decoupling of the equations, the number of unknowns may be halved (n equations in n unknowns). The coefficients of the resulting matrix involve only the electrical distances (in wavelengths) between all patches taken by pairs and the orientation of the patch surface normals. The unknown fields may be found by inverting the resulting matrix and multiplying the inverted matrix by a column matrix representing the incident field at each patch. The surface fields are then summed in integrals like Eqs. (11.9) and (11.10) to obtain the scattered field, which then may be inserted in
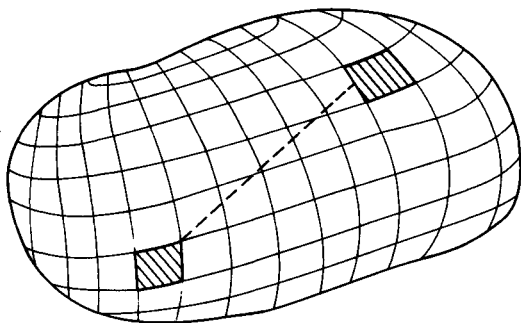
FIG. 11.22 The method of moments divides the body surface into a collection of discrete patches.

Eq. (11.1) to compute the RCS. Equation (11.2) and the two-dimensional counterparts of Eqs. (11.9) and (11.10) must be used for two-dimensional geometries, of course.

The method of moments has become a powerful tool in the prediction and analysis of electromagnetic scattering, with applications in antenna design as well as RCS prediction. The method has three limitations, however.

First, because computer memory and processing time both increase rapidly with the electrical size of the object, MOM is economically restricted to objects not much more than a few wavelengths, or perhaps a few dozen wavelengths, in size. As such, MOM is not a useful tool for predicting the RCS of, say, a jet fighter in the beam of a radar operating at 10 GHz. The second limitation is that MOM yields numbers, not formulas, and is therefore a numerical experimental tool. Trends may be established only by running a numerical experiment repeatedly for small parametric changes in the geometry or configuration of an object or in the angle of arrival or the frequency of the incident wave. Third, the solutions for some objects may contain spurious resonances that do not actually exist, thereby reducing the confidence one may have in applying the method to arbitrary structures.

Figure 11.23 traces the broadside RCS of a perfectly conducting cube computed by means of the method of moments. Spurious resonances were suppressed in the computations by forcing the normal surface component of the magnetic field to zero. The surface of the cube was divided into 384 patches (64 patches per face), which was about the limit of the central memory of the Cyber 750 computer used in the computations. It required more than 2 h for the Cyber 750 to generate the data plotted in the figure.[26]

Approximate Methods. Approximate methods for computing scattered fields are available in both the Rayleigh and the optics regions. Rayleigh region approximations may be derived by expanding the wave equation (11.5) in a power series of the wavenumber k.[27] The expansion is quasi-static for small wavenumbers (long wavelengths compared with typical body dimensions), and higher-order terms become progressively more difficult to obtain. The RCS pattern of a Rayleigh scatterer is very broad, especially if the object has similar transverse and longitudinal dimensions. The magnitude of the echo is proportional to the square of the volume of the object and varies as the fourth power of the frequency of the incident wave.[28] Because the method of moments

is well suited to the solution of Rayleigh region problems, approximate methods for predicting the RCS of electrically small objects are not presented here.

Several approximate methods have been devised for the optics region, each with its particular advantages and limitations. The most mature of the methods are *geometric optics* and *physical optics*, with later methods attacking the problem of diffraction from edges and shadow boundaries. While the general accuracy of the optics region approximations improves as the scattering obstacle becomes electrically larger, some of them give reasonably accurate results (within 1 or 2 dB) for objects as small as a wavelength or so.

The theory of geometric optics is based on the conservation of energy within a slender fictitious tube called a *ray*. The direction of propagation is along the tube, and contours of equal phase are perpendicular to it. In a lossless medium, all the energy entering the tube at one end must come out the other, but energy losses within the medium may also be accounted for. An incident wave may be represented as a collection of a large number of rays, and when a ray strikes a surface, part of the energy is reflected and part is transmitted across the surface. The amplitude and phase of the reflected and transmitted rays depend on the properties of the media on either side of the surface. The reflection is perfect if the surface is perfectly conducting, and no energy is transmitted across the boundary. When energy can pass through the surface, transmitted rays are bent toward the surface normal in crossing a surface into an electrically denser medium (higher index of refraction) and away from the surface normal into a less dense medium. This bending of rays is known as *refraction*.

Depending on surface curvature and body material, reflected and transmitted rays may diverge from one another or they may converge toward each other. This dependence is the basis for the design of lenses and reflectors at radar wavelengths as well as optical wavelengths. The variation of the refractive index of the water molecule with wavelength is responsible for the rainbow, the result of two refractions near the front of a spherical water droplet and a single internal reflection from the rear. Secondary and tertiary rainbows are due to double and triple internal reflections.

The reduction in intensity as the rays diverge (spread away) from the point of reflection can be calculated from the curvatures of the reflecting surface and the incident wave at the *specular point*, that point on the surface where the angle of reflection equals the angle of incidence. The principal radii of curvature of the surface are measured in two orthogonal planes at the specular point, as shown in Fig. 11.24. When the incident wave is planar and the direction of interest is back toward the source, the geometric optics RCS is simply
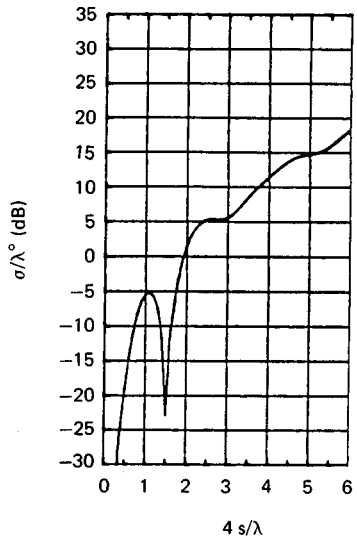
$$\sigma = \pi a_1 a_2 \tag{11.12}$$

FIG. 11.23 Broadside RCS of a perfectly conducting cube (s = edge length). (*Copyright 1985, IEEE.*[26])

The vertical axis of the graph is labeled σ/λ° (dB) with values from -30 to 35, and the horizontal axis is labeled $4 s/\lambda$ with values from 0 to 6.
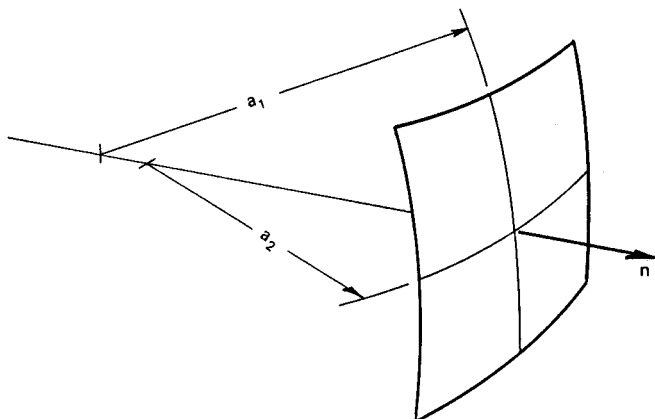
FIG. 11.24 The geometric optics RCS of a doubly curved surface depends on the principal radii of curvature at the specular point. The specular point is that point on the surface where the surface normal points toward the radar.

where a_1 and a_2 are the radii of curvature of the body surface at the specular point.

This formula becomes exact in the optical limit of vanishing wavelength and is probably accurate to 10 or 15 percent for radii of curvature as small as 2 or 3 wavelengths. It assumes that the specular point is not close to an edge. When applied to dielectric objects, the expression should be multiplied by the square of the voltage reflection coefficient associated with the material properties of the object. Internal reflections should also be accounted for, and the phase of internally reflected rays adjusted according to the electrical path lengths traversed within the body material. The net RCS then should be computed as the coherent sum of the surface reflection plus all significant internal reflections. Equation (11.12) fails when one or both surface radii of curvature at the specular point become infinite, yielding infinite RCS, which is obviously wrong. This occurs for flat and singly curved surfaces.

The theory of physical optics (PO) is a suitable alternative for bodies with flat and singly curved surface features. The theory is based on two approximations in the application of Eqs. (11.9) and (11.10), both of which are reasonably effective approximations in a host of practical cases. The first is the *far-field approximation*, which assumes that the distance from the scattering obstacle to the point of observation is large compared with any dimension of the obstacle itself. This allows one to replace the gradient of Green's function with

$$\nabla \psi = ik\psi_o \mathbf{s} \qquad (11.13)$$

$$\psi_o = e^{-ik\mathbf{r} \cdot \mathbf{s}} e^{ikR_o}/4\pi R_o \qquad (11.14)$$

where $\mathbf{r}$ is the position vector of integration patch dS and $\mathbf{s}$ is a unit vector pointing from an origin in or near the object to the far-field observation point, usually

back toward the radar.[29] R_o is the distance from the origin of the object to the far-field observation point.

The second is the *tangent plane approximation*, in which the tangential field components $\mathbf{n} \times \mathbf{E}$ and $\mathbf{n} \times \mathbf{H}$ are approximated by their geometric optics values. That is, a tangent plane is passed through the surface coordinate at the patch dS, and the total surface fields are taken to be precisely those that would have existed had the surface at dS been infinite and perfectly flat. Thus the unknown fields in the integrals of Eqs. (11.9) and (11.10) may be expressed entirely in terms of the known incident field values. The problem then becomes one of evaluating one of the two integrals and substituting the result into Eq. (11.1) to obtain the RCS.

If the surface is a good conductor, the total tangential electric field is virtually zero and the total tangential magnetic field is twice the amplitude of the incident tangential magnetic field:

$$\mathbf{n} \times \mathbf{E} = 0 \tag{11.15}$$

$$\mathbf{n} \times \mathbf{H} = \begin{cases} 2\mathbf{n} \times \mathbf{H}_i & \text{illuminated surfaces} \\ 0 & \text{shaded surfaces} \end{cases} \tag{11.16}$$

Note that the tangential components of both the electric and the magnetic fields are set to zero over those parts of the surface shaded from the incident field by other body surfaces. Other approximations may be devised for nonconducting surfaces; if the incident wavelength is long enough, for example, the surface of a soap bubble or the leaf of a tree may be modeled as a thin membrane, on which neither the electric nor the magnetic fields are zero.

The integral is easy to evaluate for flat metallic plates because the phase is the only quantity within the integral that varies, and it varies linearly across the surface. The result for a rectangular plate viewed in a *principal plane* is

$$\sigma = 4\pi \left| \frac{A \cos \theta}{\lambda} \cdot \frac{\sin (k\,\ell \sin \theta)}{k\,\ell \sin \theta} \right|^2 \tag{11.17}$$

where A is the physical area of the plate, θ is the angle between its surface normal and the direction to the radar, and ℓ is the length of the plate in the principal plane containing the surface normal and the radar line of sight. A more general physical optics formula is available for the bistatic scattering of a polygonal plate with an arbitrary number of sides.[30,31]

A rectangular plate has a pair of orthogonal principal planes, and the edge length ℓ in Eq. (11.17) is that lying in the plane of measurement. If we designate w as the width of the plate in the opposite plane, the area of the plate is $A = \ell w$. To evaluate the maximum sidelobe levels of the plate RCS in the principal plane of measurement, we may replace the numerator of the $\sin (x)/x$ term in Eq. (11.17) by unity. Normalizing with respect to the square of the width of the plane in the plane orthogonal to the measurement plane, we find the maximum sidelobe levels to be

$$\frac{\sigma}{w^2} = \frac{1}{\pi \tan^2 \theta} \tag{11.18}$$

Note that this result is *independent of the radar wavelength.*

The frequency independence of the principal-plane sidelobes is illustrated in Fig. 11.25. For viewing angles away from normal incidence, the plate edges are the dominant sources of echo, and the sin $(x)/x$ pattern is the result of the individual edge contributions changing phase with respect to each other as the aspect angle changes. Noting from Table 11.1 that the radar echoes of straight edges perpendicular to the line of sight are independent of frequency, the result of Eq. (11.18) is to be expected.
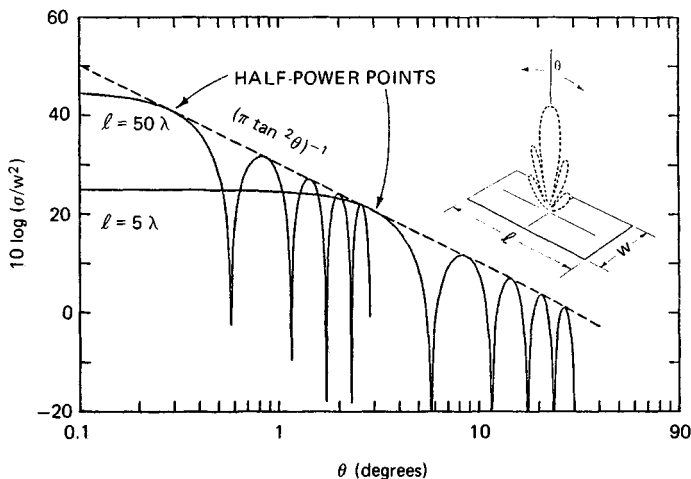
FIG. 11.25 The amplitudes of the principal-plane sidelobes of the RCS of a flat rectangular plate are independent of frequency. (*Courtesy of Walter W. Lund, Jr., The Boeing Company.*)

The physical optics formula for the RCS of a circular metallic disk is

$$\sigma = 16\pi \left| \frac{A \cos \theta}{\lambda} \cdot \frac{J_1(kd \sin \theta)}{kd \sin \theta} \right|^2 \tag{11.19}$$

where A is the physical area of the disk, d is its diameter, and $J_1(x)$ is the Bessel function of the first kind of order 1. Equations (11.17) and (11.19) both reduce to the value listed in Table 11.1 for normal incidence.

The integral is somewhat more complicated to evaluate when the surface is singly or doubly curved. An exact evaluation can be performed for a circular cylinder and a spherical cap viewed along the axis of symmetry, but not for a truncated cone or a spherical cap seen along other than the axis of symmetry. Even so, the exact evaluation for the cylinder includes fictitious contributions from the shadow boundaries at the sides of the cylinder that do not appear in a *stationary phase approximation.*[32]

The amplitude of the elemental surface patch contributions changes slowly over the surface of integration while the phase changes much more rapidly. As such, the net contribution in regions of rapid phase change is essentially zero and may be ignored. As the specular regions are approached, on the other hand, the phase variation slows down and then reverses as the specular point is crossed. This results in a nonzero specular contribution to the integral. The phase varia-

tion near the shadow boundaries is rapid; hence surface contributions there are ignored in a stationary phase evaluation, but an exact evaluation includes them because the shadow boundaries are the limits of integration. Because the actual surface field distributions do not suddenly drop to zero as the shadow boundary is crossed, as assumed by the theory, the shadow boundary contributions are spurious.[33,34] Therefore, a stationary phase approximation of the physical optics integral over closed curved surfaces tends to be more reliable than an exact evaluation.

With this in mind, the stationary phase result for a circular cylinder is

$$\sigma = ka\ell^2 \left| \frac{\sin (k\ell \sin \theta)}{k\ell \sin \theta} \right|^2 \qquad (11.20)$$

where a is the radius of the cylinder, ℓ is its length, and θ is the angle off broadside incidence. Equation (11.20) includes only the contribution from the curved side of the cylinder and not its flat ends, which may be included by using the prescription of Eq. (11.19). Equation (11.20) may be used to estimate the RCS of a truncated right circular cone if the radius a is replaced by the mean radius of the cone and ℓ is replaced by the length of the slanted surface.

While the theory of physical optics offers a significant improvement over geometric optics for flat and singly curved surfaces, it suffers from other drawbacks. Although one obtains the proper result for most of the illuminated surface, the physical optics integral yields false contributions from the shadow boundaries, as noted above. Moreover, the theory shows no dependence on the polarization of the incident wave and yields different results when the receiver and the transmitter are interchanged. These effects contradict observed behavior. Finally, it errs by wider margins as the direction of observation moves farther away from the specular direction. As illustrated in Fig. 11.11, the theory is quite accurate at broadside incidence (the specular case), but the agreement between measurement and prediction becomes progressively worse as the scattering angle moves away from that direction. Keller's *geometrical theory of diffraction* (GTD) offers an improvement in both the polarization dependence and the predicted values in the wide-angle regions.[35,36]

GTD is a ray-tracing method that assigns an amplitude and phase to fields diffracted at smooth shadow boundaries and at surface discontinuities. Because the latter are much more significant in backscattering computations than the former, we focus here on edge diffraction. The theory assumes that a ray striking an edge excites a cone of diffracted rays, as in Fig. 11.26. The half angle of this *diffraction cone* is equal to the angle between the incident ray and the edge. Unless the point
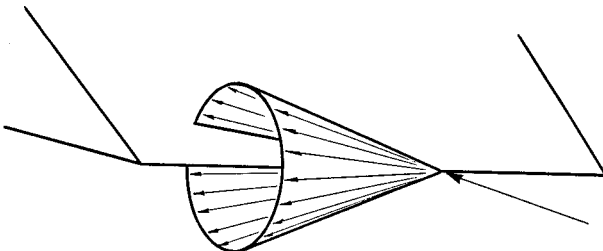
FIG. 11.26 The Keller cone of diffracted rays.

of observation lies on the diffraction cone, no value is assigned the diffracted field. The scattering direction in backscattering problems is the reverse of the direction of incidence, whence the diffraction cone becomes a disk, and the scattering edge element is perpendicular to the line of sight.

The amplitude of the diffracted field is given by the product of a *diffraction coefficient* and a *divergence factor*, and the phase depends on the phase of the edge excitation and on the distance between the observation point and the diffracting edge element. Two cases are recognized, depending whether the incident field is polarized parallel or perpendicular to the edge.

The diffracted field is given by the formula

$$E_d = \frac{\Gamma e^{iks} e^{i\pi/4}}{\sqrt{2\pi ks} \, \sin \beta} (X \mp Y) \tag{11.21}$$

where Γ is a divergence factor, X and Y are diffraction coefficients, β is the angle between the incident ray and the edge, and s is the distance to the observation point from the point of diffraction. The difference of the two diffraction coefficients is used when the incident electric field is parallel to the edge (TM polarization) and the sum when the incident magnetic field is parallel to the edge (TE polarization).

The divergence factor accounts for the decay in amplitude as the rays spread away from the edge element and includes the effects of the radius of the edge if it is curved, as at the end of a truncated cylinder, and the radius of curvature of the incident phase front.[37] The divergence factor for a two-dimensional edge (of infinite length) illuminated by a plane wave is $\Gamma = 1/s$. The diffraction coefficients are

$$X = \frac{\sin (\pi/n)/n}{\cos (\pi/n) - \cos [(\phi_i - \phi_s)/n]} \tag{11.22}$$

$$Y = \frac{\sin (\pi/n)/n}{\cos (\pi/n) - \cos [(\phi_i + \phi_s)/n]} \tag{11.23}$$

where ϕ_i and ϕ_s are the angles of the planes of incidence and scattering, as measured from one face of the wedge, and n is the exterior wedge angle normalized with respect to π; see Fig. 11.27. The three-dimensional result for an edge of finite length ℓ may be obtained by inserting Eqs. (11.22) and (11.23) in Eq. (11.21), using Eq. (11.21) for V_s/V_0 in Eq. (11.2), and then inserting Eq. (11.2) in Eq. (11.3).

Figures 11.28 and 11.29 compare measured and GTD-predicted RCS patterns of a right circular cone frustum. The theory replicates most of the pattern features for both polarizations but fails in three different aspect angle regions. These aspects are the specular directions of the flat surfaces at either end of the frustum (0 and 180° on the charts) and near the specular flash from the slanted side at 80°. The failure is due to a singularity in the diffraction coefficient Y along the reflection boundary, and a similar singularity occurs in the diffraction coefficient X along the shadow boundary, a situation encountered in forward scattering.

The singularities are overcome in the *physical theory of diffraction* (PTD) formulated by P. Ia. Ufimtsev.[39,40] (Although these publications may be difficult to find, we cite them here for completeness.) Like Keller, Ufimtsev relied on the (exact) canonical solution of the two-dimensional wedge problem, but he distin-
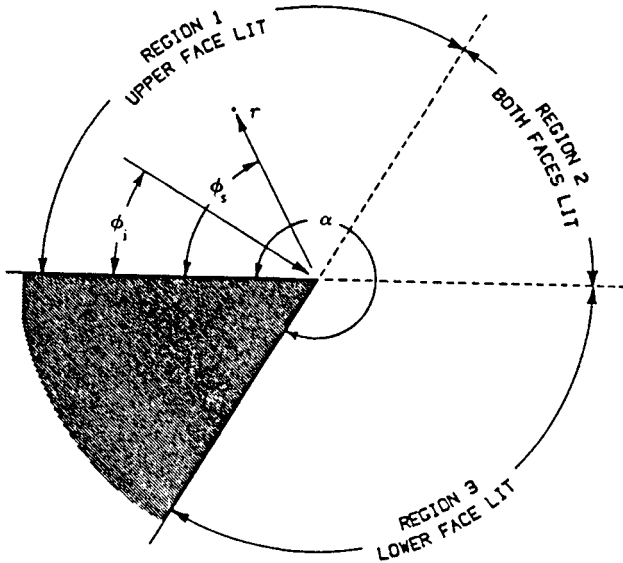
FIG. 11.27   Angles of incidence and scattering for wedge geometry.

guished between "uniform" and "nonuniform" induced surface currents. The uniform currents are the surface currents assumed in the theory of physical optics, and the nonuniform currents are associated with the edge itself (filamentary currents). The PTD result for two-dimensional problems may be represented as a linear combination of TM and TE polarizations,

$$E_s = E_0 f \frac{e^{ik\rho} e^{i\pi/4}}{\sqrt{2\pi k \rho}} \tag{11.24}$$

$$H_s = H_0 g \frac{e^{ik\rho} e^{i\pi/4}}{\sqrt{2\pi k \rho}} \tag{11.25}$$

where ρ is the distance to the far-field observation point and f and g are

$$f = \begin{cases} (X - Y) - (X_1 - Y_1) & 0 \le f_i < a\ 2\ p \\ (X - Y) - (X_1 - Y_1) - (X_2 - Y_2) & \alpha - \pi \quad \phi_i \le p \\ (X - Y) - (X_2 - Y_2) & \pi \quad \phi_i \le a \end{cases} \tag{11.26}$$

$$g = \begin{cases} (X + Y) - (X_1 + Y_1) & 0 \le f_i \le a\ 2\ p \\ (X + Y) - (X_1 + Y_1) - (X_2 + Y_2) & \alpha - \pi \quad \phi_i \le p \\ (X + Y) - (X_2 + Y_2) & \pi \quad \phi_i \le a \end{cases} \tag{11.27}$$

The subscripted coefficients are known as the *physical optics diffraction coefficients,*

$$X_1 = - \tan \left[(\phi_s - \phi_i)/2 \right] \tag{11.28}$$
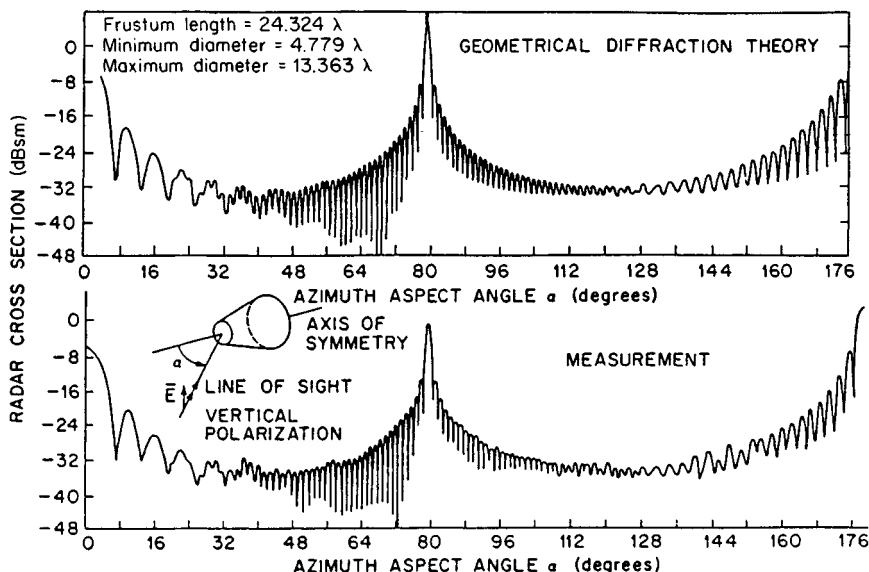
FIG. 11.28 RCS of a cone frustum, vertical polarization. (*Copyright 1966, IEEE.*[38])

$$Y_1 = - \tan \left[(\phi_s + \phi_i)/2 \right] \qquad (11.29)$$

$$X_2 = \tan \left[(\phi_s - \phi_i)/2 \right] \qquad (11.30)$$

$$Y_2 = - \tan \left[\alpha - (\phi_s + \phi_i)/2 \right] \qquad (11.31)$$

Because the PO diffraction coefficients depend on whether the upper face, the lower face, or both faces of the wedge are illuminated by the incident wave, the diffraction coefficients are combined differently in the three recognizable sectors defined in Eqs. (11.26) and (11.27). And because surface terms have been suppressed explicitly by the subtraction of the PO coefficients, the effects of surface currents (as distinguished from filamentary edge currents) must be accounted for independently. Those surface terms may be obtained, for example, by using geometrical optics or, paradoxically, the theory of physical optics after the edge terms themselves have been computed.

GTD and PTD are both based on the exact solution of the two-dimensional wedge problem, for which the directions of incidence and scattering are perpendicular to the edge. When extended to the case of oblique incidence, the direction of observation must lie along a generator of the Keller cone depicted in Fig. 11.26. If the edge is straight and of finite length, as in the three-dimensional world, Eq. (11.3) provides an approximation of the RCS. If the edge is curved, it may be regarded as a collection of infinitesimally short segments butted together, and the scattered fields may be computed via an integration of incremental fields diffracted by each element of the edge. This is the concept introduced by Mitzner,[41] and the summation of the fields diffracted by the edge elements implies an integral around the edge contour. (Although Mitzner's most significant results are embedded in a government document of limited distribution, we include this source in our references because of its significance.)

However, Mitzner sought the fields scattered in arbitrary directions, not just those along the local Keller cones, and for this purpose he developed his concept
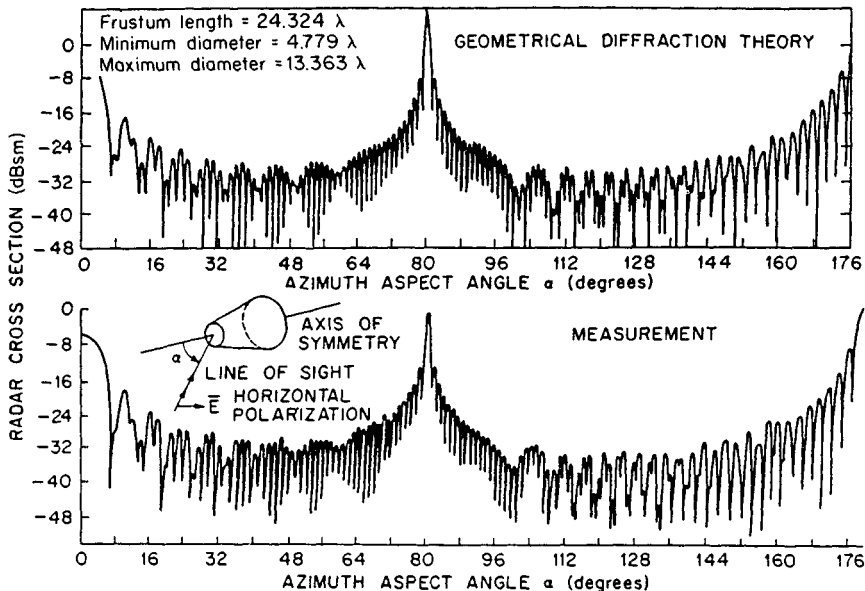
FIG. 11.29 RCS of a cone frustum, horizontal polarization. (*Copyright 1966, IEEE.*[38])

of the *incremental length diffraction coefficient*. Extending the example provided by Ufimtsev, he devised a set of diffraction coefficients for arbitrary directions of incidence and scattering. Not unexpectedly, those coefficients are more complicated than the X's and Y's appearing in Eqs. (11.22) and (11.23), and (11.28) through (11.31).[30,42]

Mitzner expressed his result as the diffracted electric-field components parallel and perpendicular to the plane of scattering in terms of the components of the incident electric field parallel and perpendicular to the plane of incidence. As such, the diffraction coefficients may be expressed as three separate pairs representing parallel-parallel, perpendicular-perpendicular, and parallel-perpendicular (or perpendicular-parallel) combinations. One member of each pair is due to the total surface current on the diffracting edge (including the assumed filamentary edge currents), and the other is due to the uniform physical optics currents. Mitzner subtracted one member of each pair from the other, thereby retaining the contributions from the filamentary currents alone.

The results have identically the form of Ufimtsev's expressions, in which the PO coefficients are subtracted from the non-PO coefficients. Thus, Mitzner's expression for the scattered field contains only the contributions from the filamentary edge currents. In applying his theory to scattering objects, therefore, the contributions of nonfilamentary induced surface currents must be accounted for separately, just as in Ufimtsev's physical theory of diffraction. When the directions of incidence and scattering become perpendicular to an edge, the perpendicular-parallel terms disappear and Mitzner's diffraction coefficients then reduce identically to Ufimtsev's.

Undertaking what he called a more rigorous evaluation of the fields induced on a wedge, Michaeli duplicated Mitzner's result for the total surface currents,

confirming Mitzner's prior development, but he did not explicitly remove the PO surface-current contributions.[43] Thus, like Keller's X and Y, Michaeli's diffraction coefficients become singular in the transition regions of the reflection and shadow directions. Michaeli later investigated the removal of the singularities, the cleverest of which was the use of a skewed coordinate system along the wedge surfaces.[44,45]

While these methods of evaluating the fields scattered by edge elements may be applicable to smooth unbounded edges, they do not account for the discontinuities at corners where the edges turn abruptly in other directions. An attack on the problem has been suggested by Sikta et al.[46]

When applying these approximate high-frequency methods of estimating the fields scattered by complex objects, it is necessary to represent the object as a collection of surfaces having relatively simple mathematical descriptions. The actual surface profiles may be approximated by segments that have conveniently simple mathematical descriptions, such as flat plates, truncated spheroids, and truncated conic sections. The total RCS may be formed by summing the field contributions of the individual segments using the methods described above or whatever other tools are available. It is important to sum the field strengths of the individual contributions, complete with phase relationships, before squaring to obtain the total RCS as given by Eq. (11.1). This is tantamount to forming the coherent sum

$$\sigma = |\sum_p \sqrt{\sigma_p} e^{i\phi_p}|^2 \qquad (11.32)$$

where σ_p is the RCS of the pth contributor and ϕ_p is its relative phase angle, accounting for the two-way propagation of energy from the radar to the scattering feature and back again. If all phase angles are equally likely, one may form instead the noncoherent sum

$$\sigma = \sum_p o_p \qquad (11.33)$$

The noncoherent RCS is meaningful only if a change in the aspect angle or a sweep in the instantaneous radar frequency does indeed result in a uniform distribution of phase angles. It is the average RCS formed over a time interval long enough to ensure the equal likelihood of all phase angles.

11.4 RCS MEASUREMENT TECHNIQUES

RCS measurements may be required for any of several reasons, ranging from scientific inquiry to verification of compliance with product specifications. There are no formal standards governing intrumentation and measurement methods, but informal standards of good measurement practice have been recognized for decades. Depending on the size of the test object, the frequencies to be used, and other test requirements, measurements may be made in indoor test facilities or on outdoor ranges. Because one is seldom interested in the RCS of an object for only one aspect angle, all static test ranges use turntables or rotators to vary the target aspect angle. Although the purpose of testing often governs how the measure-

CHAPTER 12
GROUND ECHO

Richard K. Moore
The University of Kansas

12.1 INTRODUCTION

Radar ground return is described by σ^0, the differential scattering cross section, or scattering coefficient (scattering cross section per unit area), rather than by the total scattering cross section σ used for discrete targets.[1] Since the total cross section σ of a patch of ground varies with the illuminated area and this is determined by the geometric radar parameters (pulse width, beamwidth, etc.), σ^0 was introduced to obtain a coefficient independent of these parameters.

Use of a differential scattering cross section implies that the return from the ground is contributed by a large number of scattering elements whose phases are independent. This is primarily because of differences in distance that, although small fractions of total distance, are many wavelengths. Superposition of power is possible for the computation of average returns. If this condition is not applicable to a particular ground target, the differential-scattering-cross-section concept has no meaning for that target. For example, a very-fine-resolution radar might be able to resolve a part of a car; the smooth surfaces on the car would not be properly represented by σ^0. On the other hand, a coarser radar might look at many cars in a large parking lot, and a valid σ^0 for the parking lot could be determined.

If a region illuminated at one time by a radar contains n scattering elements and the above criterion is satisfied so that power may be added, the radar equation becomes

$$P_r = \sum_i^n \frac{P_{ti}G_{ti}A_{ri}\sigma_i}{(4\pi R_i^2)^2} = \sum_i^n \frac{P_{ti}G_{ti}A_{ri}(\sigma_i/\Delta A_i)\Delta A_i}{(4\pi R_i^2)^2}$$

Here ΔA_i is an element of surface area, and P_{ti}, G_{ti}, and A_{ri} are values of P_t, G_t, and A_r appropriate for an element at the location of ΔA_i. The factor in parentheses in the numerator of the right-hand expression is the incremental scattering cross section for element i, but this concept is meaningful only in an average. Thus the average power returned is given by

$$P_r = \sum_{i}^{n} \frac{P_{ti} G_{ti} A_{ri} \sigma^0 \Delta A_i}{(4\pi R_i^2)^2}$$

Here σ^0 has been used to denote the average value of $\sigma_i / \Delta A_i$. In this formulation, we may pass in the limit from the finite sum to the integral given by

$$\overline{P}_r = \frac{1}{(4\pi)^2} \int_{\text{Illuminated area}} \frac{P_t G_t A_r \sigma^0 dA}{R^4} \tag{12.1}$$

The bar over P_r implies the average value. This integral is not really correct, for there is a minimum size for real, independent scattering centers. Nevertheless, the concept is widely used and is applicable as long as the illuminated area is large enough to contain many such centers.

Figure 12.1 illustrates the geometry associated with Eq. (12.1). Note that, for a rectangular pulse, P_t is either zero or the peak transmitter power but, for other pulse shapes, the variation with t (or R) is significant. Actual pulses are often approximated by rectangular pulses with widths equal to their half-power widths. Real pulses cannot be rectangular after passing through real receiver bandwidths. The transmitting-antenna gain and receiving-antenna aperture are functions of the elevation and azimuth angles:

$$G_t = G_t(\theta,\phi) \qquad A_r = A_r(\theta,\phi) \tag{12.2a}$$

The differential scattering cross section itself is a function of both *look angle* and ground location:

$$\sigma^0 = \sigma^0(\theta,\phi,\text{location}) \tag{12.2b}$$

The integral of Eq. (12.1) must be inverted when σ^0 is measured. With narrow beams and short pulses the inversion is relatively easy, but with the wider beams and longer pulses used in many measurements the values obtained are sometimes poorly defined.

Some authors[2] use a scattering cross section per unit projected area rather than per unit ground area. Figure 12.2 illustrates by a *side view* the difference between ground area and projected area. The ground area is proportional to $\Delta\rho$, and the projected area is smaller. Thus,

$$\sigma^0 A = \gamma d(\text{projected area}) = \gamma \cos\theta dA \tag{12.3}$$

or $\qquad\qquad \sigma^0 = \gamma \cos\theta$

Since both γ and σ^0 are called scattering coefficients, readers of the literature must be especially careful to determine which is being used by a particular author.

Radar astronomers use a different σ:[3]

$$\sigma = \frac{\text{total return power from entire surface}}{\text{power returned from perfect isotropic sphere of same radius}} \tag{12.4}$$
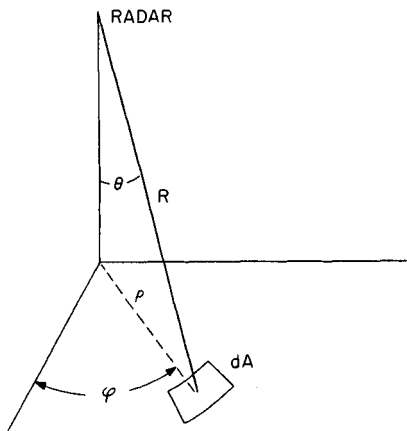
FIG. 12.1 Geometry of the radar equation.
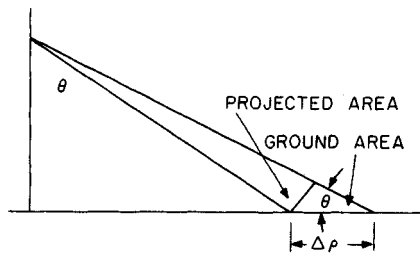
FIG. 12.2 Ground area and projected area.

The resulting value for σ is usually much smaller than σ^0 for the planet at vertical incidence and is larger than the values of σ^0 near grazing incidence (return from the limb of the planet).

Relative Importance of Theory and Empiricism. The theory of radar ground return has been the subject of many publications. The various theories, insofar as they can be confirmed by experiment, provide bases for judging the effects of variations in the dielectric properties of the ground, of the roughness of the ground and nature of vegetative or snow cover, of radar wavelength, and of angle of incidence. Viewed as aids to insight, radar ground-return theories can be extremely useful.

The validity of any ground-return theory must depend on the mathematical model used to describe the surface, as well as on the approximations required to obtain answers. Even the simplest ground surface, the sea, is extremely difficult to describe accurately; it is homogeneous to beyond the skin depth, contains relatively modest slopes, and (except for spray) has no part above another part of the surface. At grazing angles, shadowing of one wave by another might occur. Land surfaces are much more difficult to describe: Imagine an adequate mathematical description of the shape of a forest (when every leaf and pine needle must be described). Furthermore, land surfaces are seldom homogeneous either horizontally or with depth.

Since a true mathematical description of the ground surface appears out of the question, empirical measurements are necessary to describe the radar return from natural surfaces. The role of theory is to aid in interpreting these measurements and to suggest how they may be extrapolated.

Available Scattering Information. Prior to 1972 the lack of coordinated research programs over the necessary long period resulted in only one really usable set of measurements, that at Ohio State University.[2,4] Since that time extensive measurements have been made from trucks and helicopters by the University of Kansas,[6,7] a group in the Netherlands,[8] and several groups in France.[9] These measurements concentrated especially on vege-

tation, with the Kansas measurements also including some work on snow and extensive work on sea ice. Most of these measurements were in the 10 to 80° range of incidence angles. Measurements near vertical are scarcer, while well-controlled experiments near grazing are very scarce indeed.

Airborne measurements are necessary to make larger scattering areas accessible. Although airborne programs for special purposes have been legion, curves of scattering coefficient versus angle for a known homogeneous area are scarce. The work at the MIT Radiation Laboratory[10] was early work by Philco Corporation.[11] Goodyear Aerospace Corporation,[12] General Precision Laboratory,[13] and the U.S. Naval Research Laboratory (NRL)[14-16] programs were important early. More recently, the Canada Centre for Remote Sensing (CCRS) has made numerous airborne scatterometer measurements,[17] especially over sea ice. The Environmental Research Institute of Michigan (ERIM),[18] CCRS, the European Space Agency (ESA),[19] and the Jet Propulsion Laboratory (JPL)[20] used imaging synthetic aperture radars (SARs) for some scattering measurements, but most were not well calibrated.

Results of most of these measurements are summarized in Ulaby, Moore, and Fung.[21] More complete summaries of the earlier work and near-grazing studies are in Long.[22] Many applications summaries are also in the "Manual of Remote Sensing."[23] Readers requiring more detailed information should consult these books.

12.2 PARAMETERS AFFECTING GROUND RETURN

Radar return depends upon a combination of system parameters and ground parameters:

Radar system parameters [Eqs. (12.1) and (12.2*a* and *b*)]

Wavelength

Power

Illuminated area

Direction of illumination (both azimuth and elevation)

Polarization

Ground parameters

Complex permittivity (conductivity and permittivity)

Roughness of surface

Inhomogeneity of subsurface or cover to depth where attenuation reduces wave to negligible amplitude

Different wavelengths are sensitive to different elements on the surface. One of the earliest known and most striking directional effects is the *cardinal-point* effect in return from cities: Radars looking in directions aligned with primary street grids observe stronger regular returns than radars at other angles. When radars are looking at a normal-incidence angle, horizontally polarized waves are reflected better by horizontal wires, rails, etc., than are vertically polarized waves.

If the geometry of two radar targets were the same, the returns would be

stronger from the target with higher complex permittivity because larger currents (displacement or conduction) would be induced in it. Because identical geometries with differing permittivities do not occur in nature, this distinction is not easy to measure. Effective permittivity for ground targets is very strongly influenced by moisture content, since the relative permittivity of liquid water is from about 60 at X band to about 80 at S band and longer wavelengths whereas most dry solids have permittivities less than 8. Attenuation is also strongly influenced by moisture, since wet materials usually have higher conductivity than the same materials dry. Figures 12.3 and 12.4 show the effect of moisture content on properties of plants and of soil. The high permittivity of plants with much moisture means that radar return from crops varies as the plants mature, even when growth is neglected.
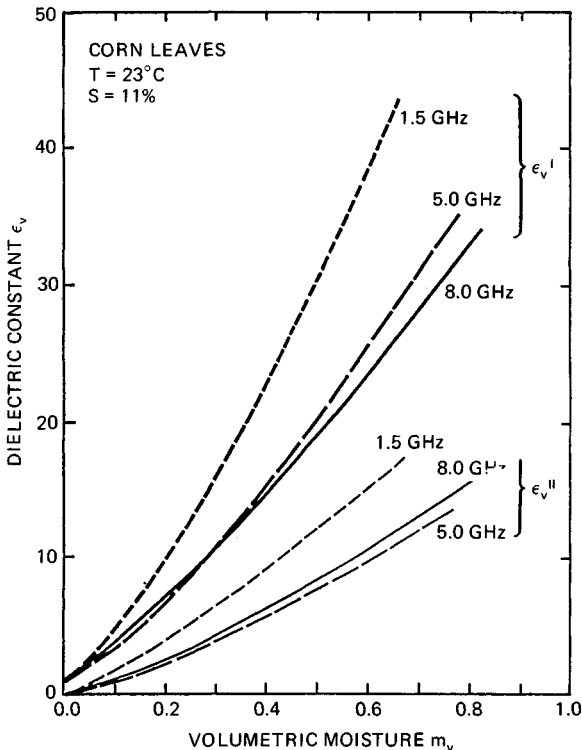
FIG. 12.3 Measured moisture dependence of the dielectric constant of corn leaves at 1.5, 5.0, and 8.0 GHz. S is the salinity of water content in parts per thousand, $\epsilon_v = \epsilon_v{}' - j\epsilon_v{}''$ is the complex dielectric constant in Fm^{-1}, and m_v is the volumetric moisture content in $kg \cdot m^{-3}$. (*After Ulaby, Moore, and Fung.*[21])

The roughness of surfaces (especially natural ones) is difficult to describe mathematically but easy to understand qualitatively. Thus it is easy to see that a freshly plowed field is rougher than the same field after rain and wind have been at work on it. A forest is inherently rougher than either a field or a city. The dif-
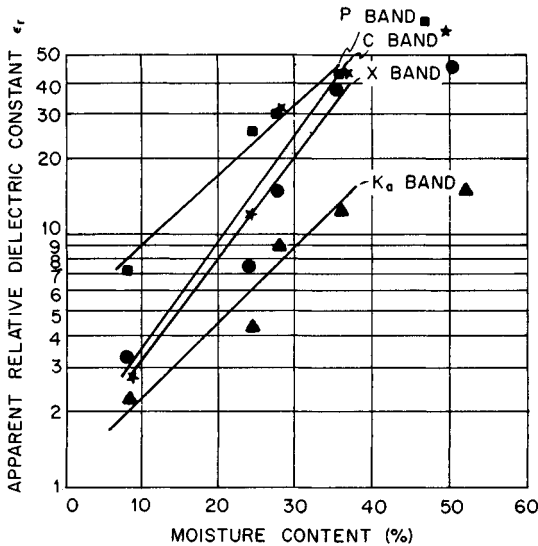
FIG. 12.4 Apparent relative dielectric constant versus moisture content (Richfield silt loam). (*After Lundien.*[24])

ference between the roughness of a city with flat walls interspersed with windowsills, with curbs, cars, and sidewalks, and the roughness of natural areas is harder to see.

Surfaces that are relatively smooth tend to reflect radio waves in accordance with the Fresnel-reflection direction,* and so they give strong backscatter only when the look angle is nearly normal to the surfaces. Rough surfaces, on the other hand, tend to reradiate nearly uniformly in all directions, and so they give relatively strong radar returns in any direction.

The problem of radar scatter is complicated because waves penetrate significant distances into many surfaces and vegetation canopies, and internal reflection and scatter contribute to the return. Measurements of attenuation for field crops[25,26] and grasses[27] show that most of the return is from the upper layers, with some contribution by the soil and lower layers if the vegetation is not very dense. Most of the signal returned from trees is usually from the upper and middle branches when the trees are in leaf,[28–32] although in winter the surface is a major contributor to the signal.

12.3 THEORETICAL MODELS AND THEIR LIMITATIONS

Descriptions of a Surface. Many theoretical models for radar return from the ground assume a rough boundary surface between air and an infinite homogeneous half space. Some include either vertical or horizontal homogeneities in the ground properties and in vegetative or snow covers.

---

*Angle of reflection equals angle of incidence.

Surface descriptions suitable for use in mathematical models are necessarily greatly idealized. Few natural grounds are truly homogeneous in composition over very wide areas. Descriptions of their detailed shape must be simplified if they are to be handled analytically, although computers permit the use of true descriptions. Very few surfaces have ever been measured to the precision appropriate for centimeter-wavelength radars; even for these there is no assurance that scattering boundaries do not exist within a skin depth beneath the surface. Surfaces containing vegetation and conglomerate rocks almost completely defy description.

Statistical descriptions of surfaces are used for most theories, since a theory should be representative of some kind of surface class, rather than of a particular surface, and since exact description is so difficult. The statistical descriptions themselves must be oversimplified, however. Many theories assume isotropic statistics, certainly not appropriate for plowed fields or gridded cities. Most theories assume some kind of model involving only two or three parameters (standard deviation, mean slope, correlation distance, etc.), whereas natural (or human-made) surfaces seldom are so simply described. The theories for vegetation and other volume scatterers have more parameters.

Simplified Models. Early radar theories for ground return assumed, as in optics, that many targets could be described by a Lambert-law variation of intensity; that is, the differential scattering coefficient varies as $\cos^2 \theta$, with θ the angle of incidence. This "perfectly rough" assumption was soon found wanting, although it is a fair approximation for the return from many vegetated surfaces over the midrange of angles of incidence.

Clapp[10] described three models involving assemblies of spheres, with different spacings and either with or without a reflecting ground plane. These models yield variations from σ^0 independent of angle through $\sigma^0 \propto \cos \theta$ to $\sigma^0 \propto \cos^2 \theta$. Since the sphere models are highly artificial, only the resulting scatter laws need be considered. Most targets give returns that vary more rapidly over part of the incidence-angle regime than these models, although forests and similar rough targets of some depth sometimes give such slowly varying returns.

Since these rough-surface models usually fail to explain the rise in return near vertical incidence, other simplified models combine Lambert's law and other rough-surface scattering models with specular reflection at vertical incidence, and a smooth curve is drawn between the specular value and the rough-surface prediction.

Specular reflection is defined as reflection from a smooth plane and obeys the Fresnel reflection laws.[33] At normal incidence, the specular-reflection coefficient is therefore

$$\Gamma_R = \frac{\eta_g - \eta_0}{\eta_g + \eta_0}$$

where η_0, η_g are the intrinsic impedances of air and earth, respectively. The fraction of total incident power specularly reflected from a rough surface is[5]

$$e^{-2(2\pi\sigma_h/\lambda)^2}$$

where σ_h = standard deviation of surface height variations
    λ = wavelength

Since this proportion is down to 13.5 percent when $\sigma_h = \lambda/2\pi$ and to 1.8 percent when $\sigma_h = \lambda/(2\pi\sqrt{2})$, significant specular reflection is seldom found for the centimeter wavelengths usually used for radar. Nevertheless, a simplified model like this is convenient for some purposes.

Observation of reflected sunlight from rippled water, from roads, and from other smooth surfaces leads to the postulation of a facet theory.[34,35] The only sunlight reaching the observer from smooth surfaces such as water is that from facets for which angle of incidence equals angle of reflection. Thus the observed light may be described by methods of *geometric optics*.

When geometric optics is used to describe radar scatter, the surface of the ground is represented by small flat-plane segments. Radar return is assumed to occur only for facets oriented normal to the radar (normal orientation is required for backscatter so that the reflected wave returns to the source). Thus, if the slope distribution of such facets is known, the fraction normal to a given diverging beam can be established, and from this the return can be obtained. Geometric optics assumes zero wavelength, and so the results of such a theory are wavelength-independent, clearly not in accord with observation.

The facet model for radar return is extremely useful for qualitative discussions, and so modification to make it fit better with observation is appropriate. Two kinds of modification may be used, separately or jointly: considering the actual reradiation pattern of finite-size facets at finite wavelengths[36] and considering the effect of wavelength on establishing the effective number of facets.[37] Thus the scatter from a facet may actually occur in directions other than that requiring that angle of incidence equal angle of reflection. Figure 12.5 illustrates this. For large facets (compared with wavelength) most of the return occurs almost at normal incidence, whereas for small facets the orientation may be off normal by a considerable amount without great reduction in scatter. As the wavelength is increased, the category of a given facet changes from *large* to *small*; eventually the facet is smaller than a wavelength, and its reradiation pattern shape remains almost isotropic from that point. Many facets that would be separate at, say, a 1-cm wavelength are combined at a 1-m wavelength; the result may be a transition from rough- to smooth-surface behavior. Figure 12.6a shows a number of facets of different sizes contributing to a radar return.

INFINITE FLAT   WIDE FACET   NARROW FACET
FACET
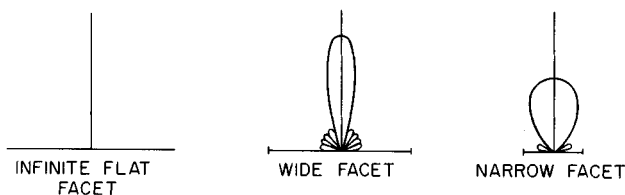
FIG. 12.5 Normal-incidence reradiation patterns of facets.

ROUGH SURFACE WITH TANGENTS

RETURN FROM INDIVIDUAL FACETS
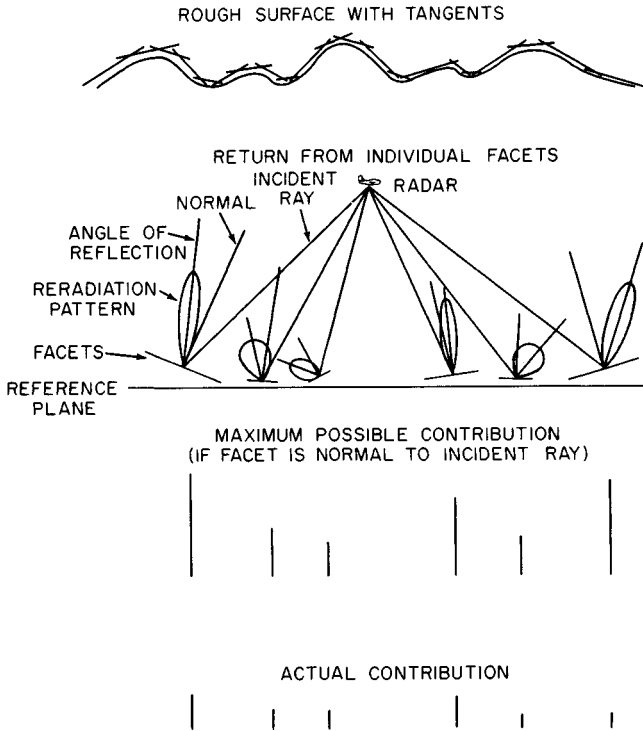
FIG. 12.6a Facet model of a radar return.

Physical Optics Models. Theories based on applications of the Kirchhoff-Huygens principle have been thoroughly developed.[21,36,38–40] The Kirchhoff approximation is that the current flowing at each point in a locally curved (or rough) surface is the same as would flow in the same surface if it were flat and oriented tangent to the actual surface. This assumption permits construction of scattered fields by assuming that the current over a rough plane surface has the same magnitude as if the surface were smooth, but with phase perturbations set by the differing distances of individual points from the mean plane. For surfaces assumed to be azimuthally isotropic, the usual approach yields integrals of the form

$$\frac{1}{\cos^3 \theta} \int e^{-(2k\sigma_h \cos \theta)^2 [1 - \rho(\xi)]} J_0(2k\xi \cos \theta) \xi \, d\xi$$

where $\rho(\xi)$ = spatial autocorrelation function of surface heights
θ = angle with vertical
σ_h = standard deviation of surface heights
k = $2\pi/\lambda$
J_0 = first-order, first-kind Bessel function

The autocorrelation function of height with distance is seldom known for terrain, although it can be determined on a large scale by analysis of contour maps,[41] and it has been found for some areas by careful contouring at close intervals and subsequent analysis. Because of lack of knowledge of actual autocorrelations, most theory has been developed with artificial functions that are chosen more for their integrability than for their fit with nature; selection among them has been on the basis of which ones yield the best fit between theoretical and experimental scatter curves.

The correlation function first used[42] was gaussian:

$$\rho(\xi) = e^{-\xi^2/L^2} \tag{12.5}$$

where L is the *correlation length*. Not only is this a function that makes the integral analytically tractable, but it also gives exactly the same results as geometric optics.[43] Since it fails, like geometric optics, to explain frequency variation, it cannot be a truly representative correlation function, although it gives a scattering curve that fits several experimental curves near the vertical. The next most frequently used function is the exponential:

$$\rho(\xi) = e^{-|\xi|/L} \tag{12.6}$$

This has some basis in contour-map analysis;[41] the results fit both earth and lunar radar return over a wider range of angles than the gaussian[41,44] (but sometimes not as well near vertical). Furthermore, it has the merit that it exhibits frequency dependence. Resulting expressions for power (scattering coefficient) variations appear in Table 12.1.

TABLE 12.1   Scattering Coefficient Variation

Correlation coefficient	Power expression	Reference
$e^{-\xi^2/L^2}$	$\dfrac{K}{\sin \theta} e^{-(L^2/2\sigma_h^2) \tan^2 \theta}$	42
$e^{-|\xi|/L}$	$\dfrac{K\theta}{\cos^2 \theta \sin \theta} \left(1 + A \dfrac{\sin^2 \theta}{\cos^4 \theta}\right)^{-3/2}$	33

Small-Perturbation and Two-Scale Models.   Recognition that existing models were inadequate for describing ocean scatter led to recognition that resonance of the signal with small structures on the surface has a powerful influence on the strength of the signal received.[45,46] Thus a small-perturbation method originally proposed by Rice[47] became the most popular way to describe ocean scatter. Its application to land scatter was not far behind.

The term *Bragg scatter* is often used to describe the mechanism for the small-perturbation model. The idea comes from the concept illustrated in Fig. 12.6b.

A single sinusoidal component of a complex surface is shown with an incoming radar wave at angle of incidence θ. The radar wavelength is λ, and the surface-component wavelength is Λ. When the signal travels an extra distance
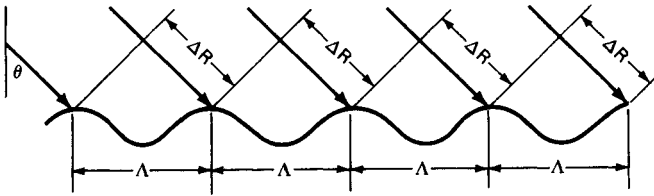
FIG. 12.6b In-phase addition for Bragg scattering; $\Delta R = n\lambda/2$.

$\lambda = 2\Delta R$ between the source and two successive wave crests, the phase difference between the echoes from successive crests is 360°; so the echo signals all add in phase. If this condition is satisfied for a particular Λ and θ, it fails to be satisfied for others. Thus, this is a resonant selection for a given θ of a particular component of the surface Λ. The strength of the received signal is proportional to the height of this component and to the number of crests illuminated by the radar. If the surface has an underlying curvature, the number of illuminated crests satisfying the resonance criterion may be limited by the length of the essentially flat region; otherwise it is limited by the radar resolution.

The theoretical expression for the scattering coefficient is[48]

$$\sigma_{pq}^{0} = 8k^4\sigma_1^2 \cos^4\theta|\alpha_{pq}|^2 W(2k\sin\theta,0) \tag{12.7}$$

where p,q = polarization indices (H or V)

$k = 2\pi/\lambda$ (the radar wavenumber)

$\alpha_{HH} = R_1$ (Fresnel reflection coefficient for horizontal polarization)

$$\alpha_{VV} = (\epsilon_r - 1)\frac{\sin^2\theta - \epsilon_r(1 + \sin^2\theta)}{[\epsilon_r\cos\theta + (\epsilon_r - \sin^2\theta)^{1/2}]^2}$$

where ϵ_r is the relative permittivity $\epsilon' - j\epsilon''$ and $\alpha_{VH} = \alpha_{HV} = 0$.

$W(2k\sin\theta, 0)$ is the *normalized roughness spectrum* (the Fourier transform of the surface autocorrelation function). It may be written as $W(K, 0)$, where K is the wavenumber for the surface. In terms of the wavelength on the surface Λ,

$$K = 2\pi/\Lambda$$

Thus the component of the surface that satisfies the Bragg resonance condition is

$$\Lambda = \lambda/2\sin\theta \tag{12.8}$$

The meaning of this is that the most important contributor to a surface return is the component of surface roughness with wavelength Λ. Even though other components may be *much* larger, the Bragg resonance makes this component more important. On the ocean this means that tiny ripples are more important than waves that are meters high; the same applies for land-surface scatter.

As originally developed, this theory was for perturbations to horizontal flat surfaces, but it was soon modified to handle surfaces with large-scale roughness. The large-scale roughness was assumed to cause a *tilting* of the flat surface to which the small-perturbation theory could be applied. The principal problem with this approach is deciding where in the spectrum lies the boundary between the larger components that do the tilting and smaller components that are Bragg-

resonant. Many papers have been written to describe the evolution of this theory; for a complete summary, the reader is referred to Fung's development in Ref. 49.

Other Models. The theory for volume scatter has led to many papers and continues to evolve. For a review of some of the approaches the reader should consult Fung's summary in Ref. 50 and papers by Kong, Lang, Fung, and Tsang. These models have been used reasonably successfully to describe scatter from vegetation,[51] snow,[52] and sea ice.[53] Models of straight vegetation such as wheat in terms of cylinders have had some success.[2] Corner-reflector effects have been used to describe strong returns from buildings at nonnormal incidence angles.[54] Other specialized models have been used for particular purposes.

Regardless of the model used and the approach applied to determining the field strength, theoretical work only guides understanding. Actual earth surfaces are too complex to be described adequately in any of the models, and the effects of signals that penetrate the ground and are scattered therein are too little known to permit its evaluation.

12.4 FADING OF GROUND ECHOES

The amplitude of ground echoes received by radars on moving vehicles fluctuates widely because of variations in phase shift for return from different parts of the illuminated area. In fact, even fixed radars frequently observe fluctuations in ground echoes because of motions of vegetation, automobiles, etc.

Regardless of the model used to describe a ground surface, signals are, in fact, returned from different positions not on a plane. As a radar moves past a patch of ground while illuminating it, the look angle changes, and this changes the relative distances to different parts of the surface; the result is that relative phase shift is changed. This is the same kind of relative-phase-shift change with direction that is present for an antenna array and results in the antenna pattern. For ground echo the distance is doubled; so the pattern of an echoing patch of length L has lobes of width $\lambda/2L$. This compares with λ/L for an antenna of the same cross-range length. Because the excitation of the *elements* of the scattering array is random, the scattering pattern in space also is random.

This fading phenomenon is usually described in terms of the doppler shift of the signal. Since different parts of the target are at slightly different angles, the signals from them experience slightly different doppler shifts. The doppler shift, of course, is simply the rate of change of phase due to motion. Thus the total rate of change of phase for a given target is

$$\omega = \omega_c + \omega_{di} = \frac{d\phi i}{dt} = \frac{d}{dt}(\omega_c t - 2kR_i) \qquad (12.9)$$

where ω_c = carrier angular frequency
ω_{di} = doppler angular frequency for ith target
ϕ_i = phase for ith target
R_i = range from radar to ith target

The doppler shift can be expressed in terms of the velocity vector **v** as

$$\omega_{di} = -2k\frac{dR_i}{dt} = -2k\,\mathbf{v}\cdot\frac{\mathbf{R}_i}{R_i} = -2kv\cos(\mathbf{v},\mathbf{R}_i) \qquad (12.10)$$

Hence the total field is given by

$$E = \sum_i A_i \exp\left\{j\left[\omega_c t - \int_0^t 2k\,\mathbf{v}\cdot\frac{\mathbf{R}_i}{R_i}dt - 2kR_{i0}\right]\right\} \qquad (12.11)$$

where A_i is the field amplitude of the ith scatterer and R_{i0} is the range at time zero.

The only reason the scalar product is different for different scatterers is the different angle between the velocity vector and the direction to the scatterer. This results in a different doppler frequency for each scatterer. If we assume the locations to be random, as most theories do, the received signal is the same as one coming from a set of oscillators with random phases and unrelated frequencies. This same model of a group of randomly phased, different-frequency oscillators is used to describe noise; *thus the statistics of the fading signal and the statistics of random noise are the same.*

This means that the envelope of the received signal is a random variable with its amplitude described by a Rayleigh distribution. Such distributions have been measured for many ground-target echoes.[15] Although the actual distributions vary widely, no better description can be given for relatively homogeneous targets.

When a target is dominated by one large echo (such as a metal roof oriented to give a strong return), the distribution is better described by that for a sine wave in noise. If the large echo is considerably stronger than the mean of the remaining contributors to the return, this approaches a normal distribution about the value for the large echo. In practice, the distribution from large targets may be more complicated than either of the simple models described.

For reference, the two distributions are given:[55]

$$p(v)dv = \frac{v}{\psi_0}e^{-v^2/2\psi_0}dv \qquad\qquad \text{(Rayleigh)}$$

$$p(v)dv = \frac{v}{\psi_0^{1/2}}e^{-(v^2+a^2)/2\psi_0}I_0\left(\frac{av}{\psi_0}\right)dv \qquad \text{(sine wave + noise)}$$

where v = envelope voltage
ψ_0 = mean square voltage
a = sine-wave peak voltage
$I_0(x)$ = Bessel function, first kind, zero order, imaginary argument

Fading-Rate Computations. Doppler frequency calculation is the easiest way to find fading rates. To compute the signal amplitude returned with a particular range of doppler shifts, all signals having such shifts must be summed. This requires knowledge of the contours of constant doppler shift (isodops) on the scattering surface. These contours must be established for each particular geometric arrangement. A simple example is presented here: horizontal motion over a plane earth. This is typical of an aircraft in ordinary cruising flight.

Consider travel in the y direction, with z vertical, and the altitude (fixed) $z = h$. Then

$$\mathbf{v} = \mathbf{1}_y v$$
$$\mathbf{R} = \mathbf{1}_x x + \mathbf{1}_y y - \mathbf{1}_z h$$

where $(\mathbf{1}_x, \mathbf{1}_y, \mathbf{1}_z)$ are unit vectors. Hence

$$v_r = \mathbf{v} \cdot \frac{\mathbf{R}}{R} = \frac{vy}{\sqrt{x^2 + y^2 + h^2}}$$

where v_r is the relative speed. Curves of constant relative speed are also curves of constant doppler shift. The equation of such a curve is

$$x^2 - y^2 \frac{v^2 - v_r^2}{v_r^2} + h^2 = 0$$

This is a hyperbola. The limiting curve for zero relative speed is a straight line perpendicular to the velocity vector. Figure 12.7 shows such a set of constant-doppler-shift contours.

The spectrum of fading can be calculated by a slight rearrangement of the radar equation (12.1). Thus, if $W_r(f_d)$ is the power received between frequencies f_d and $f_d + df_d$, the radar equation becomes

$$W_r(f_d)df_d = \frac{1}{(4\pi)^2} \int \frac{P_t G_t A_r \sigma^0 \, dA}{R^4} = \frac{df_d}{(4\pi)^2} \int \frac{P_t G_t A_r \sigma^0}{R^4} \left(-\frac{dA}{df_d} \right) \qquad (12.12)$$

<div style="text-align:center">Illuminated area
between f_d and $f_d + df_d$</div>

This is an integral in which the area element between f_d and $f_d + df_d$ is expressed in terms of coordinates along and normal to the isodops. Such coordinates must be established for each particular case.

Figure 12.8 shows the geometry for horizontal travel. The coordinate σ is along the isodop, and η is normal to it. We can express Eq. (12.12) in terms of these coordinates as

$$W_r f_d = \frac{d\eta}{df_d}\left[\frac{\lambda^2}{(4\pi)^3} \right] \int_{\text{strip}} \left[\frac{P_t G^2 \sigma^0 d\xi}{R^4} \right] \qquad (12.13)$$

Note that P_t, the transmitted power, is nonzero in the integral only for the time it illuminates the ground. In pulse radars, only that part of the ground area providing signals back to the radar at a particular time can be considered to have finite P_t, and so the range of frequencies that can be present is limited by the pulse, as well as by the antennas and the maximum velocity.

Another example is shown in Fig. 12.9. This is the small illuminated area for a narrow-beam, short-pulse system. Here we can make linear approximations with-
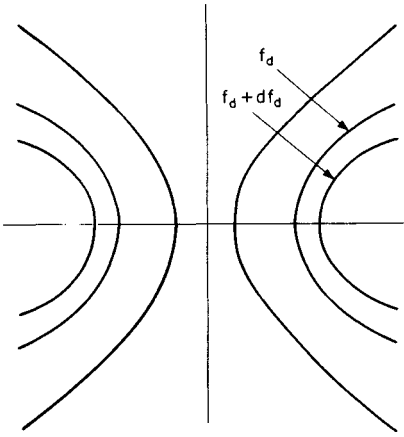
FIG. 12.7   Contours of constant doppler frequency shift on a plane earth due to horizontal motion.
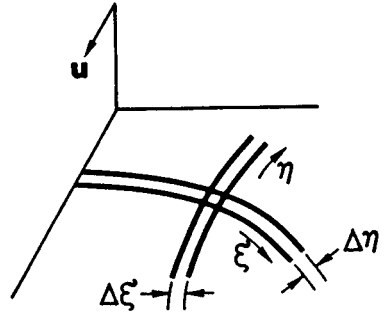
FIG. 12.8   Geometry of complex fading calculations. (*From Ulaby, Moore, and Fung.*[21])
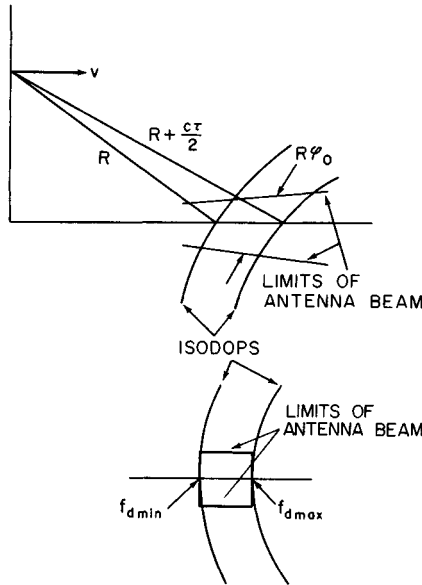
FIG. 12.9   Geometry of doppler-shift calculations for an airborne search radar.

out too much error. A pulse of length τ is transmitted from an antenna of beamwidth ϕ_0. For the simple illustration given here, we assume the pulse to be transmitted directly ahead of the horizontally moving vehicle. We may simplify the problem by assuming a rectangular illuminated area, $R\phi_0$ by $c\tau/(2\sin\theta)$. Furthermore, the curvature of the isodops may be neglected, and so the doppler fre-

quency is assumed to be the same for all maximum-range points and the same for all minimum-range points. With this assumption,

$$f_{d\text{max}} = \frac{2v}{\lambda} \sin \theta_{\text{max}}$$
$$f_{d\text{min}} = \frac{2v}{\lambda} \sin \theta_{\text{min}}$$

Thus the total width of the doppler spectrum is

$$\Delta f_d = \frac{2v}{\lambda}(\sin \theta_{\text{max}} - \sin \theta_{\text{min}})$$

For short pulses and angles away from vertical, this is

$$\Delta f_d \approx \frac{2v}{\lambda}\Delta\theta \cos \theta$$

In terms of pulse length, it becomes

$$\Delta f_d = \frac{vc\tau}{2h\lambda} \frac{\cos^3 \theta}{\sin \theta} \tag{12.14}$$

If the angular difference across the illuminated rectangle is small enough so that σ^0 is essentially constant, the doppler spectrum is a rectangle from f_{min} to f_{max}.

In practice, antenna beams are not rectangular. The result is that the doppler spectrum for a side-looking radar like that of the example is not rectangular but rather has the shape of the antenna along-track pattern. Thus, if the antenna pattern in the along-track direction is $G = G(\beta)$, with β the angle off the beam center, we can express β in terms of the doppler frequency f_d as

$$\beta = f_d\lambda/2v$$

and the spectrum is

$$W(f_d) = \frac{\lambda^3 P_t\sigma^0 r_x}{2(4\pi)^3 R^3}G^2\left[\frac{\lambda f_d}{2v}\right]$$

where r_x is the horizontal resolution in the range direction. Of course, the half-power beamwidth may be used as an approximation, resulting in the bandwidth given by Eq. (12.13).

Effect of Detection.   The effect of detecting narrowband noise has been treated extensively in the literature. Here it is necessary only to show the postdetection spectrum of the preceding example and to consider the number of independently fading samples per second. Figure 12.10 shows the spectrum before and after detection. If square-law detection is assumed, the post-detection spectrum is the self-convolution of the predetection spectrum. Only the part that passes the low-pass filters in a detector is shown in the figure. The rectangular *RF spectrum* has become a triangular *video spectrum*.
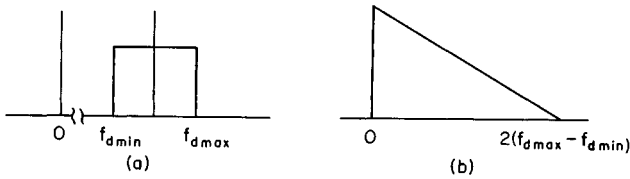
FIG. 12.10 Spectrum of fading from a homogeneous small patch (*a*) before and (*b*) after detection.

This spectrum describes the fading of the detector output for a CW radar. For a pulse radar, the spectrum is sampled by the PRF (pulse repetition frequency). If the PRF is high enough so that the entire spectrum can be reproduced (the PRF is higher than the Nyquist frequency, $2\Delta f_d$), the diagram indicated is that of the spectrum of the samples of a received pulse at a given range. Figure 12.11 shows a series of actual pulses, followed by a series of samples at range R_1. The spectrum of Fig. 12.10 is the spectrum of the envelope of samples at R_1 (after low-pass filtering). The spectrum of fading at a different range (or vertical angle) is different, in accord with Eq. (12.13).

For many purposes, the number of *independent* samples is important, since these may be treated by using the elementary statistics of uncorrelated samples. For continuous integration, the effective number of independent samples is[55]

$$ N = \frac{\overline{P_e^2}T}{2\int_0^T \left[1 - \frac{x}{T}\right] R_{sf}(x)dx} \tag{12.15} $$

where $\overline{P}_e$ is the mean envelope power, T is the integration (averaging) time, and $R_{sf}(t)$ is the autocovariance function for the detected voltage. For many practical purposes, if N is large, it may be approximated by

$$ N \approx BT \tag{12.16} $$

where B is the effective IF bandwidth. For the effect of short integration time, see Ref. 56.

Fading samples can, of course, also be independent because motion of the vehicle causes the beam to illuminate a different patch of ground. Thus, in a particular case, the independent-sample rate may be determined either by the motion of the illuminated patch over the ground or by the doppler effect, or by some combination of the two.

The number of independent samples determines the way in which the Rayleigh or other distributions may be applied. Thus, if 100 pulses give only 10 independent samples, the variance of the mean obtained by integrating these pulses is much greater than would be true if all 100 pulses were independent.

Doppler-based systems, such as doppler navigators and synthetic aperture radar systems, depend on the predetection spectrum for their operation, since they are coherent and do not use ordinary detection.
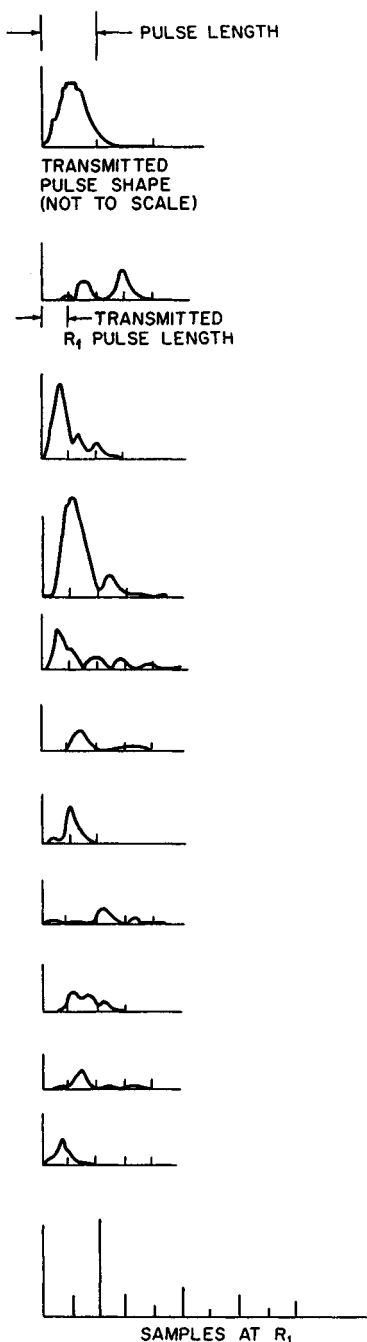
FIG. 12.11 Fading for successive pulses of a radar with ground target.

Moving-Target Surfaces. Sometimes clutter has internal motion. This can occur when fixed radars are used to observe movement of the sea and the land. On land, clutter motion is usually due to moving vegetation, although moving animals and machines create similar effects. The radar return from an assembly of scatterers like those of Fig. 12.8 can change because of motion of the individual scatterers just as it changes because of motion of the radar. Thus, if each scatterer is a tree, the waving of the trees as the wind blows causes relative phase shifts between the separate scatterers; the result is fading. For a fixed radar, this may be the only fading observed, except for very slow fading due to changes in refraction. For a moving radar, this motion of the target changes the relative velocities between target element and radar, so that the spectrum is different from that for a fixed surface. The width of the spectrum due to vehicle motion determines the ability of the radar to detect this target motion.

12.5 MEASUREMENT TECHNIQUES FOR GROUND RETURN

Special-purpose instrumentation radars and modified standard radars may be used to determine the ground return. Since the ground return is almost invariably due to scattering, these systems are termed *scatterometers*. Such systems may use CW signals with or without doppler processing, but they may also use both pulse and FM techniques. Scatterometers capable of measuring response over a wide range of frequencies are called *spectrometers*.[57] Various antenna patterns from pencil beams to fan beams may be used.

CW and FM-CW Systems. The simplest scatterometer uses a stationary CW radar. Such systems are not very flexible, but they are discussed here in some detail to illustrate calibration techniques that also apply to the more complex systems.
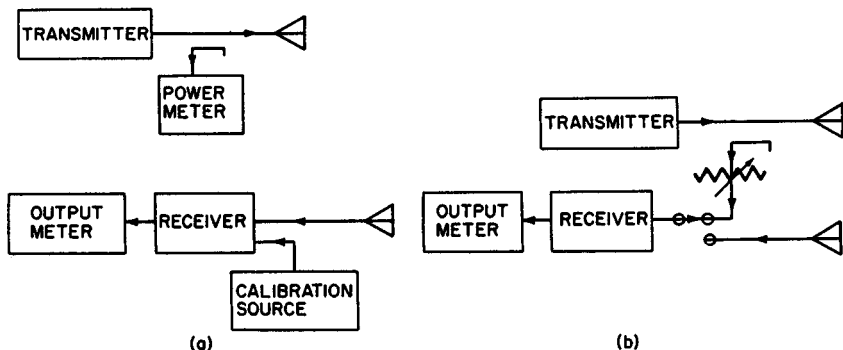
FIG. 12.12 CW-scatterometer-system block diagram. (*a*) Separate transmitter and receiver calibration. (*b*) Calibration of the ratio of received to transmitted power.

The CW scatterometer is shown in block form in Fig. 12.12. To evaluate σ^0 the ratio of transmitted to received power is required. The system in Fig. 12.12*a* measures transmitter power and receiver sensitivity separately. The transmitter feeds an antenna through a directional coupler so that a portion of the energy may be fed to a power meter. The receiver operates from a separate antenna (electrically isolated). The output of the receiver is detected, averaged, and displayed on a meter, oscilloscope, or other display or recorder. Its sensitivity must be checked by use of a calibration source. The calibrated signal may be fed through the receiver at a time when the transmitter is off. Figure 12.12*b* shows a similar arrangement in which the signal from the transmitter is attenuated a known amount and used to check the receiver. By comparing the output from the attenuated transmitter signal with that received from the ground, the scattering cross section may be determined without actually knowing the transmitted power and the receiver gain.

The calibrations shown in Fig. 12.12 are incomplete without knowledge of the antenna patterns and absolute gains. Since accurate gain measurements are difficult, absolute calibrations may be made by comparing received signals (with proper relative calibration) from the target being measured and from a *standard target*. Standard targets may be metal spheres, Luneburg-lens reflectors, metal plates, corner reflectors, or active radar calibrators (ARCs—actually repeaters).[58] Of the passive calibrators, the Luneburg-lens reflector is best, since it has a large cross section for its volume and has a very wide pattern so that alignment is not critical. Luneburg-lens reflectors are used for making strong radar targets of small vessels, and they may be obtained from companies that supply that market. For discussion of the relative merits of different passive calibration targets, see Ulaby, Moore, and Fung.[59]

The ideal receiver would respond linearly to its input, so that a single calibration at one input level would suffice for all levels. The usual receiver, however, has some nonlinearities due to detector properties and to saturation of its amplifiers by large signals. Figure 12.13 shows a typical input-output curve for a receiver. Two equal increments in input signal (Δ^i), as shown, produce different increments in output because of the nonlinearity of this curve. For this reason, receiver calibration must be performed over a range of input levels, and the nonlinearities must be compensated for in the data processing.

CW scatterometers depend on antenna beams to discriminate different angles

OUTPUT
Δ_2^0
Δ_1^0

$|\ \Delta_1^i\ |\ \Delta_2^i\ |$ INPUT
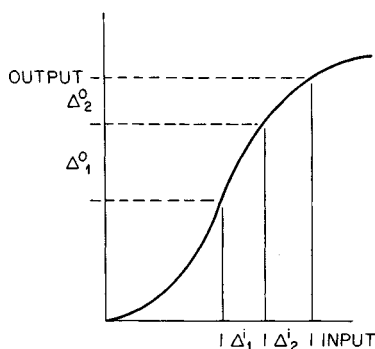
FIG. 12.13 Typical receiver input-output curve. Illustrated is the effect of non-linearity.

of incidence and different targets. Usually assumptions are made that the antenna pattern has constant gain within the actual 3 dB points and zero gain outside, but this clearly is not an accurate description. If large targets appear in the locations illuminated by the side of the main lobe or the minor lobes, their signals may contribute so much to the return that it is significantly changed. Since this changed signal is charged to the direction of the major lobe by the data reduction process, the resulting value for σ^0 is in error. Responses at vertical incidence frequently cause trouble, for vertical-incidence signals are usually fairly strong. Thus the antenna pattern must be accurately known and taken into account in the data analysis. A pattern with strong minor lobes may be simply inadmissible.

The scattering coefficient is determined by applying

$$P_r = \frac{P_t \lambda^2}{(4\pi)^3} \int\limits_{\substack{\text{Illuminated} \\ \text{area}}} \frac{G_t^2 \sigma^0 dA}{R^4}$$

The integration is over whatever area is illuminated significantly, including the regions hit by the minor lobes. The usual assumption is that σ^0 is constant over the illuminated area, so that

$$P_r = \frac{P_t \lambda^2 \sigma^0}{(4\pi)^3} \int\limits_{\substack{\text{Illuminated} \\ \text{area}}} \frac{G_t^2 dA}{R^4} \qquad (12.17)$$

This assumption would be true only if the antenna confined the radiated energy to a very small spread of angles and to a fairly homogeneous region. The resulting expression is

$$\sigma^0 = \frac{(4\pi)^3 P_r}{P_t \lambda^2 \int\limits_{\substack{\text{Illuminated} \\ \text{area}}} (G_t^2/R^4) dA} \qquad (12.18)$$

Note that only the ratio of transmitted to received power is required, and so the technique of Fig. 12.12b is justified. Sometimes R, G_t, or both are assumed constant over the illuminated area, but such an approximation to Eq. (12.18) should be attempted only after checking its validity for a particular problem.

If the result of applying the technique of Eq. (12.18) to a set of measurements indicates that σ^0 probably did vary across the significantly illuminated area, this variation may be used as a first approximation to determine a function $f(\theta)$ describing the θ variation of σ^0, and a next-order approximation then becomes

$$\sigma^0 = \frac{(4\pi)^3 P_r}{P_t \lambda^2 \int\limits_{\substack{\text{Illuminated} \\ \text{area}}} [f(\theta)G_t^2/R^4]\, dA} \qquad (12.19)$$

Proper scattering measurements demand an accurate and complete measurement of antenna gain G_t. This can be a very time-consuming and expensive process, particularly when the antenna is mounted on an aircraft or other metallic object. Nevertheless, complete patterns are a must for good scatter measurements.

Range-Measuring Systems. Radar's ability to separate returns from different ranges can be used advantageously along with directive antenna beams to simplify the scattering measurements. Most ranging scatterometers use either pulse modulation or FM, although more exotic modulations could also be used. The discussion here treats pulse systems, but since all other range-measuring systems can be reduced to equivalent pulse systems most results are general.

Figure 12.14 shows the way in which pulse measurement of range is used. Figure 12.14a shows a circular pencil beam. At angles near grazing, the illuminated patch set by the circular antenna pattern becomes rather long (the patch is an ellipse), and use of the pulse length to confine illumination to a part of the patch is helpful. Many systems that use beamwidth to set the measured area near vertical use range resolution for angles beyond, say, 60°.

Figure 12.14b shows an antenna pattern that takes better advantage of the possibilities of range measurement. A fan beam is used to illuminate a narrow strip along the ground, and the range resolution permits separating the returns from different angles by the time they return. This technique is especially effective at angles away from the vertical, for the resolution near the vertical is much poorer than near grazing. The simple approach assumes a constant gain across the beam and zero elsewhere:

$$G_t = 0 \qquad \phi_a < 2\,f_0/2 \qquad \text{or} \qquad \phi_a > f_0/2$$

$$G_t = G_o \qquad -\phi_0/2 < f_a < f_0/2$$

CIRCULAR BEAM
(a)
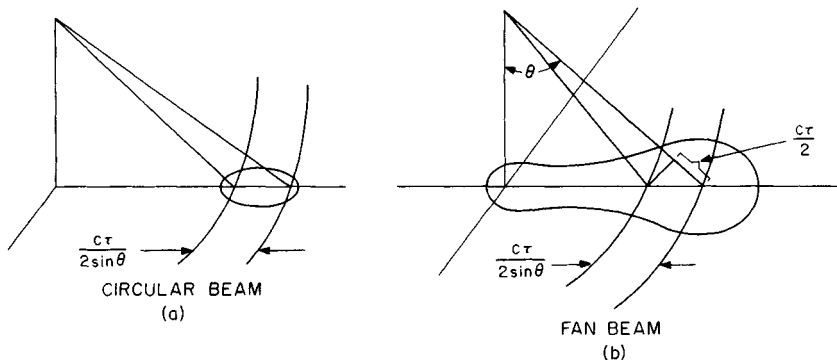
FAN BEAM
(b)

FIG. 12.14 Range resolution applied to scatterometry. (*a*) Improving one dimension of a circular-beam illumination pattern. (*b*) Use with a fan beam.

where ϕ_0 = beamwidth
        ϕ_a = transverse angle with respect to antenna axis

With the further assumption that σ^0 is essentially constant and that the difference in range across a resolution element is negligible, the expression for σ^0 becomes

$$\sigma^0 = \frac{P_r(4\pi)^3 R^3 \sin \theta}{P_t \ \lambda^2 G_0 \phi_0 \, r_R}$$

where r_R is the short-range resolution.

Janza has reported details of calibration problems with a range-measuring pulsed radar scatterometer.[60,61]

CW-Doppler Scatterometers. A convenient way to measure the scattering coefficient at many angles simultaneously is with a CW system in which the relative velocities corresponding to different angles are separated by separating their doppler frequencies. The use of a fan beam with such a system permits the simultaneous measurement of scattering coefficients at points ahead of and behind the aircraft carrying the radar. Figure 12.15 shows this. The pattern of the antenna illumination on the ground is shown intersected by two isodops (lines of constant doppler frequency), with the width of the spectrum between them shown on the diagram. The distance between them can be seen to be

$$\Delta\rho = R(\sin \theta_2 - \sin \theta_1)$$

and

$$\Delta f_d = \frac{2v}{\lambda}(\sin \theta_2 - \sin \theta_1)$$

Thus the width of the element on the ground is related to the doppler frequency bandwidth by

$$\Delta\rho = \frac{R\lambda}{2v}(\Delta f_d)$$

Where this technique is applied to the radar equation and the following are assumed:

1. σ^0 constant in the illuminated area

FAN BEAM

θ_1
θ_2

$\Delta f_d = \frac{2v}{\lambda} (\sin \theta_2 - \sin \theta_1)$
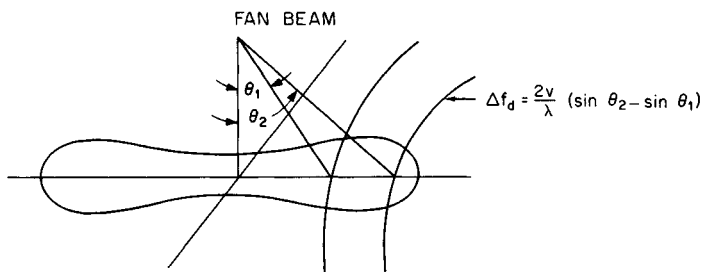
FIG. 12.15   Resolution in a fan-beam CW-doppler scatterometer.

2. Antenna gain constant over its beamwidth and zero elsewhere

3. Range variation across the small illuminated area negligible

$$P_r = \frac{P_t \lambda^2}{(4\pi)^3} \int \frac{G_t^2 \sigma^0 \, dA}{R^4} = \frac{P_t \lambda^4 \sigma^0 G_0^2 \psi_0 \Delta f_d}{2vR^2} \tag{12.21}$$

and so

$$\sigma^0 = \frac{P_r}{P_t} \frac{2vR^2}{\lambda^4 G_0^2 \psi_0 \Delta f_d} \tag{12.22}$$

Doppler scatterometers need not use fore-and-aft beams. The Seasat[62] and N-SCATT[63] spaceborne doppler scatterometers were designed with beams pointed (squinted) ahead and behind the normal to the ground track.

Independent Samples Required for Measurement Accuracy. The Rayleigh distribution describes the fading signal fairly well. If we assume a Rayleigh distribution of fading, the number of independent samples required for a given accuracy is shown in Fig. 12.16. The *range* defined in this figure is the range of mean values lying between 5 and 95 percent points on the distribution. This accuracy range is independent of any accuracy problems associated with calibration and knowledge of the antenna pattern.

The precision of the measurement depends upon the number of independent samples, not on the total number of samples. The number of independent samples can be found from Eq. (12.15) or Eq. (12.16) after suitable analysis. This analysis assumes that only doppler fading contributes to independence but motion from one cell to another also adds independent samples. Thus, the total number of such samples is approximately the product of the number calculated from Eq. (12.13) and the number of ground cells averaged. Figure 12.17 shows some examples of the effect of the angle of incidence on the number of independent samples for a horizontally traveling scatterometer with a forward-pointed beam.

FIG. 12.16   Accuracy of averages for fading signals.

Study of the results obtained in this type of analysis indicates that, in regions where the scattering coefficient does not change rapidly with angle, the widest possible angular width (obtained by a longer pulse or a wider filter for a CW-doppler system) results in the maximum number of independent samples for a given distance traveled along the ground.

Near-Vertical Problem   Most published radar return data purporting to include vertical incidence gives vertical-incidence scattering coefficients that are too small. This is a consequence of a fundamental problem in measuring near the vertical with a finite beamwidth or pulse length. Near-vertical radar returns from most targets drop off rapidly as the angle with the vertical is increased. Thus the

FIG. 12.17 Examples of the variation with angle of incidence of the number of independent samples for a scatterometer.

measuring beamwidth or pulse width usually encompasses signals from regions having values for σ^0 many decibels apart. Since the scattering coefficient varies much more rapidly near the vertical than at angles beyond 10 or 20° from the vertical, the problem is much more severe at the vertical. Furthermore, the problem is complicated at the vertical by the fact that the angular scale terminates there, so that a beam centered at the vertical illuminates weaker targets (σ^0) on both sides of its pattern, whereas a beam away from the vertical illuminates stronger signals on one side and weaker signals on the other.

Figure 12.18 shows what happens for a steeply descending curve of σ^0 versus θ. The radar return integral from Eq. (12.1) is a convolution integral; the figure shows the convolution of the beam pattern with the σ^0 curve. Clearly the average at the vertical is lower than it should be to indicate properly the variation of σ^0 near the vertical.

Figure 12.19 shows an example[64] based on the theoretical scattering coefficient for the sea derived from the spectra reported by the Stereo Wave Observation Project.[65] The effect of different beamwidths is clearly shown.

With a pulse or other range-measuring system, reported values are always in error because, as indicated above, it is almost impossible to resolve a narrow range of angles near the vertical.

Ground and Helicopter Scatterometers and Spectrometers. Many ground scattering measurements have been made with systems mounted on boom trucks and helicopters. Most of these are FM-CW systems[66,67] that use wide bandwidth to obtain extra independent samples rather than for fine resolution. Some use very wide bandwidth to obtain fine range resolution to locate sources of scattering.[68] Most have multiple-polarization capability, and some are capable of polarimetry because the phase of two received signals with orthogonal polarization can be measured.

SCATTERING COEFFICIENT

ANTENNA PATTERN

PRODUCT $G^2\sigma^0$

INTEGRAND OF POWER RETURNED
INTEGRAL

- - - RETURN THAT WOULD BE
MEASURED BY VERY NARROW
BEAM

——— RETURN MEASURED BY
ACTUAL BEAM
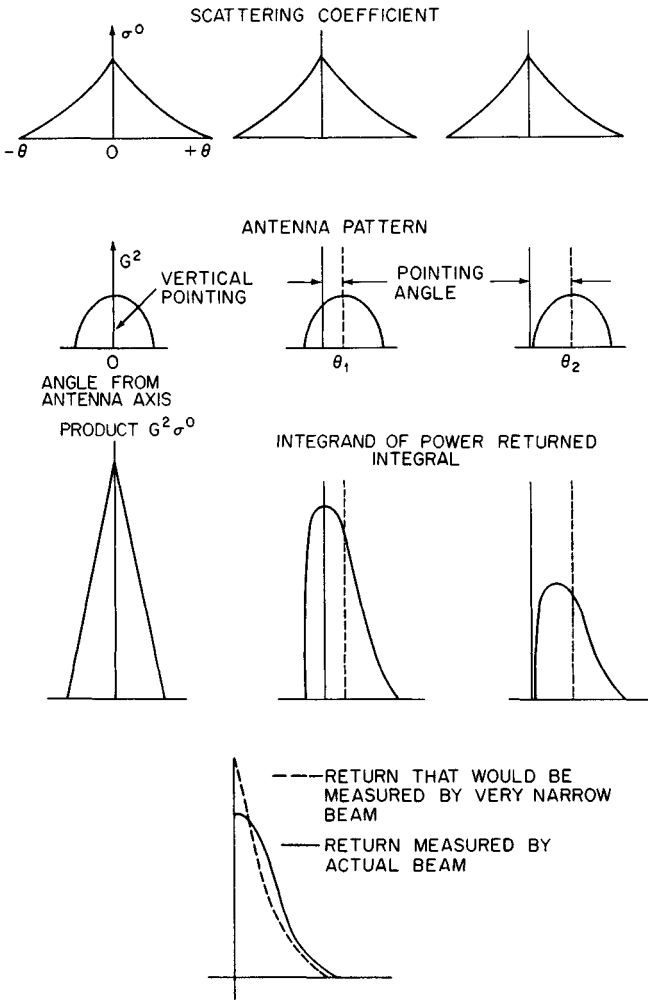
FIG. 12.18   How finite beamwidth causes a near-vertical error in measuring the scattering coefficient.

The basic elements of an FM-CW scatterometer are shown in Fig. 12.20. The swept oscillator must produce a linear sweep; this is easy with yttrium-iron-garnet (YIG)-tuned oscillators but requires linearizing circuits if tuning uses a varactor. If dual antennas are used (as shown), the overlap of the beams must be considered.[69] Single-antenna systems are sometimes used, with a circulator isolating transmitter and receiver; their performance is somewhat poorer than that of dual-antenna systems because of internal reflections and leakage through the circulator.

Two versions of the control and data-handling part of an FM-CW scatterometer are shown in Figs. 12.21 and 12.22. Figure 12.21 shows the common range-tracking scatterometer. This system can be used to measure
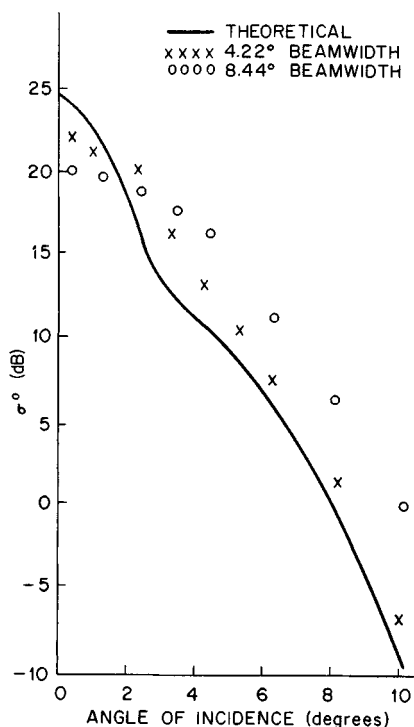
FIG. 12.19 Effect of antenna beamwidth on the measured scattering coefficient as a function of angle of incidence.

surface scattering coefficients when the distance between radar and target is changing, such as with a fixed radar observing the sea or a radar on a helicopter. If the scatterometer is mounted on a boom truck, the range tracker is not needed; but it is convenient because the range changes as the angle of incidence is changed. Figure 12.22 shows the kind of system that may be used to measure scattering from within a volume. By determining the spectrum of the return, the user can establish the scattering from different ranges. This system has been used in determining the sources of scatter in vegetation[25-27] and snow.

Ultrasonic waves in water can be used to simulate electromagnetic waves in air.[70-72] Because of the difference in velocity of propagation an acoustic frequency of 1 MHz corresponds with a wavelength of 1.5 mm. Such a wavelength is of a convenient size for many modeling measurements, and, of course, equipment in the 1-MHz region is in many ways easier to operate than equipment in the microwave region; certainly it is much easier to operate and less expensive than microwave equipment operating at a 1.5-mm wavelength.

Acoustic plane waves and electro-
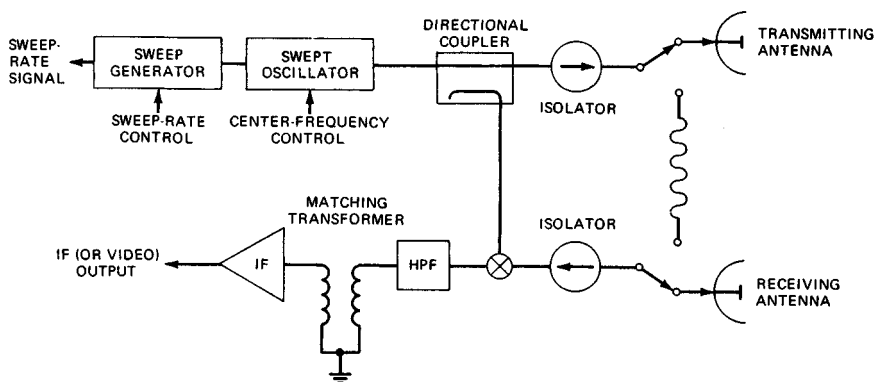
FIG. 12.20 Basic block diagram of an FM-CW scatterometer RF section.
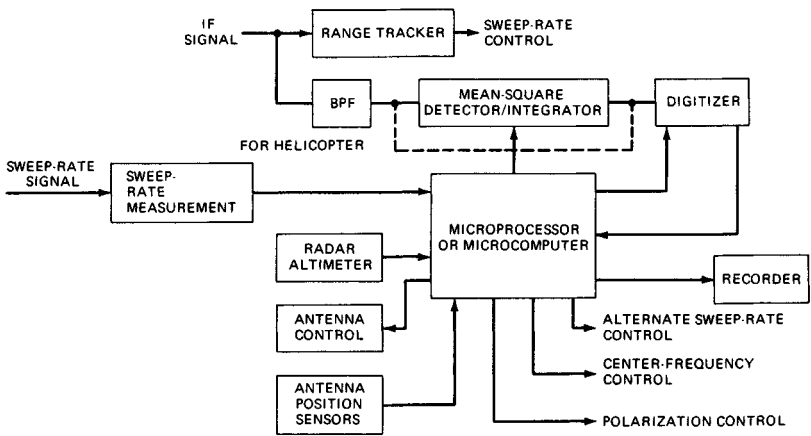
FIG. 12.21 Basic block diagram of an FM-CW range-tracking scatterometer: control and data-handling system.
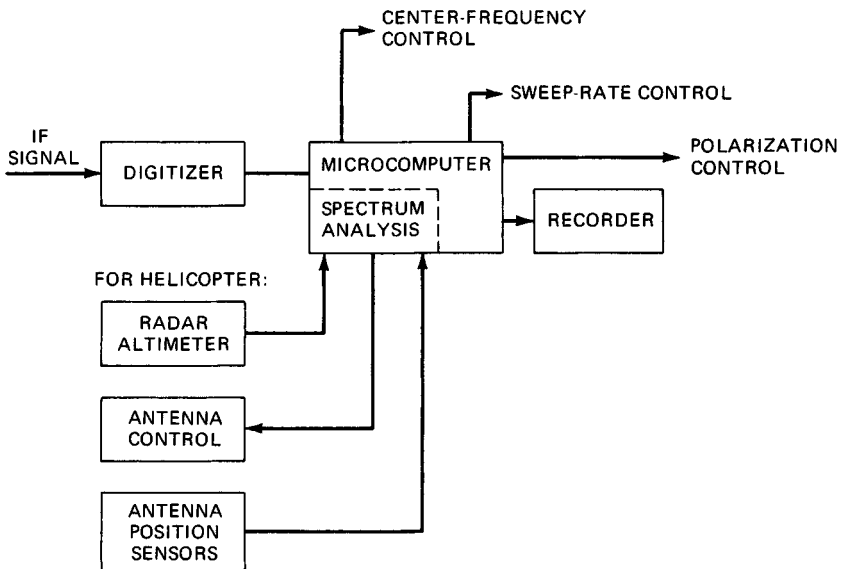
FIG. 12.22 Basic block diagram of an FM-CW range-discriminating scatterometer: control and data-handling system.

magnetic plane waves satisfy the same boundary conditions. When the scattering surfaces are not plane and when angles of incidence are rather oblique, the analogy between acoustic and electromagnetic waves is less valid.

Scattering Coefficients from Images. Radar images produced by real or synthetic aperture radars can be used for scattering coefficient measurement. Unfortunately, most such systems are uncalibrated; so the results are somewhat dubious, even on a relative basis when images are produced on different days. Relative calibration has been introduced into some systems.[12,18,20,73–75] Absolute calibration, which also serves as relative calibration in some cases, can be achieved by using strong reference targets, with the ARC repeaters especially suitable.[76] Another approach that has been used is to measure scattering from reference areas with a ground-based or helicopter system that is well calibrated and to compare the images to these measured values.[73,77]

Bistatic Measurements. Measurements of ground return when the receiver and transmitter are separated are comparatively rare. These measurements are very difficult to make from aircraft because it is necessary that both transmitter and receiver antennas look at the same ground point at the same time and that the signal be correlated with known antenna look angles. Furthermore, it is difficult to know the polarization, and the exact size and shape of the common area illuminated by the antenna beams are sometimes difficult to determine. For this reason, few bistatic measurements from aircraft have been reported in the literature.[78]

Laboratory bistatic measurements have been made by both the Waterways Experiment Station[24] and Ohio State University[2,4] groups using electromagnetic waves and by the University of Kansas[71] group using acoustic waves. Bistatic measurements of laser radiation have been made at Bell Telephone Laboratories,[79] and C-band measurements of buildings at the University of Kansas.[80]

Because of the antenna orientation problems, most electromagnetic bistatic measurements are only for forward scatter; that is, the receiver, transmitter, and target all lie in the same vertical plane. The acoustic measurements and optical measurements are easier to make over a wide range of angles and have been made with a fixed incidence angle and scatter directions covering the entire hemisphere.

Bistatic measurements call for additional calibration complications when made outside the laboratory because an absolute reference for both transmitter power and receiver sensitivity must be used. In the laboratory, however, it is possible to use techniques similar to those for monostatic measurements.

12.6 GENERAL MODELS FOR SCATTERING COEFFICIENT (CLUTTER MODELS)

Scatter measurements made during the 1970s allowed generation of models for average backscatter from large areas. In particular, these included measurements with the Skylab radiometer-scatterometer RADSCAT[83] and with truck-mounted microwave active spectrometers (MAS)[81] by the University of Kansas. Two different models were developed based on the same data, one a linear model and one a more complicated formulation. Here we present only the linear model.

These models are for *averages*, and the models do not include variations about the average. However, analysis of Shuttle Imaging Radar-B (SIR-B) data permits some estimates to be made of the variability to be expected for different sizes of illuminated footprint.

The general characteristics of radar backscatter over the range of angles of incidence have been known for decades. Figure 12.23 shows these. For like-polarized waves, one can break scatter into three angular regimes: near-vertical (the *quasi-specular region*), intermediate angles from 15 to about 80° (the *plateau region*), and near-grazing (the *shadow region*). Cross-polarized scatter does not have separate quasi-specular and plateau regions (the plateau extends to vertical), and too little is known to establish whether a shadow region exists.
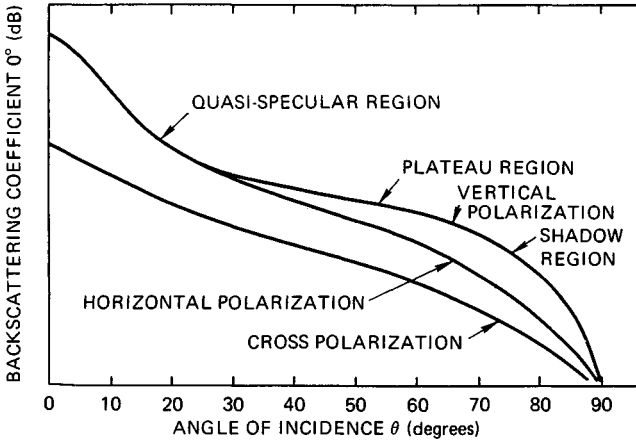
FIG. 12.23 General characteristics of scattering coefficient variation with angle of incidence. (*From Ulaby, Moore, and Fung.*[21])

For nearly every type of terrain, the measured data fits closely to the form

$$\sigma^0 = A_i e^{-\theta/\theta_i} \qquad (12.23a)$$

or
$$\sigma^0_{dB} = 10 \log A_i - 4.3434(\theta/\theta_i) \qquad (12.23b)$$

where A_i and θ_i differ for the near-vertical and midrange regions. Figure 12.24 shows an example of this variation. No theory gives exactly this result, but nearly all measurements fit such a model closely, and the model approximates most theoretical curves well over the relevant regions. This simple result means that simple clutter models may be developed and used although more complex models may be necessary for some remote-sensing applications.

The basis for the linear model[82] is a combination of the Skylab results over North America[83] and those from Kansas cropland measurements over three complete seasons with the microwave active spectrometer (MAS).[84] The 13.9-GHz Skylab RADSCAT had a ground footprint of from a 10-km circle at vertical to an ellipse of 20 by 30 km at 50°. The MAS had footprints at 50° ranging from 5.5 by 8.5 m at 1.1 GHz to 1.4 by 2.1 m at 17 GHz, but millions of measurements were averaged for the model. Because the Skylab data was at only one frequency and the responses for the two experiments were essentially the same at that
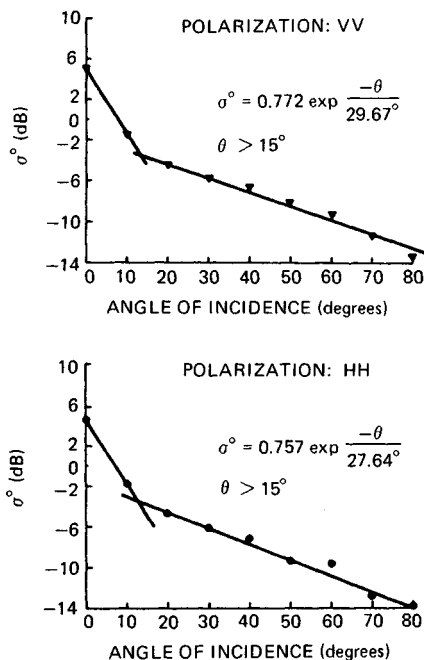
FIG. 12.24 Regression of average of all 1974 and 1975 13.8-GHz cropland data obtained with a microwave active spectrometer. (*From Moore, Soofi, and Purduski.*[82])

frequency, the frequency response shown in the model depends entirely on the MAS measurements.

The summer Skylab observations included deserts, grassland, cropland, and forests, whereas the Kansas measurements were only of cropland. However, early and late in the growing season the cropland was essentially bare, similar to the summer desert except for soil moisture content. During the height of the growing season the crops were dense enough so that scatter was similar to that from forests. Thus, the overall model seems representative of summer conditions averaged over all of North America.

The model takes the form

$$\sigma^0_{dB}(f,\theta) = A + B\theta + Cf + Df\theta \qquad 20° \le \theta \le 70° \qquad (12.24a)$$

where A, B, C, and D take on different values for different polarizations and above and below 6 GHz. The frequency response below 6 GHz is much more rapid than above 6 GHz. Moreover, at frequencies above 6 GHz the frequency response is independent of angle, so that $D = 0$. For lower frequencies, the frequency response is angle-dependent.

For angles less than 20°, only two points were available, 0° and 10°; so separate frequency regressions were run at each of these angles. The model for these angles is

$$\sigma^0_{dB}(f,\theta) = M(\theta) + N(\theta)f \qquad \theta = 0°, \ 10° \tag{12.24b}$$

The frequency responses below 6 GHz differed for the two years; so the models have separate values of the constants for 1975 and 1976. The year 1976 was very dry in Kansas; so the 1975 values are probably more representative, but both are given here. Values of the constants are in Table 12.2. Figure 12.25 shows the clutter model for the midrange of angles as a function of frequency. The figure is only for vertical polarization because results are so similar for vertical and horizontal.

Ulaby developed a different, more complex model from the Kansas vegetation data.[85] This model fits curves rather than straight lines to the measured data. For most purposes the straight-line model is adequate, and it is much easier to use.

A straight-line model for snow-covered grassland similar to that for vegetation depends on a more limited data set.[86,87] The data was for only one season in Colorado when the snow was only about 50 cm deep. This means that the signal probably penetrated to the ground surface at frequencies below about 6 GHz. Nevertheless, the model indicates the kind of results to be expected for this important situation. Table 12.3 gives the resulting constants to use in Eq. (12.24a).

Snow scatter depends strongly on the free-water content of the upper layer of snow; so scatter is much lower from the wet daytime snow (where solar melting has commenced) than for the dry nighttime snow. Hence, different models must be used for day and night; compare the day and night measurements shown in Fig. 12.26. The difference between day and night scatter from snow is even more pronounced at 35 GHz, but the model does not include 35 GHz because no data exists between 17 and 35 GHz.

Although no specific clutter model has been developed for forest, results from the Skylab RADSCAT and Seasat scatterometer show that the Amazon rain forest scatters almost independently of the angle of incidence even near vertical.[88] The mean measured value at 33° was -5.9 ± 0.2 dB at 13.9 GHz. Observations with SIR-B indicated that this lack of angular variation of σ^0 also is present at 1.25 GHz, but lack of calibration prohibited learning the level of scatter at this frequency.

TABLE 12.2 Constants for Linear Scattering Model (Summer)*

Eq.	Polarization	Angular range,°	Frequency range, GHz	Constant A or M, dB	Angle slope B or N, dB	Frequency slope C, dB/GHz	Slope correction D, dB/ (° · GHz)
12.24a	V	20–60	1–6 (1975)	−14.3	−0.16	1.12	0.0051
	V	20–50	1–6 (1976)	−4.0	−0.35	−0.60	0.036
	V	20–70	6–17	−9.5	−0.13	0.32	0.015
	H	20–60	1–6 (1975)	−15.0	−0.21	1.24	0.040
	H	20–50	1–6 (1976)	−1.4	−0.36	−1.03	
	H	20–70	6–17	−9.1	−0.12	0.25	
12.24b	V and H	0	1–6 (1975)	7.6	. . .	−1.03	
	V and H	0	1–6 (1976)	6.4	. . .	−0.73	
	V and H	0	6–17	0.9	. . .	0.10	
	V and H	10	1–6 (1975)	−9.1	. . .	0.51	
	V and H	10	1–6 (1976)	−3.6	. . .	−0.41	
	V and H	10	6–17	−6.5	. . .	0.07	

*After Moore, Soofi, and Purduski.[82]
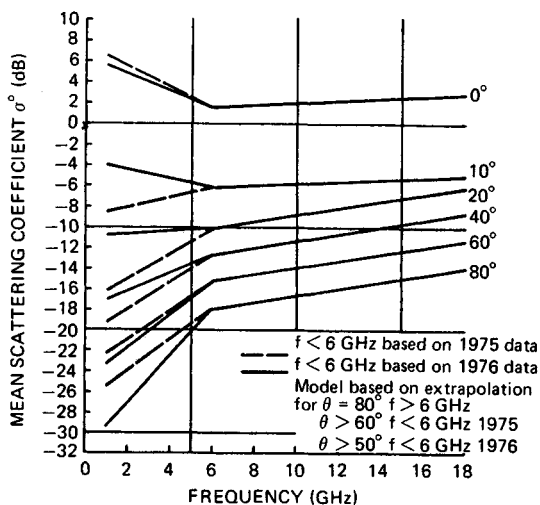
FIG. 12.25 General land-scattering-clutter model (vertical polarization). Horizontal polarization is very similar. (*From Moore, Soofi, and Purduski.*[82])

TABLE 12.3 Regression Results for Ground-Based Measurements of Snow-Covered Ground*

Time of day	Polarization	Frequency range, GHz	Constant A, dB	Angle slope B, dB/°	Frequency slope C, dB/ GHz	Slope correction D, dB/ (° · GHz)
Day	V	1–8	−10.0	−0.29	0.052	0.022
Day	V	13–17	0.02	−0.37	−0.50	0.021
Day	H	1–8	−11.9	−0.25	0.55	0.012
Day	H	13–17	−6.6	−0.31	0.0011	0.013
Night	V	1–8	−10.0	−0.33	−0.32	0.033
Night	V	13–17	−10.9	−0.13	0.70	0.00050
Night	H	1–8	−10.5	−0.30	0.20	0.027
Night	H	13–17	−16.9	−0.024	1.036	−0.0069

*After Moore, Soofi, and Purduski.[82]
NOTE: θ = 20 to 70°. Values of coefficients in this table also are considered those of the model.

The models described above are based on averages over very large areas. For this situation the variability from place to place is small, particularly in the midrange of angles. Figure 12.27 shows the mean and upper and lower decile values measured by the Skylab RADSCAT over North America. The larger variation near vertical apparently results from the effect of nearly specular reflection from water bodies. When the footprint is smaller, more variability occurs. This is shown in Fig. 12.28 from a study of the variation of scatter observed by SIR-B
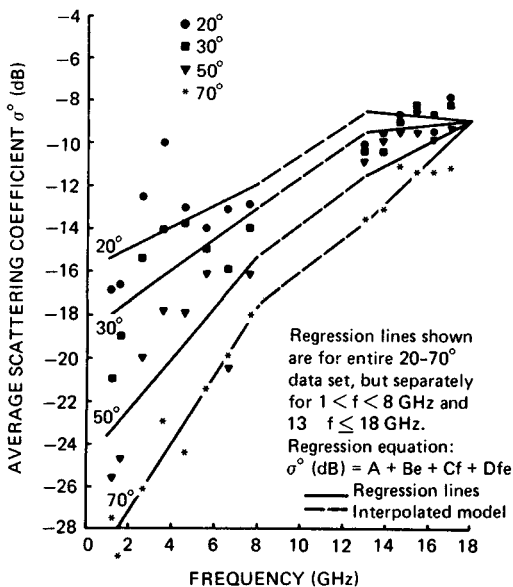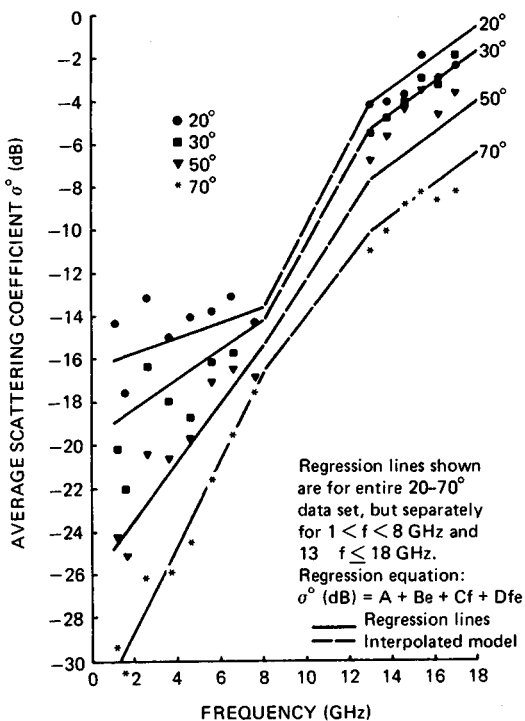
FIG. 12.26 Regressions for vertical-polarization clutter model for snow: (a) day and (b) night. Note the large differences. Horizontal polarization is similar. (*From Moore, Soofi, and Purduski.*[82])
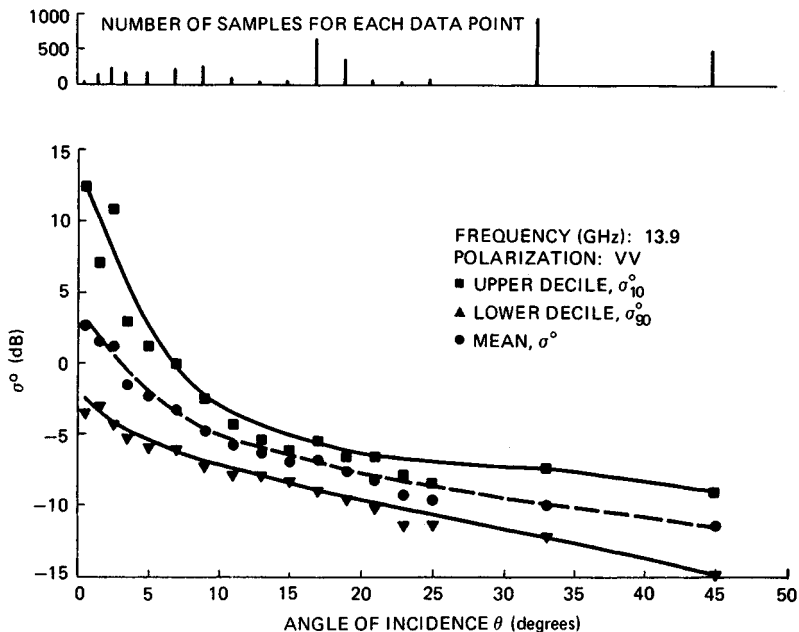
FIG. 12.27 Angular patterns of the mean, upper decile, and lower decile of Skylab scatterometer observations over North America during the summer season. (*From Moore et al., University of Kansas Remote Sensing Laboratory Technical Report 243-12, 1975.*)
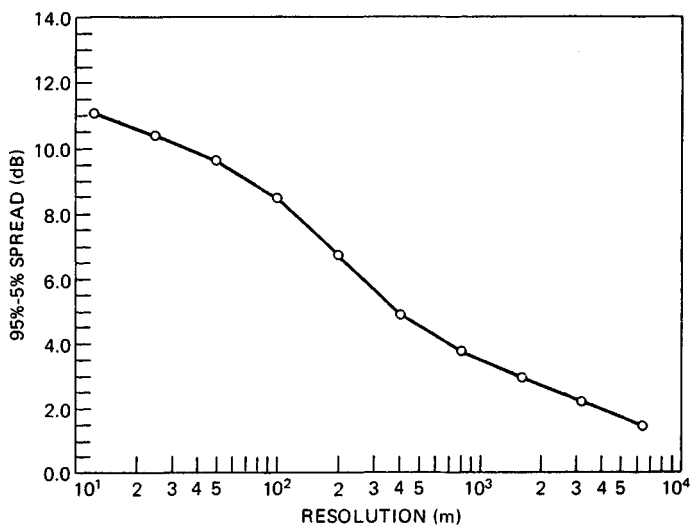
FIG. 12.28 90 percent range of pixel amplitude versus resolution.

with averages over different-sized footprints. For small footprints the scatter varies over a wide range, and system designers must account for this.

12.7 SCATTERING COEFFICIENT DATA

Numerous programs to gather scattering coefficient data existed prior to 1972, but sizable data collections with accompanying "ground truth" were rare. Since 1972, however, several major programs have changed the situation so that much information is now available. Indeed, this information is so widespread that an adequate summary of the literature is impossible. Hence, this section can only give highlights of the results and major programs. The reader should consult the three major compendia of such data for more information both on results and on bibliography[21,23,114] (note that information is spread through many chapters).

Some early scattering-coefficient-measurement programs worth mentioning include those of the Naval Research Laboratory,[15,16] Goodyear Aerospace Corporation,[12] Sandia Corporation (near-vertical data),[89,90] and particularly Ohio State University.[2,4] Since 1972 the largest program has been at the University of Kansas.[6,7,21,53,57,69,91] Extensive programs were also in France (Centre National d'Etudes Spatiales, Centre National d'Etudes des Télécommunications, Université Paul Sabatier),[9] the Netherlands,[8] Canada Centre for Remote Sensing (CCRS; especially sea ice),[17] and the University of Bern, Switzerland (snow).[92] Many of the results from these programs appear in digests of the International Geoscience and Remote Sensing Symposia (IGARSS; IEEE Geoscience and Remote Sensing Society) and journals such as *IEEE Transactions on Geoscience and Remote Sensing* and *on Ocean Engineering, International Journal of Remote Sensing, Remote Sensing of Environment,* and *Photogrammetric Engineering and Remote Sensing.*

Although calibrations for some of the older data were doubtful, summary presentations are not available for newer data. Accordingly, Fig. 12.29 shows an earlier summary based mostly on X-band data. One should be cautious in using this data, but the figure gives a feel for the overall variations. Figure 12.30 is a similar presentation for near-vertical data.[93] Calibration of the systems was good, but the antenna effect discussed in Sec. 12.5 makes the values from 0 to 5° low.

Effects of Roughness, Moisture Content, and Vegetation Cover. Scattering falls off more rapidly with angle for smooth surfaces than for rough surfaces. Since the roughness that affects radar must be measured in wavelength units, a surface smooth at long wavelengths may be rough at shorter ones. This is illustrated in Fig. 12.31,[94] which shows these effects with measurements from plowed fields. At 1.1 GHz the signal changed 44 dB between 0 and 30° for the smoothest field and only 4 dB for the roughest. At 7.25 GHz the smoothest field was rough enough to reduce the variation to 18 dB.

For most surfaces cross-polarized scatter is lower than like-polarized, often by about 10 dB. Cross-polarized scatter from smooth surfaces is much less near vertical than elsewhere. Figure 12.32[95] shows this effect. Cross-polarized returns from volume scatterers with elements that are large compared with a wavelength are stronger than for surfaces, sometimes being only 3 dB down.

Scatter depends on dielectric constant, which depends on moisture content. Thus scatter from wet soils at angles off vertical is usually much higher

CHAPTER 13
SEA CLUTTER

Lewis B. Wetzel
Naval Research Laboratory

13.1 INTRODUCTION

For an operational radar, backscatter of the transmitted signal by elements of the sea surface often places severe limits on the detectability of returns from ships, aircraft and missiles, navigation buoys, and other targets sharing the radar resolution cell with the sea. These interfering signals are commonly referred to as *sea clutter* or *sea echo*. Since the sea presents a dynamic, endlessly variable face to the radar, an understanding of sea clutter will depend not only on finding suitable models to describe the surface scattering but on knowledge of the complex behavior of the sea as well. Fortunately, a close relationship between radar and oceanography has grown up in the remote-sensing community, leading to the accumulation of a large amount of useful information about scattering from the sea and how this scattering relates to oceanographic variables.

It would seem a simple matter to characterize sea clutter empirically by direct measurement of radar returns for a wide variety of both the radar and environmental parameters that appear to affect it. Parameters relating to the radar or its operating configuration, such as frequency, polarization, cell size, and grazing angle, may be specified by the experimenter, but the environmental parameters are quite another matter—for two reasons. First, it has not always been clear which environmental variables are important. For example, wind speed certainly seems to affect clutter levels, but correlation of clutter with, say, ships' anemometer readings has not been entirely satisfactory. The state of agitation of the surface (*sea state*) appears to have a strong effect, but it is a subjective measure, and its relation to the prevailing local winds is often uncertain. Moreover, it has been found that the temperatures of the air and the sea surface can affect the way in which the measured wind speed is related to the generation of clutter-producing waves, yet the importance of these effects were unappreciated over most of the history of sea clutter measurements; so air and sea temperatures were seldom recorded. Even if the importance of an environmental parameter has been recognized, it is often difficult to measure it with accuracy under real-sea conditions, and there are practical and budgetary limits to obtaining open-ocean measurements in sufficient variety to develop any really meaningful statistical models of

clutter. Little wonder that many aspects of sea clutter remain frustratingly ill defined.

Before the late 1960s, most clutter data was collected in bits and pieces from isolated experiments, often with poor or incomplete ground truth. (For reviews of the older literature see, for example, Long,[1] Skolnik,[2] or Nathanson.[3]) Nevertheless, though much of the earlier clutter data was of limited scientific value, it did disclose some general trends, such as the tendency of clutter signal strength at low to intermediate grazing angles to increase with the grazing angle and with wind (or sea state) and generally to be greater for vertical polarization and in upwind-downwind directions.

It is commonly noted that, when viewed on an A scope, the appearance of sea clutter depends strongly on the size of the resolution cell, or *radar footprint*. For large cells it appears *distributed* in range and may be characterized by a surface-averaged cross section with relatively modest fluctuations about a mean value. As the size of the resolution cell is reduced, clutter takes on the appearance of isolated targetlike, or *discrete*, returns that vary in time. At these higher resolutions, the distributed clutter is often seen to consist of a dense sequence of discrete returns. When the discrete returns stand well out of the background, as they are seen to do for both polarizations but most clearly with horizontal polarization at small grazing angles, they are called *sea spikes* and are a common clutter contaminant in this radar operating regime.

Attempts to provide a theoretical explanation of the observed behavior of clutter signals trace essentially from the work pursued during World War II and described in the well-known MIT Radiation Laboratory book edited by Kerr.[4] Unfortunately, the scattering models developed during this period, along with most of those published over the following decade, failed to account for the behavior of sea backscatter in a very convincing way. In 1956, however, Crombie observed that at high-frequency (HF) wavelengths (tens of meters) scattering appeared to arise from a resonant interaction with sea waves of one-half of the incident wavelength, i.e., to be of the Bragg type.[5] Reinforced by the theoretical implications of various small waveheight approximations and wave tank measurements under idealized conditions, the *Bragg model* was introduced into the microwave regime by many workers in the mid-1960s.[6-8] This produced a revolution in thinking about the origins of sea clutter because it involved the sea wave *spectrum*, thus forging a link between clutter physics and oceanography in what became the field of *radio oceanography*. However, fundamental conceptual problems in applying the Bragg hypothesis in microwave scattering, along with recent questions about the validity of its predictions and the possibility of alternative scattering hypotheses, have reopened inquiry into the physical origins of sea scatter and how best to model it.[9-14] This being the case, speculation about physical models will be kept to a minimum in the sections on the empirical behavior of sea clutter. The problem of modeling sea scatter will be discussed separately in a later section.

13.2 DESCRIPTION OF THE SEA SURFACE

Close observation of the sea surface discloses a variety of features such as wedges, cusps, waves, foam, turbulence, and spray, as well as breaking events of all sizes and masses of falling water. Any or all of these might contribute to the scattering of electromagnetic waves responsible for sea clutter. The basic ocean-

ographic descriptor of the sea surface, however, is the *wave spectrum*, which, while saying little about these features, contains a great deal of information about the sea surface in general and is central to the application of the Bragg scattering hypothesis. In view of the need to understand the sea surface in order to understand sea clutter and the prominence of the Bragg hypothesis in existing clutter models, some tutorial material describing the spectral characterization of the sea surface is included below.

There are basically two types of surface waves, *capillary* and *gravity*, depending on whether surface tension or gravity is the dominant restoring force. The transition between one and the other takes place at a wavelength of about 2 cm; so the smaller capillary waves supply the surface fine structure while gravity waves make up the larger and most visible surface structures. Waves have their origin ultimately in the wind, but this does not mean that the "local" wind is a particularly good indicator of what the wave structure beneath it will be. In order to arouse the surface to its *fully developed* or *equilibrium* state, the wind must blow for a sufficient time (*duration*) over a sufficient distance (*fetch*). That part of the wave structure directly produced by these winds is called *sea*. But waves propagate, so even in the absence of local wind, there can be significant local wave motion due to waves arriving from far away, perhaps from a distant storm. Waves of this type are called *swell*, and since the surface over which the waves travel acts as a low-pass filter, *swell* components often take the form of long-crested low-frequency sinusoids.

The Wave Spectrum. The wave spectrum which provides the primary oceanographic description of the sea surface appears in several forms. If the time history of the surface elevation is monitored at a fixed point, the resulting time series may be processed to provide a *frequency spectrum* $S(f)$ of the surface elevation, where $S(f)df$ is a measure of the *energy* (i.e., square of the waveheight) in the frequency interval between f and $f + df$. Wave spectra have been measured in the open ocean primarily for gravity waves down to wavelengths of about 1 m. Open-ocean measurements of capillary waves are especially difficult to perform.[15]

For a gravity wave, the frequency f and the wavenumber K are related by the dispersion relation

$$f = (1/2\pi)(gK)^{1/2} \tag{13.1}$$

where g is the acceleration of gravity and $K = 2\pi/\Lambda$, with Λ being the wavelength. Although each individual gravity wave obeys this relation, the waves at a point on the sea surface could come from any direction; so they are characterized by a two-dimensional propagation *vector* with orthogonal components K_x and K_y, where the K to be used in Eq. (13.1) is the magnitude $K = (K_x^2 + K_y^2)^{1/2}$.

The wavenumber spectrum associated with $S(f)$ is a function of the two components of K and is commonly written as $W(K_x, K_y)$. This is called the *directional wave spectrum* and expresses the asymmetries associated with winds, currents, refraction, isolated swell components, etc. For a given source of asymmetry like the wind, various parts of the spectrum will display different directional behaviors. For example, in a fully developed sea, the larger waves will tend to move in the direction of the wind while the smaller waves will be more isotropic. Directional spectra are more difficult to measure and are obtained by a variety of experimental methods, such as an array of wave staffs to measure surface heights over a matrix of points, a multiaxis accelerometer buoy, and stereo photography,

and even by processing radar backscatter signals. However, a frequency spectrum measured at a point can contain no knowledge of wave direction; so a wavenumber spectrum $W(K)$ is often defined in terms of the frequency spectrum $S(f)$ by the relation

$$W(K) = S(f(K))(df/dK) \qquad (13.2)$$

with the relation between f and K given by Eq. (13.1). To account for the wind direction, $W(K)$ is sometimes multiplied by an empirical function of K and direction v relative to the (up)wind direction.

Oceanographers have not always been in complete agreement about the form of the frequency spectrum. Nonequilibrium wave conditions, inadequate sampling times, poor ground truth, etc., can contaminate the data set from which empirical spectra are derived. However, by careful selection of data from many sources, ensuring that only equilibrium (fully developed) sea conditions were represented and the wind was always measured at the same reference height, Pierson and Moskowitz[16] established an empirical spectrum that has proved popular and useful. It has the form

$$S(f) = Af^{-5}e^{-B(f_m/f)^4} \qquad (13.3)$$

where g is the acceleration of gravity, and $f_m = g/2\pi U$, corresponding to the frequency of a wave moving with a velocity equal to the wind speed U; A and B are empirical constants. This spectrum is illustrated in Fig. 13.1 for several wind
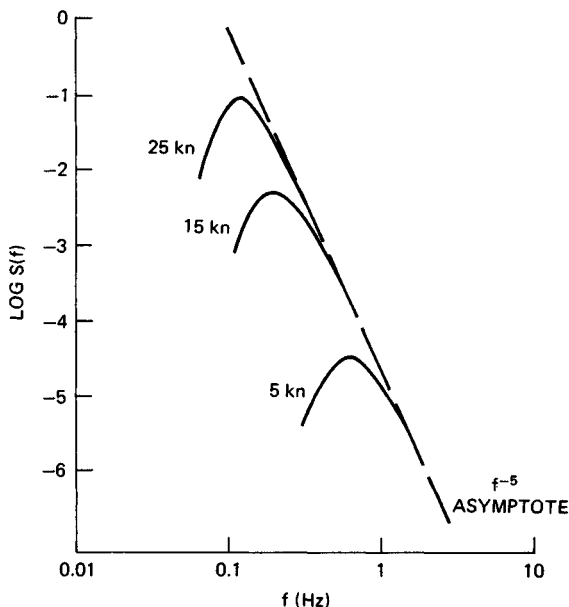
FIG. 13.1 Sea wave frequency spectra of the Pierson-Moskowitz type, representing fully developed seas.

speeds. The effect of increasing wind speed is simply to move the low-frequency cutoff to lower frequencies along the high-frequency f-minus-5 asymptote. (It should be noted that most of the oceanographers' spectra are based on measurements at relatively low frequencies and so cannot be taken seriously at frequencies above about 2 Hz. Nevertheless, these spectral forms are often used up to 20 Hz or greater in predicting radar clutter under the Bragg hypothesis.)

Converting this frequency spectrum into an isotropic wavenumber spectrum through Eq. (13.2) results in a spectrum of similar form, only with a K-minus-4 asymptote. Phillips[17] derived this asymptotic behavior on dimensional grounds, and a widely used simplification, obtained by replacing the smooth peak in Fig. 13.1 by a sharp cutoff, is generally referred to as the *Phillips spectrum* and in wavenumber space is written

$$W(K) = 0.005/K^{-4} \qquad K > g/U^2$$
$$= 0 \qquad\qquad\quad K < g/U^2 \tag{13.4}$$

where the cutoff wavenumber corresponds to the frequency f_m of the peak in Eq. (13.3). Opposed to this highly simplified form are increasingly complex spectra based on more careful empirical studies[18] as well as more sophisticated theoretical considerations.[19,20]

In discussing the characterization of the sea surface by its spectrum, it must be kept in mind that the spectrum is a highly averaged description of how the *energy* of the surface is distributed among the wavenumbers, or frequencies, of the waves present on it. Since the phases of these waves are lost, the spectrum gives no information about the morphology of the surface itself, i.e., about the complex surface features that are responsible for the scattered field. This point will be raised again in the section below on theories of sea clutter.

General Sea Descriptors. The shape of the curves in Fig. 13.1 suggests that the sea wave system has a relatively high Q; so it should be possible to get a rough idea of the behavior of the major waves on the surface by taking the values of *period* ($1/f$) and *wavelength* ($2\pi/K$) defined at the spectral peak. These values belong to a wave satisfying the dispersion relation Eq. (13.1) and having a phase velocity $C = 2\pi \times f/K$ equal to the wind speed U. By using Eq. (13.1), the period T and wavelength Λ thereby defined take the form

$$T = 0.64U \qquad \Lambda = 0.64U^2 \tag{13.5}$$

where U is in meters per second. Thus, for example, the largest waves in a fully developed sea for a 15-kn (7.5 m/s) wind will have a wavelength of about 120 ft (36 m) with a period of 5 s.

The statistical distribution of waveheights on the ocean surface is quite close to gaussian, with a mean square deviation that can be obtained by integrating the waveheight spectrum over all frequencies (or wavenumbers). For spectra resembling those in Fig. 13.1, the rms waveheight is given approximately by

$$h_{rms} = 0.005U^2 \qquad \text{m} \tag{13.6}$$

The rms waveheight contains contributions from all the waves on the surface, but very often it is the peak-to-trough height for the higher waves that is of major interest. This is certainly the case for a ship in a seaway or in the shadowing of the surface at low radar grazing angles. The *significant height*, or height of the

one-third highest waves, provides such a measure. It is denoted by $H_{1/3}$ and is taken to be about 3 times the rms height given by Eq. (13.6). For a 15-kn wind, this is only about 3 ft, but for gale-force winds of 40 kn it rises to over 20 ft, which is a rather formidable sea.

Looking at the sea, an observer might describe what he or she sees in terms of a subjective *state of the sea*, e.g., "smooth," "rough," "terrifying!" If these descriptions are listed in order of severity and assigned numbers, these numbers define a *sea state*. A similar numerical scale exists for wind speeds, the *Beaufort wind scale*, with numbers about an integer higher than the corresponding sea state. But it is seldom used in reference to sea clutter.

There are, then, two numbers commonly used to indicate the activity of the sea surface: a subjective sea state and a measured wind speed. Only when the wind has sufficient *fetch* and *duration* to excite a *fully developed* sea, can a wave height be unambiguously associated with it. The surface descriptors generally used in connection with sea clutter—sea state, wind speed, and its associated equilibrium waveheight—are given in Table 13.1, with the wind speed in knots, the significant waveheight in feet, and the duration/fetch required for a fully developed sea in hours/nautical mile. It is of interest to note that the median wind speed over the world's oceans is about 15 kn, corresponding to sea state 3.

TABLE 13.1 Sea-Surface Descriptors

Sea state	Wind speed, kn	Waveheight $H_{1/3}$, ft	Duration/fetch, h/nmi
1 (smooth)	< 7	1	1/20
2 (slight)	7–12	1–3	5/50
3 (moderate)	12–16	3–5	15/100
4 (rough)	16–19	5–8	23/150
5 (very rough)	19–23	8–12	25/200
6 (high)	23–30	12–20	27/300
7 (very high)	30–45	20–40	30/500

13.3 EMPIRICAL BEHAVIOR OF SEA CLUTTER

Sea clutter is a function of many parameters, some of them showing a complicated interdependence; so it is not an easy task to establish its detailed behavior with a great deal of confidence or precision. For example, in a proper sea clutter measurement, the polarization, radar frequency, grazing angle, and resolution cell size will have been specified. Then the wind speed and direction must be measured at a reference altitude, and if the results are to be compared with those of other experimenters, the proper *duration* and *fetch* should be present to ensure standardization to equilibrium sea conditions. Since these measured winds are related to the wind structure at the surface through the atmospheric boundary layer, the shape of this layer must be determined by measuring the air and sea temperatures. To complicate the picture still further, it is becoming increasingly clear that sea backscatter has a strong dependence on the direction of the long waves, which include *swell*, in the measurement area; so ideally the *directional wave spectrum* should be measured as well. Obviously, it is unlikely that all these

environmental parameters will be recorded with precision in every (or even *any*) sea clutter measurement; so considerable variability in the basic conditions under which sea clutter data is collected by different experimenters can be expected. It is of interest to note that in many of the reported measurements of sea clutter, particularly in the older literature, wide inconsistencies between wind speed and waveheight may be found. For example, a wind speed of 5 kn might be reported with waveheights of 6 ft, or 20-kn winds with 2-ft waves. These pairings are inconsistent with the values for an equilibrium sea described in Table 13.1 and indicate the unnoticed presence of heavy swell or highly nonequilibrium wind conditions, or both. Even with all the variables properly specified, recorded clutter data can be spread over a wide dynamic range, often as great as 40 dB at low grazing angles, so that clutter behavior is best described in terms of probability distribution functions.

Since sea clutter is generally viewed as a surface-distributed process, the basic clutter parameter is taken to be the normalized radar cross section (NRCS), σ^0, of the surface, commonly referred to as *sigma zero* and expressed in decibels relative to 1 m^2/m^2. It is obtained experimentally by dividing the measured radar cross section of an illuminated patch of the surface by a normalizing area; so differences in the definition of this area can lead to inconsistencies among various reports of NRCS measurements. Scattering from any distributed target involves the product of the transmitting and receiving system footprints integrated over the target. These footprints cover exactly the same area for a monostatic radar and will depend on the pulse and beamwidths, the range, and the grazing angle. If the footprints are assumed to be of the *cookie-cutter* type (constant amplitude falling sharply to zero at the half-power points), then the relation between the actual radar clutter cross section σ_c, as inferred from the received power via the radar equation, and the NRCS σ^0 is given by

$$\sigma^0 = \sigma_c/A_f \qquad (13.7)$$

where for a radar with an antenna beamwidth B and rectangular pulse of length T, viewing the surface at range R and grazing angle θ, the area A_f is either

$$A_f = \pi(BR)^2/4\sin\theta \qquad (13.8)$$

for beam-limited conditions [e.g., continuous-wave (CW) or long-pulse radar at high grazing angles] or

$$A_f = (c\tau/2)BR/\cos\theta \qquad (13.9)$$

for pulse-width-limited conditions (e.g., short-pulse radar at low grazing angles).

Real radars do not produce cookie-cutter footprints, however, since the antenna beam will have, say, a Bessel or gaussian profile and the pulse might be shaped. For this reason, an effective A must be obtained from a surface integration of the square of the actual amplitude profile of the footprint, which will always result in a smaller value of A than that defined by Eq. (13.8) or Eq. (13.9), and therefore in larger values of σ^0 as derived from measured values of σ_c by Eq. (13.7). Most experimenters use the half-power beamwidth in Eq. (13.8) or Eq. (13.9), with an error that is usually only 1 or 2 dB.

Dependence on Wind Speed, Grazing Angle, and Frequency. It was noted earlier that summaries of clutter measurements made before about 1970 may be found in several of the standard reference books on radar[2,3] and radar clutter.[1] Among the programs of this period, the most ambitious was that pursued in the late 1960s at the Naval Research Laboratory (NRL),[21] in which an airborne four-frequency radar (4FR), operating with both horizontal and vertical polarizations at UHF (428 MHz), L band (1228 MHz), C band (4455 MHz), and X band (8910 MHz), made clutter measurements upwind, downwind, and crosswind in winds from 5 to 50 kn for grazing angles between 5 and 90°. The system was calibrated against standard metal spheres, and wind speeds and waveheights were recorded in the measurement areas from instrumented ships.

Typically, samples of σ^0 for a given set of radar and environmental parameters are scattered over a wide range of values and in the NRL experiments were organized into probability distribution functions of the type shown in Fig. 13.2. The
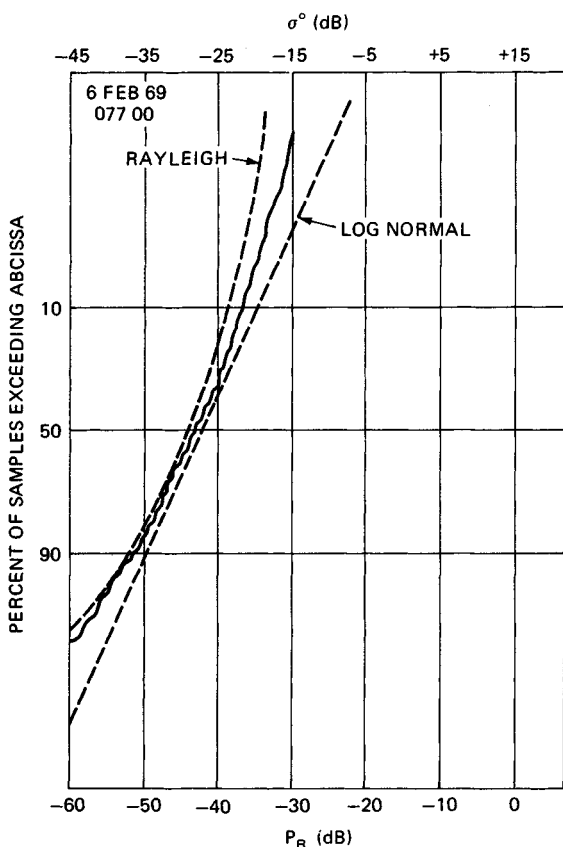
FIG. 13.2   An example of the probability distribution of sea clutter data. (*From Daley.*[21])

data, represented by the solid line, is plotted on normal probability paper with Rayleigh and log-normal distributions shown for comparison (dashed lines). The ordinate is the *percent of time by which the abscissa is exceeded*, and the abscissa is the value of σ^0 as defined by Eq. (13.7), with A taken from Eq. (13.8) or Eq. (13.9) as appropriate. This particular distribution is representative of clutter from a relatively large radar footprint (pulse length about 0.5 μs) measured at intermediate grazing angles (20 to 70°) for moderate wind speeds (about 15 kn). It is Rayleigh-like but shows a tendency toward log-normal behavior for the larger cross sections. From a detailed statistical analysis of the NRL 4FR data, Valenzuela and Laing[22] concluded that, for this data at least, the distributions of sea clutter cross sections were intermediate between the exponential (which is the power distribution corresponding to Rayleigh-distributed scattered-field amplitudes) and log-normal distributions.

Organizing the data samples into probability distributions makes the *median* (50 percent) value a convenient statistical measure of the clutter cross section. But many investigators process their data to provide the *mean* value, and since the conversion of a *median* to a *mean* requires knowledge of the probability distribution function, care must be taken to avoid ambiguity in comparing the measurements of different experimenters. The original analysis of the NRL 4FR data was based on *median* cross sections and the assumptions of the cookie-cutter antenna beam embodied in Eqs. (13.8) and (13.9).[21,23] In later presentations of this data,[24] the *median* values of σ^0 were replaced by *means*, raising them by about 1.6 dB, and the area A in Eq. (13.7) was redefined in terms of a more realistic tapered footprint, adding another 1 to 2 dB. This means that there is a difference of 3 to 4 dB between the earlier and later presentations of the same data, and since these results are widely used and quoted, it is important to ensure that the proper definition of σ^0 is being used when comparing them with clutter data that has been taken by other experimenters or in using these results in clutter predictions.

General Results.   Being the first really comprehensive collection of clutter data over a wide range of radar frequencies, the 4FR program produced many plots showing the dependence of sea clutter on grazing angle, frequency, polarization, wind direction, and wind speed. However, comparison of these plots with others made both earlier and later shows the extent of the variations to be found in sea clutter measurements reported by different investigators for exactly the same set of parameters. This is seen clearly in Fig. 13.3a and b, which compares the grazing-angle dependence of X-band clutter data for wind speeds in the neighborhood of 15 kn obtained from four sources: NRL 4FR[24] (these are *mean* results for upwind directions and include the antenna corrections mentioned above), aircraft measurements by Masuko et al.[29] (also in the upwind direction), and summaries of the older data (pre-1970) taken from books on radar systems by Skolnik[2] and Nathanson.[3] The discrepancies between the different data sets can be accounted for, at least in part, as follows. The older data set was based on published measurements from various sources, and since there is no specification of wind direction, it may be assumed that it represents some kind of average of upwind, downwind, and crosswind directions. As will be seen below, this average is about 2 to 3 dB smaller than the upwind returns. Moreover, the *early* NRL 4FR data was used liberally in the older data summaries, and it was noted above that there is a difference of 3 to 4 dB between the early and later presentations of the same NRL 4FR data, the latter being used in Fig. 13.3a and b. With these
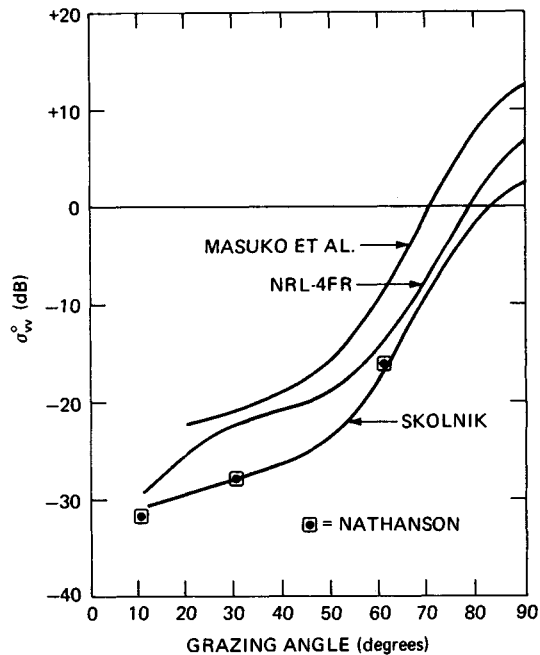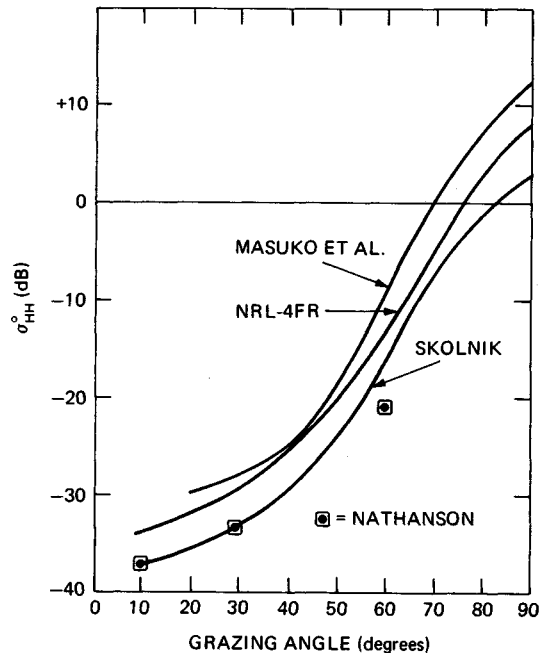
FIG. 13.3 Comparison of X-band clutter data from different sources for a nominal wind speed of 15 kn. (*a*) Vertical polarization. (*b*) Horizontal polarization. (*Based on data from Masuko et al.,*[29] *NRL 4FR,*[21] *Skolnik,*[2] *and Nathanson.*[3])

corrections, the curves would show closer agreement. Nevertheless, it is clear that uncritical use of published clutter data could lead two radar systems designers to choose sea clutter estimates almost an order of magnitude apart for the same conditions.

The NRL 4FR data set is unique in that no other program has reported measurements made over so wide a range of frequencies, grazing angles, and wind speeds at the same time. Figure 13.4 shows the trends for both vertically and horizontally polarized sea clutter over a range of grazing angles down to 5°. The curves represent the centers of ±5 dB bands which contain the major returns for the three higher frequencies (L, C, and X bands—the UHF returns were a few decibels lower) and wind speeds above about 12 kn. The major differences in sea clutter for the two polarizations are seen to lie in the range of grazing angles between about 5 and 60°, where the horizontally polarized returns are smaller. This difference is found to be emphasized at both lower wind speeds and lower frequencies. The cross sections approach each other at high angles (>50°) and, for the higher microwave frequencies, at low angles (<5°) as well. In fact, for grazing angles less than a few degrees and moderate to strong wind speeds, several observers have reported that at X band and at the higher sea states the horizontally polarized returns often exceed the vertically polarized returns.[1,25,26]

The NRL 4FR system permitted transmission and reception on orthogonal polarizations so that data could be collected for cross-polarized sea clutter. These returns tended to have a weak dependence on grazing angle and were always smaller than either of the like-polarized returns, lying in the cross-hatched region shown on Fig. 13.4.

It is informative to compare measurements by different investigators in different parts of the world under similar wind conditions. Figure 13.5 displays measurements of vertically polarized sea clutter down to a grazing angle of 20° for wind speeds of about 15 kn.from three independent experiments using airborne radars at C-, X-, and K-band frequencies.[27-29] While there is no assurance that all these measurements were made over fully developed seas, it is clear that there is a rather strong consistency among them, which reinforces the observation made in reference to Fig. 13.4 that the frequency dependence of sea clutter at intermediate grazing angles is weak at microwave frequencies from L to K band.

Dependence on Wind Speed. The relation between sea clutter and wind speed is complex and uncertain, since it has been found to depend on almost all the parameters that characterize sea clutter: frequency, grazing angle, polarization, the state of the sea surface, the direction and speed of the wind itself, and even on whether the measurements are made from an aircraft or a tower platform.[30]

A common way to organize clutter data is to seek the best straight-line fit (linear regression) between clutter cross sections in decibels and the log of the wind speed (or some other parameter). This, of course, *imposes* a power-law relation between the variables: $\sigma^0 \propto U^n$, where n is determined by the slope of the
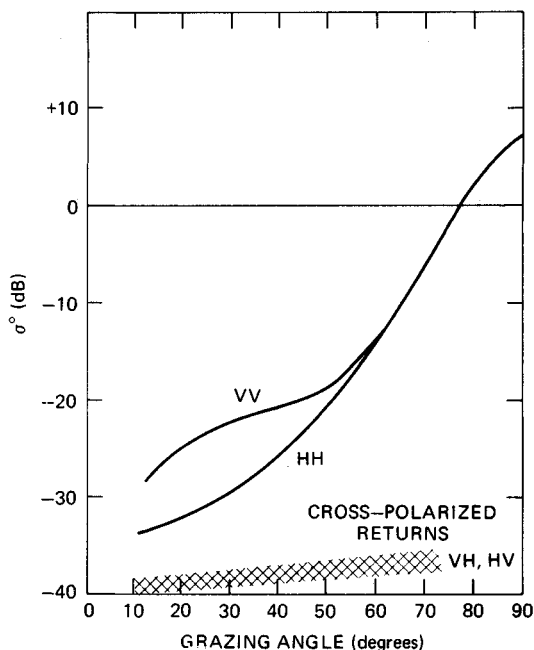
FIG. 13.4 General trends in clutter behavior for average wind speeds (about 15 kn) based on NRL 4FR data. Plots represent L-, C-, and X-band data within ±5 dB.
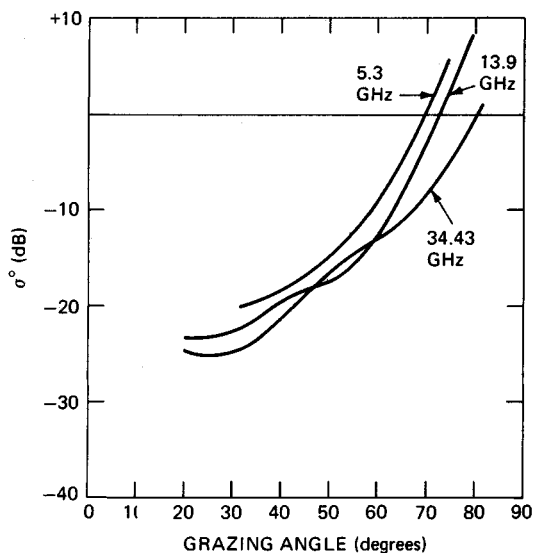
FIG. 13.5 Frequency dependence of sea clutter for wind speeds of about 15 kn: 5.3 GHz, Feindt;[27] 13.9 GHz, Schroeder;[28] 34.4 GHz, Masuko.[29]

line. An example is shown in Fig. 13.6.[31] On the other hand, the totality of the NRL 4FR results appeared to show saturation for wind speeds above about 20 kn, but the high and low- to moderate- wind-speed data was collected at different times in different places under different conditions of sea-surface development, and discrepancies between the two data sets for common wind speeds have weakened the evidence for saturation.[32] Other investigators deny that it is even possible to express wind dependence in the form of a power law, proposing the existence of a kind of threshold wind speed, below which clutter virtually vanishes and above which the clutter level rises toward a saturation value.[18] This is indicated by the curves in Fig. 13.7, where the straight lines correspond to various power laws. Once this possibility is raised, it is possible to find examples of data that appear to track such a curve while at the same time yielding a power law by linear regression, as illustrated in the tower data shown in Fig. 13.8.[31] This behavior is not uncommon.
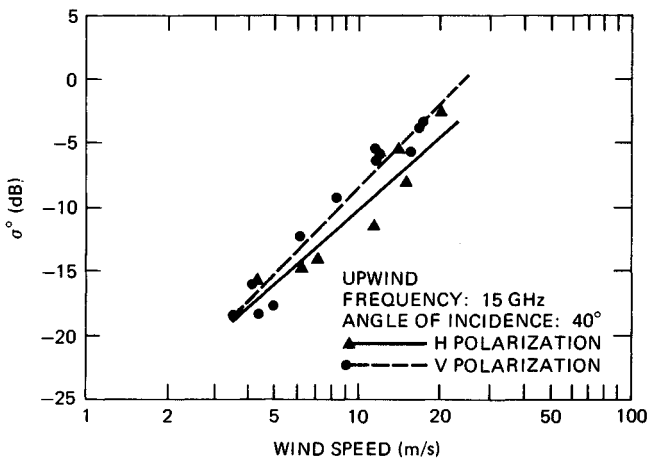
FIG. 13.6 Sea clutter from a tower platform with power-law wind-speed dependence defined by linear regression. (*From Chaudhry and Moore*,[31] © *1984, IEEE.*)

Nevertheless, the imposition of a power-law relation provides a convenient way to visualize trends in the behavior of sea clutter with wind speed. The various aircraft measurements referred to above[27–29] as well as data from a tower in the North Sea[30,31] were all treated in this way, yielding plots of σ^0 as a function of wind speed and grazing angle of the form shown in Fig. 13.9a and b. Plots of this type give information about both the wind-speed and grazing-angle dependence of sea clutter for a given frequency, polarization, and wind direction. Figure 13.9a and b is based on a blend of radiometer-scatterometer (RADSCAT) data at 13.9 GHz[29] and measurements by Masuko et al. at 10 GHz,[29] both for upwind directions. Thus they can be viewed as representative of clutter behavior in the vicinity of X band, since the difference between the two frequencies is small. However, examination of the data points underlying these linear regressions show point scatter that sometimes resembles Fig. 13.6, sometimes Fig. 13.8, and sometimes neither; so the straight lines in these figures cannot be taken too seriously. In fact, it appears that there is no simple functional dependence of
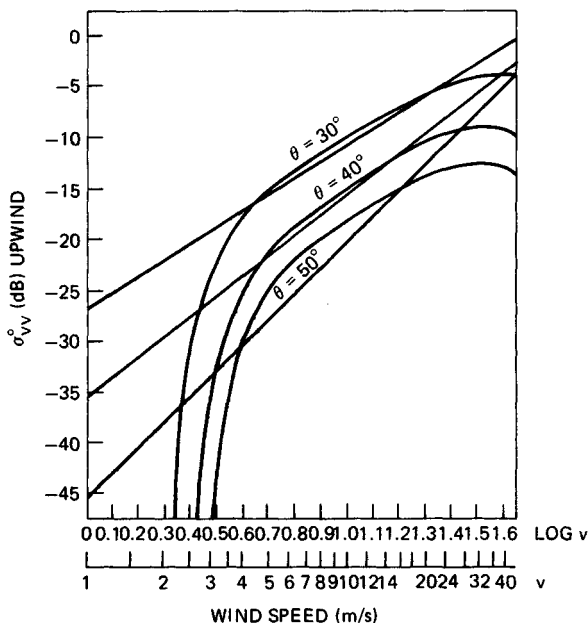
FIG. 13.7 A hypothetical wind-speed dependence of sea clutter (curved traces) compared with various power laws (straight lines). (*Derived from Pierson and Donelan.*[18])
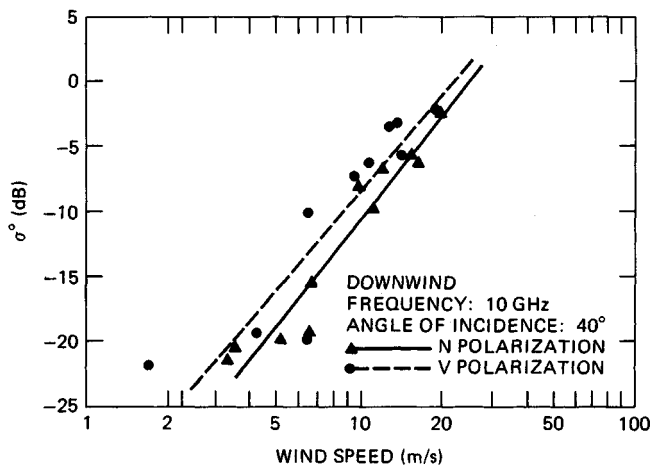
FIG. 13.8 Example of forcing a power-law fit (compare data with curves in Fig. 13.6). (*From Chaudhry and Moore,*[31] *© 1984, IEEE.*)
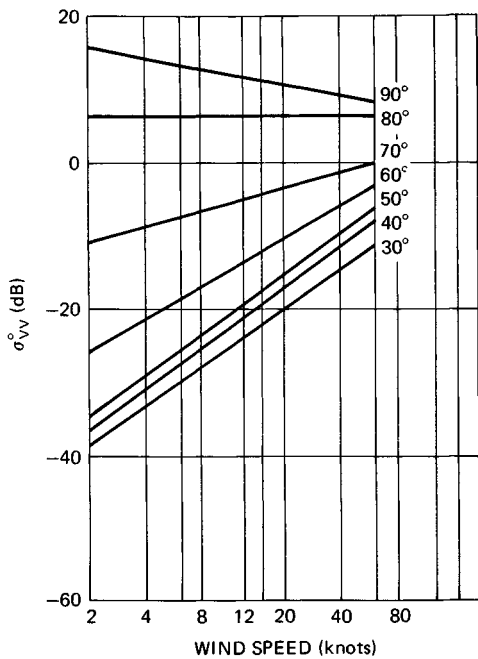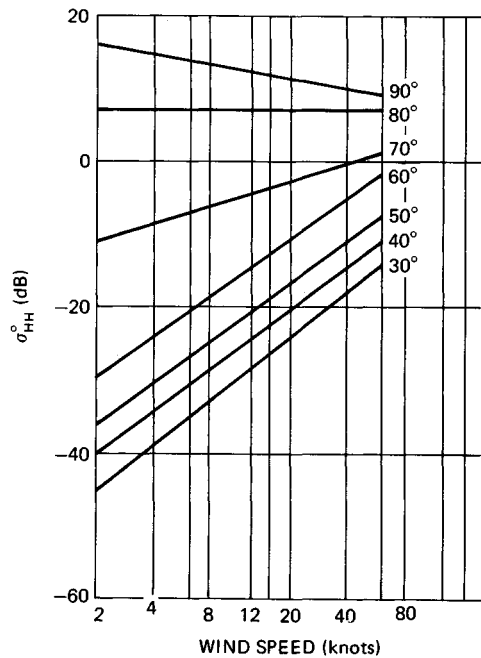
FIG. 13.9   Example of clutter behavior with wind speed and grazing angle—average of data at 10 GHz[29] *and 13.9 GHz.*[28] (*a*) Vertical polarization. (*b*) Horizontal polarization.

sea clutter on wind speed that can be established with any confidence from existing data, although most investigators would probably agree that the behavior of microwave sea clutter with wind speed at intermediate grazing angles can be roughly described as follows: for light winds (less than 6 to 8 kn) sea clutter is weak, variable, and ill defined; for intermediate winds (about 12 to 25 kn) it can be described roughly by a power law of the type found in Fig. 13.6; and for strong winds (above about 30 kn) there is a tendency for it to level off. In fact, the convergence of the lines in Fig. 13.9a and b with increasing wind speed suggests that the reflectivity of the sea surface is tending toward Lambert's law, for which there is no dependence on grazing angle, frequency, or polarization but only on surface *albedo*.

Dependence on Wind Direction. In several of the experiments referenced above, the dependence of sea backscatter on angle relative to the wind direction was found by recording the radar return from a spot on the surface while flying around it in a circle. Figure 13.10a and b gives an example of this behavior for grazing angles of about 45° and wind speeds close to 15 kn.[29] The figures contain results obtained independently by three different groups. The behavior shown here is representative of that found generally: sea clutter is strongest viewed upwind, weakest viewed crosswind, and of intermediate strength viewed downwind, the total variation being about 5 dB.

At High Grazing Angles. The top curve in Fig. 13.9a and b corresponds to clutter at a grazing angle of 90°, that is, for a radar looking straight down. On a strictly empirical basis, the clutter cross section at this angle is only weakly de
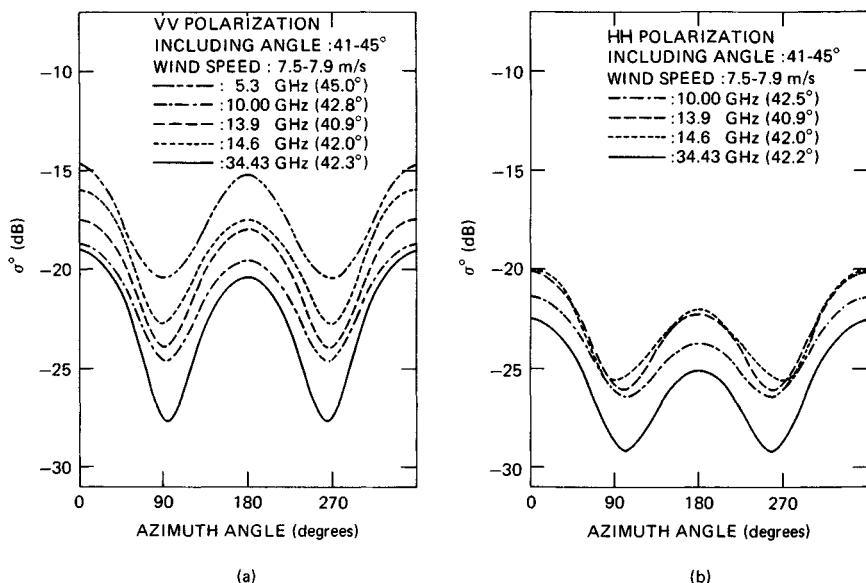
FIG. 13.10 Dependence of clutter on wind direction: nominal wind speed, 15 kn; grazing angle, 45°. (*From Masuko et al.*,[29] © *by the American Geophysical Union.*)

pendent on frequency, has a maximum of about $+15$ dB at zero wind speed (at least for the antenna beamwidths and experimental configurations reported), and falls off gradually as the wind picks up. Scattering at high grazing angles is commonly regarded as a form of specular scattering from tilted facets of the surface; so it is of interest to note that there appears to be a small range of angles in the neighborhood of 80° for which the cross section is almost completely independent of wind speed. Since these angles correspond to complements of the common rms sea slope angles of about 10°, it might be argued that as the wind increases, the clutter decrease due to increasing surface roughness is balanced at these angles by a clutter increase due to an increasing population of scattering facets. This line may therefore be regarded as the boundary separating the *specular* regime, where the cross section is decreased by surface roughness, from the *rough-surface* regime, where the cross section increases with surface roughness. It should further be noted that clutter measurements at these high grazing angles will be relatively sensitive to the averaging effects of wide antenna beamwidths, which could become a source of ambiguity in aircraft measurements at the lower radar frequencies.

At Low Grazing Angles. At low grazing angles, below mean sea slope angles of about 10°, sea clutter takes on a different character. The sharp clutter peaks known as sea spikes began to appear on A-scope presentations,[1,25,33] and the probability distributions assume a different form.[34] Figure 13.11 shows the presence of sea spikes in the time histories of returns from a fixed spot, measured from a tower in the Gulf of Mexico with a high-resolution X-band radar looking into an active sea at a 1.5° grazing angle.[33] The vertically polarized returns appear to be a bit broader, and while the horizontally polarized returns are more spiky, both polarizations display the sharp bursts characteristic of sea clutter at small grazing angles. The peak cross sections in these records are of the order of 10 m^2 and are roughly the same for the two polarizations, which is another characteristic of sea clutter at these angles. Interestingly, while the same measurements made in "calm" water looked virtually identical in every detail, peak cross sections were now only 10 cm^2, or 40 dB, less.

VERTICAL POLARIZATION

HORIZONTAL POLARIZATION
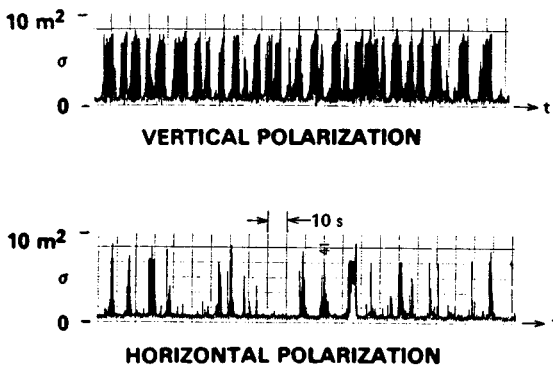
FIG. 13.11 Sea spikes at X band, 1.4° grazing angle, moderate to strong winds. Note equal amplitudes at the two polarizations. (*From Lewis and Olin.*[33])

Trizna has accumulated a considerable body of data from measurements of low-angle sea clutter using high-resolution (40-ns) shipboard radar in both the Atlantic and the Pacific oceans.[34] The probability distributions of the clutter cross sections were plotted in the manner of Fig. 13.12, which shows the distributions of horizontally polarized X-band data at a 3° grazing angle for low, medium, and high wind speeds (in order from left to right). The low-wind trace corresponds to a Rayleigh distribution, while the other straight-line segments are two-parameter Weibull distributions defined by different parameter pairs. It is clear that the behavior is different and considerably more complex than that shown in Fig. 13.2 for the higher grazing angles and wider pulses. Trizna interprets these distributions as follows: in each trace, the left-hand segment (lowest cross section) is actually receiver noise, recorded when the radar footprint lay in shadow; the middle section corresponds to distributed clutter, for reasons relating mainly to its weak dependence on resolution cell size; the right-hand section (highest cross section) describes the sea spikes, for reasons relating to the dependence on wind speed (similar to whitecap dependence) and the sheer size of the components (some individual absolute cross sections in excess of 1000 m²). For the higher wind speeds and fully developed seas encountered in the North Atlantic, the population of this sea-spike sector (the percentage of sea spikes) was found to grow as the 3.5th power of the wind speed, which, interestingly, is the same wind-speed dependence shown by the percentage of whitecaps seen on the surface.[36]
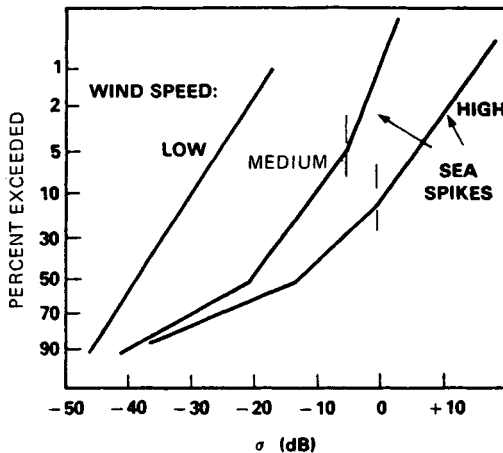
FIG. 13.12 Segmented clutter probability distributions at low grazing angles. (*Based on Trizna.*[34])

It should be kept in mind that, to the extent that the sea surface may be viewed as a stationary homogeneous process, as it generally is over the duration and spatial extent of any particular experimental event, the scattering cross section may be said to be *ergodic*, which means that the statistical results obtained by time averaging from a small cell are equivalent to a shorter time average from a larger cell, provided that the number of "samples" is the same in the two cases. For this reason, the statistical implications of experimental data can be properly compared only if the details of the sampling procedure are specified. However, the number of samples in the experimental results shown thus far have been suf-

ficiently large that the differences between, for example, Figs. 13.2 and 13.12, may be considered real and related to differences in grazing angle rather than in resolution cell size. In fact, distributions closely resembling those in Fig. 13.12 were obtained much earlier from similar measurements with considerably broader pulse widths.[35]

At Very Low Grazing Angles. There is some evidence that sea clutter might drop off more sharply below a *critical angle* in the neighborhood of a degree or so (see Long[1]). This critical angle, or *critical range* for a radar at a fixed height, has been observed from time to time since first noted in early observations of sea clutter.[4] According to Katzin,[37] the critical angle occurs as a result of interference between direct and (perfectly) reflected rays at the scattering *targets* responsible for the clutter signal. While this simple picture can account for the R-minus-7 decay sometimes observed, a critical angle often fails to materialize, and when it does, it need not show an R-minus-7 decrease with range (or the equivalent fourth-power dependence on grazing angle).[1] An alternative explanation for this behavior, applicable at the higher microwave frequencies, has been suggested by Wetzel,[12,38] based on a *threshold-shadowing* model for upwind and downwind directions that implies a sharp decrease in the average cross section for grazing angles below a few degrees. In crosswind directions, with the radar looking along the troughs of the major waves, a much milder shadowing function will apply; so there should be a clear distinction between the upwind-downwind and crosswind behavior of sea clutter at very low grazing angles.

Examples of clutter behavior at these angles may be found in independent measurements at relatively high wind speeds by Hunter and Senior off the south coast of England[39] and by Sittrop off the west coast of Norway.[40] Their results for orthogonal directions relative to the wind are shown in Fig. 13.13, along with the predictions of a conventional shadowing function[41] and the threshold-shadowing function.[38] It would appear that a combination of conventional shadowing (which goes as the first power of the grazing angle) across the wind and threshold shadowing in upwind and downwind directions accounts for the observed behavior of this very low angle clutter quite well. The decay law for low-angle clutter should therefore depend on the viewing angle relative to the wind direction; so it might occur with powers between the first and the fourth. This is just what is observed.[42] It should be remarked, however, that shadowing at low grazing angles is a complex phenomenon (see below), and the physical origin or even the existence of a critical angle is still open to question. Moreover, there is relatively little good data on very low angle clutter for other than X-band frequencies; so the general behavior of sea clutter in this angular regime remains uncertain.

At HF and Millimeter-Wave Frequencies. All the measurements described above were made at *microwave* frequencies between UHF (428 MHz) and K_a band (35 GHz). High-frequency (HF) radars usually operate in the frequency range between about 5 and 30 MHz, corresponding to wavelengths between 60 and 10 m, respectively. Since the operation of such radars takes place either by the ground wave or over ionospheric (*sky-wave*) paths spanning great ranges, the grazing angles tend to be small (between 0 and 20°). For these wavelengths and grazing angles, initial measurements by Crombie indicated that the scattering from the sea surface was the result of Bragg scatter from sea waves of one-half the radar wavelength.[4] In the years since these early measurements, there has been considerable activity in the field of HF radar and HF clutter,[43,44] and the results can be summarized as follows: For vertical polarization, the major energy of the HF clutter signal appears in spectral lines displaced to either side of the
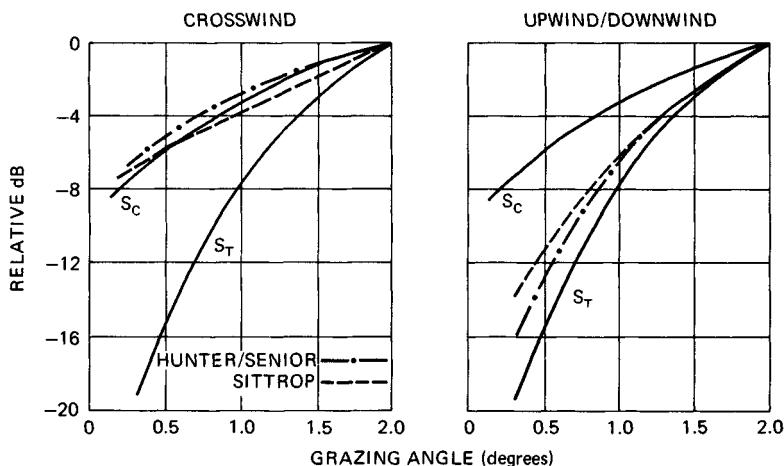
FIG. 13.13 Differential behavior of very low angle clutter for orthogonal wind directions: S_c is a conventional shadowing function;[41] S_T is a threshold-shadowing function.[38] (*Data from Hunter and Senior*[39] *and Sittrop.*[40])

carrier frequency by the frequency of sea waves having a wavelength equal to half the HF wavelength λ (in meters). The relative strengths of the plus and minus lines are determined by the proportion of advancing and receding Bragg-resonant wave components in the clutter cell. Provided the wind speed is greater than about $\sqrt{3\lambda}$ kn (with λ in meters) and the sea is fully developed, the clutter cross section σ^0 is about -27 dB and is relatively independent of wind speed and frequency. (The definition of σ^0 in HF radar is complicated by problems in properly defining antenna gains for ground-wave and sky-wave paths and by propagation effects due to the ionosphere.) The clutter spectrum tends to fill in around and between the lines as the wind picks up. For horizontal polarization (which is possible only over sky-wave paths), the cross section is much smaller and shows the characteristic fourth-power decay with decreasing grazing angle. For these HF wavelengths of tens of meters, the sea is relatively flat and the scattering laws are simple. A discussion of HF radar may be found in Chap. 24.

At the other end of the potentially useful radar spectrum, in the millimeter-wave band, the few published measurements of radar clutter lead to the conclusion that millimeter-wave backscatter behaves in much the same manner as backscatter at the lower microwave frequencies. This was suggested by the K-band curves shown in Fig. 13.5 for moderate wind speeds and further supported by some older shipboard data at frequencies between 9 and 49 GHz.[45] It should be noted that clutter signal paths lie close to the sea surface, where the atmospheric and water-vapor densities are highest. This means that at these higher frequencies the clutter signal will be strongly affected by the atmospheric absorption effects described in Chap. 2, and consequently the surface-related cross section inferred from the received signal strength in any given measurement will depend upon the path length. Moreover, the role of sea *spray* in both scattering and absorption will certainly be more important than at the lower microwave frequencies.

It is difficult to find clutter data at frequencies above K_a band, although *H*- and *V*-polarized returns at 95 GHz at a grazing angle of 1° were reported, both with values of close to -40 dB.[46,47] Interestingly, this is just the cross section

measured at this angle by a number of investigators at X band (see Ref. 12), showing a similarity between the returns at these two widely spaced frequencies. However, at lower frequencies, at L band and below, there is a noticeable tendency for the cross section to fall off with decreasing grazing angles below about 15 to 20°.

The Spectrum of Sea Clutter. The scattering features producing sea clutter are associated with a surface subject to several types of motion. The features may themselves be moving with small group or phase velocities over this surface while the surface, in turn, is moved by the orbital velocities of the larger waves passing across it. Or the scatterers might be detached from the underlying surface, as in the plumes emitted at the crests of breaking waves, and move at speeds much greater than the orbital speeds.[48] At higher radar frequencies and in strong winds, the possibility of scattering from spray, advected by the wind field above the surface, must be considered. All this complex motion shows up in a doppler shift imparted to the scattered electromagnetic wave.

Surprisingly few measurements of microwave clutter spectra for real seas have been reported in the literature, and those few that exist can be separated into aircraft measurements of the spectral shape alone[49,50] and fixed-site shore measurements showing a shift in the spectral peak.[51,52] All these studies were performed at relatively low grazing angles (less than 10°), although Valenzuela and Laing include a few measurements up to 30°. Other measurements of sea clutter spectra include those made at much lower frequencies in the HF band, as described in the last section, those made under artificial conditions in the wave tanks,[53] whose application to real-sea conditions is uncertain, and other fixed-site measurements at high resolution and short averaging times, to be discussed later.

As it turns out, microwave sea clutter spectra have a rather simple form at the lower grazing angles. Figure 13.14 illustrates typical spectral behavior at the two polarizations, based on data collected by Pidgeon for C-band clutter looking upwind at a few degrees grazing.[51] The peak frequency of the upwind spectrum appears to be determined by the orbital velocity of the largest sea waves, plus a wind-dependent velocity increment containing, but not entirely explained by, wind-induced surface currents. The orbital velocity V_{orb} is taken to be that of the major waves and is obtained in terms of significant height $H_{1/3}$ and period "T" from the expression

$$V_{orb} = \pi H_{1/3} / ``T" = 0.1U \qquad (13.10)$$

The approximate dependence on wind speed U was found by substituting $H_{1/3} = 3h_{rms}$ from Eq. (13.6), assuming a *fully developed sea*, and T from Eq. (13.5). To this there must be added a wind-drift velocity of about 3 percent of U and a fixed *scatterer* velocity (which appears to be about 0.25 m/s in the X- and C-band measurements[51,52,54]). Summing these components yields the virtual doppler velocity at the peak of the clutter spectrum for the particular case of a *vertically polarized, X- or C-band* radar looking *upwind* at *low grazing angles*:

$$V_{vir} \approx 0.25 + 0.13U \qquad m/s \qquad (13.11)$$

(As noted earlier, care must be taken whenever wind speed is used to parameterize a process that depends on waveheight. There is an unambiguous relation only for a fully developed sea in the absence of swell.) The remaining properties

of the clutter spectrum can now be discussed in terms of V_{vir}. For example, the spectral peak for *horizontal* polarization follows a similar linear dependence on U, only with a coefficient lying somewhere between 0.17 and 0.20, as reflected by the sketch in Fig. 13.14. The (half-power) width of the clutter spectrum is roughly the same for both polarizations and is equal approximately to the upwind vertical velocity given in Eq. (13.11). For look directions away from upwind, the peak doppler follows a cosine dependence very closely, going to zero at crosswind aspects and turning negative downwind. Interestingly, the *bandwidth* of the spectrum remains relatively constant.
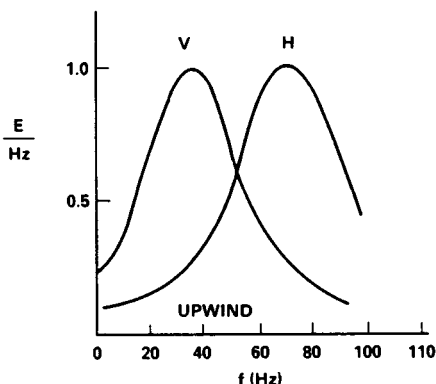
FIG. 13.14 Qualitative behavior of doppler spectra of sea clutter looking upwind at low grazing angles. (*Based on C-band measurements by Pidgeon.*[51])

The details of the clutter spectrum show little dependence on either the radar frequency or the grazing angle, at least for angles less than about 10°. In reviewing the results of measurements at four frequencies—UHF, L, C, and X bands—Valenzuela and Laing[50] noted a relatively weak tendency of clutter bandwidth to decrease with increases in frequency between the UHF and X bands and grazing angles between 5 and 30°. Since both of these variations entail a decrease in the size of the radar footprint on the surface, they might be due to a dependence on resolution cell dimensions, although the other workers found that the pulse length had little effect on clutter bandwidth for values between about 0.25 and 10 μs. The equivalence between time and space averaging in sea clutter measurements was discussed earlier, and in the case of clutter spectra the averaging times were all quite long (of the order of 10 to 20 min), which should be sufficient to stabilize the spectra for almost any resolution cell size.

Spectra obtained with *short* averaging times disclose something of the origins of the clutter spectrum. Figure 13.15 is a sequence of 0.2-s spectra obtained by Keller et al.[55] with a coherent vertically polarized X-band radar operating at a grazing angle of 35° and a resolution cell size of about 10 m². The zero-doppler reference in this figure was located arbitrarily at −16 Hz, and because of the high grazing angle the effects of both senses of the orbital velocities are seen, unlike the low-angle shadowed surface results shown in Fig. 13.14. The spread along each line is due to the small-scale wave motions on the surface, while the larger meanders are induced by the orbital velocities of the large waves moving

through the measurement cell. The wind speed was 16.5 m/s, and a doppler shift of 100 Hz corresponds to a radial velocity of 1.6 m/s. The average clutter spectrum expected for this wind speed and grazing angle, with bandwidth obtained from Eq. (13.11), is sketched on the figure. The large spectral spike appearing in the center of the display is no doubt due to a wave breaking in or close to the measurement cell. The doppler velocity for this spike suggests a peak scatterer velocity of about half the wind speed, which would correspond to the group velocity of the longest waves on the surface. Although such events are relatively rare in a fixed area of 10 m^2, they should occur quite frequently within a large surveillance cell and might often have large scattering cross sections associated with them.
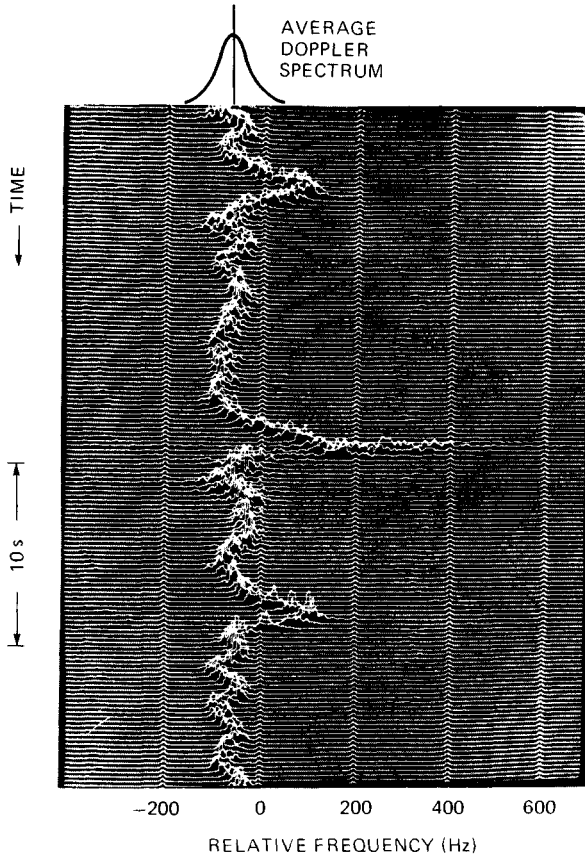
FIG. 13.15   Short-time averaged doppler spectra at X band for an intermediate grazing angle of 35°; spectra computed at 0.2-s intervals. (*From Keller et al.*[55])

Other Effects on Sea Clutter

Rain.   Evidence of the effect of rain on sea clutter is mainly anecdotal; for example, radar operators report that sea clutter tends to decrease when it starts to rain. However, there has been little in the way of reliable, quantitative experi-

mental information about the interaction between rain and wind-driven sea scatter. Laboratory measurements by Moore et al.[56] with artificial "rain" suggested that for light winds the backscatter level increased with the rain rate, while for heavy winds rain made little difference. In measurements in natural rain over Chesapeake Bay, Hansen[57] found that even a light rain (2 mm/h) changes the spectral character of sea clutter at moderate wind speeds (6 m/s) by introducing a significant high-frequency component. He also found some evidence in support of the radar operators, at least for the low grazing angles and horizontal polarizations with which most shipboard radars operate. Figure 13.16 compares the correlation function of sea clutter (X band, low grazing angle, H polarization) with and without rain for a 15-kn wind speed and a rain rate of 4 mm/h. The sharp decrease in correlation time in the presence of rain reflects the broadening of the clutter spectrum. Beyond this, there is virtually no quantitative information about the effect of rain on existing sea clutter.
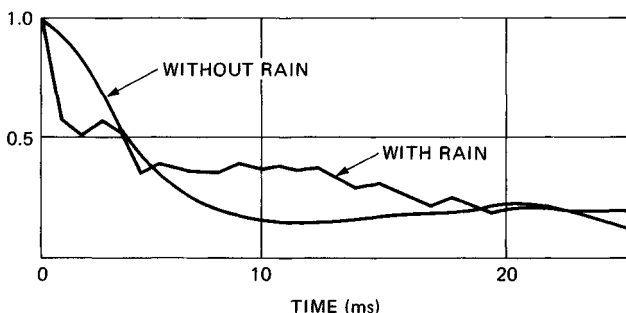
FIG. 13.16   Effect of rain on the correlation function of wind-driven sea clutter; X band, horizontal polarization, wind speed 15 kn, rain rate 4 mm/h. (*From Hansen.*[57])

The production of sea clutter by rain falling on a "calm" surface in the absence of wind was also investigated by Hansen, with the results shown in Fig. 13.17.[57] A high-resolution X-band radar (40-ns pulse, 1° beamwidth), operating at a grazing angle of about 3°, viewed the backscatter from a fixed spot on the windless surface of Chesapeake Bay as the rain steadily increased from 0 to 6 mm/h. The cross sections for vertical and horizontal polarizations were quite different for low rain rates but tended to merge at a rain rate of about 6 mm/h. The magnitude of this *splash* cross section rose to a σ^0 of about -40 dB, corresponding to wind-induced cross sections at this grazing angle for winds of about 10 kn. Further laboratory[58] and theoretical[59] studies have shown that the major scattering feature is the vertical *stalk* that emerges shortly after drop impact. Moreover, these studies suggest that the V-polarized returns from raindrop splashes should be relatively insensitive to the rain rate, while the H-polarized returns should show a strong dependence on both the rain rate and the drop-size distribution.

Propagation Effects.   Another topic in sea clutter that has been largely unexplored is the role played by *propagation effects* within the atmospheric boundary layer lying over the sea surface. The effects of atmospheric absorption have been noted above in connection with millimeter-wave clutter. However, at very low grazing angles the ray paths joining the radar to the surface become very sensi-
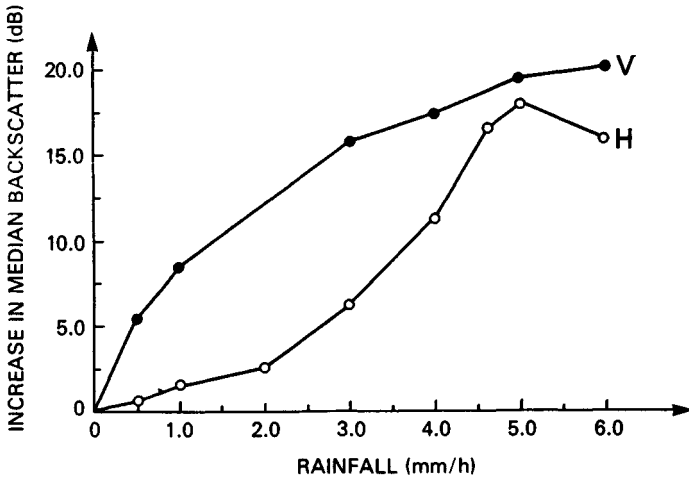
FIG. 13.17 Sea clutter produced by rain splashes alone on a calm surface. (*From Hansen.*[57])

tive to refractive inhomogeneities in the atmospheric boundary layer. Over distances approaching and beyond the conventional optical horizon, such perturbations could produce strong focus-defocus variations along the illumination profile[60] or a general rise in the local grazing angle.[38] Figure 13.18 gives an experimental example of the effect of *ducting* on very low angle sea clutter.[42] Since the grazing angle given as the abscissa is actually a plot of inverse range, the lifting of the cross section by ducting over an order-of-magnitude span of ranges is very likely due to a rise in the mean grazing angle produced by refraction in the evaporative layer.[38] Such effects should be suspected whenever the radar propagation path extends beyond the *optical* horizon.
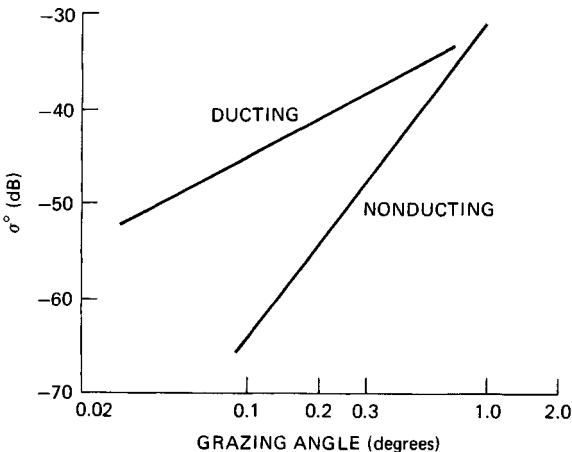
FIG. 13.18 Effect of ducting on low-angle clutter; wind speed about 10 kn.

Shadowing. The possibility of shadowing must be seriously considered whenever the sea is viewed at grazing angles smaller than the rms slope angle of the sea surface. Some examples were discussed earlier in connection with the behavior of sea clutter at low grazing angles in Fig. 13.13. In fact, the sharp falloff of the *nonducting* data in Fig. 13.18 gives further evidence of the *threshold shadowing* mentioned there. However, the common idea of shadowing, along with all existing theories of a shadowed surface, rests on the geometrical optics concept of a sharp transition between light and darkness. By considering the implications of diffraction at the wave peaks, it is possible to determine the domain of radar frequencies and wind speeds over which the concepts of geometrical optics may be applied. This was done by Wetzel,[12] who showed in detail how diffraction, rather than shadowing, controls propagation into and out of the troughs of the waves under many of the usual frequencies and wind speeds encountered in practical radar operations at low grazing angles. For example, shadowing will take place at K_a band for any winds above 15 kn, yet will hardly ever occur at L-band frequencies.

Contaminants. The idea of pouring oil on troubled waters is a familiar one: the angry surface will smooth and subside. In another age, the survival-gear locker of every sailing ship would contain a bottle of oil to quiet the sea in a storm. Although the effectiveness of this procedure has always been somewhat controversial, there is no question that oil can produce a *slick* of smooth water at relatively low wind speeds. In fact, biological oils, produced by bacteria, algae, and plankton, can be found everywhere on the world's oceans and form natural slicks in those regions that combine the greatest oil concentration with the lowest wind speeds, e.g., close to continental shorelines.[61] Human-made contaminants can, of course, have the same effect. A layer of oil only 1 molecule thick will significantly affect the ability of the surface to support wave motions, but this layer must be continuous. The adjacent molecules then sense each other and form a film that is resistant to horizontal compression. The surface elasticity is changed, a type of longitudinal viscosity is introduced, and the surface becomes stabilized against the growth of short waves up to several inches in length.[62,63]

To the extent that radar sea clutter is produced by small-scale surface roughness (at grazing angles less than about 80°), the presence of oil on the surface should lead to a measurable decrease in clutter cross section. But, as noted above, the reduction of small wave motions requires the existence of a *continuous* monolayer; slick formation is a go–no-go process, and so slicks will tend to have relatively sharp boundaries. In operating the NRL 4FR system as a synthetic aperture radar to obtain images of the slicks produced by oil spills, Guinard found that the slicks were well defined, that it took very little oil to maintain a visible slick, that vertical polarization provided much greater contrast than did horizontal, and that the slicks were quenched by winds and currents.[64] Although signal strength was not recorded in this imaging experiment, later measurements at X and L bands by others[65] indicated that at the higher grazing angles (about 45°) the clutter reduction produced by the types of oil occurring in *natural* slicks was rather small, of the order of a few tenths of a percent. Since slicks are dispersed by the wind and associated wave action at wind speeds greater than about 10 kn, the effect of natural slicks on clutter may not be clear because they tend to occur in the regime of low wind speeds where the sea surface is already ill defined.

The celebrated *sun glitter* measurements by Cox and Munk[66] gave a quantitative measure of the effect of contaminants on the surface slopes in open water, showing that the wind-generated component of the rms slope of "oiled" waters is significantly smaller than that of "clean" water. The heavy human-made oils used in their experiment were effective in suppressing small-scale waves over a range of wind

speeds well beyond those which would normally disperse the lighter natural oils; so the effect of oil spills on sea clutter should be expected to extend to the higher wind speeds. In fact, at these higher wind speeds the depression of radar backscatter by such oils at X and K_a bands can reach 10 to 20 dB at intermediate grazing angles between 30 and 60°.[67,68]

Currents. The most obvious effect of a current on sea clutter would be a shift in the peak of the doppler spectrum, similar to the contribution of the 3 percent wind-drift current mentioned in connection with Eq. (13.8). Another effect is related to the fact that the excitation of the surface-wave system depends on the *apparent* wind; so there can be significant differences in waveheight according as the wind is blowing with or against the current. According to Eq. (13.6), waveheight is proportional to the square of the wind speed; so in the Gulf Stream, for example, with a current of 4 kn flowing north, a 15-kn northerly blowing against the current will raise a sea 3 times as high as a 15-kn southerly blowing with the current. Even with no wind the presence of strong current shears can produce highly agitated surfaces. Shipboard observers have reported bands of roaring breakers passing by on an otherwise-smooth surface, presumably produced by powerful surface-current shears associated with large-amplitude internal waves.[69] In a more subtle way, currents are held responsible for synthetic aperture radar (SAR) images which contain the expression of bottom topography in shallow waters.[70] In each of the examples cited above, the current produces a change in the surface roughness, which can be expected to give rise to a change in sea clutter cross section.

Combined Effects. Some idea of the complexity within a clutter scene due to *other effects* may be obtained from Fig. 13.19, which shows a digitized PPI display of clutter in the Sargasso Sea, under light wind conditions and near a thermal oceanic front.[71] Although all possible contributing effects were not identified, observers noted the presence of human-made detritus organized by the currents at the edges of the thermal front, slicks probably of both natural and artificial origin, fronds of seaweed close to the surface, and the presence of light and variable winds. The dynamic range of the digitized (false-color) PPI was 30 dB, and some of the clutter contrasts, across what are obviously extremely sharp boundaries, were almost this great.

13.4 THEORIES OF SEA CLUTTER

The sea surface is so rich in potential scattering structures that in seeking to understand the phenomenology experimenters and theorists alike have proposed and have found support for almost any imaginable model. However, aside from providing an intellectual basis for "understanding" sea clutter phenomena, a theory of sea clutter should serve the practical purpose of providing accurate a priori predictions of all aspects of clutter behavior under all possible environmental conditions. At present, the theory of sea clutter does neither of these tasks very well and must be thought of as a book with the final chapters still to be written.

Before discussing the current theories of sea clutter, it is important to distinguish them from other so-called sea clutter models that are designed to provide a predictive capability. Some such models organize large quantities of empirical data by finding a multiple linear-regression formula relating the clutter cross section to a variety of parameters, such as grazing angle, wind speed, frequency, etc., all measured concurrently.[1,40] Even a multiparameter matrix tabulation

CHAPTER 14

CW AND FM RADAR

William K. Saunders
Formerly of Harry Diamond Laboratories

14.1 INTRODUCTION AND ADVANTAGES OF CW

The usual concept of radar is a pulse of energy being transmitted and its round-trip time being measured to determine target range. Fairly early it was recognized that a continuous wave (CW) would have advantages in the measurement of the doppler effect and that, by some sort of coding, it could measure range as well.

Among the advantages of CW radar are its apparent simplicity and the potential minimal spread in the transmitted spectrum. The latter reduces the radio interference problem and simplifies all microwave preselection, filtering, etc. A corollary is the ease in the handling of the received waveform, as minimum bandwidth is required in the IF circuitry. Also, with solid-state components peak power is usually little greater than average power; CW then becomes additionally attractive, particularly if the required average power is within the capability of a single solid-state component.

Another very apparent advantage of CW (unmodulated) radar is its ability to handle, without velocity ambiguity, targets at any range and with nearly any conceivable velocity. With pulse doppler or moving-target indication (MTI) radar this advantage is bought only with considerable complexity. An unmodulated CW radar is, of course, fundamentally incapable of measuring range itself. A modulated CW radar has all the unwanted compromises, such as between ambiguous range and ambiguous doppler, that are the bane of coherent pulsed radars. (See Chaps. 15 to 17.)

Since CW radar generates its required average power with minimal peak power and may have extremely great frequency diversity, it is less readily detectable by intercepting equipment. This is particularly true when the intercepting receiver depends on a pulse structure to produce either an audio or a visual indication. Police radars and certain low-level personnel detection radars have this element of surprise. Even a chopper receiver, in the simplest video version, may not give warning at sufficient range to prevent consequences.

It should not be concluded that CW radar has all these advantages without corresponding disadvantages. Spillover, the direct leakage of the transmitter and its accompanying noise into the receiver, is a severe problem. This was recognized fairly early by Hansen[1] and Varian[2] and others. In fact, the history of CW

radar shows a continuous attempt to devise ingenious methods to achieve the desired sensitivity in spite of spillover.

14.2 DOPPLER EFFECT

Complete descriptions of the doppler phenomenon are given in most physics texts, and a discussion emphasizing radar is to be found in Skolnik[3] (chap. 3, pp. 68–69).

When the radar transmitter and receiver are colocated, the doppler frequency f_d obeys the relationship

$$f_d = \frac{2v_r f_T}{c}$$

where f_T = transmitted frequency
c = velocity of propagation, 3×10^8 m/s
v_r = relative (or radial) velocity of target with respect to radar

Thus when the relative velocity is 300 m/s, the doppler frequency at X band is about 20 kHz. Alternatively, 1 ft/s corresponds to 20 Hz at this frequency. Scaling is a convenient way to handle other microwave frequencies or velocities.

As in a pulse radar, a CW radar that uses a rapid rate of frequency modulation, in order to sample the doppler, must have this rate twice the highest expected doppler frequency if an unambiguous reading is to be obtained. If the rate falls below the doppler frequency itself, there are problems of blind speeds as well as ambiguities. (A blind speed is defined as a relative velocity that renders a target invisible.) This will be discussed in more detail in Sec. 14.10; see also Sec. 15.3.

14.3 UNMODULATED CW RADAR

Spectral Spreading. The following discussion concerns the larger CW radars used for target illumination in semiactive systems, for acquisition, or for warning. A highly simplified diagram is shown in Fig. 14.1. Insofar as the primary operation of the equipment is concerned, the transmitted signal may be considered an unmodulated CW, although small-amplitude amplitude modulation (AM) or small-deviation frequency modulation (FM) is sometimes employed to provide coding or to give a rough indication of range. The modulation frequency is chosen to lie above the doppler band of interest, and the circuitry is designed to degrade the basic noise performance as little as possible.

A spectrum-spreading problem is also posed by conical scan. In this case, the scan frequency will lie, it is hoped, below any doppler of interest, with the result that the conical-scan frequency will appear as small-amplitude sidebands on the doppler frequency when it is recovered in the equipment. In the material that follows, these secondary issues will be largely ignored, and an unmodulated CW transmission plus a receiver that introduces no intentional modulation will be assumed.
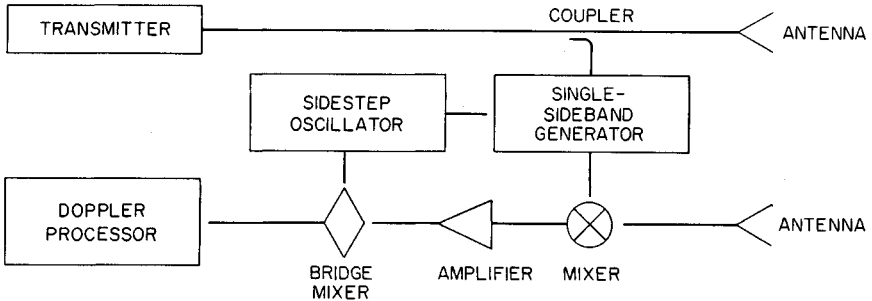
FIG. 14.1   Basic diagram of a CW radar.

Noise in Sources.   The primary noise problem is in the microwave source itself. All klystrons, triodes, solid-state sources, etc., generate measurable noise sidebands in a band extending beyond any conceivable doppler frequency. Unless the source is unacceptably bad, these noise sidebands may be subdivided into pairs: those whose phase relationship to each other and the main carrier line is such as to represent amplitude modulation and those corresponding to a small index of frequency or phase modulation. The AM components, offset a given frequency from the carrier, are usually many decibels below the corresponding FM components. Moreover, balanced mixers, limiters, and other design techniques may be used to suppress AM noise. FM noise therefore is usually of greater concern in CW radar.

The FM noise of a good klystron amplifier driven by a klystron oscillator having either an active or a passive FM stabilizer is about 133 dB below the carrier in a 1-Hz bandwidth* 10 kHz removed from the carrier. The noise power decreases approximately as $1/f^2$ at larger offsets. The corresponding AM noise is 150 to 160 dB below the carrier. As the local-oscillator signal of a CW radar is usually derived from the transmitter, it too will show comparable amounts of noise. With noise this low relative to the carrier there is no concern that a significant portion of the target's energy will be lost in any practical filtering operation. What is of concern is the energy of the noise components carried into the receiver on the unwanted spillover and clutter signals. Moreover, should vibrations cause variations in the length or lengths of the spillover path the problem is more severe, since such effects introduce spectral lines that may fall directly into the doppler band of interest.

Noise from Clutter.   Clutter, unwanted reflections from the ground, rain, etc., reflects the transmitted power and its noise sidebands back to the receiver. Suppose that in the foreground there is an industrial area having a clutter cross section of 0.1 m² per square meter of ground illuminated. Consider only an area located 2½ to 3½ km from the radar ($R \approx 35$ dB, where the range R is in decibels relative to 1 m) and illuminated with an antenna having 0.1 rad of beamwidth. There is a reflecting cross section σ of approximately

---

*In this chapter a 1-Hz bandwidth has been chosen as the reference. There has been no consistent practice in the literature, since it has been customary to use either the bandwidth of the system in question or that of the measuring equipment.

$$3 \times 10^3 \times 0.1 \times 10^3 \times 0.1 = 3.0 \times 10^4\,\mathrm{m}^2 \text{ or } 45\text{ dB}$$

Suppose that there is present a target at a greater range and that a +22-dB signal-to-noise ratio in a 1-kHz bandwidth is needed to have a suitable probability of detection with an acceptable probability of false alarm (this includes allowance for a +10-dB receiver noise figure). If the target produces a 10-kHz doppler signal, the noise of the clutter signal must not exceed −144 dBm* + 22 dB = −122 dBm at the 10-kHz offset. If a transmitter is assumed in which the noise sidebands in a 1-kHz bandwidth are 103 dB below the carrier, the clutter signal must not exceed −19 dBm. Assume that the antenna gain is +30 dB and the transmitted power +60 dBm at X band. We have

$$\text{(power)} \quad (G^2) \quad (\lambda^2) \quad (64\pi^3) \quad (\sigma) \quad (R^4)$$

$$-19 \text{ dBm} > +60 \text{ dBm} + 60 \text{ dB} - 30 \text{ dB} - 33 \text{ dB} + 45 \text{ dB} - 140 \text{ dB}$$

$$> -38 \text{ dBm, or a } 19 \text{ dB favorable margin}$$

It should be noted that we have assumed a very quiet transmitter and clutter centered about a 3-km range. (This ignores correlation; see below.)

It is convenient to express the FM-noise sidebands on the transmitter in another manner. Consider a single modulating frequency, or line, f_m in the noise having a peak frequency deviation ΔF_p. By the frequency-modulation formulas, the carrier has a peak amplitude of $J_0(\Delta F_p/f_m)$, and each of the nearest sidebands a peak amplitude of $J_1(\Delta F_p/f_m)$.[4] If the arguments are small, as they must be for our computations, the Bessel functions may be approximated by

$$J_0(X) \sim 1$$
$$J_1(X) \sim \frac{X}{2}$$

Hence the power ratio between the carrier and one of the first harmonic sidebands is $\Delta F_p^2/4f_m^2$. For the ratio of the power in both sidebands to that in the carrier (the quantity usually of interest),

$$\text{FM noise power} = \frac{\Delta F_p^2}{2f_m^2} = \frac{\Delta F_{\text{rms}}^2}{f_m^2} \tag{14.1}$$

A 1-Hz peak deviation at a 10-kHz rate represents a double-sideband noise ratio of $\frac{1}{2}(\frac{1}{10^4})^2$ or −83 dB with respect to the carrier. The −103-dB (double sideband at 10 kHz) transmitter used as an example above has a peak deviation of 0.1 Hz. These numbers are all equivalent only when referred to a particular bandwidth, 1 kHz in this case. (The description in hertz is convenient only when the noise is expressed in the bandwidth of interest in a particular radar.)

The concept of a peak signal is properly associated only with a sine wave. With random noise the rms description is more meaningful. However, a noise power at a given frequency and bandwidth is equivalent to that which would be produced by a sine wave having a certain peak or rms value.[5]

To carry the computations further, the correlation effect must be discussed.

---

*Thermal noise in a 1-kHz band at 20°C.

When a transmitter is producing FM noise, it may be thought to be modulating in frequency at various rates and small deviations. Consider, for example, a particular one of these modulating frequencies. If it is a low frequency and the delay associated with the spillover or clutter is short, the returning signal finds the carrier at nearly the same frequency that it had at the time of transmission; that is, the decorrelation is small. Higher frequencies in the noise spectrum have greater decorrelation. Moreover, the effect is periodic with range: For any given sinusoidal modulating frequency the FM noise produced will increase as a function of range out to a given range and will then decrease. The zeros occur at the ranges $R = nc/2f_m$, where f_m is the frequency of the sinusoidal modulating component, c is the velocity of light, and n is any integer.

However, in general, one deals with noise rather than a sinusoidal component. For this reason, the discrete zeros indicated are seldom of interest, and for ease in the computations the signal is assumed to be decorrelated at a frequency f_l approaching $f_l = c/8R$. Some frequencies higher than f_l cause no problems at particular ranges, but nearby ones do so. Moreover, as the formulas below will show, the deviation of the recovered signal can be twice as great as that of the transmitter. (The returning signal may be swinging up while the transmitter is swinging down.)

The peak voltage of the first harmonic sideband in the IF spectrum of an FM signal mixed with itself after a time delay $T = 2R/c$ is[6]

$$v_{p1} = J_1\left(2\frac{\Delta F_p}{f_m} \sin \pi f_m T\right)$$

and the peak voltage of the carrier is

$$v_{p0} = J_0\left(2\frac{\Delta F_p}{f_m} \sin \pi f_m T\right)$$

In both formulas ΔF_p is the peak frequency deviation of the carrier. As before, $J_1(X) \sim x/2$, $J_0(X) \sim 1$, $X<1$. Hence the ratio of the power in a single sideband to that in the carrier is

$$\frac{P_s}{P_c} = \left(\frac{\Delta F_p}{f_m} \sin \pi f_m T\right)^2$$

and in the pair of sidebands

$$\frac{P_{2s}}{P_c} = 2\left(\frac{\Delta F_p}{f_m} \sin \pi f_m T\right)^2 \tag{14.2}$$

The maximum value of this is $2(\Delta F_p/f_m)^2$, which is in agreement with the maximum deviation of the IF being twice as large as that of the transmitted frequency [Eq. (14.1)].

For smaller values of $f_m T$ the double-sideband power ratio is

$$\frac{P_{2s}}{P_c} = 2(\pi \Delta F_p T)^2 \qquad \pi f_m T < 1 \tag{14.3}$$

This is an interesting formula as it shows that when ΔF_p is constant, as it is with many klystrons, the correlated noise power is independent of frequency and directly dependent on range.

A convenient curve which gives the ratio of noise at the receiver to measured noise on the transmitter (Fig. 14.2) is based on the approximations of formulas (14.2) and (14.3). The dotted portion of the curve reflects only the approximations formula (14.2).
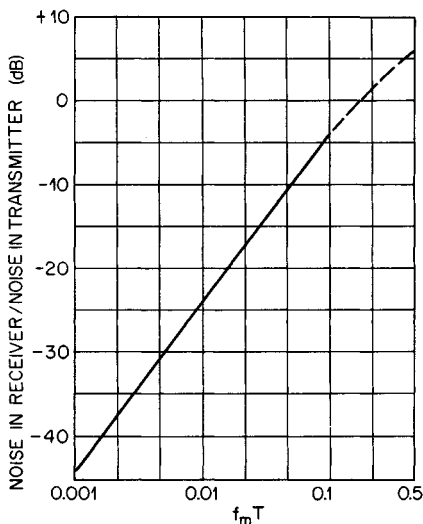
FIG. 14.2 Noise suppression by the correlation effect.

The problem started above can now be completed. The center of the clutter was taken at 3 km or a T of 2×10^{-5} s. The frequency of interest was 10^4 Hz. Hence $f_m T = 0.2$, which is beyond the region of noise correlation.

An unmodulated CW radar must contend with clutter almost down to zero range. Without the correlation effect, this would generally be impossible. For a given antenna beam the width of the illuminated clutter area decreases as R but there is a $1/R^4$ in the radar equation. The result is that the clutter return varies as $1/R^3$. The correlation effect shows that for a fixed noise frequency the correlated noise sidebands decrease as R^2 [Eq. (14.3), with $T = 2R/c$]. Hence there is an apparent rate of increase of $1/R$. The integral representing the clutter power appears to diverge, but two factors so far ignored have a decisive influence at very short ranges. The first is that the intersection of a beam emitted from an antenna of finite height and the earth is the interior of a hyperbola and not a sector, as implied above. The second is that at close ranges the clutter is in the Fresnel region of the antenna and the far-field gain formula no longer applies. In a more careful analysis, by using either of these factors, the integral may be shown to be convergent.

Shreve[7] has derived a formula for the double-sideband noise power. He took the boundary of the Fresnel region $R_F = D^2/\lambda$ as the lower limit of the integral.

His formula for the above parameters yields a value of -117 dBm for the correlated noise power from the clutter. This is for the extreme case of a single antenna at exact ground level looking into very severe clutter (0.1 m²/m²).

A more practical way to look at the problem is to note that clutter from very short ranges and spillover are almost equivalent phenomena. For a ground-based CW radar to operate at maximum sensitivity, two antennas must be employed; this reduces both the spillover and the near-in clutter since no close-in point can be in the main lobes of both transmitting and receiving beams. Moreover, as described below, spillover cancellation (and hence near-in clutter cancellation) is usually employed.

The discussion above assumes that the local oscillator employed in the radar is either derived from the transmitter or locked to it with a servo which has a frequency response sufficiently high to cover the doppler and noise band of interest.

Microphonism. Microphonism can cause the appearance of additional noise sidebands on the spillover and occasionally on the clutter signals. If the structures are sufficiently massive, the microphonism is greatest at the lower frequencies, where it can be counteracted by a feedthrough servo. To this end, however, it is most important that microwave components employed in the feedthrough nulling as well as in the remainder of the microwave circuitry be as rigid as possible.[8] It is customary to use a milled-block form of construction. In the rare cases where a single antenna plus duplexer or a pair of nested antennas has been used in an airborne high-power CW radar, the mechanical design problems have been all but insurmountable. Even in a ground-based radar, fans, drive motors, motor-generator sets, rotary joints, cavitation in the coolant, etc., are very troublesome.

Scanning and Target Properties. In addition to the spectral spreading caused by transmitter noise and by microphonism, there is a spreading of the CW energy by the target and by the scanning of the antenna. Generally, the spreading by even a rapidly scintillating aircraft target does not produce appreciable energy outside a normal doppler frequency bandwidth. The filter is usually set by the acquisition problem or the time on target rather than by the intrinsic character of the return signal. Rapid antenna scan, however, can cause an appreciable broadening of the spectrum produced by the clutter. Were it not for the particular shape of the typical antenna beam, the transients produced by clutter while scanning would be far more serious.

An approximate analysis assumes a gaussian two-way gain for the antenna, $G^2 = e^{-2.776\theta^2/\theta_B^2}$, where θ is measured from the axis of the beam and θ_B is the beamwidth between the half-power points of the antenna. We shall discuss a two-way pattern down 3 dB at $\pm\frac{1}{2}°$ ($\theta_B = 1°$). If the antenna scans 180° a second, we shall need the Fourier transform of e^{-at^2} with $a = 9 \times 10^4$. This has the form $Ae^{-\omega^2/3.6\times10^5}$, which is down to $\frac{1}{1000}$ (60 dB) of its peak when $\omega^2/(3.6 \times 10^5) = 6.9$, $\omega \approx 1575$, and $f \approx 250$ Hz.

Actual antenna patterns produce somewhat less favorable transients than the gaussian shape. Limiting in the receiver is equivalent to altering the shape of the beam.[9] For any antenna pattern there is a definite limitation on the scanning speed of a narrow-beam antenna. Actually, mechanical limitations usually prevent trouble except with the very slowest targets, but with nonmechanical scanning methods degradations may occur.

14.4 *SOURCES*

Master Oscillator Power Amplifier (MOPA) Chains.   Requirements peculiar to CW radar are the use of extremely quiet tubes throughout the transmitter chain, very quiet power supplies, and, often, stabilization to reduce the total noise of the system. In theory, any of the methods for measuring FM and AM noise to be discussed in Sec. 14.5 might be modified to produce a noise-quieting servo. Practical considerations have resulted in a wide variety of additional schemes. The simplest is the introduction of a high-Q cavity between a klystron driver and the power amplifiers. Q's of 20,000 to 100,000 are normally employed. The action of the cavity is primarily that of an additional reactive element directly in parallel with the cavity in the klystron. With a high-quality reflex klystron having an FM noise 110 dB below the carrier in a 1-Hz band spaced 10 kHz from the carrier, use of the high-Q cavity as a passive stabilizer reduces the corresponding FM noise 130 to 135 dB below the carrier. The cost is a power loss of about 11 dB. The 20 to 25 dB noise improvement is obtained at most frequencies of interest. No noticeable improvement is made in the AM noise level at doppler frequencies by this technique.

It should be remembered that this results only in a stable driver, any noise generated in the power amplifier being unaffected. And, as noted above, unless the local oscillator is generated from the output of the power amplifier, which, of course, cannot be done in a pulse doppler system, this noise is uncorrelated. Fortunately, good power amplifiers driven by highly regulated supplies add extremely small amounts of excess, or additive, noise. (See Fig. 14.8.)

It is to be noted that the illuminator for the basic Hawk surface-to-air missile system used a magnetron as the transmitter rather than a MOPA chain. Cost, availability, lower weight, and lower high-voltage requirements were all factors in the choice. Later versions of Hawk use the inherently quieter klystrons. Banks[10] provides a definitive overview of the Hawk illuminator including both the noise degeneration loop (feedthrough nulling) and the transmitter stabilization microwave circuitry. An interesting feature of the latter is a spherical cavity that is far stiffer under vibration than the usual cylindrical cavities.

Active Stabilization.   All the schemes for active stabilization on a MOPA chain depend on the use of a high-Q cavity as the reference element. The cavity must be isolated from the tube so that it functions as a measuring device without introducing the pulling that results from the frequency dependence of its susceptance.

For reflection and transmission cavities, useful equations adapted from Grauling and Healy[11] are given below. For a matched reflection cavity,

$$\Gamma = \frac{Z - 1}{Z + 1} = \frac{j\delta Q}{1 + j\delta Q} = \frac{j2\delta Q_L}{1 + j2\delta Q_L}$$

where Γ = reflection coefficient
     δ = $(f - f_0)/f_0$
     Q = unloaded Q of cavity
     Q_L = loaded Q of cavity
     Z = normalized impedance looking into cavity

The transmission cavity has similar characteristics except that both the carrier (V_c) and sidebands (V_{sb}) are passed. For a transmission cavity,

$$V_0 = \frac{V_1}{3 + j2\delta Q} = V_c + V_{sb}$$

(Both coupling coefficients have been assumed equal to unity.) V_1 is the input voltage and V_0 the output voltage. With a little algebra, it is seen that the frequency-dependent terms of Γ and V_{sb} are similar in form. In stabilization systems, $f - f_0$ is kept small and

$$\text{Reflection:} \quad \Gamma \approx j2\delta Q$$

$$\text{Transmission:} \quad V_{sb} \approx 2(- 2j\delta Q)$$

One might expect that the stabilization would be equally effective in reducing regardless of the frequency. This ignores two factors: the cavity has only a finite linear range, and larger values of f_m may produce sidebands that lie outside this range; for stability the servo that follows the cavity must have a response that rolls off at the higher frequencies.

The simplest stabilization bridges result directly from the character of the transmission and reaction cavities. The transmission-cavity bridge might have the arrangement of Fig. 14.3.
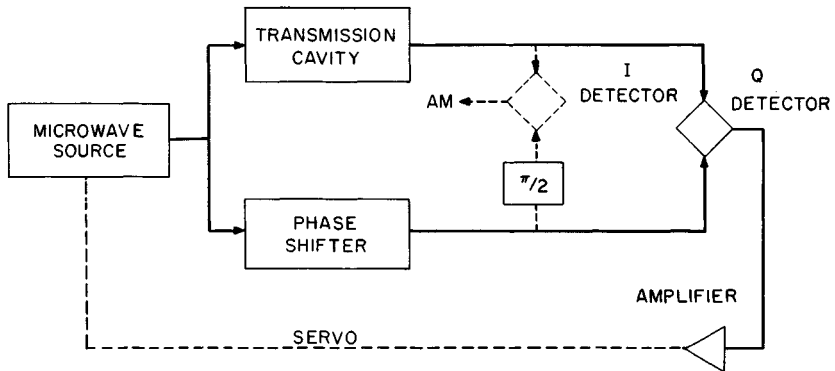
FIG. 14.3 Transmission bridge, video version.

The phase shifter is set so that the Q, or quadrature, detector receives signals in phase quadrature* and, to first order, is sensitive only to FM. Should there be a requirement for AM stabilization, it would normally be only at very low frequencies such as those introduced by the power supplies. Such a requirement could be met easily by adding a $\pi/2$ phase shift and an I (coherent amplitude-sensitive) detector, as indicated by the dotted blocks in Fig. 14.3. The discussion of the servo constants will be postponed until the end of this section.

---

*The most sensitive technique for adjusting the quadrature detector is to introduce intentional AM and null this by adjustment of the phase shifter. Maximizing the response to FM yields a less exact adjustment.

An obvious disadvantage of the transmission-cavity bridge is that the carrier is not suppressed in the microwave circuitry. Since the total input power is limited by fear of crystal damage and of exceeding the linear range in the mixing process, the intelligence signal power at a relatively low level is in competition with the thermal noise generated by the crystals.

The reflection cavity has an advantage in that a sizable portion of the carrier power is absorbed if the transmitter is kept tuned to the frequency of the cavity. This eliminates much of the saturation problem.

A particularly attractive arrangement was proposed by Marsh and Wiltshire[12] (Fig. 14.4). It is the basis of the earliest successful FM noise-measuring instruments and has been employed in stabilization as well.[13] It is the only bridge that removes most of the carrier power to avoid saturation of the mixer crystals. The key to its operation is a balancing element which matches exactly the reflection from the cavity at resonance.
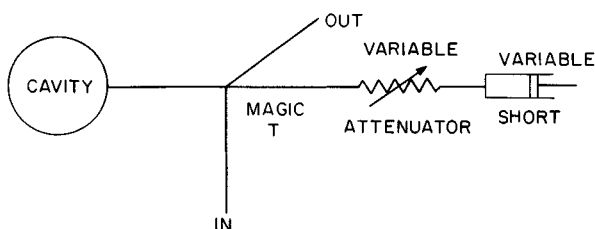
FIG. 14.4   Marsh and Wiltshire microwave bridge.

At resonance the cavity is nearly a perfect absorber, and the residual reflection is tuned out with the variable short and variable attenuator. As the frequency varies, the cavity produces a reactive component which alters the balance. The result, at least for small deviations, is a double-sideband suppressed-carrier signal. With care, the carrier may be suppressed as much as 40 dB with manually or statically balanced bridges and as much as 60 dB if either the cavity or the source is electronically or thermally tuned. The result is that 2 W of power may be handled in the manually tuned version, and up to 1000 W in the servo-tuned equipment. The balance of the circuitry is shown in Fig. 14.5.
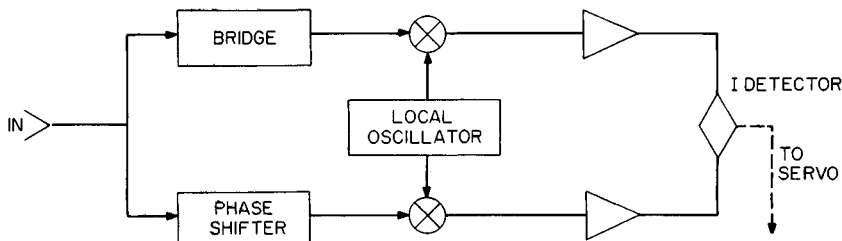
FIG. 14.5   IF circuitry for a Marsh and Wiltshire bridge (shown in Fig. 14.4).

As is seen, a properly phased second sample of the transmitter is processed in a parallel path and serves as the reinjected carrier to recover the sidebands at the *I* detector. Since the lower path is nondispersive and since the sidebands are

small (a requisite for all the above quieting schemes), the large signal may be regarded as an essentially pure carrier in the reinjection process. The LO must be reasonably quiet, and the phase delay of the amplifiers must be matched.

Figure 14.6 gives the servo-loop gain, and Fig. 14.7 shows the three resulting curves: *A*, the FM noise on the free-running oscillator; *B*, the FM noise of the stabilized oscillator; and *C*, the expected theoretical improvement based on the servo gain and the noise analysis.
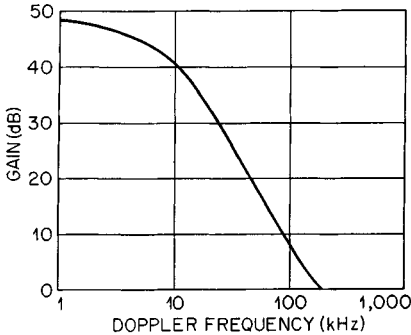
FIG. 14.6   Frequency control-loop gain.
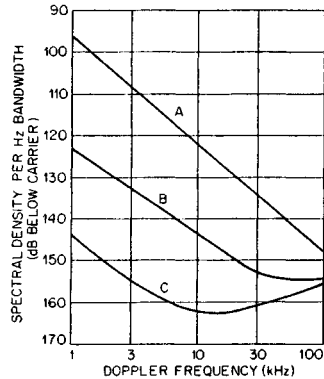
FIG. 14.7   Transmitter noise spectra: *A*, free-running FM; *B*, closed-loop FM; *C*, free-running FM divided by control-loop gain.

For details of the construction and a discussion of the results, see the first edition (1970) of the handbook or Ref. 12.

Stabilization of Power Oscillators.   The active methods described above can be used for the stabilization of power oscillators as well as for the stabilization of drivers. The measurement bridges are unchanged, but servo circuits must be altered to operate at high voltage and, in some cases, supply considerable power, since multicavity klystrons and crossed-field devices such as magnetrons and amplitrons can be modulated only through their high-voltage high-current supplies. Moreover, the fact that these are essentially "stiff" devices imposes more stringent requirements on the design of the servo.

14.5   NOISE MEASUREMENT TECHNIQUE

Two basic types of noise measurement are of interest to the designer: primary-noise measurements to be made on drivers or power oscillators and additive-, or excess-, noise measurements to be made on amplifiers, multipliers, rotary joints, etc.

Although microwave cavities, such as used in the Marsh-Wiltshire bridge, were widely employed at one time, commercial instruments generally avoid them.[14] They accomplish this by comparing, in a phase detector, the source under test with either an external nearly duplicate source or an internal source supplied by a portion of the test equipment. If nearly duplicate sources are used, one is assured that at least one of them (not necessarily always the same one) is

at least 3 dB quieter at each offset frequency than the phase noise indicated by the instrumentation. By using three essentially duplicate sources and measuring the phase noise generated by each pair at all the desired offset frequencies, one derives three sets of measurements. This leads to three equations with three unknowns, and the phase noise of each of the three sources can be derived as a function of frequency. If one of the internal sources supplied by the instrumentation is used, there is a distinct limitation owing to the phase-noise characteristic of that particular internal source. In general, this is the paramount limitation since the noise floor of the phase detection circuitry is usually well below that of the internal or, for that matter, the external reference sources. This assumes that the AM noise on both sources used in the test is well below the phase-modulation noise. The only safe course is first to measure the AM noise on any unknown source with a simple amplitude detector available with the instrumentation.

The instruments can provide a servo voltage to hold the two sources at the same frequency and in quadrature at the phase detector. If the sources are such that neither is readily voltage-tunable, then one source is chosen at a typical IF frequency away from the other, and an IF oscillator is locked, in the mean, to the difference frequency. This technique was originally employed in military test equipment designed to measure noise on radars in the field.[15,16] The instruments provide a wide range of internal frequencies through a combination of synthesizer techniques at the lower frequencies and combs going up to 18 GHz. The latter are created by using the harmonics of a step-recovery diode multiplier.

The signal coming from the phase detector is filtered to remove microwave frequencies and is then amplified in a low-noise baseband amplifier. The resulting phase noise can be measured by any of a variety of methods, including spectrum analyzers and analog wave analyzers. The most accurate and convenient method for the measurement of the lower frequency noise is the fast Fourier transform (FFT). The method is too time-consuming for analysis of the far-out phase noise.

With computer control of all of the components of the test equipment almost any desired measurement can be made, adjusted for filter shape, and printed out. There is even an option to remove spurs (spurious frequencies), occurring during the measurements, from the calculations and from the data. All this comes at a considerable cost, not all of which is monetary. One's ability to understand any laboratory technique and its inherent reliability are usually inversely related to complexity. For example, given enough equipment, each generating a plethora of internal signals, one is almost guaranteed spurs in any sensitive measurement. If the ultimate aim is a quiet transmitter, one attributes such spurs to gremlins in the test equipment at one's peril. A well-shielded screen room with a minimum of (well-understood) instrumentation in that screen room eliminates many variables.

In general, modern commercial instrumentation is vastly superior to the earlier cavity bridges for making routine measurements on low-power sources. It can measure phase noise very close to the carrier if the servo is tailored to force the two sources to track in the mean, without effectively locking them together, at the offset frequency of interest. By knowing the servo's characteristics, the instrumentation can adjust the data to reflect the phase noise actually present. The technique is limited, at small offset frequencies, by thermal noise that competes with the low values of phase deviation permitted by the servo. Even with this limitation, one is considerably better off than with a cavity bridge whose sensitivity falls off rapidly at small offsets. Without the aid of commercial phasebridge instrumentation, it would have been difficult to develop the crystal sources having much reduced phase noise at close-in frequencies. (These have been key components in long-range airborne radars that are required to detect

crossing targets immersed in clutter.) Nor are commercial instruments limited in their ability to measure phase noise at the larger offsets. They appear to have just two significant limitations. Cavity bridges are superior for development work on state-of-the-art sources, especially those that are difficult or expensive to produce in pairs, and in the measurement of high-power transmitters such as the Hawk illuminator. Compare curve *I* of Fig. 14.8, measured in the early 1960s, with curve *P* of the same figure, which is the measurement floor of typical commercial instrumentation.

An alternative to both cavity and source comparison techniques is the use of a delay line to provide a primary reference for the measurement of phase noise. A method that eliminates the noise contributed by the local oscillator is suggested in the appendix of Ref. 16. Unfortunately, the accuracy of any phase-noise measurement that depends on a delay line is proportional to the length of the delay. Long delays imply difficulty in maintaining sufficient signal amplitude to make satisfactory measurements. Incidentally, as noted above, many measurement techniques can be altered to provide a valid source-stabilization method. The delay line is an exception. When one attempts to servo-out phase noises at the
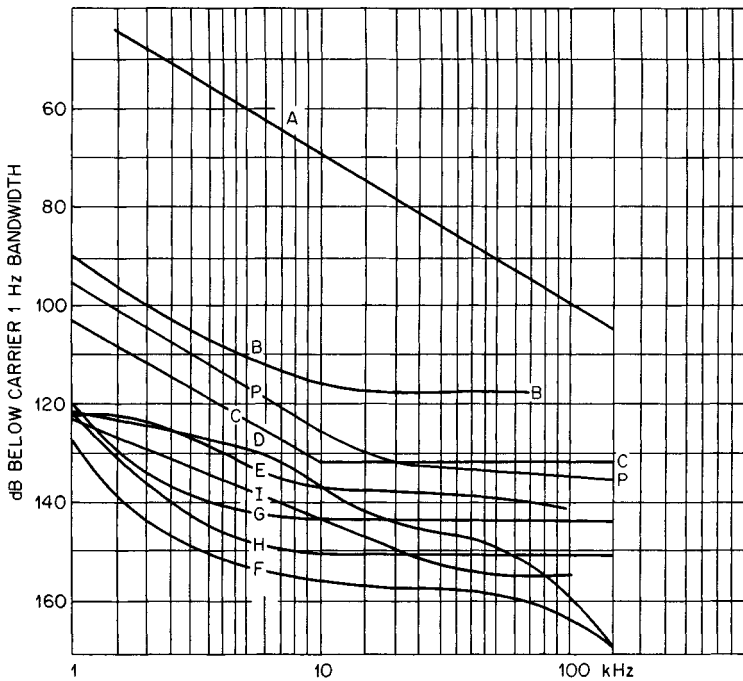
FIG. 14.8 FM noise in microwave sources. *A*, voltage-controlled *LC* oscillator multiplied to X band; *B*, crystal-controlled oscillator, step-recovery multiplier, to X band (*courtesy of D. Leeson*); *P*, noise floor at X band of 11729B/8640B combination (*courtesy of Hewlett-Packard*[14]); *C*, crystal oscillator (ST cut) multiplied to X band (*courtesy of Westinghouse Corporation*[18]); *D*, compact X-band klystron CW amplification (*Hughes*); *E*, compact X-band klystron pulsed amplification (*Hughes*); *F*, X-band klystron CW amplification (*Varian*); *G*, X-band klystron pulsed amplification (*Varian*); *H*, S-band electrostatically focused klystron amplifier (*Litton*); *I*, curve *B* in Fig. 14.7. Note that curves *D* to *H* are additive-noise measurements.

higher offset frequencies, one runs into the Nyquist restriction. For any fixed delay length, there is a corresponding offset frequency where the servo gain must go to zero or the whole system will become unstable.

Except for multipliers and dividers, the measurement of additive or excess phase noise* on components such as power amplifiers is considerably easier than the similar measurements on sources. All that is required is one moderately quiet source, a phase shifter, a phase detector, a suitable wave analyzer, and a method of calibration. The commercial instruments, described above, provide all this and much more. The reason that the source phase noise is not critical to the measurement is that it is common practice to add sufficient coaxial or microwave delay to equalize the two paths to the phase detector. Figure 14.2 indicates what such equalization (i.e., correlation) buys. As above, the amplitude modulation introduced by the source or the component under test must be checked first. For very demanding measurements, such as shown in curve *F* of Fig. 14.8, it might be well to consult the 1970 edition of this handbook and Ref. 17 of this chapter. For such work, a screen room is a must.

The measurement of multipliers is considerably more difficult since the two signals arriving at the phase detector must have the same frequency. This implies that two similar multipliers must enter the circuitry, and one has most of the problems associated with the measurement of sources. The only problem one is spared is the phase locking, which is usually required when working with sources. Fortunately, well-designed multipliers usually add little phase noise to a radar (above that to be expected from the increased FM deviation produced by the multiplication process). When 100 sources, consisting of a crystal oscillator plus a multiplier chain, were supplied by a subcontractor to a military radar program, the only ones that were unable to meet an extremely severe specification were ones that had substandard crystals in the oscillator.[16]

Because of the similarity of methods, measurement of noise in pulsed transmitters will merely be sketched. The measurement of pulsed sources is intrinsically much more difficult than the measurement of CW sources. The pulse structure produces very substantial AM that inevitably conflicts in direct and indirect ways with any attempt to measure the FM. In fact, it is possible to measure FM only up to half the repetition frequency, and then only by the use of rather sharp filters placed immediately following the *Q* detector. Measurements of −100 dB with respect to the carrier in a 1-Hz band 10 kHz from the carrier require excellent technique.

Similar problems occur in the measurement of the additive noise produced by pulse amplifiers. At Harry Diamond Laboratories some added sensitivity has been obtained by producing another pulse spectrum as similar as possible to that produced by the transmitter and subtracting this.

A suitable device for switching low-level signals is a PIN diode modulator, but with it there is difficulty in obtaining an exact reproduction of the pulse shape produced by a high-power amplifier. On the other hand, the rather large phase perturbation produced by the PIN diode modulator on the leading edge of the pulse is repeated pulse to pulse and produces spectral energy only at multiples of the repetition frequency, where measurements are impossible in any case.

Typical additive-noise measurements made on a variety of FM and CW sources at the Harry Diamond Laboratories[17] and elsewhere are shown in Fig. 14.8. The considerable improvement since 1970 in crystal-oscillator multiplier

*The adjective *residual* sometimes appears in the literature. It is not an apt choice in radar, where the source is usually the element that sets the phase-noise floor.

chains, especially below 5 kHz, can be seen by comparing curve B (1970) with curve C (1988). Even better performance is to be expected as research in this very important area continues. Although the curves are given to 150 kHz at most, one is often interested in FM noise out to $1/\tau$ (where τ is the pulse length). Solid-state sources, unlike klystrons, have white FM noise at the higher frequencies.[19] This noise folds down in the operation of a pulse doppler radar. The locked source method of measurement, mentioned on page 14.12, can be conveniently altered to measure the total folded noise. The servo is designed to remove noise from the phase detector in the correct proportion to account for the correlation effect. The output of the phase detector is then chopped at the radar's pulse repetition frequency and duty factor. The folding, thus produced, accurately reproduces the radar's demands on its source. The required $1/f^2$ frequency response of the servo is a convenient and stable choice. It is superior to a strictly narrowband loop followed by a shaped amplifier even for CW measurements.

14.6 RECEIVERS

RF Amplification. Although it is apparently attractive, low-noise RF amplification of the received signal has not been extensively employed in CW radar. Transistor amplifiers with excellent noise figures are available to K_u band. Traveling-wave tubes are expensive. In many cases, the determining factor in deciding against a low-noise RF amplifier has been the presence of spillover noise, clutter signals, and signals produced by electronic countermeasures that equal or exceed the noise contributed by conventional front ends.

Few modern receivers have been designed for CW target illuminators since many installations avoid the use of a separate receiver. With space in the nose of an aircraft at a premium, it is usual for airborne missile systems to employ a common antenna for both the tracking radar and the illuminator.[71] In some shipboard weapon systems, the illuminator does not require a receiver since the illuminator antenna is pointed to the direction of the target from information obtained by the tracking radar of the weapon control system rather than have the illuminator track the target itself.

Generation of the Local-Oscillator Signal. To realize adequate signal-to-noise performance, it is customary to perform the first amplification in a CW radar at an intermediate frequency such as 30 MHz. To obtain the necessary coherent local oscillator signal, various types of sidestep techniques are employed. These include modulators, balanced modulators, single-sideband generators (SSG), or phase-locked oscillators.

The SSG is probably the most cumbersome, since the suppression of the carrier and unwanted sidebands is seldom better than 20 dB and filtering must be employed to suppress these signals further. The balanced modulator is much simpler and suppresses the carrier as well as the SSG. The filter needed for further carrier suppression usually suppresses the unwanted sideband to the desired level without added poles. The simple modulator is scarcely less complex than the balanced modulator and requires a sharper filter for the necessary added carrier suppression. Phase locking eliminates the need for high-frequency filtering altogether but requires a skillfully designed servo loop to impress the transmitter's FM noise faithfully on the local oscillator. It may also require a search mechanism to pull in initially. All these methods require the use of an oscillator at the intermediate frequency. The stability required
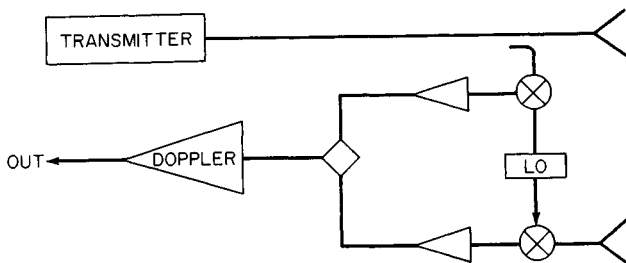
FIG. 14.9 Balanced receiver with a floating LO.

of this oscillator is not excessive since FM is contributed only in the ratio of the IF frequency to microwave frequency.

An alternative approach to the problem of an IF offset is the free-floating local oscillator employed by Harris et al.[13] and by O'Hara and Moore.[8] This has a similarity to the method used to introduce the local oscillator in the Marsh and Wiltshire bridge and requires twin IF amplifiers. The basic diagram is shown in Fig. 14.9. The simple appearance of the figure is deceptive. The local oscillator must normally be positioned by the AFC to hold the signal in the IF bands. As shown, the system folds the doppler frequencies. To avoid this, either quadrature techniques must be employed or a second sidestep be introduced into the reference channel. The latter is unattractive as it destroys the symmetry that may be required to assure uniformity of time delays to cancel the FM noise of the LO. Even in the simplest version the symmetry is far from complete, as the signal channel must handle signals over a wide range of amplitudes while the reference channel carries a signal of uniform amplitude.

IF Amplifier. Traditional low-noise IF amplifiers are usually employed. Because of the levels of clutter signals, ECM, and spillover signals that must be carried by the IF amplifier, it is usual to restrict the gain to no more than 40 dB. This establishes the noise figure and raises the signal to a value where microphonism is less serious without risking levels where saturation and the attendant intermodulation are problems.

Subcarriers. Although doppler filtering may be carried out at slightly higher levels, it is desirable to reject the signal produced by clutter and by spillover at the lowest level possible. Unfortunately, sufficiently high Q's are not available, even in quartz filters, to make it possible to reject clutter at, say, 30 MHz without diminishing the lower doppler frequencies as well.

The simplest method is to mix the signal from the IF amplifier with the signal used in the sidestep. This reduces the spillover signal to dc and the clutter signal to dc and very low frequencies. A multipole filter will suppress those unwanted signals with minor suppression of the very lowest dopplers. Unfortunately, this process folds the spectrum so that incoming targets are indistinguishable from outgoing targets and the random-noise sidebands accompanying each appear in the baseband amplifier. Even if one is prepared to accept the ambiguity, the 3 dB loss in the signal-to-noise ratio (SNR) is a matter of concern in a high-power radar.

There are two alternatives, both of which have been extensively employed. The first is a subcarrier band for the doppler intelligence which does not extend

to dc but is centered at a frequency where either quartz or electromechanical filters have sufficient Q's to permit sharp filtering. (Values of 0.1 to 0.5 MHz or 1 to 5 MHz are suitable ranges for quartz filters; 0.1 to 0.5 MHz is proper for electromechanical filters.)

The second alternative is quadrature detection.[20] A suitable block diagram for this technique is shown in Fig. 14.10. A single 90° phase shift can be substituted for the +45° and −45° shifts at the constant frequency coming from the oscillator. The plus and minus 45° in the two signal paths are required to maintain a semblance of balance over a wide band of frequencies. A phasor diagram of the system (simplified by omitting the IF sidestep) is shown in Fig. 14.11. If the output from mixer 1 is
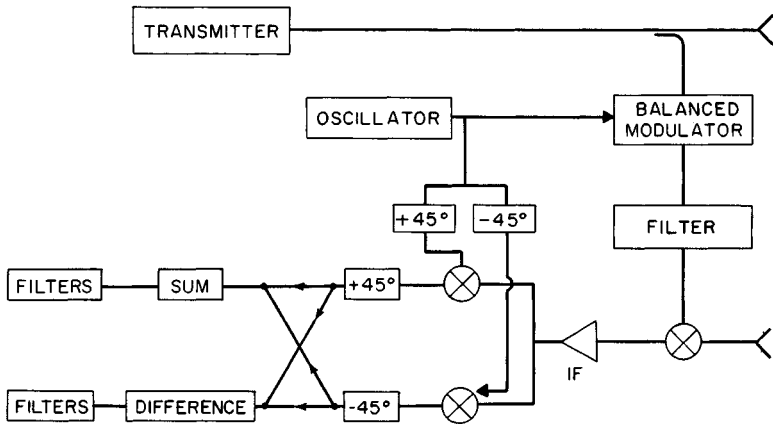
FIG. 14.10 Quadrature receiver.
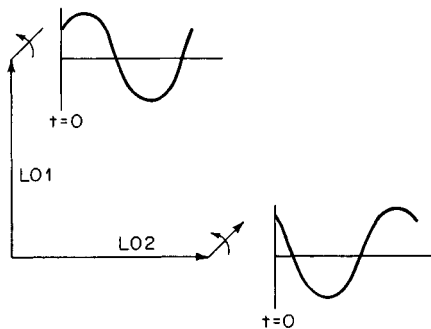
FIG. 14.11 Phasor diagram for a quadrature receiver.

advanced 90° and added to that of mixer 2, the signals will sum in the combiner and disappear in the differencer. A corresponding diagram would show that receding targets reinforce in the differencer and cancel in the combiner.

An advantage of the quadrature system is that the filter bands are completely symmetrical and all filter elements may be identical. Moreover, high-pass filters

having steep slopes to reject clutter are somewhat easier to design near dc than they are at low IF. A disadvantage is the requirement to maintain balanced operation over the full range of doppler frequencies in the two second mixers and in the two 45° phase shifters in order to eliminate false targets.

Amplification.   After the undesired clutter and spillover signals have been removed, substantial amplification can be achieved either at the second subcarrier frequency or, in the case of folded or quadrature systems, at the doppler frequency itself. It is customary to add additional filtering against the unwanted signals as interstage networks between the stages. The only requirement that must be met is that the combination of the amplification and the total filtering is such that the amplitude of the unwanted signals nowhere approaches the saturation level.

Doppler Filter Banks.   Ideally no nonlinear operation occurs in the signal processing and amplification. There is still a coherent signal, and the band narrowing buys decibel for decibel in improved signal-to-noise ratio. Steinberg[21] has shown that, given a fixed doppler band, one pays no penalty in false-alarm rate for subdividing it. In a radar, then, it would be desirable to have the final doppler bandwidth limited only by the time on target. This might be possible in a rapidly scanning radar, but with tracking radars or illuminators the indicated bandwidths would be unrealistically narrow. Moreover, the target itself seldom produces a clearly defined doppler but, rather, a spread of frequencies by the scintillation and glint effects. Bandwidth may also have to be allowed for the coding frequencies which may accompany the doppler, such as those injected by conical scan. A typical circuit for an X-band radar might have a suitable bank of adjacent two-pole filters, each 1000 Hz wide, or an equivalent set of digital filters produced by an FFT.

Following each filter are a detector and a postdetection integrator whose time constant is matched to the time on target or, in the case of a tracker, the demands of the servo data rate. A threshold level is set in the circuitry following each detector; when this is exceeded, a voltage is generated and held until such time as it is read. In acquisition the threshold circuits are normally scanned by some type of readout mechanism. This is fundamentally a computer-type operation.

Doppler Trackers.   Doppler filter banks are satisfactory for acquisition and for track-while-scan radars. They are not commonly used in tracking radars or illuminators to improve *SNR* since the use of a doppler tracker (speedgate) is far less complex. The usual speedgate circuit is identical with the AFC circuit in an FM radio. A voltage-controlled oscillator (VCO) is used to beat the signal to be analyzed to a convenient intermediate frequency. A narrowband amplifier at this frequency performs the filtering operation. The VCO is in turn controlled by the output of a discriminator connected to the amplifier. The input to the speedgate can either be the full doppler band, as in the folded or quadrature receiver, or be a subcarrier containing the full doppler intelligence. Although some clutter filtering may take place in the speedgate, earlier removal of unwanted signals is preferable. This is particularly important with the folded receiver in certain airborne situations in which the clutter has a substantial spread because mixer nonlinearities produce harmonics of the unwanted signals that may fall directly on the target signal in the speedgate.

Once in track, the speedgate follows the proper doppler component. The response is limited only by the bandwidth of the servo, which is designed to follow

the expected target maneuvers. To acquire track, intelligence may be passed on an open-loop basis from the doppler filter bank, if one is available, or the VCO may have a sawtooth or triangular voltage applied to produce a programmed search. Search is stopped when the output registers the desired target. Coding signals may be employed to aid in the detection and the stopping.

It is usually necessary to restrict the VCO from moving to a frequency that will lock the speedgate on spillover or clutter. With ground-based systems the problem may be simply solved by fixed-limit stops placed on the search voltage. Airborne systems having clutter signals that vary in frequency require more sophisticated solutions.

Constant False-Alarm Rate (CFAR). A constant false-alarm rate in the presence of variable levels of noise is usually a requirement placed on any modern radar. It is very easily achieved in CW radars by the use of filter banks or FFTs. The energy reaching the filter banks is restricted either by automatic gain control (AGC) or, when feasible, by limiting, and the thresholds in the circuitry following the filter banks are properly set with respect to the level in the total band. In a typical setting technique, random noise is injected into the amplifier that drives the filter banks, and each threshold is set to achieve the desired false-alarm rate. The level of noise is then varied and the threshold rechecked. If the limiting is proper, the false-alarm rate should not change. However, target signals in the absence of noise are unaffected, as they do not change the total energy present in the broad doppler spectrum sufficiently to change the AGC level or reach the limiting level. Similar remarks apply to the speedgate as well.

14.7 MINIMIZATION OF FEEDTHROUGH

All major ground-based CW radars have two antennas to minimize spillover. Isolation may be improved further by the use of various absorbers or of an intentional feedthrough path that is adjusted in phase and amplitude to cancel spillover energy. In free space such solutions are all that would be required. When a radar scans across a rough ground plane, however, the energy reflected to the receiver antenna does not remain constant. A dynamic canceler is required. A diagram and description of one such device are given in Ref. 10. Other descriptions are to be found in Refs. 8 and 22.

All dynamic cancelers depend on synthesizing a proper amplitude and phase of a signal taken from the transmitter and using this to buck out the spillover signal. To achieve independence of the servos, the vector is synthesized in orthogonal rectangular coordinates. Figure 14.12 is a typical arrangement for use with a CW radar in which the local oscillator is derived by sidestepping the transmitter. Slight modification[8,13] is needed when the basic radar has the balanced method (Fig. 14.9) of generating the offset required for the first IF amplifier.

The servo amplifiers have response from dc to some frequency well below the doppler band of interest. They respond to the slow variations in the feedthrough signal without damage to the dopplers. For complete details of the mechanical design, see Ref. 8.

Harmer and O'Hara[22] show a variant of the equipment that may be used with a single antenna plus a duplexer. This would be very attractive, especially for an airborne radar that must fit into a small radome. Unfortunately, experience has shown that there is a limit to the transmitter power that may be employed in such
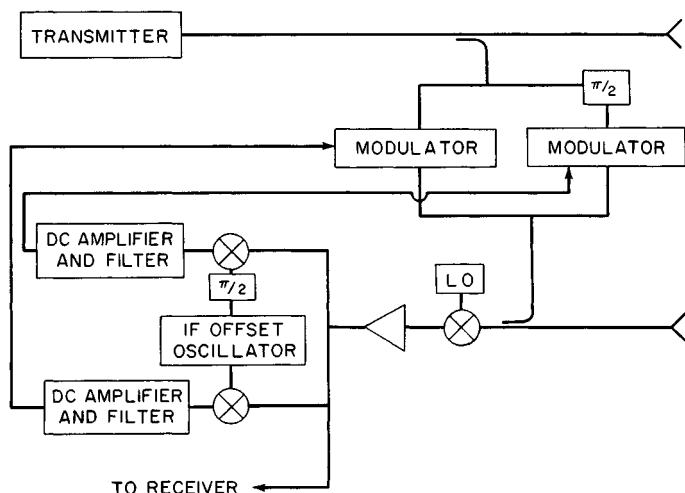
14.12 Feedthrough nulling bridge.

an arrangement. Beyond a modest level of power, the servo is unable to cancel out the −20 dB reflection from the antenna or duplexer sufficiently to prevent receiver degradation.

It should be noted that microwave-feedthrough cancellation is of principal value in preventing saturation and in minimizing the effects of AM noise. Because of the correlation effect, FM noise produced by spillover tends to cancel in the receiver. Near-in AM and FM noise produced by clutter is also beneficially reduced by the spillover servo, since, in nulling out the carrier, it automatically removes both sidebands, whatever their origin, as long as the decorrelation interval is short. Clutter signals from long ranges have both AM and FM noise that is essentially decorrelated, and feedthrough nulling of these signals may increase their deviation by a factor of 2 or their power by a factor of 4. See Eq. (14.3).

14.8 MISCELLANEOUS CW RADARS

There are several small CW radars for applications that require equipment of modest sensitivity. In all these the homodyne technique is employed, the transmitter itself serving as a local oscillator. The transmitter signal reaches the first mixer either by a direct connection or, more frequently, by controlled leakage.

CW Proximity Fuzes. The basic proximity fuze[23,24] is a CW homodyne device whose only range sensitivity is in the rise of the doppler voltage signals as ground is approached or in the behavior of the signal when the antenna pattern intercepts an aircraft. Commonly a single element is used as both oscillator and mixer-detector.

Characteristically, proximity fuzes use a common antenna for transmitting and receiving and hence suffer from a large leakage problem. The situation is tolerable only in the VHF band where the signals returned from the target (terrain or aircraft) are very large. Frequently a projectile body is used as an end-fed an-

tenna although separate transverse dipole or loop antennas have been employed to avoid a null in the forward direction.

The principal problems with the device are those associated with requirements of small size, long shelf life, low cost, and reliability under high acceleration. Because of the very light weight of all solid-state circuitry with integrated components, complex circuits may be built that will allow proximity fuzes to withstand accelerations in excess of 100,000 g.

Police Radars.   This is a straightforward application of the CW homodyne radar technique. Controlled leakage is used to supply the required LO signal to a single crystal mixer. The amplification takes place at the doppler frequency. At 10,525 MHz, one of the frequencies currently approved by the Federal Communications Commission (FCC), 50 m/h corresponds to 1570 Hz, which is in a convenient range.

A squelch circuit is used to prevent random or noisy signals from reaching the counter. Three amplifier levels relative to the squelch yield suitable gains for the detection of short-, medium-, or maximum-range automobiles. The output signal from the doppler amplifier is clipped, differentiated, and integrated. Each pulse from the differentiator makes a fixed contribution to the integrated signal, and the higher the frequency the greater the output. This dc value actuates a meter or a recording device marked directly in velocity. A tuning fork may be used to calibrate the equipment. Some equipments offer a burst mode which determines the speed of the vehicle before it can be altered.

14.9 FM RADAR

The material to follow is on the homodyne FM radar, i.e., a CW radar in which a microwave oscillator is frequency-modulated and serves as both transmitter and local oscillator. For additional material on FM radars, see Refs. 3 and 6. An excellent introduction to FM in general is contained in Ref. 4, Chap. 12.

There are three approaches to the analysis of this type of radar: the phasor diagram, the time-frequency plot, and Fourier analysis. One should have some facility with each. Perhaps the most useful attack for an FM radar having modest deviation is the phasor diagram. To construct the diagram, a large phasor is drawn to represent the carrier. This is taken as a reference and is considered stationary; higher frequencies are represented by phasors rotating counterclockwise and lower frequencies by phasors rotating clockwise. In applying the phasor method to FM homodyne radar, the instantaneous phase of the local oscillator (i.e., that of the transmitter) is taken as the reference phasor, and the returning signal or signals as the small phasor or phasors. The output from the mixer is proportional to the projection of the small phasor or phasors on the large one.

For example, consider an altimeter with triangular frequency modulation. In its phasor diagram (Fig. 14.13) the small phasor will, except at the turnarounds, swing either clockwise or counterclockwise at a uniform rate. If the swing is short (i.e., the range to the ground is short), then, depending on the phase, either of two situations results: Fig. 14.14a or b. In Fig. 14.14a twice as many cycles of difference frequency will be developed in unit time as in Fig. 14.14b. This leads to the so-called critical-distance problem in an FM altimeter. The situation will be covered more fully below; here the interest is in the phasor diagram and what it reveals.
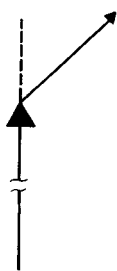
FIG. 14.13 Phasor diagram for an FM-CW radar.

(a)   (b)
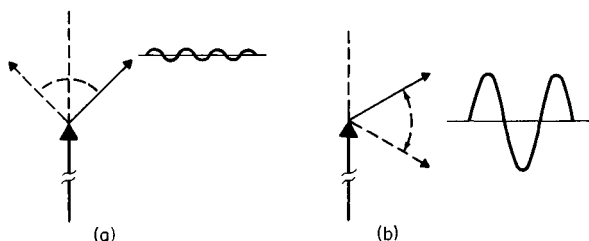
FIG. 14.14 Phasor diagrams showing critical distance

The second method is the drawing of an instantaneous-frequency diagram. In these diagrams a curve is drawn in a time-frequency plane to represent each of the signals of interest. A typical plot, that for a sinusoidally modulated altimeter, is shown in Fig. 14.15. Curve *A* represents the frequency-time history of the transmitter (and local oscillator) and curves *B* and *C* that of returns from two different ranges. Note that the vertical distance between curves (e.g., curves *D* and *E*) yields a heuristic picture of the average frequency behavior of the difference signal from the mixer. This is somewhat naïve. Both the transmitted signal and the returned signal are periodic waves, as is their difference. Hence there cannot be a continuum of difference frequencies; there can be only harmonics of the fundamental modulation frequency. Diagrams such as Fig. 14.15 are most useful when the different frequencies indicated are several multiples of the repetition frequency. In this event, the many harmonic lines act almost like a continuum. Such a diagram would not be useful to discover the step error shown in the phasor diagram above.

Finally, there are mathematical approaches limited originally to those systems
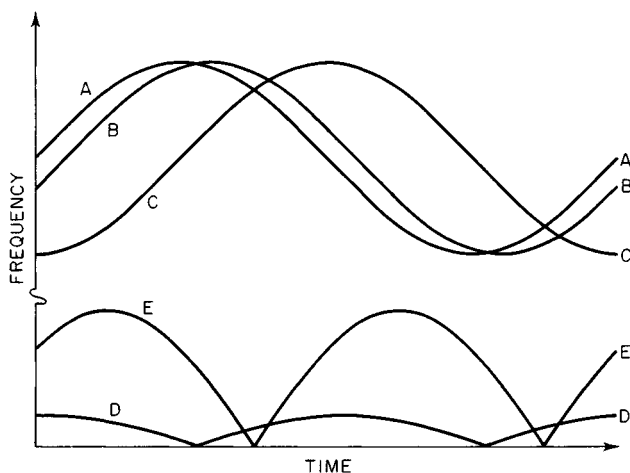
FIG. 14.15 Schematic diagram for a sinusoidally modulated FM.

employing one or more sinusoidal modulations. There exist exact analyses of triangular, sawtooth, dual triangular, dual sawtooth, and combinations of some of these with noise,[25] but heuristic techniques are usually a necessary starting point.

14.10 SINUSOIDAL MODULATION

Suppose one transmits an FM wave of the form

$$\mu_s = U_s \sin \left(\Omega_0 t + \frac{\Delta\Omega}{\omega_m} \sin \omega_m t \right)$$

where ω_m = modulation frequency
Ω_0 = carrier frequency
$\Delta\Omega/\omega_m$ = modulation index

An echo from a point target will have the form

$$\mu_e = U_e \left\{ \sin \left[\Omega_0 (t - T) + \frac{\Delta\Omega}{\omega_m} \sin \omega_m (t - T) \right] + \phi \right\}$$

where ϕ = arbitrary phase angle produced on reflection
T = time delay of echo

To introduce the effect of doppler we let T be time-dependent: $T = T_0 + 2vt/c$, where v is the velocity of the echoing object and c the velocity of light. After the usual trigonometric manipulation, the difference μ_i takes the form

$$\mu_i = U_i \cos \left[\Omega_0 \left(T_0 + \frac{2vt}{c} \right) - \phi + D \cos \omega_m \left(t - \frac{T}{2} \right) \right]$$

$$D = \frac{2\Delta\Omega}{\omega_m} \sin \frac{\omega_m T}{2}$$

The reflection phase ϕ may generally be disregarded and μ_i expanded in a Fourier series.[6]

$$\mu_i = U_i \left(J_0(D) \cos \Omega_0 \left(T_0 + \frac{2vt}{c} \right) + \sum_{n \text{ odd}}^{\infty} (-1)^{(n+1)/2} J_n(D) \left\{ \sin \left[n\omega_m \left(t - \frac{T}{2} \right) \right. \right. \right.$$

$$\left. + \Omega_0 \left(T_0 + \frac{2vt}{c} \right) \right] - \sin \left[n\omega_m \left(t - \frac{T}{2} \right) - \Omega_0 \left(T_0 + \frac{2vt}{c} \right) \right] \right\}$$

$$+ \sum_{n \text{ even}}^{\infty} (-1)^{n/2} J_n(D) \left\{ \cos \left[n\omega_m \left(t - \frac{T}{2} \right) + \Omega_0 \left(T_0 + \frac{2vt}{c} \right) \right] \right.$$

$$\left. \left. \right) \cos \left[n\omega_m \left(t - \frac{T}{2} \right) - \Omega_0 \left(T_0 + \frac{2vt}{c} \right) \right] \right\} \right)$$

CHAPTER 15
MTI RADAR

William W. Shrader
V. Gregers-Hansen
Equipment Division
Raytheon Company

15.1 INTRODUCTION TO MTI RADAR

The purpose of moving-target indication (MTI) radar is to reject signals from fixed or slow-moving unwanted targets, such as buildings, hills, trees, sea, and rain, and retain for detection or display signals from moving targets such as aircraft. Figure 15.1 shows a pair of photographs of a PPI (plan position indicator) which illustrate the effectiveness of a properly working MTI system. The distance from the center to the edge of the PPI is 40 nmi. The range marks are at 10-nmi intervals. The picture on the left is the normal video display, showing the fixed-target returns. The picture on the right shows the MTI clutter rejection. The camera shutter was left open for three scans of the antenna; thus aircraft show up as a succession of three returns.

MTI radar utilizes the doppler shift imparted on the reflected signal by a moving target to distinguish moving targets from fixed targets. In a pulse radar system this doppler shift appears as a change of phase of received signals between consecutive radar pulses. Consider a radar which transmits a pulse of RF energy that is reflected by both a building (fixed target) and an airplane (moving target) approaching the radar. The reflected pulses return to the radar a certain time later. The radar then transmits a second pulse. The reflection from the building occurs in exactly the same amount of time, but the reflection from the moving aircraft occurs in less time because the aircraft has moved closer to the radar in the interval between transmitted pulses. The precise time that it takes the reflected signal to reach the radar is not of fundamental importance. What is significant is whether the time changes between pulses. The time change, which is of the order of a few nanoseconds for an aircraft target, is determined by comparing the phase of the received signal with the phase of a reference oscillator in the radar. If the target moves between pulses, the phase of the received pulses changes.

Figure 15.2 is a simplified block diagram of one form of a coherent MTI system. The RF oscillator feeds the pulsed amplifier, which transmits the pulses. The RF oscillator is also used as a phase reference for determining the phase of reflected signals. The phase information is stored in a PRI (pulse repetition interval) memory for the period between transmitted pulses, and it is also subtracted

15.1
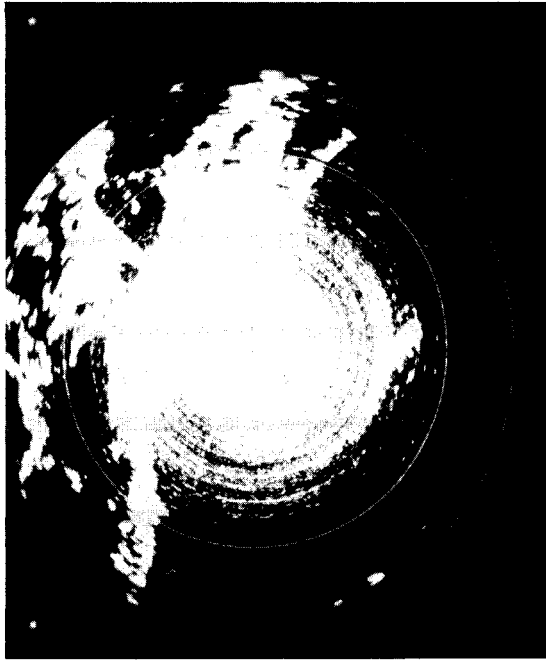
(a)
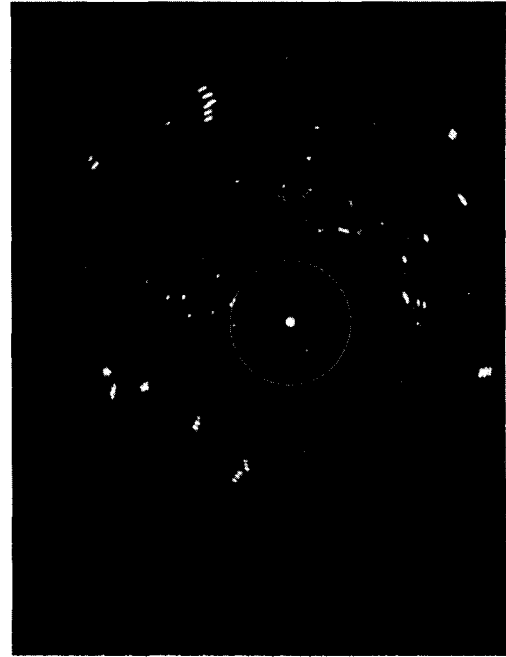
(b)

FIG. 15.1 (a) Normal video. (b) MTI video. These PPI photographs show how effective an MTI system can be. Aircraft appear as three consecutive blips in the right-hand picture because the camera shutter was open for three revolutions of the antenna. The PPI range is 40 nmi.
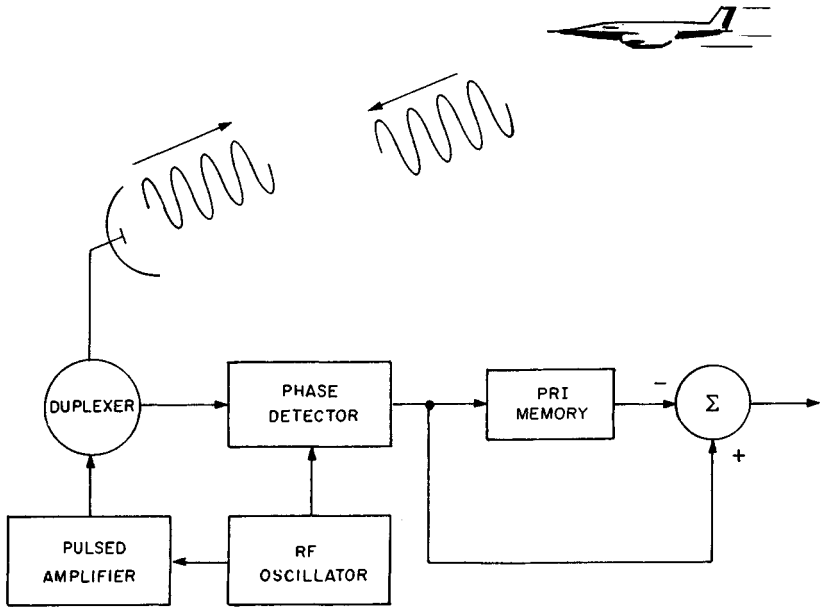
FIG. 15.2 Simplified block diagram of a coherent MTI system.

from the phase information from the previous transmitted pulse. There is an output from the subtractor only when a reflection has occurred from a moving target.

Moving-Target Indicator (MTI) Block Diagram. A block diagram of a complete MTI system is shown in Fig. 15.3. This block diagram represents an MTI system that uses a pulsed oscillator. It is not as sophisticated as MTI systems to be described later, but most of the practical considerations applying to any MTI system can be understood by examining this block diagram. The frequencies and 2500-μs interpulse period are typical for a 200-mi L-band radar.

The transmitter shown employs a magnetron. Because a magnetron is a pulsed oscillator that has no phase coherence between consecutive pulses, a phase reference must be established for each transmitted pulse. This is done by taking a sample of the transmitted pulse at a directional coupler, mixing this pulse with the stalo (stabilized local oscillator) and then using this pulse to phase-lock the coho (coherent oscillator). The coho then becomes the reference oscillator for the received signals. (The stability requirements for the coho and stalo will be described in Sec. 15.11.) The lock-pulse amplifier is gated off just before the end of the transmitted pulse because a magnetron emits a certain amount of noise during the fall of the high-voltage pulse applied to it, and this noise can prevent perfect locking of the coho.

The received signals are mixed with the stalo and amplified in a linear-limiting amplifier. (In some implementations the limiting is not deliberately provided. However, receiver saturation occurs at some signal level, and thus limiting inadvertently exists.)

The received signals are then compared in phase with the coho in a phase de-
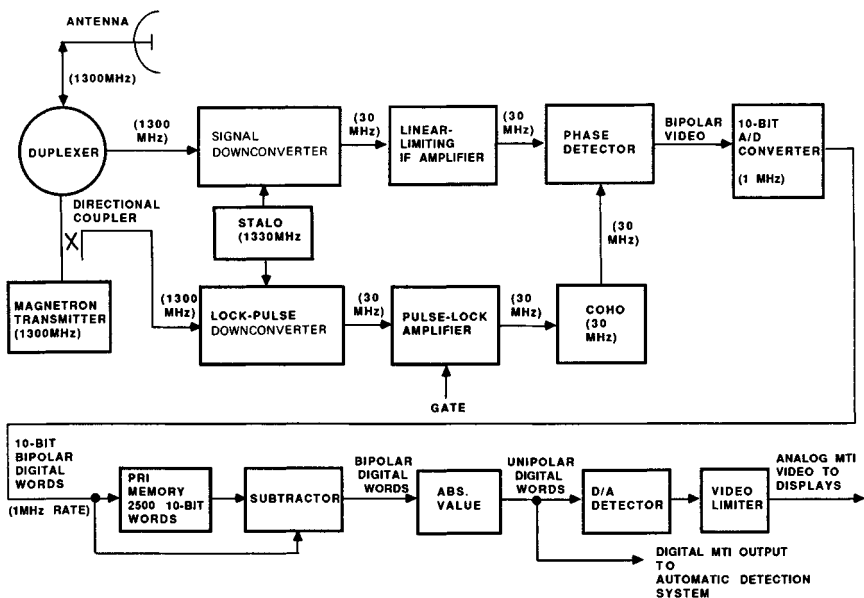
FIG. 15.3   MTI system block diagram.

tector. The output of the phase detector is a function of the relative phase of the signal and the coho, and it is also a function of the amplitude of the signal. At the output of the phase detector, the signal phase and amplitude information has been converted into bipolar video. The bipolar video received from a single transmitted pulse may appear as sketched in Fig. 15.4. If the point target is moving and if there is also a moving target in the region of strong clutter return, the superimposed bipolar video from several transmitted pulses may appear as in Fig. 15.5.

The remainder of the block diagram in Fig. 15.3 shows what is necessary for detecting the moving targets so that they may be displayed on a PPI or sent to an automatic target extractor. The bipolar video is converted to digital words in an analog-to-digital (A/D) converter. The A/D output is stored in a PRI memory and also subtracted from the memorized A/D output from the previous transmitted pulse.
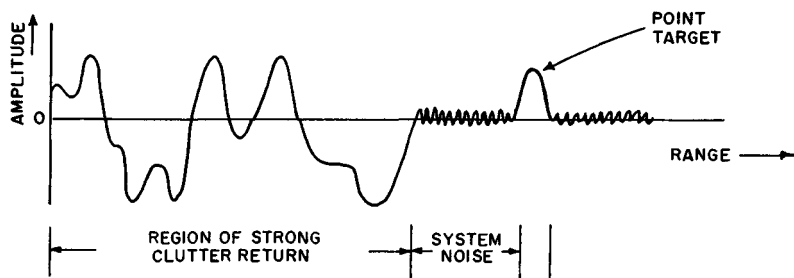
FIG. 15.4   Bipolar video: single sweep.
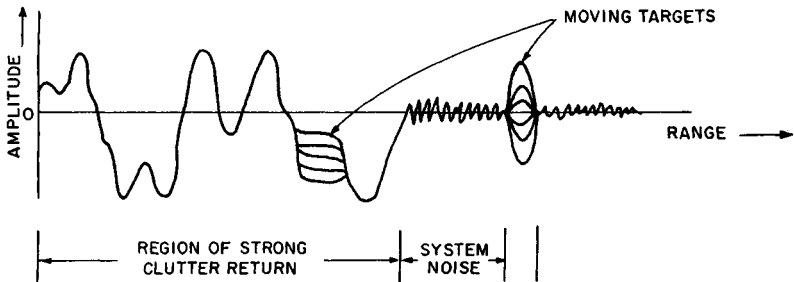
FIG. 15.5   Bipolar video: several sweeps.

The output of the subtractor is a digital bipolar signal that contains moving targets, system noise, and a small amount of clutter residue if clutter cancellation is not perfect. The absolute value of the signal is then converted to analog video in a digital-to-analog (D/A) converter for display on a PPI. The digital signal may also be sent to automatic target detection circuitry. The dynamic range (peak signal to rms noise) is limited to about 20 dB for a PPI display.

Moving-Target Detector (MTD) Block Diagram.   In the moving-target detector (MTD) the basic MTI principle, as described above, is enhanced by increasing the linear dynamic range of the signal processor, using a number of parallel doppler filters followed by constant-false-alarm-rate (CFAR) processing, and adding one or more high-resolution clutter maps to suppress point clutter residues. With these additions a complete signal-processing system is obtained for suppressing clutter returns in a modern surveillance radar. A typical implementation of such an MTD processing system is shown in Fig. 15.6.

The MTD radar transmits a group of N pulses at a constant pulse repetition frequency (PRF) and at a fixed radar frequency. This set of pulses is usually referred to as the coherent processing interval (CPI) or pulse batch. Sometimes one or two additional fill pulses are added to the CPI in order to suppress range-ambiguous clutter returns, as might occur during periods of anomalous propagation. The returns received during one CPI are processed in the bank of N-pulse finite-impulse-response (FIR) filters. Then the radar may change PRF and/or RF
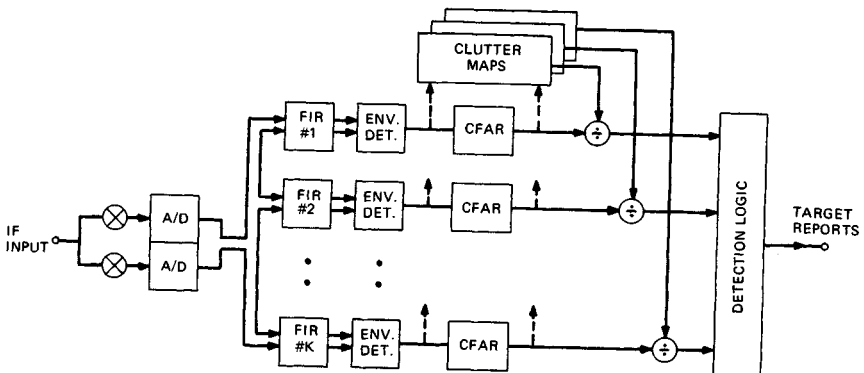
FIG. 15.6   MTD block diagram.

frequency and transmit another CPI of N pulses. Since most search radars are ambiguous in doppler, the use of different PRFs on successive coherent dwells will cause the target response to fall at different frequencies of the filter passband on the successive opportunities during the time on target, thus eliminating blind speeds.

Each doppler filter is designed to respond to targets in nonoverlapping portions of the doppler frequency band and to suppress sources of clutter at all other doppler frequencies. This approach maximizes the coherent signal integration in each doppler filter and provides clutter attenuation over a larger range of doppler frequencies than achievable with a single MTI filter. Thus one or more clutter filters may suppress multiple clutter sources located at different doppler frequencies. An example of the use of an MTD doppler filter bank against simultaneous land and weather clutter (Wx) is illustrated in Fig. 15.7. It can be seen that filters 3 and 4 will provide significant suppression of both clutter sources.
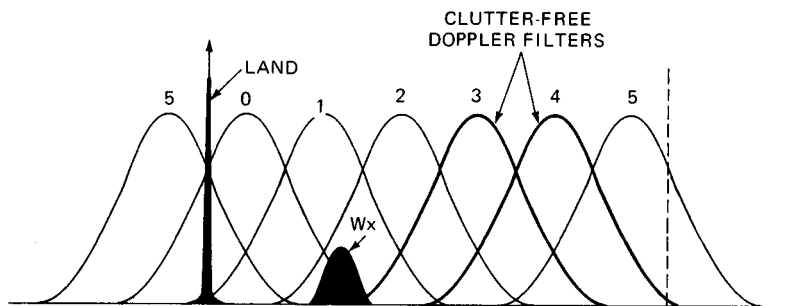
FIG. 15.7 Suppression of multiple clutter sources by using a doppler filter bank.

The output of each doppler filter is envelope-detected and processed through a cell-averaging CFAR processor to suppress residues due to range-extended clutter which may not have been fully suppressed by the filter.

As will be discussed later in this chapter, the conventional MTI detection system relies on a carefully controlled dynamic range in the IF section of the radar receiver in order to ensure that clutter residues at the MTI output are suppressed to the level of the receiver noise or below. This limited dynamic range, however, has the undesirable effect of causing additional clutter spectral broadening, and achievable clutter suppression is consequently reduced.

In the MTD one or more high-resolution clutter maps are used to suppress the clutter residues, after doppler filtering, to the receiver noise level (or, alternatively, to increase the detection threshold above the level of the residues). This in turn eliminates the need to restrict the IF dynamic range, which can then be set to the maximum value supported by the A/D converters. Thus, a system concept is obtained that provides a clutter suppression capability that is limited only by the radar system stability, the dynamic range of the receiver-processor, and the spectrum width of the returns from clutter. The concept of a high-resolution digital clutter map to suppress clutter residues is related to earlier efforts to construct analog area MTI systems using, for example, storage tubes.

In subsequent sections specific aspects of the design of an MTD system will be discussed. Thus Sec. 15.8 will discuss the design and performance of doppler filter banks, and a detailed discussion of clutter maps will follow in Sec. 15.14.

15.2 CLUTTER FILTER RESPONSE TO MOVING TARGETS

The response of an MTI system to a moving target varies as a function of the target's radial velocity. For the MTI system described above, the response, normalized for unity noise power gain, is shown in Fig. 15.8. Note that there is zero response to stationary targets and also to targets at ±89, ±178, ±267,... knots. These speeds, known as blind speeds, are where the targets move 0, ½, 1, 1½,... wavelengths between consecutive transmitted pulses. This results in the received signal being shifted precisely 360° or multiples thereof between pulses, which results in no change in the phase-detector output. The blind speeds can be calculated:

$$V_B = k\frac{\lambda f_r}{2} \qquad k = \pm 0, 1, 2, \ldots \tag{15.1}$$

where V_B is the blind speed, in meters per second; λ is the transmitted wavelength, in meters; and f_r is the PRF, in hertz. A convenient set of units for this equation is

$$V_B\,(\text{kn}) = k\,\frac{0.29 f_r}{f_{\text{GHz}}} \qquad k = \pm 0, 1, 2, \ldots \tag{15.2}$$

where f_r is the PRF, in hertz; and f_{GHz} is the transmitted frequency, in gigahertz. Note from the velocity response curve that the response to targets at velocities midway between the blind speeds is greater than the response for a normal receiver.

The abscissa of the velocity response curve can also be labeled in terms of
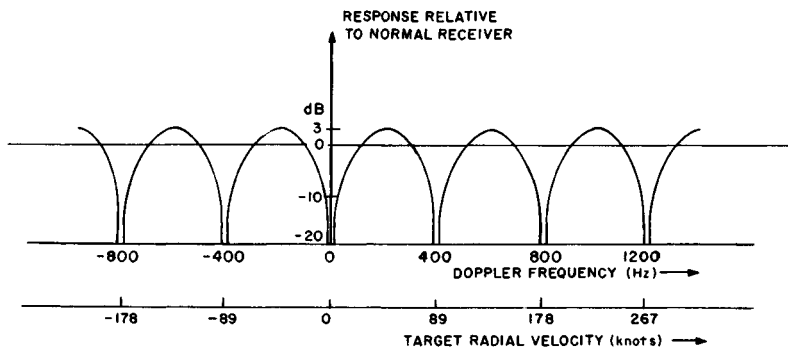
FIG. 15.8 MTI system response for 1300-MHz radar operating at 400 pps.

doppler frequency. The doppler frequency of the target can be calculated from

$$f_d = \frac{2V_R}{\lambda} \qquad (15.3)$$

where f_d is the doppler frequency, in hertz; V_R is the target radial velocity, in meters per second; and λ is the transmitted wavelength, in meters. It can be seen from Fig. 15.8 that the doppler frequencies for which the system is blind occur at multiples of the pulse repetition frequency.

15.3 CLUTTER CHARACTERISTICS

Spectral Characteristics.   The spectrum of a pulsed transmitter transmitting a simple rectangular pulse of length τ is shown in Fig. 15.9. The spectral width of the (sin $U)/U$ envelope is determined by the transmitted pulse width, the first nulls occurring at a frequency of $f_0 \pm 1/\tau$. The individual spectral lines are separated by a frequency equal to the PRF. These spectral lines fall at precisely the same frequencies as the blind speeds in Fig. 15.8. Thus a canceler will, in theory, fully reject signals with an ideal spectrum, as shown here. In practice, however, the spectral lines in clutter signals are broadened by motion of the clutter (such as windblown trees) and by motion of the antenna in a scanning radar. Barlow[1] stated that the returns from clutter have a gaussian spectrum, which may be characterized by its standard deviation σ_v. This spectral spread prevents perfect cancellation of clutter in the MTI system.
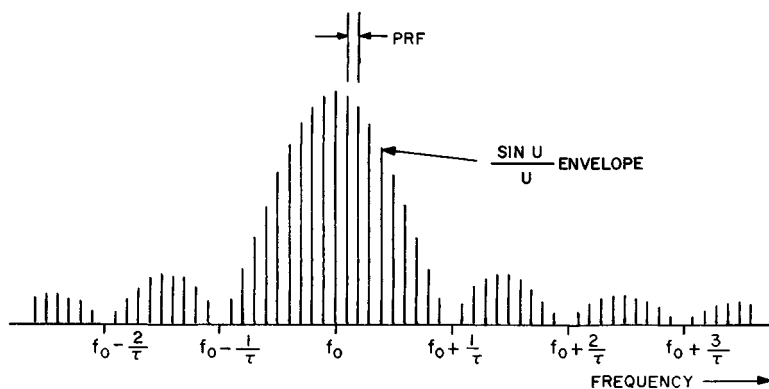
FIG. 15.9 Pulse transmitter spectrum.

Table 15.1 gives the standard deviation σ_v of the clutter spectrum in meters per second. More sophisticated and detailed clutter spectrum models exist,[6] but the gaussian model is usually an adequate model for understanding system limitations and obtaining good performance predictions.

Nathanson and Reilly[7] have shown that the clutter spectral width of rain is primarily due to a turbulence component and a wind-shear component (change in

TABLE 15.1   Summary of Standard Deviations of the Clutter Spectrum*

Source of clutter	Wind speed, kn	σ_v, m/s	Reference
Sparse woods	Calm	0.017	Barlow[1]
Wooded hills	10	0.04	Goldstein,[3] pp. 583–585
Wooded hills	20	0.22	Barlow[1]
Wooded hills	25	0.12	Goldstein,[3] pp. 583–585
Wooded hills	40	0.32	Goldstein,[3] p. 583
Sea echo		0.7	Wiltse et al.,[4] p. 226
Sea echo		0.75–1.0	Goldstein,[3] pp. 580–581
Sea echo	8–20	0.46–1.1	Hicks et al.,[5] p. 831
Sea echo	Windy	0.89	Barlow[1]
Chaff		0.37–0.91	Goldstein,[3] p. 472
Chaff	25	1.2	Goldstein,[3] p. 472
Chaff		1.1	Barlow[1]
Rain clouds		1.8–4.0	Goldstein,[3] p. 576
Rain clouds		2.0	Barlow[1]

*From Barton.[2]

wind velocity with altitude). Their measurements indicate that good average values are $\sigma_v = 1.0$ m/s for turbulence and $\sigma_v = 1.68$ m/(s/km) for wind shear. A convenient equation is $\sigma_v = 0.04R\theta_{el}$ m/s for the effects of wind shear, provided the rain fills the vertical beam, where R is the range to the weather, in nautical miles; and v_{el} is the one-way half-power vertical beamwidth, in degrees. Thus, for example, σ_v of rain viewed at 25 nmi with a vertical beamwidth of 4° would have a $\sigma_v = 4.1$ m/s, of which the shear component is dominant. Rain and chaff also have an average velocity, in addition to the spectral spread noted above, which must be taken into account when designing an MTI system.

The clutter spectral width in meters per second is independent of the radar frequency. The standard deviation of the clutter power spectrum σ_c, in hertz, is

$$\sigma_c = \frac{2\sigma_v}{\lambda} \quad \text{Hz} \tag{15.4}$$

where λ is the transmitted wavelength, in meters; and σ_v is the clutter standard deviation, in meters per second.

Antenna scanning also causes a spread of the clutter power spectrum because of amplitude modulation of the echo signals by the two-way antenna pattern.[2] The resulting clutter standard deviation is

$$\sigma_c = \frac{\sqrt{\ln 2}}{\pi} \times \frac{f_r}{n} = 0.265 \frac{f_r}{n} \quad \text{Hz} \tag{15.5}$$

where f_r is the PRF and n is the number of hits between the one-way 3 dB points of the antenna pattern. This equation was derived for a gaussian beam shape but is essentially independent of the actual beam shape or aperture illumination function used.

The clutter spectral spread due to scanning, normalized to the PRF, is

$$\sigma_c T = \frac{0.265}{n} \qquad (15.6)$$

where $T = 1/\text{PRF}$ is the interpulse period.

Amplitude Characteristics.   To predict the performance of an MTI system, the amplitude of the clutter signals with which a target must compete should be known. The amplitude of the clutter signals is dependent on the size of the resolution cell of the radar, the frequency of the radar, and the reflectivity of the clutter. The expected radar cross section of clutter can be expressed as the product of a reflectivity factor and the size of the volume or area of the resolution cell.

For surface clutter, as viewed by a surface-based radar,

$$\bar{\sigma} = A_c \sigma^0 = R\theta_{az}\frac{c\tau}{2}\sigma^0 \qquad (15.7)$$

where $\bar{\sigma}$ is the average radar cross section, in square meters; A_c is the area of clutter illuminated, in square meters; R is the range to clutter patch, in meters; θ_{az} is the one-way half-power azimuthal beamwidth, in radians; c is the speed of propagation, 300 million m/s; τ is the half-power radar pulse length (after the matched filter), in seconds; and σ^0 is the average clutter reflectivity factor, in square meters per square meter.

For clutter that is airborne, such as chaff or rain,

$$\bar{\sigma} = V_c\eta = R\theta_{az}H\frac{c\tau}{2}\eta \qquad (15.8)$$

where V_c is the volume of clutter illuminated, in cubic meters; H is the height of clutter, in meters (if clutter fills the vertical beam, then $H = R\theta_{el}$, where θ_{el} is the elevation beamwidth); and η is the clutter reflectivity factor, in square meters per cubic meter.

It should be noted that, for land clutter, σ^0 can vary considerably from one resolution cell to the next. A typical distribution of σ^0, taken from Barton,[8] is shown in Fig. 15.10. Typical values for σ^0 and η taken from the same reference are given in Table 15.2. Because of the imprecision in predicting σ^0 and η, these equations do not include an antenna beam-shape factor. For the measurement of the reflectivity of rain, references on radar meteorology present more precise equations.[9]

In addition to distributed clutter targets, there are many targets that appear as *points*, such as radio towers, water tanks, and buildings. These point targets typically have a radar cross section of 10^3 to 10^4 m^2.

Figure 15.11*a* shows a PPI display of all clutter observed with a surveillance radar with a 1.3° by 2-μs resolution cell in the mountainous region of Lakehead, Ontario, Canada. (The PPI range is set for 30 nmi.) Clutter that exceeds the minimum-discernible signal (MDS) level of the radar by 60 dB is shown in Fig. 15.11*b*. Note that the clutter in Fig. 15.11*b* is very spotty in character, including the strong fixed-point targets and returns from extended targets. It is significant that the extended targets are no longer very extended. The face of a mountain at 10 mi from 5 to 7 o'clock is only a line. If the MTI system were incapable of displaying an aircraft while it was over the mountain face, it would display the aircraft on the next scan of the antenna because the aircraft would have moved either farther or nearer. The PPI does not have a resolution that approaches the resolution of the signal-processing
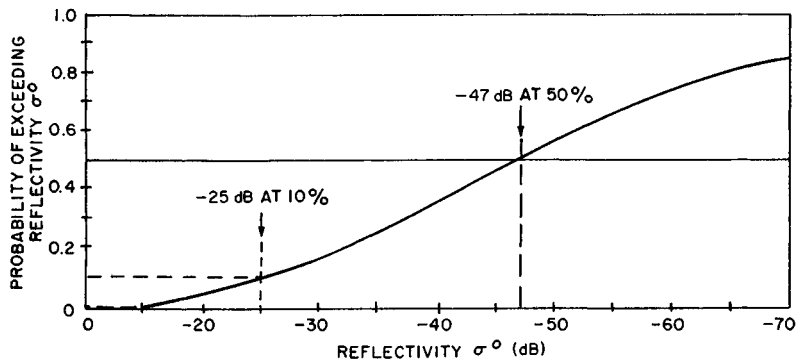
FIG. 15.10 Distribution of reflectivity for ground clutter typical of heavy clutter at S band.

circuits of this radar. Thus the apparent extended clutter has many weak areas not visible in these photographs, where targets could be detected by virtue of an MTI radar's interclutter visibility (defined in Sec. 15.4).

15.4 DEFINITIONS

Improvement Factor (I). The MTI improvement factor I is defined as "the signal-to-clutter ratio at the output of the clutter filter divided by the signal-to-clutter ratio at the input of the clutter filter, averaged uniformly over all target radial velocities of interest."[10] This definition accounts for both the clutter attenuation and the average noise gain of the MTI system. It is therefore a measure of the MTI system response to clutter relative to the average MTI system response to targets. An equivalent definition of improvement factor is $I = r_i/r_o$, where r_i is the input ratio of clutter to noise and r_o is the output ratio of clutter residue to noise. The use of I is encouraged instead of older terms, such as *cancellation ratio* and *clutter attenuation*, because these terms have not been consistently used in the literature and are not always normalized to the average canceler noise gain.

Signal-to-Clutter Ratio Improvement (I_{SCR}). For a system employing multiple doppler filters, such as the MTD, each filter will have a different improvement factor against the same clutter source. In this case it is preferable to define the performance against clutter in terms of the signal-to-clutter improvement (I_{SCR}) versus target doppler shift. This quantity is not included in the *IEEE Dictionary*,[10] but common usage defines the I_{SCR}, at each target doppler frequency, as the ratio of the signal-to-clutter ratio obtained at the output of the doppler filter bank (including all filters) to the signal-to-clutter ratio at the input of the filter bank. It should be noted that the signal-to-clutter improvement of any one filter is equal to the product of the MTI improvement factor of the filter as defined earlier and the coherent gain of the filter at the particular doppler frequency. The coherent gain of a doppler filter is equal to the increase in signal-to-thermal-noise ratio between the input and the output of the filter due to the coherent summation of individual target returns.

TABLE 15.2 Typical Values of Clutter Reflectivity*

Clutter	Reflectivity, λ, m η, $(m)^{-1}$	Conditions	Band λ, m	Clutter parameters for typical conditions			
				L 0.23	S 0.1	C 0.056	X 0.032
Land (excluding point clutter)	$\sigma^0 = \dfrac{0.00032}{\lambda}$ (worst 10 percent)	. . . .	σ^0 dB =	-29	-25	-22	-20
Point clutter	$\sigma = 10^4$ m^2	. . . .	σ m^2 =	10^4	10^4	10^4	10^4
Sea (Beaufort scale K_B, angle E)	σ^0 dB = $-64 + 6K_B +$ $(\sin E)$dB $- \lambda$ dB	Sea state 4 (6-ft waves, rough); $E = 1°$	σ^0 dB =	-51.5	-47.5	-44.5	-42.5
Chaff (for fixed weight per unit volume)	$\eta = 3 \times 10^{-8}\lambda$	. . . .	η (m)$^{-1}$ =	7×10^{-9}	3×10^{-9}	1.7×10^{-9}	10^{-9}
Rain (for rate r, mm/h)	$\eta = 6 \times 10^{-14}r^{1.6}\lambda^{-4}$ (matched polarization)	$r = 4$ mm/h	η (m)$^{-1}$ =	2×10^{-10}	5×10^{-9}	5×10^{-8}	5×10^{-7}

*From Barton.[8]

(a)   (b)
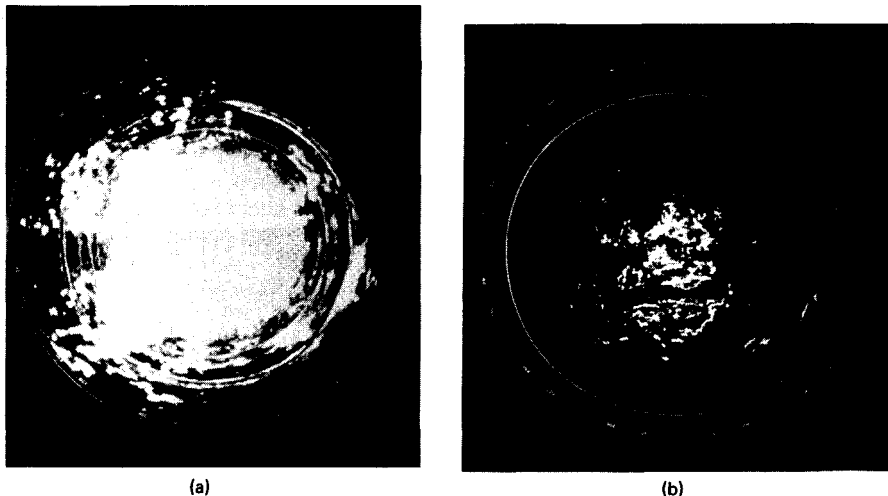
FIG. 15.11   PPI display, 30-nmi range, of (*a*) all clutter at a mountainous site and (*b*) clutter that exceeds the system noise level by 60 dB.

Subclutter Visibility (SCV).   The subclutter visibility (SCV) of a radar system is a measure of its ability to detect moving-target signals superimposed on clutter signals. A radar with 20 dB SCV can detect an aircraft flying over clutter whose signal return is 100 times stronger. The *IEEE Dictionary*[10] defines the subclutter visibility as "the ratio by which the target echo power may be weaker than the coincident clutter echo power and still be detected with specified detection and false alarm probabilities. Target and clutter powers are measured on a single pulse return and all target radial velocities are assumed equally likely." The SCV of two radars cannot necessarily be used to compare their performance while operating in the same environment, because the target-to-clutter ratio seen by each radar is proportional to the size of the radar resolution cell and may also be a function of frequency. Thus a radar with a 10-μs pulse length and a 10° beamwidth would need 20 dB more subclutter visibility than a radar with a 1-μs pulse and a 1° beamwidth for equal performance in a specified clutter environment.

The subclutter visibility of a radar, when expressed in decibels, is less than the improvement factor by the clutter visibility factor V_{oc} (see definition below).

Interclutter Visibility (ICV).   The interclutter visibility (ICV) of a radar is a measure of its capability to detect targets between points of strong clutter by virtue of the ability of the radar to resolve the areas of strong and weak clutter. A radar with high resolution makes available regions between points of strong clutter where the target-to-clutter ratio will be sufficient for target detection even though the SCV of the radar (based on average clutter) may be relatively low. A low-resolution radar averages the clutter over large resolution cells, most of which will contain one or more strong point targets, and thus the radar will have very little ICV. Because of the ICV capability of high-resolution radars, they tend to perform better in a clutter environment than would be predicted by using the average clutter amplitude characteristics of Sec. 15.3.[11] To achieve ICV, a mechanism must be furnished to provide CFAR operation

against the residue from strong clutter. This CFAR is provided in the typical MTI system by IF limiting or, in the MTD implementation, through the use of high-resolution clutter maps.

Filter Mismatch Loss. The maximum signal-to-noise ratio available from an N-pulse filter is N times the signal-to-noise ratio of a single pulse, assuming all pulses have equal amplitude. When weighting is applied to reject clutter and control the filter sidelobes, the peak output signal-to-noise ratio is reduced. The filter mismatch loss is the amount by which the peak-output signal-to-noise ratio is reduced by the use of the weighting. A three-pulse MTI filter using binomial weights has a filter mismatch loss of 0.51 dB. The mismatch loss for the binomial-weighted four-pulse canceler is 0.97 dB.

Clutter Visibility Factor (V_{oc}). This factor is "the predetection signal-to-clutter ratio that provides stated probabilities of detection and false alarm on a display; in moving-target-indicator systems, it is the ratio after cancelation or doppler filtering."[10] The clutter visibility factor is the ratio by which the target signal must exceed the clutter residue so that target detection can occur without having the clutter residue result in false-target detections. The system must provide a threshold that the targets will cross and the clutter residue will not cross.

15.5 IMPROVEMENT FACTOR CALCULATIONS

Using Barton's approach (Ref. 2, pp. 210–219), the maximum improvement factor I against zero-mean clutter with a gaussian-shaped spectrum, for different implementations of the finite-impulse-response binomial-weight MTI canceler (see Sec. 15.7), is

$$I_1 \approx 2\left(\frac{f_r}{2\pi\sigma_c}\right)^2 \tag{15.9}$$

$$I_2 \approx 2\left(\frac{f_r}{2\pi\sigma_c}\right)^4 \tag{15.10}$$

$$I_3 \approx \frac{4}{3}\left(\frac{f_r}{2\pi\sigma_c}\right)^6 \tag{15.11}$$

where I_1 is the MTI improvement factor for the single-delay coherent canceler; I_2 is the MTI improvement factor for the dual-delay coherent canceler; I_3 is the MTI improvement factor for the triple-delay coherent canceler; σ_c is the rms frequency spread of the gaussian clutter power spectrum, in hertz; and f_r is the radar repetition frequency, in hertz. When the values of σ_c for scanning modulation [Eq. (15.5)] are substituted in the above equations for I, the limitation on I due to scanning is

$$I_1 \approx \frac{n^2}{1.39} \tag{15.12}$$

$$I_2 \approx \frac{n^4}{3.84} \tag{15.13}$$

$$I_3 \approx \frac{n^6}{16.0} \tag{15.14}$$

These relationships are shown graphically in Fig. 15.12. This derivation assumes a linear system. That is, it is assumed that the voltage envelope of the echo signals, as the antenna scans past a point target, is identical to the two-way antenna voltage pattern. This assumption of a linear system may be unrealistic for some practical MTI systems with relatively few hits per beamwidth, as discussed in Sec. 15.10.

The scanning limitation does not apply to a system that can step-scan, such as a phased array. Note, however, that sufficient pulses must be transmitted to initialize the filter before useful outputs may be obtained. For example, with a three-pulse binomial-weight canceler, the first two transmitted pulses initialize the canceler, and useful output is not available until after the third pulse has been transmitted. Feedback or infinite impulse response (IIR) filters would not be used with a step-scan system because of the long transient settling time of the filters.

The limitation on I due to internal-clutter fluctuations can be determined by substituting the appropriate value of σ_c into Eqs. (15.9) to (15.11). By letting $\sigma_c = 2\sigma_v/\lambda$, where σ_v is the rms velocity spread of the clutter, the limitation on I can be plotted for different types of clutter as a function of the wavelength λ and the pulse repetition frequency f_r. This is done for one-, two-, and three-delay binomial-weight cancelers in Figs. 15.13 to 15.15. The values of V_B given are the first blind speed of the radar (or where the first blind speed V_B would be for a staggered PRF system if staggering were not used). The improvement factor
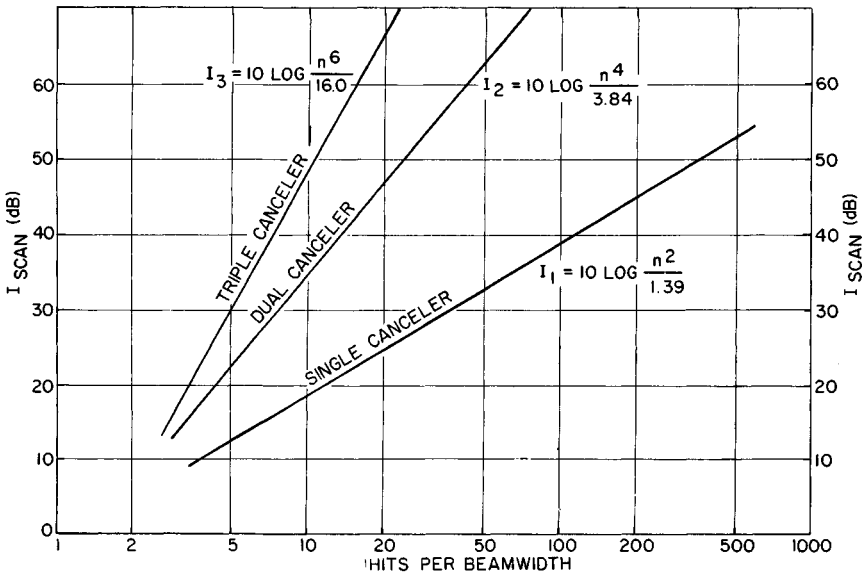
FIG. 15.12 Theoretical MTI improvement factor due to scan modulation; gaussian antenna pattern; n = number of pulses within the one-way half-power beamwidth.
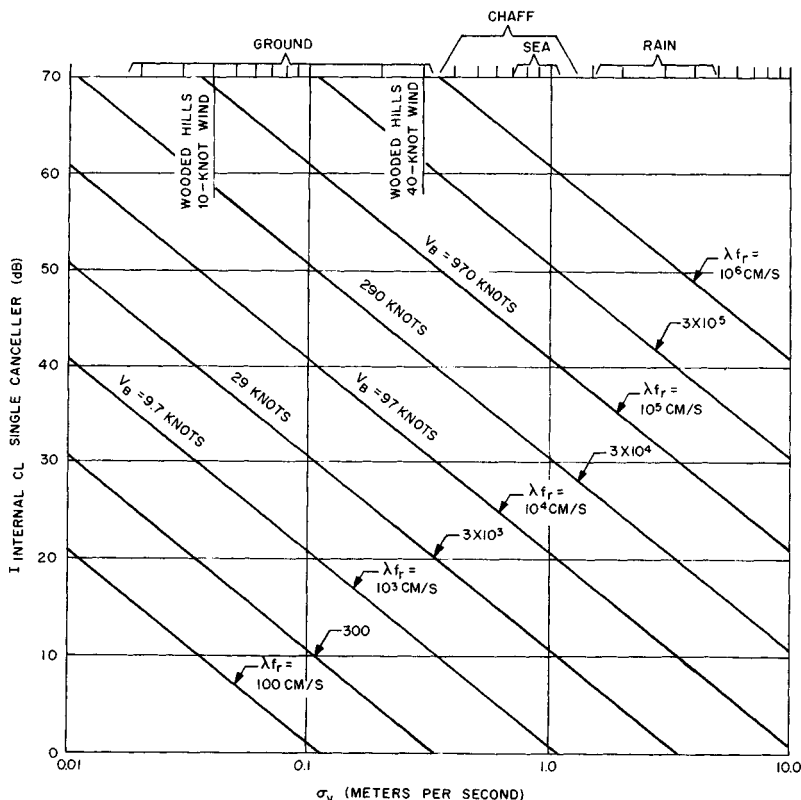
FIG. 15.13 MTI improvement factor as a function of the rms velocity spread of clutter for a two-pulse binomial-weight canceler.

shown in these figures for rain and chaff is based on the assumption that the average velocity of the rain and chaff has been compensated for so that the returns are centered in the canceler rejection notch. Unless such compensation is provided, the MTI offer little or no improvement for rain and chaff.

Two further limitations on I are the effect of pulse-to-pulse repetition-period staggering combined with clutter spectral spread from scanning and internal-clutter motion. These limitations, plotted in Figs. 15.16 and 15.17, apply to all cancelers, whether single or multiple. (The derivation of these limitations and a means of avoiding them by the use of time-varying weights are given in Sec. 15.9.)

15.6 OPTIMUM DESIGN OF CLUTTER FILTERS

The statistical theory of detection of signals in gaussian noise provides the required basis for the optimum design of radar clutter filters. Such theoretical results are important to the designer of a practical MTI or MTD system, in that
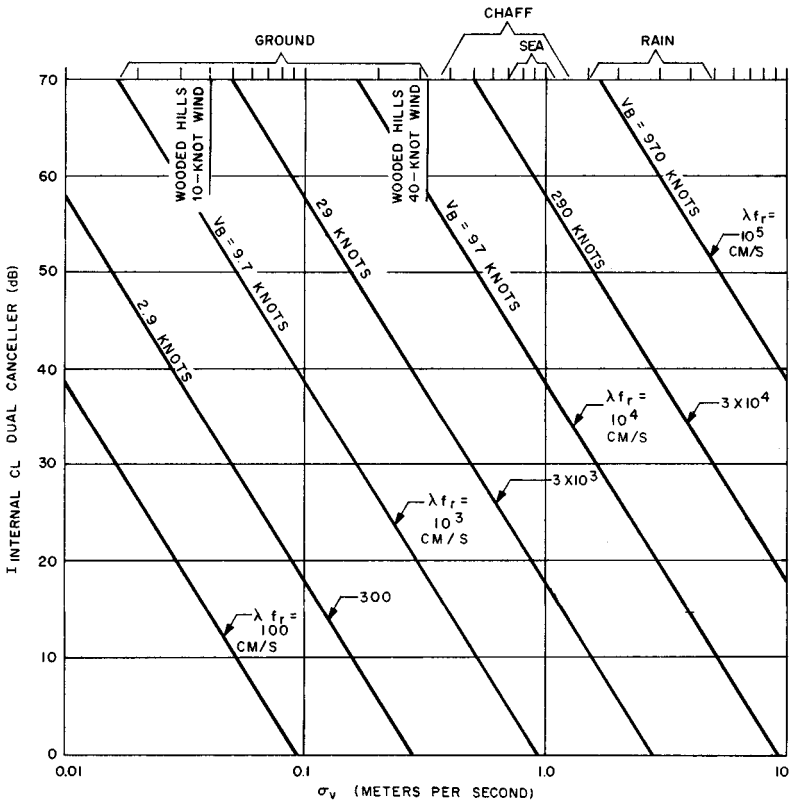
FIG. 15.14 MTI improvement factor as a function of the rms velocity spread of clutter for a three-pulse binomial-weight canceler.

they establish upper bounds on the achievable performance in a precisely specified clutter environment. It should be noted, however, that owing to the extreme variability of the characteristics of real clutter returns (power level, doppler shift, spectrum shape, spectral width, etc.) any attempt to actually approximate the performance of such optimum filters for the detection of targets in clutter requires the use of adaptive methods. The adaptive methods must estimate the unknown clutter statistics and subsequently implement the corresponding optimum filter. The design of such adaptive MTI systems is discussed in Sec. 15.13.

For a single radar pulse with a duration of a few microseconds, the doppler shift due to aircraft target motion is a small fraction of the signal bandwidth, and conventional MTI and pulse doppler processing are not applicable. It is well known that the classical single-pulse "matched" filter provides optimum radar detection performance when used in a white-noise background. Against clutter returns which have the same spectrum as the transmitted radar pulse, the matched filter is no longer optimum, but the potential improvement in the output signal-to-clutter ratio by designing a modified optimized filter is usually insignificant.

When the duration of the transmitted radar signal, whether CW or a repetitive
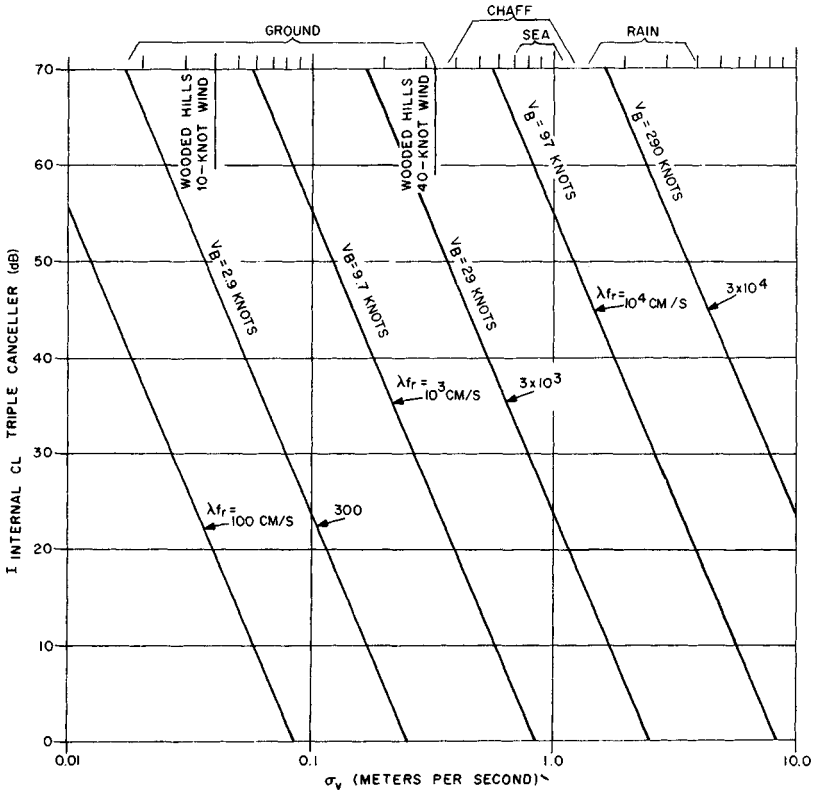
FIG. 15.15 MTI improvement factor as a function of the rms velocity spread of clutter for a four-pulse binomial-weight canceler.

train of N identical pulses, is comparable with or larger than the reciprocal of anticipated target doppler shifts, the difference between a conventional white-noise matched filter (or coherent integrator) and a filter optimized to reject the accompanying clutter becomes significant. The characteristics of the clutter are characterized by the covariance matrix Φ_C of the N clutter returns. If the power spectrum of the clutter is denoted $S_C(f)$ and the corresponding autocorrelation function is $R_C(t_i - t_j)$, then the elements of Φ_C are given by

$$\Phi_{ij} = R_C(t_i - t_j) \tag{15.15}$$

where t_i is the transmission time of the ith pulse. For example, for a gaussian-shaped clutter spectrum we have

$$S_C(f) = P_C \frac{1}{\sqrt{2\pi}\, \sigma_f} \exp\left[- \frac{(f - f_d)^2}{2\sigma_f^2} \right] \tag{15.16}$$

where P_C is the total clutter power, σ_f is the standard deviation of the clutter
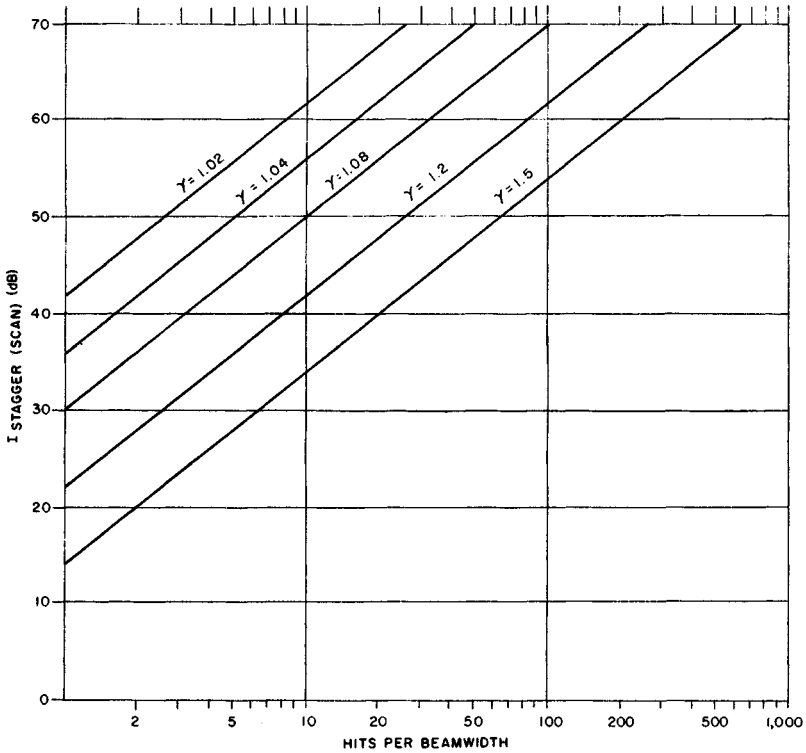
FIG. 15.16 Approximate MTI improvement factor limitation due to pulse-to-pulse repetition-period staggering and scanning (all canceler figurations). I(dB) = 20 log [2.5n/ (γ − 1)]; γ = maximum period/minimum period.

spectral width, and f_d is the average doppler shift of the clutter. The corresponding autocorrelation function is

$$R_C(\tau) = P_C \exp(-4\pi\sigma_f^2\tau^2) \exp(-j2\pi f_d\tau) \qquad (15.17)$$

For two pulses separated in time by the interpulse period T the complex correlation coefficient between two clutter returns is

$$\rho_T = \exp(-4\pi\sigma_f^2 T^2) \exp(-j2\pi f_d T) \qquad (15.18)$$

The second factor in this expression represents the phase shift caused by the doppler shift of the clutter returns.

For a known target doppler shift the received target return can be represented by an N-dimensional vector:

$$s = P_S\mathbf{f} \qquad (15.19)$$

where the elements of the vector $\mathbf{f}$ are $f_i = \exp[j2\pi f_s t_i]$. On the basis of this de-
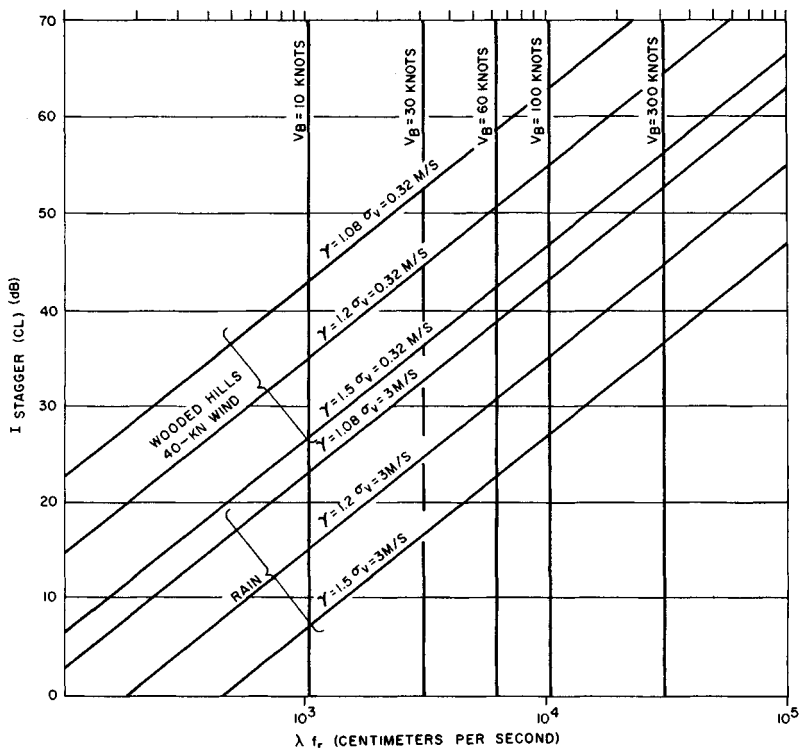
FIG. 15.17 Approximate MTI improvement factor limitation due to pulse-to-pulse staggering and internal-clutter motion (all canceler configurations). $I(dB) = 20 \log [0.33/(\gamma - 1)(\lambda f_r/\sigma_v)]$; γ = maximum period/minimum period.

scription of signal and clutter it has been shown[12] that the optimum doppler filter will have weights given by

$$w_{opt} = \Phi_C^{-1} s \qquad (15.20)$$

and the corresponding signal-to-clutter improvement is

$$I_{SCR} = \frac{w_{opt}{}^T s \cdot s^{T*} w_{opt}{}^*}{w_{opt}{}^T \Phi_C w_{opt}{}^*} \qquad (15.21)$$

where the asterisk denotes complex conjugation and superscript T is the transposition operator. An example where the optimum performance is determined for the case of clutter at zero doppler having a wide gaussian-shaped spectrum and a normalized width of $\sigma_f T = 0.1$ is shown in Fig. 15.18. In this case a coherent processing interval of CPI = nine pulses was assumed, and the limitation due to thermal noise was ignored by setting the clutter level at 100 dB above noise.

It should be kept in mind that Eq. (15.21) for the optimum weights will yield a different result for each different target doppler shift, so that a large number of
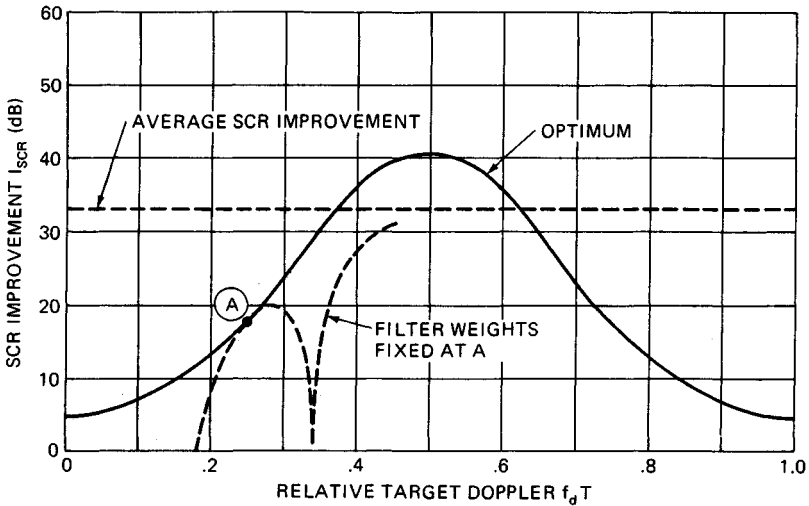
FIG. 15.18 Optimum signal-to-clutter ratio improvement (I_{SCR}) for gaussian-shaped clutter spectrum and a CPI of nine pulses; clutter-to-noise ratio, 100 dB.

parallel filters would be needed to approximate the optimum performance even when the clutter characteristics are known exactly. As an example, the response of the optimum filter designed for one particular target doppler frequency labeled as point A in Fig. 15.18 is shown in a broken line. At approximately ±5 percent from the design doppler the performance starts to fall significantly below the optimum.

Also shown in Fig. 15.18 is a horizontal line labeled "average SCR improvement." This indicates the level corresponding to the average of the optimum SCR curve across one doppler interval and may be considered as a figure of merit for a multiple-filter doppler processor somewhat analogous to the MTI improvement factor defined for a single doppler filter. In Fig. 15.19 the optimum average I_{SCR} has been computed for several different values of the CPI as a function of the normalized spectrum width. These results may be used as a point of reference for practical doppler processor designs as discussed in Sec. 15.8. Note that for $\sigma_f T \approx 1$ the average SCR improvement is due only to the coherent integration of all the pulses in the CPI.

The implementation of a single MTI filter will result in a performance below that shown in Fig. 15.19. Further, it can be shown that the average SCR improvement calculated for a single filter is equal to the MTI improvement factor as defined in Sec. 15.4. The basis for obtaining the optimum MTI filter is again the covariance matrix of the clutter returns as given by Eq. (15.15). As shown by Capon,[13] the weights of the optimum MTI are found as the eigenvector corresponding to the smallest eigenvalue of the clutter covariance matrix and the MTI improvement factor is equal to the inverse of the smallest eigenvalue.

In Fig. 15.20 the improvement factor of an MTI using the optimum weights is compared with the binomial coefficient MTI for different values of the relative clutter spectral spread and shown as a function of the number of pulses in the CPI. These results again assume a gaussian-shaped clutter spectrum. For typical
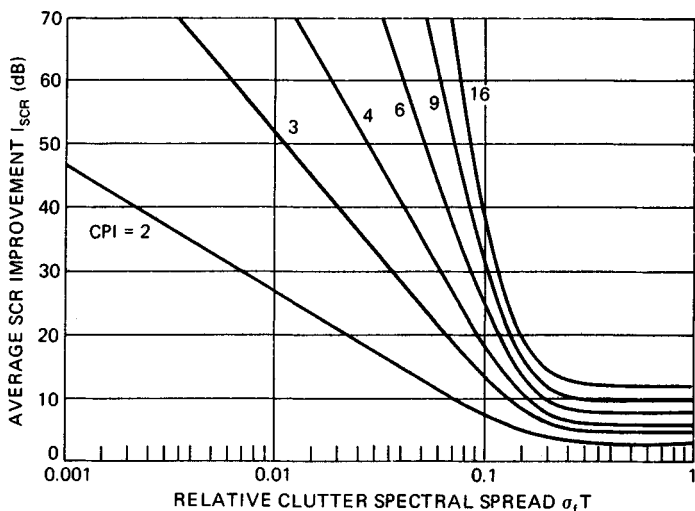
FIG. 15.19 Reference curve of optimum average SCR improvement for a gaussian-shaped clutter spectrum.
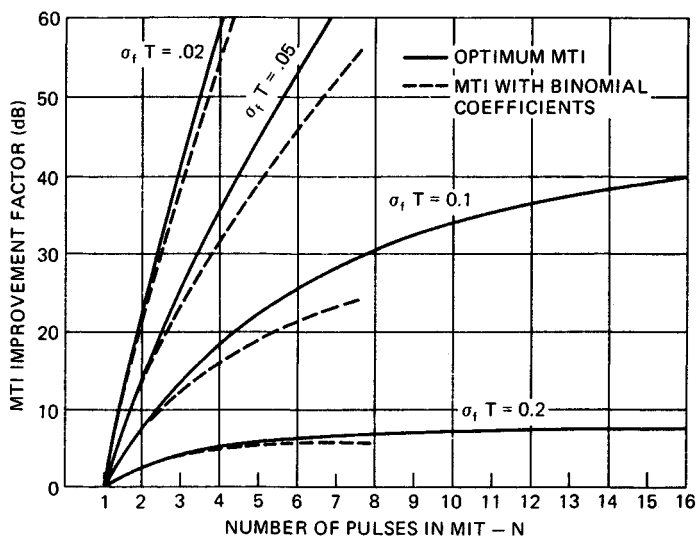
FIG. 15.20 Comparison of MTI improvement factor of binomial-weight MTI and optimum MTI against a gaussian-shaped clutter spectrum.

numbers of pulses in the MTI (three to five) the binomial coefficients are remarkably robust and provide a performance which is within a few decibels of the optimum. Again, it should be noted that any attempt to implement an MTI canceler, which performs close to the optimum, would require the use of adaptive techniques which estimate the clutter characteristics in real time. If the estimate is in

error, the actual performance may fall below that of the binomial-weight MTI canceler.

15.7   MTI CLUTTER FILTER DESIGN

The MTI block diagram shown in Fig. 15.3 and discussed in detail in Sec. 15.2 uses a single-delay canceler. It is possible to utilize more than one delay and to introduce feedback and/or feedforward paths around the delays to change the MTI system response to targets of different velocities. Multiple-delay cancelers have wider clutter rejection notches than single-delay cancelers. The wider rejection notch encompasses more of the clutter spectrum and thus increases the MTI improvement factor attainable with a given clutter spectral distribution.

When a number of single-delay feedforward cancelers are cascaded in series, the overall filter voltage response is $k2^n \sin^n (\pi f_d T)$, where k is the target amplitude, n is the number of delays, f_d is the doppler frequency, and T is the interpulse period.[15] The cascaded single-delay cancelers can be rearranged as a transversal filter, and the weights for each pulse are the binomial coefficients with alternating sign: 1, -1 for two pulses; 1, -2, 1 for three pulses; 1, -3, 3, -1 for four pulses, etc. Changes of the binomial feedforward coefficients and/or the addition of feedback modify the filter characteristics. Within this chapter, reference to *binomial-weight cancelers* refers to cancelers with the $2^n \sin^n (\pi f_d T)$ transfer function.

Figures 15.21 to 15.23 represent typical velocity response curves obtainable from one-, two-, and three-delay cancelers. Shown also are the canceler configurations assumed, with appropriate Z-plane pole-zero diagrams. The Z plane is the comb-filter equivalent of the S plane,[16,17] with the left-hand side of the S plane transformed to the inside of the unit circle centered at $Z = 0$. Zero frequency is at $Z = 1 + j0$. The stability requirement is that the poles of the Z transfer function lie within the unit circle. Zeros may be anywhere.

These velocity response curves are calculated for a scanning radar system with 14.4 hits per beamwidth. An antenna beam shape of $(\sin U)/U$, terminated at the first nulls, was assumed. The shape of these curves, except very near the blind speeds, is essentially independent of the number of hits per beamwidth or the assumed beam shape.

The ordinate, labeled "response," represents the single-pulse signal-to-noise response of the MTI receiver relative to the signal-to-noise response of a normal linear receiver for the same target. Thus all the response curves are normalized with respect to the average gain for the given canceler configuration. The intersection at the ordinate represents the negative decibel value of I, the MTI improvement factor for a point clutter target processed in a linear system.

Because these curves show the signal-to-noise response for each output pulse from the MTI canceler, the inherent loss incurred in MTI processing due to the reduction of the effective number of independent pulses integrated[18] is not apparent. This loss may vary from 1½ to almost 3 dB, depending upon the number of pulses on target. In addition, if quadrature MTI channels (see Sec. 15.12) are not employed, there is an additional loss of 1½ to 3 dB, again depending upon the number of pulses on target.

The abscissa of these curves, V/V_B, represents the ratio of target velocity V to the blind speed $V_B = \lambda f_r/2$, where λ is the radar wavelength and f_r is the average
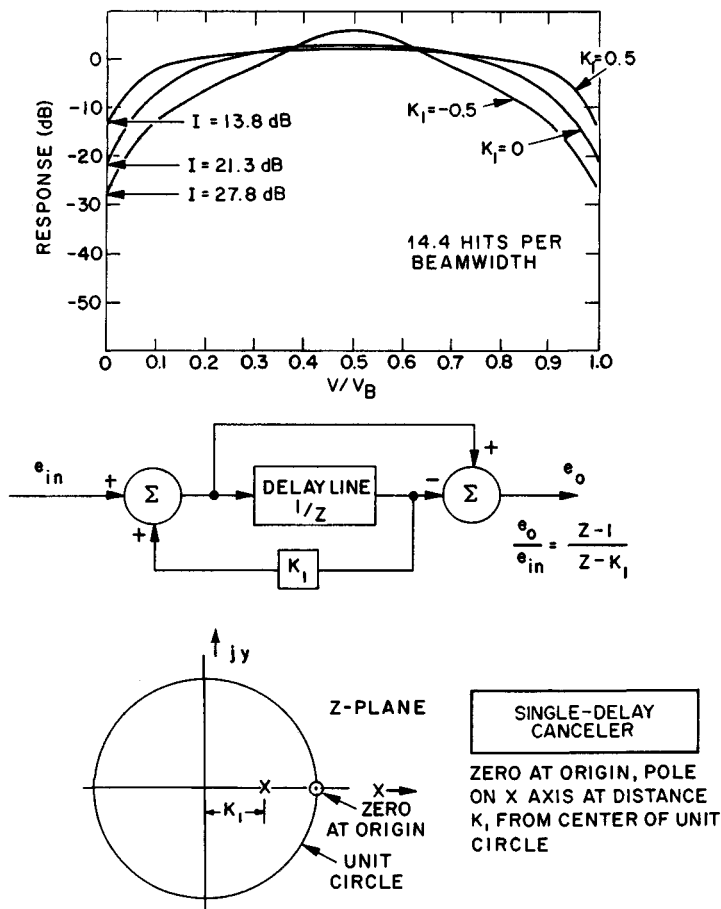
FIG. 15.21   One-delay canceler.

PRF of the radar. The abscissa can also be interpreted as the ratio of the doppler frequency to the average PRF of the radar.

The canceler configurations shown are not the most general feedforward, feedback networks possible but, rather, are practical configurations easy to implement. Many configurations are computationally equivalent. More flexibility in locating zeros and poles is achieved with delays in pairs as shown for the second and third delays of the triple canceler. (In this configuration, the zeros are constrained to the unit circle.)

The triple-canceler configuration is such that two of the zeros can be moved around the boundary of the unit circle in the Z plane. Moving the zeros gives a 4 or 5 dB increase in the MTI improvement factor for specific clutter spectral spreads, as compared with keeping all three zeros at the origin.[19]

It is interesting to note the width of the rejection notches for the different

The graph shows response curves with the following annotations:
- $K_1 = 0.5, K_2 = 0$
- $K_1 = 0.2, K_2 = 0.6$
- $K_1 = 0, K_2 = 0$
- $K_1 = 0.2, K_2 = -0.6$

14.4 HITS PER BEAMWIDTH

$I = 28.5$ dB
$I = 37.7$ dB
$I = 39.9$ dB
$I = 46.0$ dB

$$\frac{e_0}{e_{in}} = \frac{(Z-1)^2}{Z^2-(K_1+K_2)Z+K_1}$$

DUAL-DELAY CANCELER

TWO ZEROS AT ORIGIN

POLES POSITIONED AS SHOWN
DISTANCE FROM CENTER OF
UNIT CIRCLE = $\sqrt{K_1}$
DISTANCE FROM ORIGIN $\sqrt{1-K_2}$

$$\theta = \text{ARC COS} \frac{K_1+K_2}{2\sqrt{K_1}}$$

POLES SHOWN FOR $K_1 = 0.5$, $K_2 = 0$
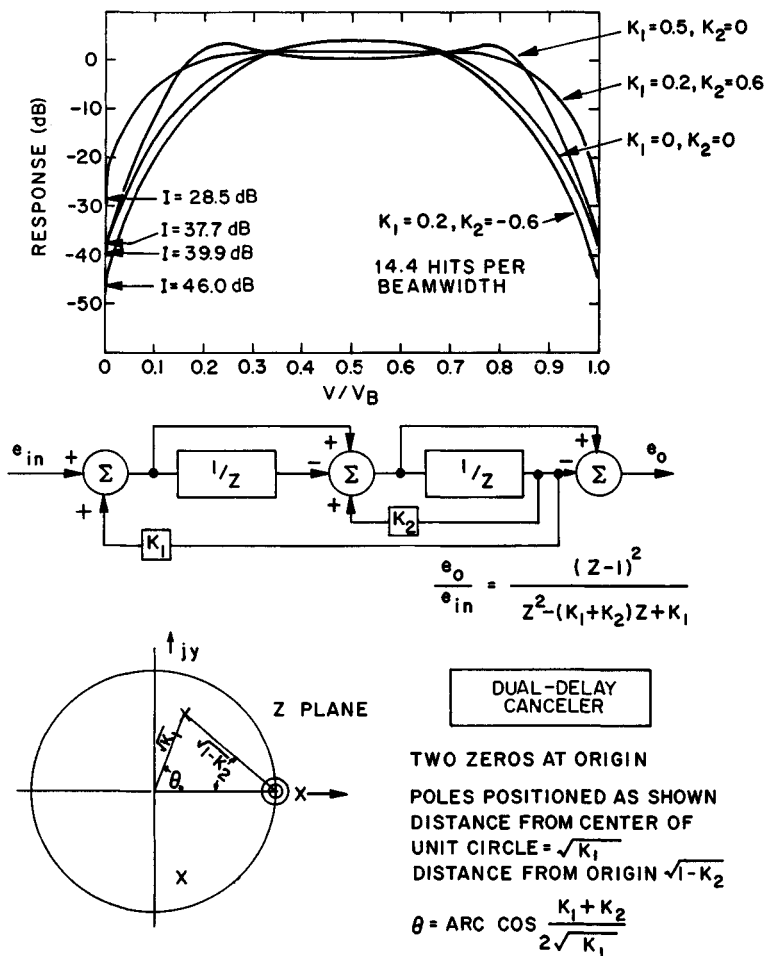
FIG. 15.22 Two-delay canceler.

binomial-weight canceler configurations. If the -6 dB response relative to normal response is used as the measuring point, the rejection is 24 percent of all target dopplers for the single canceler, 36 percent for the dual canceler, and 45 percent for the triple canceler. Consider the dual canceler, for example. Eliminating 36 percent of the dopplers means limiting the system to a long-term average of 64 percent single-scan probability of detection. Feedback can be used to narrow the rejection notch without much degradation of I. If feedback is used to increase the improvement factor, the single-scan probability of detection becomes worse.

Figure 15.24 shows the effect of feedback on I. These curves are calculated for a $(\sin U)/U$ antenna pattern terminated at the first nulls. The no-feedback curves shown here are almost indistinguishable from the theoretical curves derived for a
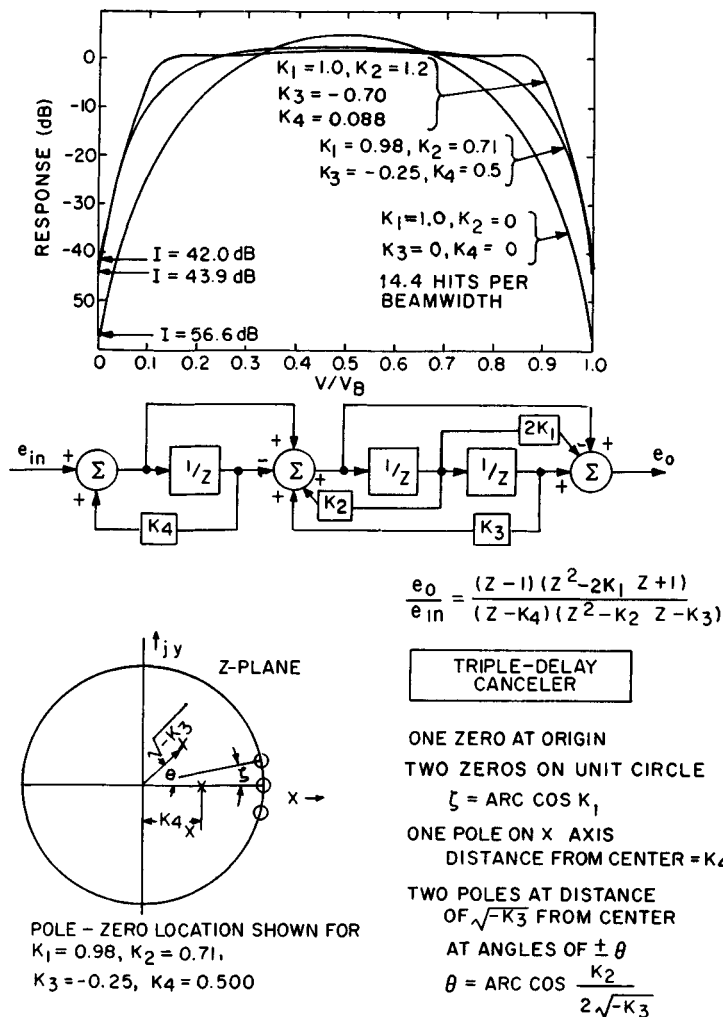
FIG. 15.23 Three-delay canceler.

gaussian pattern shown in Fig. 15.12. (One of the curves showing the effect of feedback on the triple canceler is not straight because two of the three zeros are not at the origin but have been moved along the unit circle the optimum amount for 14 hits per beamwidth. Thus, at 40 hits per beamwidth, these two zeros are too far removed from the origin to be very effective.)

In theory, it is possible to synthesize almost any velocity response curve with digital filters.[16] For each pair of poles and pair of zeros on the Z plane, two delay sections are required. The zeros are controlled by the feedforward paths, and the poles by the feedback paths.

Velocity response shaping can be accomplished by the use of feedforward only,
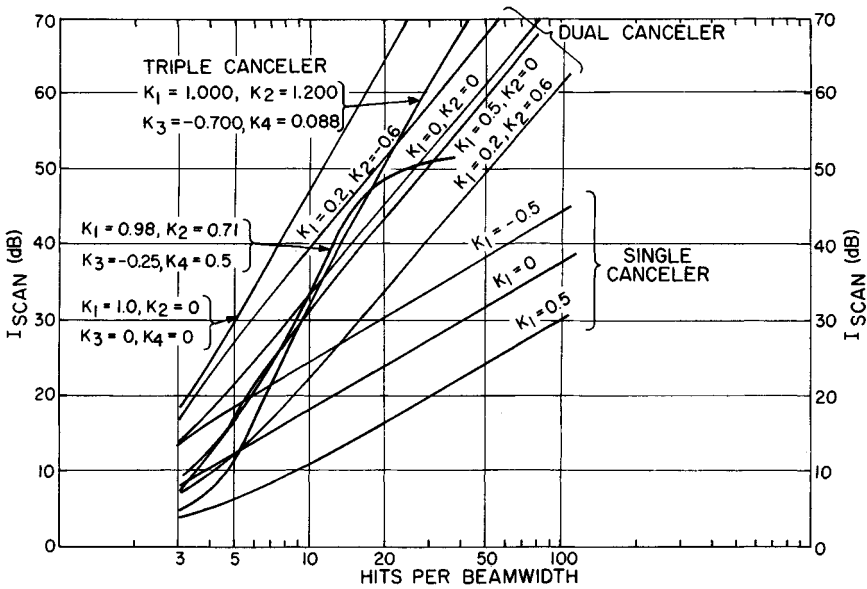
FIG. 15.24 Improvement factor limitation due to scanning for cancelers with feedback, computer-calculated for an assumed antenna pattern of (sin U)/U terminated at the first nulls.

without the use of feedback. The principal advantage of not using feedback is the excellent transient response of the canceler, an important consideration in a phased array or when pulse interference noise is present. If a phased array radar uses a feedback canceler, many pulses may have to be gated out after the beam has been repositioned before canceler transient ringing has settled to a tolerable level. An initialization technique has been proposed[20] to alleviate this problem, but it provides only partial reduction in the transient settling time. If feedforward only is used, only three or four pulses have to be gated out after moving the beam. The disadvantage of using feedforward for velocity response shaping is that an additional delay must be provided for each zero used to shape the response. Also, an inherent loss in improvement factor capability is caused by using zeros to shape the velocity response. This may or may not be significant, depending on the clutter spectral spread and the number of zeros available for cancellation. Figure 15.25 shows the velocity response and Z-plane diagram of a feedforward-only, shaped-response four-pulse canceler. Also shown are the velocity responses of a five-pulse feedforward canceler and a three-pulse feedback canceler. For the cancelers shown, the improvement factor capability of the three-pulse canceler is about 4 dB better than the shaped-response four-pulse feedforward canceler, independent of clutter spectral spread.

The five-pulse canceler response shown is a linear-phase[21] MTI filter described by Zverev.[22] The four zeros are located on the Z-plane real axis at $+1.$, $+1.$, -0.3575, and -2.7972. Much of the literature on filter synthesis describes linear-phase filters, but for MTI applications linear phase is of no importance. Almost identical filter responses can be obtained with nonlinear-phase filters that require fewer pulses, as shown in Fig. 15.25. Because only a fixed number of pulses is available during the time on target, none should be wasted. Thus one should choose the nonlinear-phase filter that uses fewer pulses.
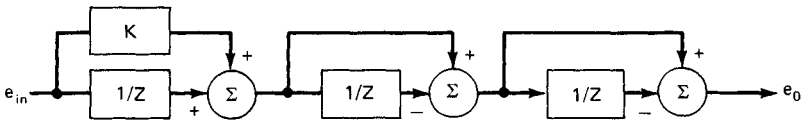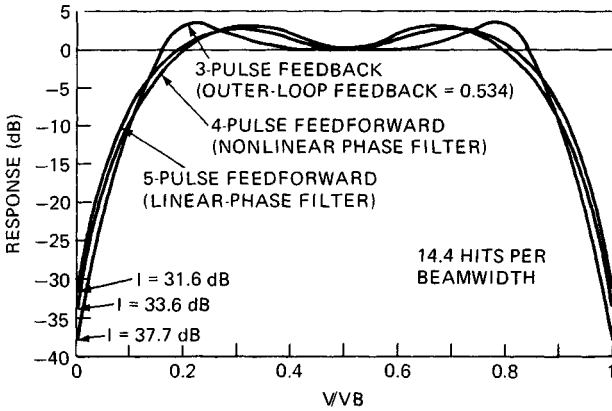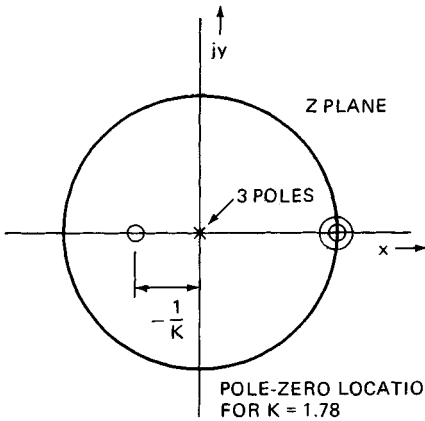
FIG. 15.25 Shaped-velocity-response feedforward cancelers compared with three-pulse feedbac
canceler. See text for five-pulse canceler parameters.

15.8 CLUTTER FILTER BANK DESIGN

As discussed in Sec. 15.1, the MTD uses a waveform consisting of coherent processing intervals (CPIs) of N pulses at the same PRF and RF frequency. The PRF and possibly the RF are changed from one CPI to the next. With this constraint only finite-impulse-response (FIR) filter designs are realistic candidates for the filter bank design. (Feedback filters require a number of pulses to settle after either the PRF or the RF is changed and thus would not be practical.)

The number of pulses available during the time when a surveillance radar beam illuminates a potential target position is determined by system parameters and requirements such as beamwidth, PRF, volume to be scanned, and the required data update rate. Given the constraint of the number of pulses on target, one must decide how many CPIs should occur during the time on target and how many pulses per CPI. The compromise is usually difficult. One wishes to use more pulses per CPI to enable the use of better filters, but one also wishes to have as many CPIs as possible. Multiple CPIs (at different PRFs and perhaps at different RF frequencies) improve detection and can provide information for true radial velocity determination.[23]

The design of the individual filters in the doppler filter bank is a compromise between the frequency sidelobe requirement and the degradation in the coherent integration gain of the filter. The number of doppler filters required for a given length of the CPI is a compromise between hardware complexity and the straddling loss at the crossover between filters. Finally the requirement of providing a high degree of clutter suppression at zero doppler (land clutter) sometimes introduces special design constraints.

When the number of pulses in a CPI is large (≥ 16), the systematic design procedure and efficient implementation of the fast Fourier transform (FFT) algorithm is particularly attractive. Through the use of appropriate weighting functions of the time-domain returns in a single CPI, the resulting frequency sidelobes can be readily controlled. Further, the number of filters (= the order of the transform) needed to cover the total doppler space (= the radar PRF) can be chosen independently of the CPI, as discussed below.

As the CPI becomes smaller (≤ 10), it will become important to consider special designs of the individual filters to match the specific clutter suppression requirements at different doppler frequencies in order to achieve better overall performance. While some systematic procedures are available for designing FIR filters subject to specific passband and stopband constraints, the straightforward approach for small CPIs is to use an empirical approach in which the zeros of each filter are adjusted until the desired response is obtained. An example of such filter designs is presented below.

Empirical Filter Design.   An example of an empirical filter design for a six-pulse CPI follows. (The six pulses per CPI may be driven by system considerations, such as time on target.) Because the filter will use six pulses, only five zeros are available for the filter design: the number of zeros available is the number of pulses minus one. The filter design process consists of placing the zeros to obtain a filter bank response that conforms to the specified constraints. The example that follows was produced with an interactive computer program with which the zeros could be moved until the desired response was obtained. The assumed filter requirements are as follows:

1. Provide a response of −66 dB in the clutter rejection notch (relative to the peak target response) of the moving-target filters.

2. Provide a response of −46 dB for chaff rejection at velocities between ±20 percent of the ambiguous doppler frequency range.

3. Owing to hardware limitations, only five filters will be implemented.

4. Three of the five filters will reject fixed clutter and respond to moving targets. Two filters will respond to targets at zero doppler and its ambiguities. (With good fixed clutter rejection filters, it takes two or more coherent filters to cover the gap in response at zero velocity.)

With the above considerations, a filter bank can be constructed.

Figure 15.26a shows the filter designed to respond to targets in the middle of the doppler passband. The sidelobes near zero velocity are 66 dB down from the peak, thus providing good clutter rejection for clutter within 5 percent of zero doppler. The −46 dB sidelobe provides chaff rejection to ±16 percent. Because of the constraint of only 5 zeros available, this filter could not provide −46 dB rejection to ±20 percent.
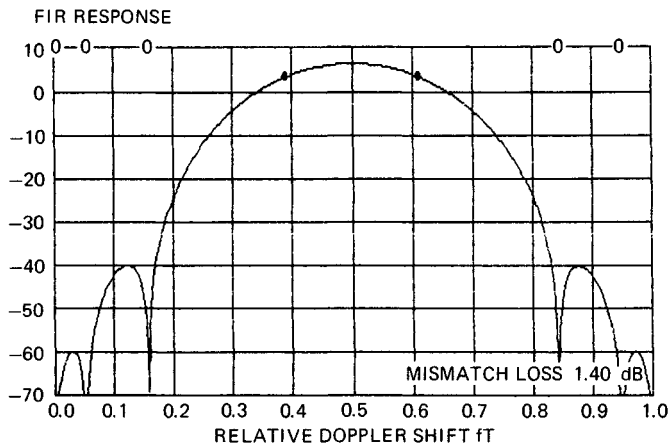
FIG. 15.26a Six-pulse filter for targets at $fT = 0.5$.

Figure 15.26b shows the filter that responds to targets as near as possible to zero doppler, while having zero-doppler response of −66 dB. Two zeros are placed near zero, providing −66 dB response to clutter at zero. The filter sidelobes between 0.8 and 1.0 doppler provide the specified chaff rejection of 48 dB. A mirror image of this filter is used for the third moving doppler filter. (The mirror-image filter has coefficients that are complex conjugates of the original filter coefficients.)

Figure 15.26c shows the first filter designed for response at zero doppler. Considerations here are that the straddling loss of the filter bank be minimized (this dictates the location of the peak), that the response to chaff at 0.8 doppler be down 46 dB, and that the mismatch loss be minimized. Minimizing the mismatch loss is accomplished by permitting the filter sidelobes between 0.3 and 0.8 to rise as high as needed (lower sidelobes in this range increase the mismatch loss). The second zero-doppler filter is the mirror image of this one.
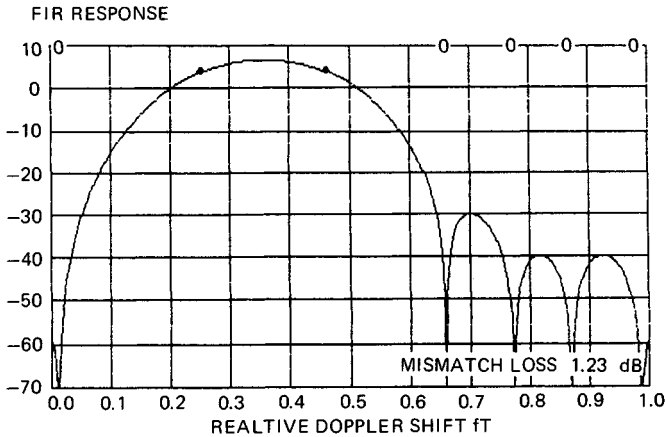
FIR RESPONSE

FIG. 15.26*b*   Six-pulse filter for targets at *fT* = 0.3 that rejects fixed clutter.
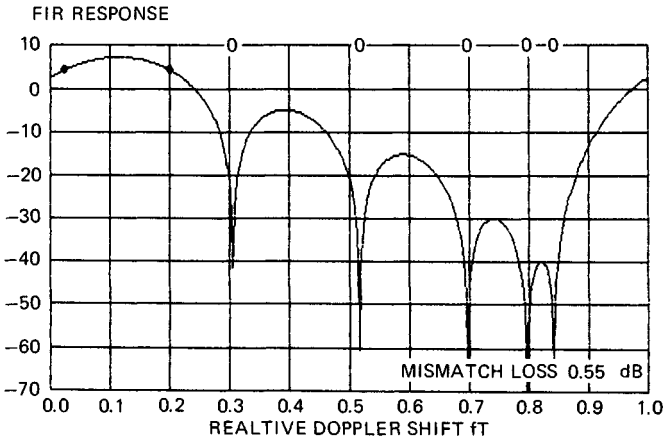
FIR RESPONSE

FIG. 15.26*c*   Six-pulse filter that responds to targets at zero doppler but rejects chaff at *fT* = 0.8.

Figure 15.26*d* shows the composite response of the filter bank. Note that the filter peaks are fairly evenly distributed. The dip between the first zero-doppler filter and the first moving doppler filter is larger than the others, primarily because, under the constraints, it is impossible to move the first doppler filter nearer to zero velocity.

Chebyshev Filter Bank.   For larger number of pulses in the CPI a more systematic approach to filter design is desirable. If a doppler filter design criterion is chosen that requires the filter sidelobes outside the main response to be below a specified level (i.e., providing a constant level of clutter suppression), while simultaneously minimizing the width of the filter response, a filter design based on the Dolph-Chebyshev distribution provides the optimum solution. Properties and design procedures based on the Dolph-Chebyshev distribution can be found in
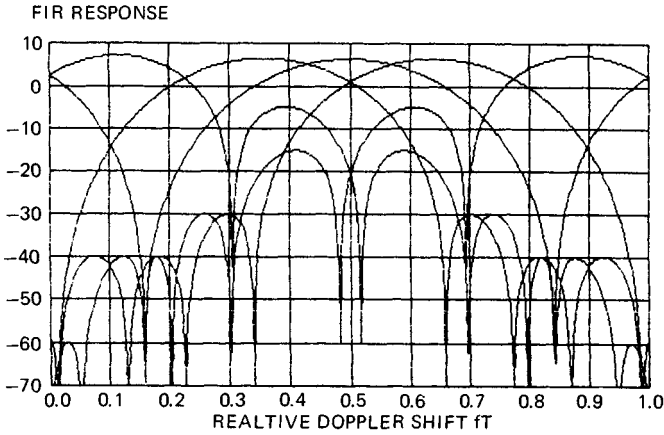
FIR RESPONSE

FIG. 15.26d   Composite response of the bank for five six-pulse filters.
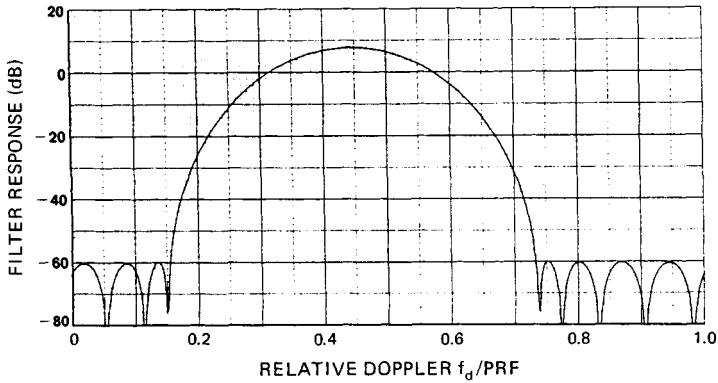
FIG. 15.27   Chebyshev FIR filter design with 68 dB doppler sidelobes.

the antenna literature. An example of a Chebyshev filter design for a CPI of nine pulses and a sidelobe requirement of 68 dB is shown in Fig. 15.27. The peak filter response can be located arbitrarily in frequency by adding a linear-phase term to the filter coefficients.

The total number of filters implemented to cover all doppler frequencies is a design option trading straddling loss at the filter crossover frequencies against implementation complexity. An example of a complete doppler filter bank implemented with nine uniformly spaced filters is shown in Fig. 15.28. The performance of this doppler filter bank against the clutter model considered in Fig. 15.18 is shown in Fig. 15.29. This graph shows the signal-to-clutter ratio improvement against clutter at zero doppler as a function of target doppler frequency. Only the response of the filter providing the largest improvement is plotted at each target doppler. For comparison the optimum curve from Fig. 15.18 is shown by a broken line and thus provides a direct assessment of how well the
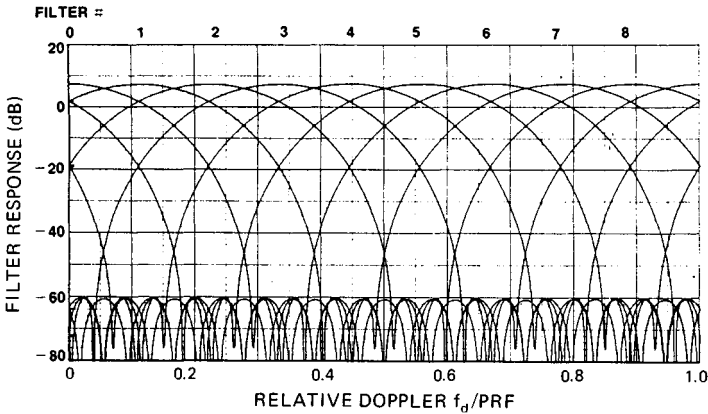
FIG. 15.28   Doppler filter bank of 68 dB Chebyshev filters. CPI = nine pulses.

Chebyshev filter design performs against a given clutter model. Also shown is the average SCR improvement for both the optimum and the Chebyshev filter bank.

Finally, Fig. 15.30 shows the average SCR improvement of the 68 dB Chebyshev doppler filter bank as well as the optimum curve (from Fig. 15.19) as a function of the relative spectrum spread of the clutter. Owing to the finite number of filters implemented in the filter bank, the average SCR improvement will change by a small amount if a doppler shift is introduced into the clutter returns. This effect is illustrated by the cross-hatched region, which shows upper and lower limits on the average SCR improvement for all possible clutter doppler shifts. For a smaller number of filters in the doppler filter bank this variation would be larger.
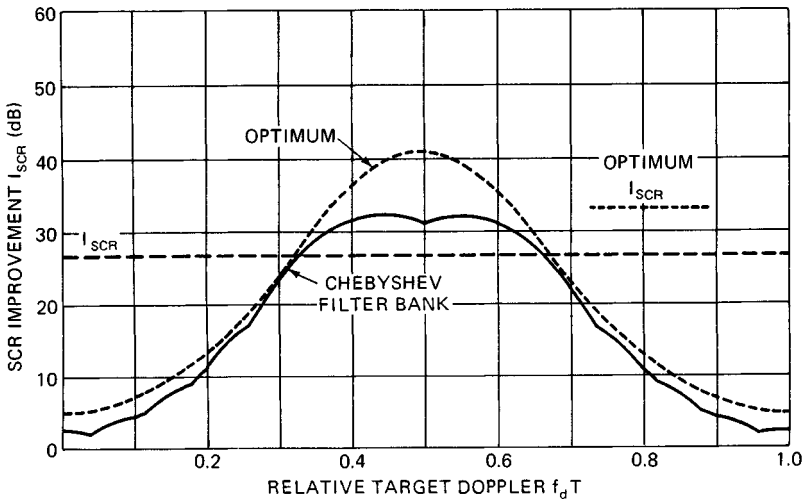
FIG. 15.29   SCR improvement of 68 dB Chebyshev doppler filter bank compared with the optimum.
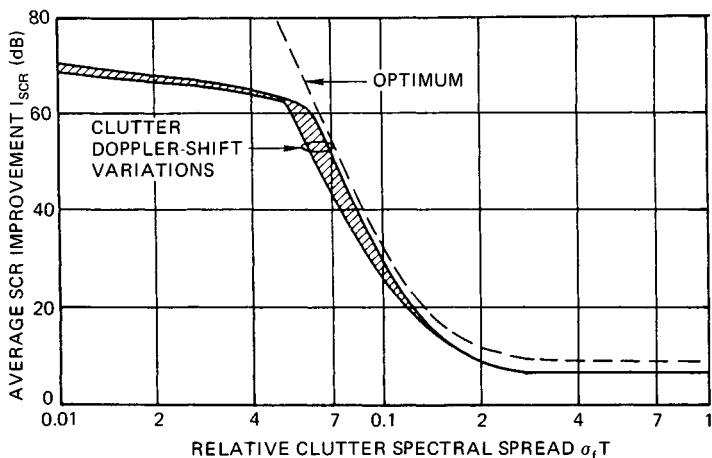
FIG. 15.30 Average SCR improvement for the 68 dB Chebyshev filter bank shown in Fig. 15.28. CPI = nine pulses. Optimum is from Fig. 15.19.

Fast Fourier Transform Filter Bank. For a large number of parallel doppler filters, hardware implementation can be significantly simplified through the use of the FFT algorithm. The use of this algorithm constrains all filters in the filter bank to have identical responses, and the filters will be uniformly spaced along the doppler axis. The number of filters implemented for a given size of the CPI can, however, be varied. For example, a larger number of filters can be realized by extending the received data with extra zero values (also known as zero padding) after the received returns have been appropriately weighted in accordance with the desired filter response (e.g., Chebyshev).

15.9 STAGGERED PRF

Stagger Design Procedures. The interval between radar pulses may be changed to shift the target velocities to which the MTI system is blind. The interval may be changed on a pulse-to-pulse, dwell-to-dwell (each dwell being a fraction of the beamwidth), or scan-to-scan basis. Each approach has advantages. The advantages of the scan-to-scan method are that the radar system is easier to build, and multiple-time-around clutter is canceled in a power amplifier MTI system. The transmitter stabilization necessary for good operation of an unstaggered MTI system costs money and weight. To stabilize the transmitter sufficiently for pulse-to-pulse or dwell-to-dwell stagger operation is considerably more difficult. Pulse-to-pulse staggering is used with MTI processing, while dwell-to-dwell staggering is used with filter bank processing.

For many MTI applications pulse-to-pulse staggering is essential. For example, if a binomial-weighted three-pulse canceler which has 36 percent–wide rejection notches is employed and if scan-to-scan pulse staggering is used, 36 percent of the desired targets would be missing on each scan owing to doppler considerations alone. This might be intolerable for some applications. With pulse-to-pulse staggering, good response can be obtained on all dopplers of interest on each scan. In addition, better velocity response can be obtained at some dopplers than

CHAPTER 16
AIRBORNE MTI

Fred M. Staudaher
Naval Research Laboratory

16.1 SYSTEMS USING AIRBORNE MTI TECHNIQUES

Airborne search radars were initially developed for the detection of ships by long-range patrol aircraft. During the latter part of World War II, airborne early-warning (AEW) radars were developed by the U.S. Navy to detect low-flying aircraft approaching a task force below the radar coverage of the ship's antenna. The advantage of the airborne platform in extending the maximum detection range for air and surface targets is apparent when one considers that the radar horizon is 12 nmi for a 100-ft antenna mast compared with 123 nmi for a 10,000-ft aircraft altitude.

Loss of picket ships due to kamikaze attacks led to the concept of the autonomous airborne detection and control station. This type of system was further developed as a barrier patrol aircraft for continental air defense.

The carrier-based E-2C aircraft (Fig. 16.1) uses AEW radar as the primary sensor in its airborne tactical data system. These radars with their extensive field of view are required to detect small aircraft targets against a background of sea and land clutter. Because of their primary mission of detecting low-flying aircraft, they cannot elevate their antenna beam to eliminate the clutter. These considerations have led to the development of airborne MTI (AMTI)[1-3] radar systems similar to those used in surface radars[1,4-6] discussed in the preceding chapter.

Airborne MTI radar systems have also been utilized to acquire and track targets in interceptor fire control systems. In this application the system has to discriminate against clutter only in the vicinity of a prescribed target. This allows the system to be optimized at the range and angular sector where the target is located. MTI is also used to detect moving ground vehicles by reconnaissance and tactical fighter-bomber aircraft. Because of the low target velocity, higher radar frequencies are employed to obtain a significant doppler shift. Since a strong clutter background is usually present, these systems can effectively utilize noncoherent MTI techniques.

The environment of high platform altitude, mobility, and speed coupled with restrictions on size, weight, and power consumption presents a unique set of

FIG. 16.1 E-2C airborne early-warning (AEW) aircraft showing rotodome housing the antenna.

problems to the designer of airborne MTI systems. This chapter will be devoted to considerations unique to the airborne environment.

16.2 COVERAGE CONSIDERATIONS

Search radars generally require 360° azimuthal coverage. This coverage is difficult to obtain on an aircraft since mounting an antenna in the clear presents major drag, stability, and structural problems. When extensive vertical coverage is required, the aircraft's planform and vertical stabilizer distort and shadow the antenna pattern. Analysis of tactical requirements may show that only a limited coverage sector is required. However, this sector usually has to be capable of being positioned over the full 360° relative to the aircraft's heading because of the requirements for coverage while reversing course, large crab angles when high winds are encountered, need to position ground track in relation to wind, nontypical operating situations, and operational requirements for coverage while proceeding to and from the station.

16.3 PLATFORM MOTION AND ALTITUDE EFFECTS ON MTI PERFORMANCE

MTI discriminates between airborne moving targets and stationary land or sea clutter. However, in the airborne case the clutter moves with respect to the airborne platform. It is possible to compensate for the mean clutter radial velocity

by using techniques such as TACCAR (time-averaged-clutter coherent airborne radar).

As shown in Fig. 16.2, the apparent radial velocity of the clutter is $V_r = -V_g \cos \alpha$, where V_g is the ground speed of the platform and α is the angle subtended between the line of sight to a point on the earth's surface and the aircraft's velocity vector. Figure 16.3 shows the loci of constant radial velocity along the surface. In order to normalize the figure, a flat earth is assumed, and the normalized radial velocity $V_n = V_r/V_g$ is presented as a function of azimuth angle ψ and normalized ground range R/H, where H is the aircraft's altitude.

Instead of a single clutter doppler frequency corresponding to a constant radial velocity (V_B in Fig. 16.2) determined by the antenna pointing angle α_0, the radar sees a continuum of velocities. This results in a frequency spectrum at a particular range whose shape is determined by the antenna pattern that intersects the surface, the reflectivity of the clutter, and the velocity distribution within the beam. Furthermore, since V_r varies as a function of range at a particular azimuth ψ, the center frequency and spectrum shape vary as a function of range and azimuth angle ψ_0.

When the antenna is pointing ahead, the predominant effect is the variation of the center frequency corresponding to the change in α_0 with range. When the antenna is pointing abeam, the predominant effect is the velocity spread across the antenna beamwidth. These are classified as the slant-range effect and the platform-motion effect, respectively.
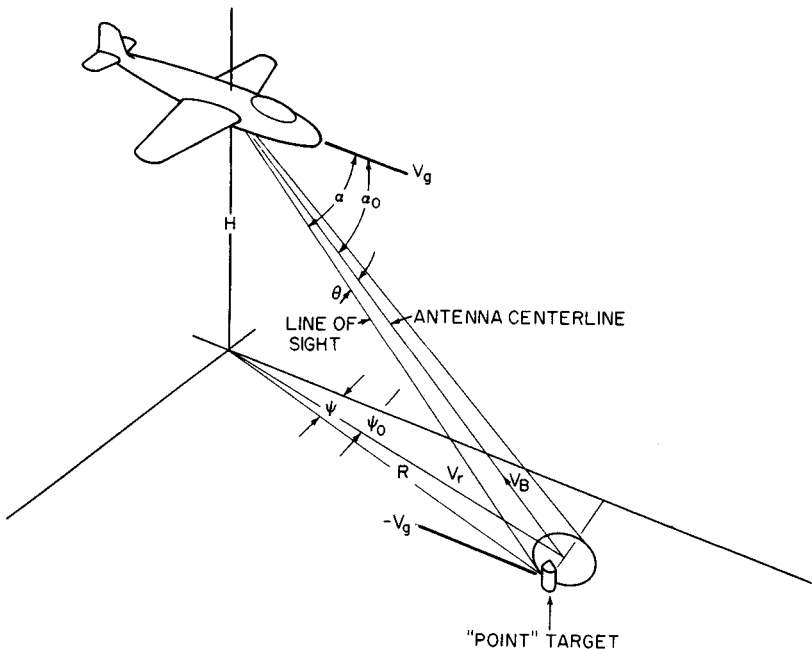
FIG. 16.2 Defining geometry: α_0 = antenna pointing angle; α = line-of-sight angle; θ = angle from antenna centerline; V_g = aircraft ground speed; V_r = radial velocity of point target; V_B = radial velocity along antenna centerline (boresight); ψ_0 = antenna azimuth angle; ψ = azimuth angle; R = ground range to point target; H = aircraft height.
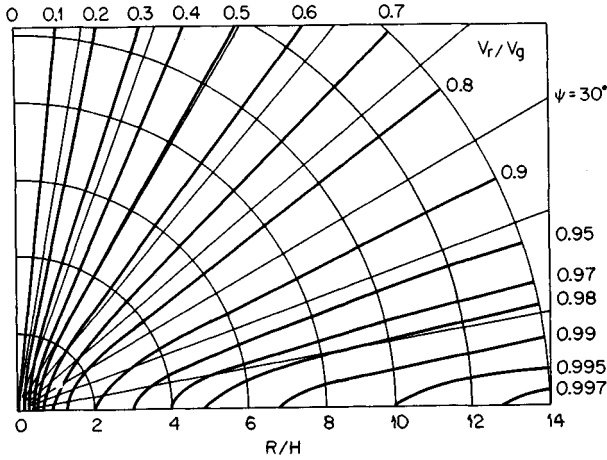
FIG. 16.3 Loci of constant normalized radial velocity V_r/V_g as a function of aircraft range-to-height ration R/H and azimuth angle ψ.

Effect of Slant Range on Doppler Offset. The antenna boresight velocity V_B is the ground-velocity component along the antenna centerline (boresight) and is given as $-V_g \cos \alpha_0$. If the clutter surface were coplanar with the aircraft, this component would be equal to $-V_g \cos \psi_0$ and would be independent of range. The ratio of the actual boresight velocity to the coplanar boresight velocity is defined as the normalized boresight-velocity ratio:

$$VBR = \frac{\cos \alpha_0}{\cos \psi_0} = \cos \phi_0 \qquad (16.1)$$

where ϕ_0 is the depression angle of the antenna centerline from the horizontal. Figure 16.4 shows the variation of the normalized boresight-velocity ratio as a
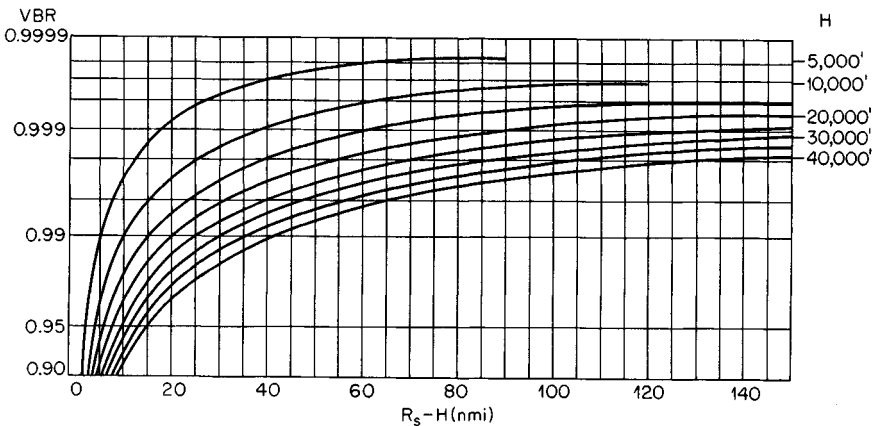
FIG. 16.4 Normalized boresight-velocity ratio VBR as a function of the difference between slant range R_s and aircraft altitude H for different aircraft altitudes.

function of slant range for a curved earth and different aircraft altitudes. The variation is fairly rapid for slant ranges less than 15 mi.

It is desirable to center the clutter spectrum in the notch (i.e., minimum-response region) of the AMTI filter in order to obtain maximum clutter rejection. This can be accomplished by offsetting the IF or RF frequency of the radar signal by an amount equal to the average doppler frequency of the clutter spectrum. Since the clutter center frequency varies with range and azimuth when the radar is moving, it is necessary for the filter notch to track the doppler-offset frequency, using an open- or closed-loop control system such as TACCAR, described below.

TACCAR. The MIT Lincoln Laboratory originally developed TACCAR to solve the AMTI radar problem. After many other approaches, it was recognized that if one used the clutter return rather than the transmit pulse to phase-lock the radar to the clutter filter, one could center the clutter in the filter stopband. The clutter phase varies from range cell to range cell owing to the distribution of the location of the scatterers in azimuth. Hence it is necessary to average the return for as long an interval as possible. Other processing features, such as phase comparison cancellation, were included in this radar (AN/APS-70). Today TACCAR is used to describe the centering of the returned clutter spectrum to the zero filter frequency. Since the technique compensates for drift in the various system elements and biases in the mean doppler frequency due to ocean currents, chaff, or weather clutter, it is used in shipboard and land-based radars as well as airborne radar.

A functional block diagram of an airborne radar employing TACCAR is shown in Fig. 16.5. The clutter error signal is obtained by measuring the pulse-to-pulse phase shift $\omega_d T_p$ of the clutter return. This provides a very sensitive error signal. The averaged error signal controls a voltage-controlled coherent master oscillator (COMO), which determines the transmitted frequency of the radar. The COMO is slaved to the system reference oscillator frequency via the automatic frequency control (AFC) loop shown in Fig. 16.5. This provides a stable reference in the absence of clutter. An input from the aircraft inertial navigation system and the
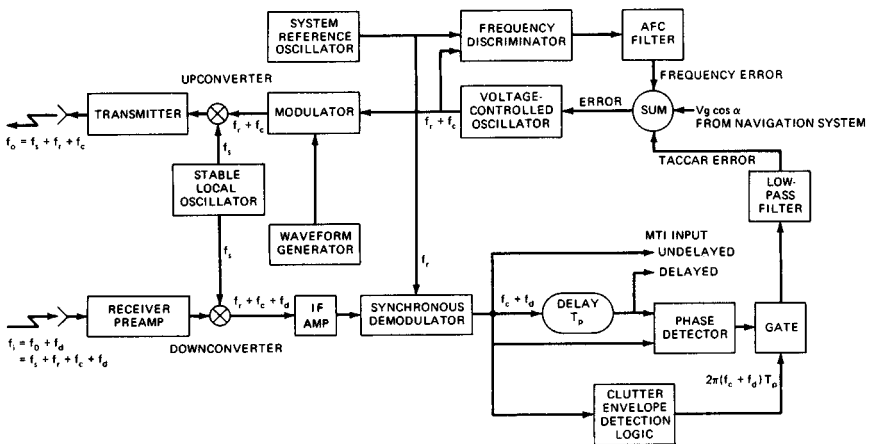
FIG. 16.5 Block diagram of a radar illustrating the signal flow path of the TACCAR control loop.
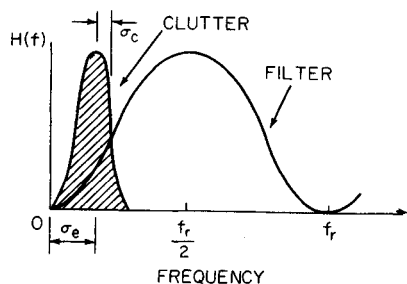
FIG. 16.6 Effect of doppler-offset error; f_r = PRF.

antenna servo provide a predicted doppler offset. These inputs allow the TACCAR system to provide a narrow-bandwidth correction signal.

Because of the noisy nature of the clutter signal, the need to have the control system bridge regions of weak clutter return, and the requirement not to respond to the doppler shift of a true target, the control system usually tracks the azimuth variation of a specific radar range interval. The maximum range of this interval is chosen so that clutter will be the dominant signal within the interval. The minimum range is chosen to exclude signals whose average frequency differs substantially from the frequency in the region of interest. For some applications it may be necessary to use multiple control loops, each one covering a specific range interval, or to vary the offset frequency in range. At any particular range the filter notch is effectively at one frequency and the center frequency of the clutter spectrum at another. The difference between these frequencies results in a doppler-offset error as shown in Fig. 16.6. The clutter spectrum will extend into more of the filter passband, and the improvement factor will be degraded.

Figure 16.7 shows the improvement factor for single- and double-delay cancelers as a function of the ratio of the notch-offset error to the pulse repetition frequency (PRF) for different clutter spectral widths. Fortunately, the platform-motion spectrum is narrow in the forward sector of coverage where offset error is
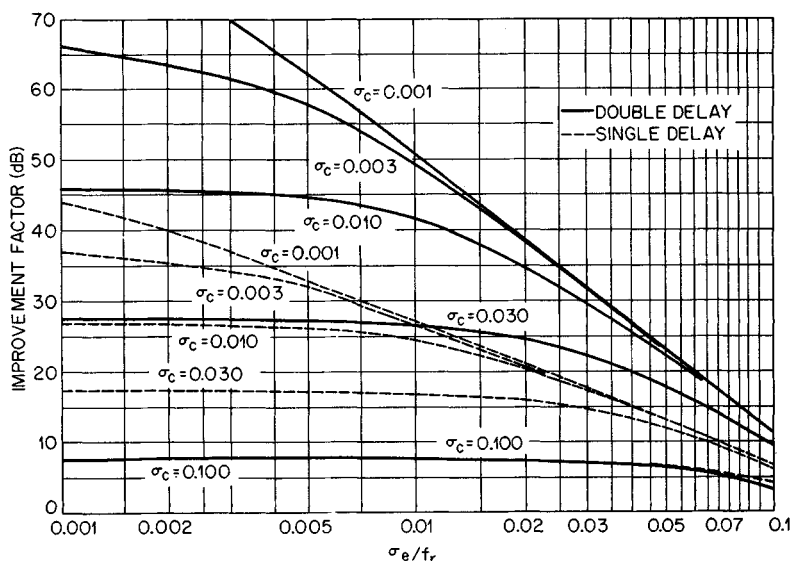
FIG. 16.7 Improvement factor I versus normalized doppler offset σ_e as a function of clutter spectrum width σ_c.

maximum. An offset error of one-hundredth of the PRF would yield 26 dB improvement for a double canceler with an input clutter spectrum whose width was 3 percent of the PRF. If the radar frequency were 10 GHz, PRF 1 kHz, and ground speed 580 kn, the notch would have to be held within 0.29 kn or $0.005V_g$.

Because of these requirements and the width of the platform-motion spectrum, stagger PRF systems must be chosen primarily on the basis of maintaining the stopband rather than flattening the passband. Similarly, higher-order delay-line filters (with or without feedback) are synthesized on the basis of stopband rejection. The limiting case is the narrowband filter bank where each individual filter consists of a small passband, the balance being stopband.

Platform-Motion Effect.   To an airborne radar a clutter scatterer appears to have a radial velocity that differs from the antenna-boresight radial velocity at the same range by

$$
\begin{aligned}
V_e &= V_r - V_B \\
&= V_g \cos \alpha_0 - V_g \cos \alpha \\
&= V_g[\cos \alpha_0 - \cos (\alpha_0 + \theta)] \\
&= V_x \sin \theta + 2V_y \sin^2 \frac{\theta}{2}
\end{aligned}
\tag{16.2}
$$

for small values of θ and depression angle ϕ_0, where V_x is the horizontal component of velocity perpendicular to the antenna boresight and V_y is the component along the antenna boresight. θ is the azimuthal angle from the antenna boresight, or intersection of the vertical plane containing the boresight with the ground. The corresponding doppler frequency, when α_0 is a few beamwidths from ground track, is

$$
f_d = \frac{2V_x}{\lambda} \sin \theta \approx \frac{2V_x}{\lambda}\theta
\tag{16.3}
$$

This phenomenon results in a platform-motion clutter power spectrum which is weighted by the antenna's two-way power pattern in azimuth. The true spectrum may be approximated by a gaussian spectrum,

$$
H(f) = e^{-\frac{1}{2}(f_d/\sigma_{pm})^2} = e^{-(V_x\theta/\lambda\sigma_{pm})^2} \approx G^4(\theta)
\tag{16.4}
$$

$G^4(\theta)$, the two-way power pattern of the antenna, is 0.25 when $\theta = \theta_a/2$, where θ_a is the half-power beamwidth which can be approximated by λ/a, a being the effective horizontal aperture width. Thus

$$
e^{-\frac{1}{2}(V_x/a\sigma_{pm})^2} = 0.25
$$

or

$$
\sigma_{pm} = 0.6\frac{V_x}{a}
\tag{16.5}
$$

where V_x and a are in consistent units. This value is lower than ones derived by other authors.[4,5] However, it agrees with more exact analysis of antenna radiation patterns and experimental data analyzed by the author.

A more exact value of the parameter σ_{pm} may be obtained by matching a two-

way power pattern of interest with the gaussian approximation at a specific point on the pattern, determining the standard deviation of θ by using statistical techniques, or fitting the pattern and using numerical methods. The calculation of the improvement factor I can be performed by averaging the resultant residue power, obtained by summing the signal phasors at specific values of θ, from null to null of the antenna pattern.

Figure 16.8 shows the effect of platform motion on the MTI improvement factor as a function of the fraction of the aperture displaced in the plane of the aperture per interpulse period T_p. A 5.4 percent displacement would reduce the double-delay improvement factor to 30 dB. This corresponds to a speed of 540 ft/s if the system has a PRF of 1000 Hz and a 10-ft antenna aperture. For a single-delay system the displacement would have to be held to 1.1 percent for a 30 dB performance limit.
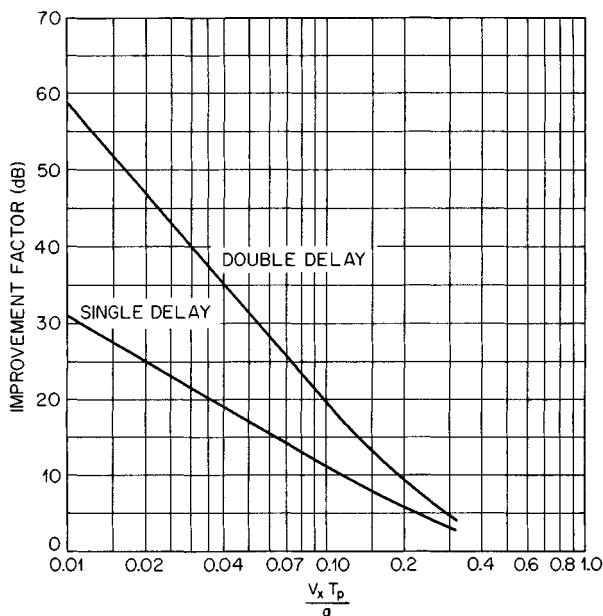
FIG. 16.8 Effect of platform motion on the MTI improvement factor as a function of the fraction of the horizontal antenna aperture displaced per interpulse period, $V_x T_p/a$.

16.4 PLATFORM-MOTION COMPENSATION ABEAM

The deleterious effects of platform motion can be reduced by physically or electronically displacing the antenna phase center along the plane of the aperture. This is referred to as the displaced phase center antenna (DPCA) technique.[7-11]

Physically Displaced Phase Center Antenna. In physical DPCA,[10,11] the

apertures of two side-looking antennas are aligned parallel with the aircraft longitudinal axis. Their phase centers are separated by the distance d. If the aircraft is moving at ground speed V_g, then the phase centers move $V_g T_p$ during the interpulse period T_p.

In Fig. 16.9a the first pulse is transmitted and received on the forward antenna A_1. The second pulse is transmitted and received on the rear antenna A_2 during the next interpulse period. If $d = V_g T_p$, then the antenna used on the first pulse, A_1, will coincide with the antenna A_2 used on the second pulse. On a two-pulse-pair basis the signals received at A_1 and A_2 make it appear as if the antenna were stationary. There is actually a displacement with respect to the transmitter, but the signal path difference will be the same pulse to pulse. This will appear as a negligible range error.

Since it is difficult to change the spacing between antennas, the displacement is set by the design speed and PRF limits. Then the PRF is varied during operation to maintain the proper alignment.

If the antenna is not aligned with the flight path and if d is not equal to $V_g T_p$, then an error occurs between A_1 and A_2 as shown in Fig. 16.9b. The result is as though the aircraft were flying at a speed and heading such that the displacement $A_1 - A_2'$ occurs during an interpulse period. The TACCAR circuits could center the resultant spectrum at zero-doppler frequency. However, the cancellation will correspond to a value $V_x T_p = 2\epsilon$ in Fig. 16.8. If $2\epsilon/T_p$ is small enough, then the sidelobe clutter spectrum will be in the filter notch and will be canceled.

FIG. 16.9 Physical DPCA defining geometry. (a) Perfect motion compensation, where A_1 and A_2 are the antenna phase centers for pulse 1 and the primed quantities are for pulse 2. (b) Imperfect motion compensation due to displacement error and alignment error α.

The two-antenna scheme is difficult to mechanize, and additional errors can occur if the antennas are mounted one above the other because of antenna field variations. These variations are caused by the difference in physical location due to vertical displacement, the effect of the different near-field environments, and fabrication errors. Furthermore, the PRF is effectively cut in half by having to receive during the unique alternate transmission-path configurations. A single-antenna scheme is possible by using an array with multiple feed structures that utilize a common set of elements and a switching network as shown in Fig. 16.10. The top row of switches connects the elements to the corporate feed or to the dummy loads. The bottom row of switches connects the corporate feed to the appropriate elements. The left six elements are active in the configuration illustrated. When the switches are placed in the alternate configuration, the right six elements are active. The subarrays are displaced from each other by two elements in the example.

Another variant is to have separate corporate feeds for the left and right subarrays of the antenna.[11] One of the subarrays or a central group of elements is

FIG. 16.10 Switching network to synthesize displaced subarrays, within an antenna array.

used as the transmit array to avoid high-power switching circuits. The separation of the subarrays d must equal $2V_gT_p$ to compensate for the transmit phase-center displacement. This allows $A_1 - A'_2$ to be paired and then A'_1 to be paired with A''_2, where A_1 is displaced from A_2 by d. This allows cancellation to be made every interpulse period, maintaining the effective PRF equal to the basic PRF.

The two-way patterns from the antenna at A_1 and A_2 must be nearly identical; otherwise, cancellation will be degraded. This degradation may be calculated by measuring the two antenna patterns, $G_1(\theta)$ and $G_2(\theta)$, and then calculating the correlation coefficient

$$\rho = \frac{[\int G_1{}^2(\theta)G_2^{*2}(\theta)d\theta]^2}{\int [G_1(\theta)]^4 d\theta \int [G_2(\theta)]^4 d\theta} \qquad (16.6)$$

The resultant cancellation ratio is then

$$CR = 10 \log [1/(1 - \rho)] \qquad (16.7)$$

If $G_2(\theta)$ is nearly identical to $G_1(\theta)$, then ρ is approximately equal to 1 and the cancellation ratio is large. When measuring G_1 and G_2, the array must be displaced for the second measurement to ensure that each subarray is in the same physical position on the antenna range.

Electronically Displaced Phase Center Antenna. Figure 16.11a shows the pulse-to-pulse phase advance of an elemental scatterer as seen by the radar receiver. The amplitude E_1 of the received signal is proportional to the two-way antenna field intensity. The phase advance is

$$2\eta = 2\pi f_d T_p = \frac{4\pi V_x T_p \sin \theta}{\lambda} \qquad (16.8)$$

where f_d = doppler shift of scatterer [Eq. (16.3)]
T_p = interpulse period
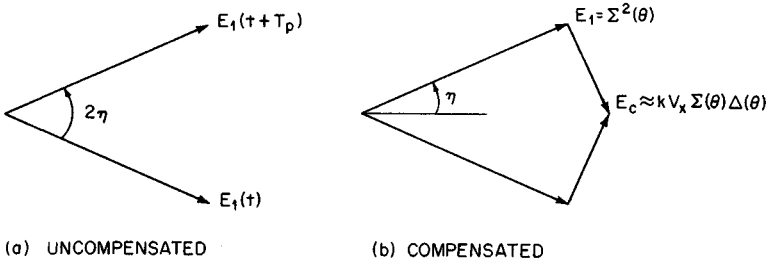
FIG. 16.11 Phasor diagram showing the return from a point scatterer due to platform motion.

Figure 16.11b shows a method of correcting for the phase advance η. An idealized correction signal E_c is applied, leading the received signal by 90° and lagging the next received signal by 90°. For exact compensation the following relation would hold:

$$E_c = E_1 \tan \eta = \Sigma^2(\theta) \tan\frac{2\pi V_x T_p \sin \theta}{\lambda} \tag{16.9}$$

This assumes a two-lobe antenna pattern similar to that in a monopulse tracking radar. Two receivers are used, one supplying a sum signal $\Sigma(\theta)$ and the other a difference signal $\Delta(\theta)$. The difference signal is used to compensate for the effects of platform motion.

If the system is designed to transmit the sum pattern $\Sigma(\theta)$ and receive both $\Sigma(\theta)$ and a difference pattern $\Delta(\theta)$, then at the design speed the received signal $\Sigma(\theta)\Delta(\theta)$ can be applied as the correction signal. The actual correction signal used to approximate E_c is $k \Sigma(\theta)\Delta(\theta)$, where k is the ratio of the amplification in the sum and difference channels of the receiver.

A uniformly illuminated monopulse array[12] has the difference signal Δ in quadrature with the sum and has the amplitude relationship

$$\Delta(\theta) = \Sigma(\theta) \tan \left(\frac{\pi W}{\lambda} \sin \theta\right) \tag{16.10}$$

where W is the distance between the phase centers of the two halves of the antenna. Hence a choice of $W = 2V_x T_p$ and $k' = 1$ would ideally result in perfect cancellation.

In practice, a sum pattern is chosen based on the desired beamwidth, gain, and sidelobes for the detection system requirements. Then the difference pattern $\Delta(\theta)$ is synthesized independently, based on the relationship required at design radar platform speed and allowable sidelobes. The two patterns may be realized by combining the elements in separate corporate-feed structures.

Figure 16.12 shows the idealized improvement factor as a function of normalized aperture movement for a double-delay canceler. The improvement factor shown is the improvement factor for a point scatterer averaged over the null-to-null antenna beamwidth. In one case the gain ratio k' is optimized at each value of pulse-to-pulse displacement. In the other compensated case the optimum gain ratio k is approximated by the linear function of interpulse platform motion kV_x.
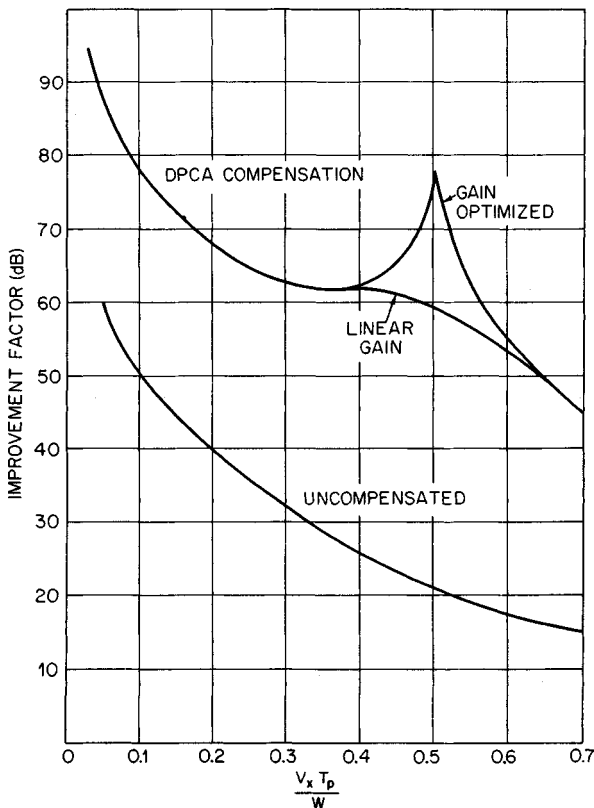
FIG. 16.12 MTI improvement factor I for DPCA compensation as a function of the fraction of the horizontal phase center separation W that the horizontal antenna aperture is displaced per interpulse period, $V_x T_p/W$. $W = 0.172a$.

A block diagram of the double-delay system is shown in Fig. 16.13. Since the transmitted pattern $\Sigma(\theta)$ appears in both channels, it is not shown. A single-delay system would not have the second delay line and subtractor. The normally required circuitry for maintaining coherence, gain and phase balance, and timing is not shown. The speed control V_x is bipolar and must be capable of reversing the sign of the $\Delta(\theta)$ signal in each channel when the antenna pointing angle changes from the port to the starboard side of the aircraft.

The hybrid amplifier shown has two input terminals which receive $\Sigma(\theta)$ and $j\Delta(\theta)$ and amplify the $\Delta(\theta)$ channel by kV_x relative to the $\Sigma(\theta)$ channel. The output terminals produce the sum and difference of the two amplified input signals. Since DPCA compensates for the complex signal, both amplitude and phase information must be retained. Therefore, these operations usually occur at RF or IF. Digital compensation can be used if synchronous detection and analog-to-digital (A/D) conversion are performed and the components are treated as complex phasors. Furthermore, the operations must be linear until the sum signal and difference signals have been processed by the hybrid amplifier. After this
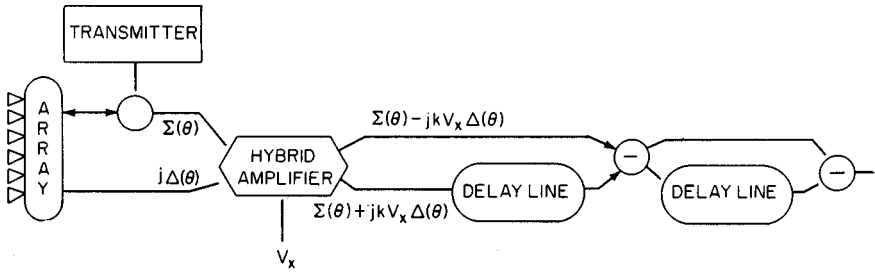
FIG. 16.13 Simplified double-delay DPCA mechanization.

single-pulse combination, the actual double cancellation can be performed by any of the processing techniques outlined in Chap. 15.

Power in the Antenna Sidelobes. Airborne systems are limited in their ability to reject clutter due to the power returned by the antenna sidelobes. The full 360° azimuthal pattern sees velocities from $-V_g$ to $+V_g$. The compensation circuits offset the velocity by an amount corresponding to the antenna boresight velocity V_B, but the total range of doppler frequencies corresponding to $2V_g$ is obtained because of echoes received via the sidelobes. For airborne systems with low and medium PRFs, these doppler frequencies can cover several multiples of the PRF so that the sidelobe power is folded into the filter. This limitation is a function of the antenna pointing angle, the MTI filter response, and the sidelobe pattern. If the sidelobes are relatively well distributed in azimuth, a measure of performance can be obtained by averaging the power returned by the sidelobes.

The limiting improvement factor due to sidelobes is

$$I_{\text{sl limit}} = \frac{K \int_{-\pi}^{\pi} G^4(\theta)\, d\theta}{\int_{\text{sl}} G^4(\theta) d\theta} \tag{16.11}$$

where the lower integral is taken outside the main-beam region. Main-beam effects would be included in the platform-motion improvement factor. The constant K is the noise normalization factor for the MTI filter. ($K = 2$ for single delay and 6 for double delay.) $G^4(\theta)$ is the two-way power of the antenna in the plane of the ground surface.

The DPCA performance described in the preceding subsection can be analyzed on the basis of radiation patterns or the equivalent aperture distribution function.[8] If the radiation pattern is used, the composite performance may be obtained either by applying the pattern functions over the entire 360° pattern or by combining the improvement factors for the DPCA main-beam and the sidelobe regions in the same manner as parallel impedances are combined:

$$\frac{1}{I_{\text{total}}} = \frac{1}{I_{\text{sl}}} + \frac{1}{I_{\text{DPCA}}} \tag{16.12}$$

If the aperture distribution is used, the sidelobe effects are inherent in the analysis. Care must be taken since if the array or reflector function is used with-

out considering the weighting of the elemental pattern or the feed distribution, the inherent sidelobe pattern can obscure the main-beam compensation results.

16.5 SCANNING-MOTION COMPENSATION

Figure 16.14a shows a typical antenna main-beam radiation pattern and the response of a point scatterer for two successive pulses when the antenna is scanning. It is seen that the signals returned would differ by $\Delta G^2(\theta)$. This results in imperfect cancellation due to scanning. The average effect on the improvement factor can be obtained by integrating this differential effect over the main beams:

$$I_{\text{scan}} = \frac{2\int_{-\theta_0}^{\theta_0} |G(\theta)|^2 d\theta}{\int_{-\theta_0}^{\theta_0} |G(\theta + T_p\dot{\theta}) - G(\theta)|^2 d\theta} \quad \text{for single-delay cancellation} \quad (16.13a)$$

$$I_{\text{scan}} = \frac{6\int_{-\theta_0}^{\theta_0} |G(\theta)|^2 d\theta}{\int_{-\theta_0}^{\theta_0} |G(\theta + T_p\dot{\theta}) - 2G(\theta) + G(\theta - T_p\dot{\theta})|^2 d\theta} \quad \text{for double-delay cancellation}$$

$$(16.13b)$$

where θ_0 = null of main beam
$\quad G(\theta)$ = two-way voltage pattern

In order to treat scanning motion in the frequency domain, the apparent clutter velocity seen by the scanning antenna is examined to determine the doppler frequency. Each element of an array or incremental section of a continuous aperture can be considered as receiving a doppler-shifted signal due to the relative

(a) RADIATION   (b) APERTURE   (c) PHASOR COMPENSATION
    PATTERN
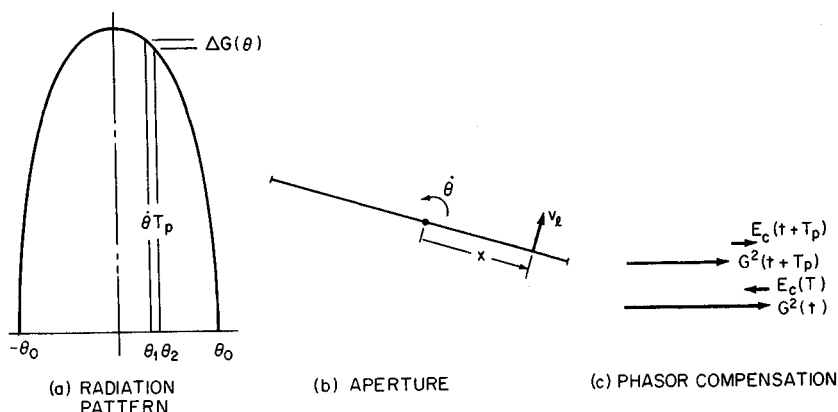
FIG. 16.14 Antenna scanning effects. (a) As seen by the antenna radiation pattern, due to the apparent change in azimuth of the scatterer, $\theta_2 - \theta_1 = \dot{\theta} T_p$. (b) As seen by the aperture illumination function, due to the apparent motion, $v_l = x\dot{\theta}$, of the scatterer relative to the antenna element at position x. (c) Step-scan compensation of two received phasors.

motion of the clutter. The power received by the element is proportional to the two-way aperture power distribution function $F_2(x)$ at the element.

In addition to the velocity seen by all elements because of the motion of the platform, each element sees an apparent clutter velocity due to its rotational motion, as illustrated in Fig. 16.14b. The apparent velocity varies linearly along the aperture. Hence the two-way aperture distribution is mapped into the frequency domain. The resulting power spectrum due to the antenna scanning is

$$H(f) \not\approx F_2\left(\frac{\lambda f}{2 \dot\theta}\right) \qquad 0 \le f \le \frac{a \dot\theta}{\lambda} \tag{16.14}$$

where $\dot\theta$ = antenna rotation rate
$\quad a$ = horizontal antenna aperture

This spectrum can be approximated by a gaussian distribution with standard deviation

$$\sigma_c = 0.265\frac{f_r}{n} = 0.265\frac{\dot\theta}{\theta_a} \approx 0.265\frac{a\,\dot\theta}{\lambda} \tag{16.15}$$

where λ and a are in the same units, θ_a is the one-way half-power beamwidth, and n is the number of hits per beamwidth. The approximation $\theta_a \approx \lambda/a$ is representative of antenna distribution yielding acceptable sidelobe levels.

It can be seen that the differential return is

$$\Delta G^2(\theta) \not\approx \frac{dG^2(\theta)}{d\theta}\Delta\theta = \frac{dG^2(\theta)}{d\theta}\dot\theta T_p \tag{16.16}$$

This suggests[7,13] that a correction signal in the reverse sense to $\Delta G^2(\theta)$ be applied, as shown in Fig. 16.14c. Half the correction is added to one pulse and half subtracted from the other, so that

$$\text{Correction signal} = \frac{\Delta G^2(\theta)}{2} = \frac{\dot\theta T_p}{2}\frac{d\Sigma^2(\theta)}{d\theta} \tag{16.17}$$

$$= \dot\theta T_p \Sigma(\theta)\frac{d\Sigma(\theta)}{d\theta}$$

where $\Sigma^2(\theta)$ was substituted for $G^2(\theta)$. The radar transmits a sum pattern $\Sigma(\theta)$ and receives on the difference pattern $\Delta(\theta)$, so that the received signal is proportional to the product of the two. If the signal received on the difference pattern is used as the correction, we have

$$E_c = \Delta(\theta)\Sigma(\theta) \tag{16.18}$$

By comparing Eqs. (16.17) and (16.18), we see that, for E_c to approximate the correction signal, the difference patterns should be
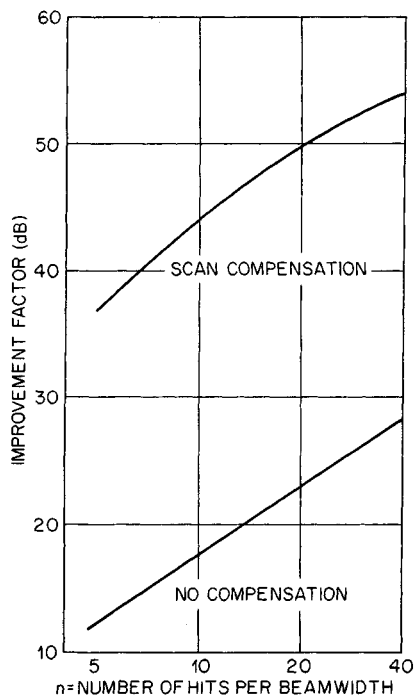
FIG. 16.15 MTI improvement factor for a step-scan compensation of a single-delay canceler as a function of the number of hits per beamwidth. The antenna pattern is (sin x)/x.

$$\Delta(\theta) = \dot{\theta} T_p \frac{d\Sigma(\theta)}{d\theta} \qquad (16.19)$$

The derivative of the sum pattern is similar to a difference pattern in that it is positive at the main-beam null, $-\theta_0$, and decreases to zero on the antenna centerline and then goes negative until θ_0.

By referring to Fig. 16.13, one observes that the mechanization for scan compensation is fundamentally similar to the DPCA mechanization except that the difference signal is applied in phase with the sum signal and amplified by an amount determined by the antenna rotation per interpulse period.

The signals required, if the transmission signal $\Sigma(\theta)$ that appears in each channel is neglected, are $\Sigma(\theta) \pm \dot{\theta} T_p \Delta(\theta)$, where [is the ratio of the amplification in the two channels chosen to maximize the clutter rejection. The required difference-pattern slope is determined by the derivative of the scan pattern, which differs from the DPCA criterion. This technique is known as step-scan compensation because the system electronically points the antenna slightly ahead of and behind boresight each pulse so that a leading and lagging pair are taken from successive returns to obtain the effect of the antenna remaining stationary.

Figure 16.15 shows the improvement obtained by Dickey and Santa[7] for single-delay cancellation.

Compensation-Pattern Selection. Selection of the compensation pattern depends on the level of system performance required, the type of MTI filtering used, the platform velocity, scan rate, and the characteristics required by normal radar parameters such as resolution, distortion, gain, sidelobes, etc. For instance, an exponential pattern and its corresponding difference pattern are excellent for single-delay-cancellation DPCA but are unsatisfactory when double-delay cancellation is used. This is because the single-delay canceler requires the best match between the actual pattern and the required pattern near boresight, whereas double cancellation requires the best match on the beam shoulder. Step-scan compensation usually requires the difference-pattern peaks to be near the nulls of the sum pattern to match.

Grissetti et al.[13] have shown that for step-scan compensation the improvement factor for single-delay cancellation increases as a function of the number of hits at

20 dB/decade; for the first-derivative*-type step-scan compensation, at the rate of 40 dB/decade; and with first- and second-derivative compensation, at the rate of 60 dB/decade. Hence, for a ground-based system that is limited by scan rate, one should improve the compensation pattern rather than use a higher-order MTI canceler. However, airborne systems are primarily limited by platform motion and require both better cancelers and compensation for operation in a land clutter environment. In the sea clutter environment the system is usually dominated by the spectral width of the velocity spectrum or platform motion rather than scanning. The applicability of DPCA or step-scan compensation in the latter case is dependent on the particular system parameters.

16.6 SIMULTANEOUS PLATFORM MOTION AND SCAN COMPENSATION

In AMTI systems having many hits per scan, scanning is a secondary limitation for an uncompensated double canceler. However, the performance of a DPCA system is significantly reduced when it is scanned. This is due to the scanning modulation on the difference pattern used for platform-motion compensation.

Since the DPCA applies the difference pattern in quadrature to the sum pattern to compensate for phase error and step scan applies the difference pattern in phase to compensate for amplitude error, it is possible to combine the two techniques by properly scaling and applying the difference pattern both in phase and in quadrature. The scaling factors are chosen to maximize the improvement factor under conditions of scanning and platform motion.

The relationships for a double-delay (three-pulse) AMTI are shown in the phasor diagram, Fig. 16.16. The phase advance between the first pair of pulses, received by the sum pattern Σ, is

$$2_{\eta 1} = \frac{4\pi T_p}{\lambda}\left[V_x\left(\sin\theta_2 - \sin\frac{\omega_r T_p}{2}\right) + V_y\left(\cos\frac{\omega_r T_p}{2} - \cos\theta_2\right)\right] \quad (16.20)$$

and the phase advance between the second pair of pulses is

$$2_{\eta 2} = \frac{4\pi T_p}{\lambda}\left[V_x\left(\sin\theta_2 + \sin\frac{\omega_r T_p}{2}\right) + V_y\left(\cos\frac{\omega_r T_p}{2} - \cos\theta_2\right)\right] \quad (16.21)$$

where θ_2 is the direction of the clutter cell with respect to the antenna pointing angle when the second pulse is received and ω_r is the antenna scan rate. The subscripts on the received signals Σ_i and Δ_i indicate the pulse reception sequence.

The difference pattern Δ is used to generate an in-phase correction for scanning motion and a quadrature correction for platform motion. This process yields the set of resultant signals R_{ij}, where the subscript i denotes the pulse pair

---

*The compensation required by $\Delta G^2(\theta)/2$ can be determined from a Taylor's series expansion of $G^2(\theta)$. In the preceding discussion we used the first derivative. Using higher-order terms gives an improved correction signal.
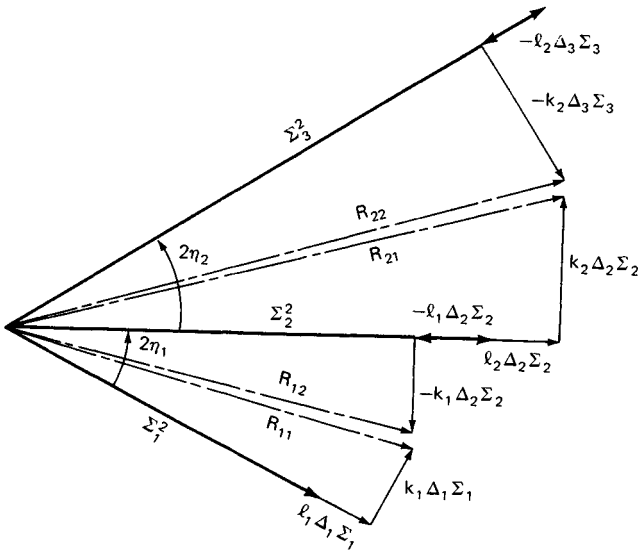
FIG. 16.16   Phasor diagram for simultaneous scanning and motion compensation.
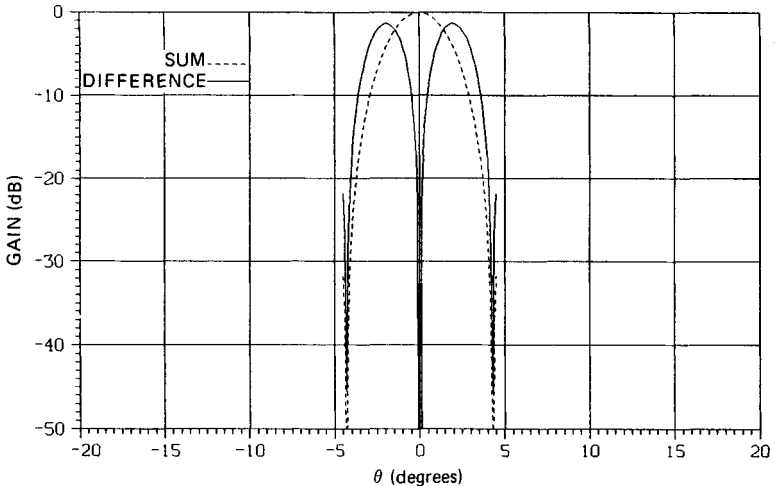
FIG. 16.17   Sum and difference patterns used to determine DPCA performance.

and the subscript j denotes the component of the pair. Since η_1 does not equal η_2, different weighting constants are required for each pulse pair. The values of k_1 for the quadrature correction of the first pulse pair, k_2 for the quadrature correction for the second pulse pair, $[_1$ for the in-phase correction for the first pulse pair, and $[_2$ for the in-phase correction for the second pulse pair are optimized by minimizing the integrated residue power over the significant portion of the antenna pattern, usually chosen between the first nulls of the main beam.

Figure 16.17 shows the sum and difference main-beam patterns for an aperture
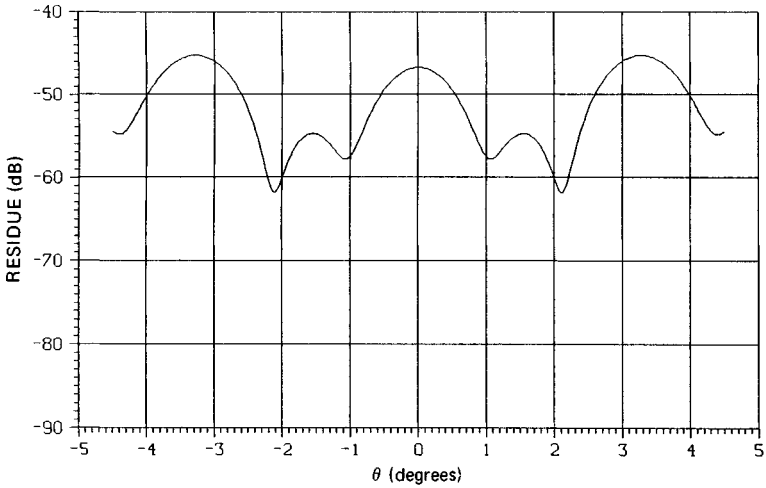
FIG. 16.18 DPCA clutter residue versus angle for normalized displacement $V_n = 0.04$ and normalized scanning motion $W_n = 0.04$.

20 wavelengths long. Figure 16.18 shows the residue for the case when the fraction of the horizontal aperture width a traveled per interpulse period T_p, $V_n = V_x T_p/a$, is equal to 0.04 and when the number of wavelengths that the aperture tip rotates per interpulse period, $W_n = a\omega_r T_p/2\lambda$, is equal to 0.04. The corresponding improvement factor is 52 dB.

The improvement factor is shown in Fig. 16.19 for a range of normalized platform motion V_n as a function of normalized scanning displacements W_n. The
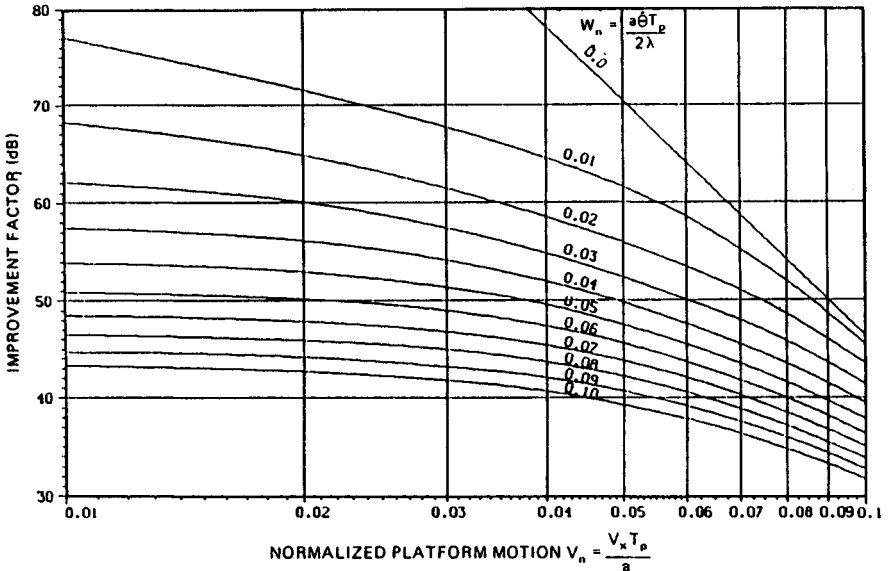
FIG. 16.19 DPCA improvement factor versus normalized platform motion V_n as a function of normalized scanning motion W_n.
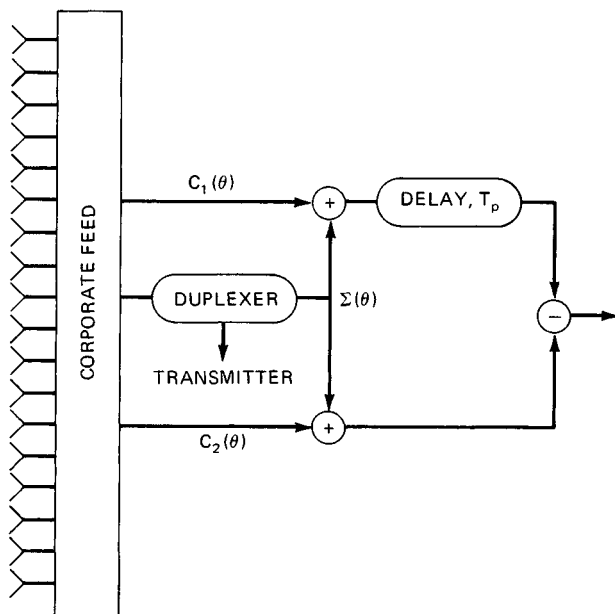
FIG. 16.20   Optimized DPCA phase compensation.

nonscanning case is shown as $W_n = 0$. The improvement factors were computed for the 20-wavelength aperture patterns shown in Fig. 16.17.

Andrews[14] has developed an optimization procedure for platform-motion compensation that rotates the phasors directly rather than by using a quadrature correction. The procedure determines the antenna feed coeficients for two compensation patterns, one of which, $C_1(\theta)$, is added to the sum pattern $\Sigma(\theta)$ and fed to the undelayed canceler path and the other, $C_2(\theta)$, is added to the sum pattern and fed to the delayed path as shown in Fig. 16.20. The procedure was developed for a single-delay canceler and a nonscanning antenna. Andrews used the procedure to minimize the residue power over the full antenna pattern, which includes the main-beam and sidelobe regions.

16.7   PLATFORM-MOTION COMPENSATION, FORWARD DIRECTION

The previous sections discussed the compensation for the component of platform motion parallel to the antenna aperture. TACCAR removes the average component of platform motion perpendicular to the aperture. Wheeler Laboratories (now Hazeltine Corporation) developed the Coincident Phase Center Technique (CPCT)[15] to remove the spectral spread due to the velocity component perpendicular to the aperture and due to the component parallel to the aperture. Removal of the component parallel to the aperture uses the DPCA pattern synthesis technique described in Ref. 8, which creates two similarly shaped illumination

functions whose phase centers are physically displaced. Removal of the component perpendicular to the aperture is accomplished by a novel extension of this concept.

The first term of Eq. (16.2) for spectral width due to platform motion approaches zero as the antenna points ahead. However the second term of Eq. (16.2) dominates as the antenna approaches within a few beamwidths of the aircraft's ground track. In this region

$$f_d \approx \frac{4V_y}{\lambda} \sin 2 \frac{\theta}{2} \approx \frac{V_y\theta^2}{\lambda} \qquad (16.22)$$

which yields a single-sided spectrum that is significantly narrower than the spectrum abeam. For moderate platform speeds and lower-frequency (UHF) radars this effect is negligible, and compensation is not required.

When it is necessary to compensate for this effect, the phase center of the antenna must be displaced ahead of the aperture and behind the aperture for alternate receive pulses so that the phase centers are coincident for a moving plat-

(a)

(b)
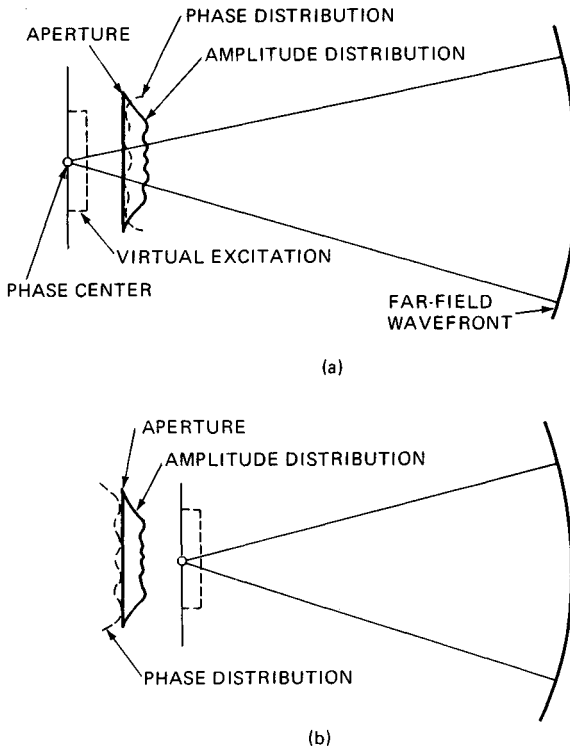
FIG. 16.21 CPCT concept showing displacement of the phase center (*a*) behind the physical aperture and (*b*) ahead of the physical aperture. (*Courtesy of Hazeltine Inc.*[15])

form. This technique can be extended to more than two pulses by using the necessary phase-center displacements for each pulse. In order to maintain the effective PRF, the displacement must compensate for the two-way transmission path. To accomplish this displacement, near-field antenna principles are utilized. A desired aperture distribution function is specified. The near-field amplitude and phase are calculated at a given distance from the origin. If this field is used as the actual illumination function, a virtual aperture is created with the desired distribution function at the same distance behind the physical antenna. Figure 16.21a[15] shows the phase and amplitude distribution required to form a uniform virtual distribution displaced behind the physical aperture. It can be shown that if the phase of the illumination function is reversed $\phi' = -\phi$, the desired virtual distribution function is displaced ahead of the aperture as shown in Fig. 16.21b.

In practice, performance is limited by the ability to produce the required illumination function. As the displacement increases, a larger physical aperture size is required to produce the desired virtual aperture size owing to beam spreading. This can be seen in Fig. 16.21. The effectiveness of the correction varies with elevation angle since the actual displacement along the line of sight varies with elevation angle. This effect is more pronounced at higher aircraft speeds and higher radar frequencies. A change in the magnitude of the correction factor or even the compensation pattern with range, height, and velocity could be utilized to retain performance.

Figure 16.22 illustrates the theoretical MTI performance of a CPCT system as a function of beam-pointing direction and interpulse motion normalized to the interpulse motion used to design the compensation patterns. (*Cancellation ratio*
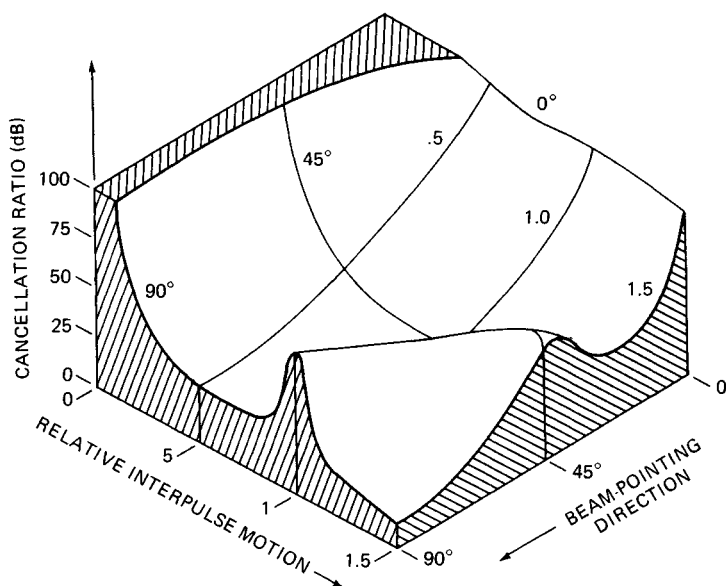
FIG. 16.22 CPCT cancellation ratio, in decibels, as a function of relative interpulse motion and beam-pointing direction. (*Courtesy of Hazeltine Inc.*[15])

is defined as the ratio of input clutter power to output clutter residue power.) The peak on the 90° axis is typical of the optimized DPCA performance illustrated in Figure 16.12.

16.8   SPACE-TIME ADAPTIVE MOTION COMPENSATION

Several methods have been described to compensate for antenna motion. All these techniques are applied in the radar design phase for a specific set of operational parameters. Controls (usually automatic) are provided to adjust weights for operational conditions around the design value.

The development of digital radar technology and economical high-speed processors allows the use of dynamic space-time adaptive array processing,[16] whereby a set of antenna patterns that displace the phase center of the array both along and orthogonal to the array are continually synthesized to maximize the signal-to-clutter ratio. Spatial adaptive array processing combines an array of signals received at the same instant of time that are sampled at the different spatial locations corresponding to the antenna elements. Temporal adaptive array processing combines an array of signals received at the same spatial location (e.g., the output of a reflector antenna) that are sampled at different instances of time, such as several interpulse periods for an adaptive MTI. Space-time adaptive array processing combines a two-dimensional array of signals sampled at different instances of time and at different spatial locations.

A basic block diagram of a radar incorporating space-time adaptive array processing is shown in Fig. 16.23. Circuits for auxiliary functions such as pulse compression, clutter gating, synchronization, and TACCAR are not shown. With the exception of the interchange of the corporate-feed and duplexing functions, the transmit channel is identical to that of any other radar. An individual duplexer is placed between each corporate-feed output and its corresponding antenna element. Provision could be included for electronic beam steering using high-power phase shifters or transmit modules with low-power beam steering.

On receive, each duplexer output is sent to its own digital receiver and adaptive processing module (APM), which provides a weighted undelayed signal that is combined with the outputs of the other adaptive processing modules to form an undelayed antenna beam. The weighted signal received on the previous pulse is combined with the corresponding outputs of the adaptive processing modules to form a delayed antenna beam. The two beams are then subtracted to produce the single-delay AMTI output. The output is then sent to the automatic detector for further processing and display. The output is also returned to the adaptive processing modules.

Figure 16.24 shows the block diagram of a typical digital receiver. The signal received from a single antenna element is amplified and converted to IF. The IF signal is further amplified and converted to baseband by using the synchronous demodulators. One of the bipolar video outputs, I, is the component that is in phase with the reference oscillator. The other bipolar video output, Q, is in quadrature with the reference oscillator. The two bipolar video signals are sampled for each range cell and converted to digital representation by the A/D converters. The output logic formats the I and Q values for transfer to the adaptive processing module. The I, Q pair of numbers is a phasor representing the instantaneous phase and amplitude of each range cell in rectangular coordinates.

Figure 16.25 shows a block diagram of the adaptive processing module used
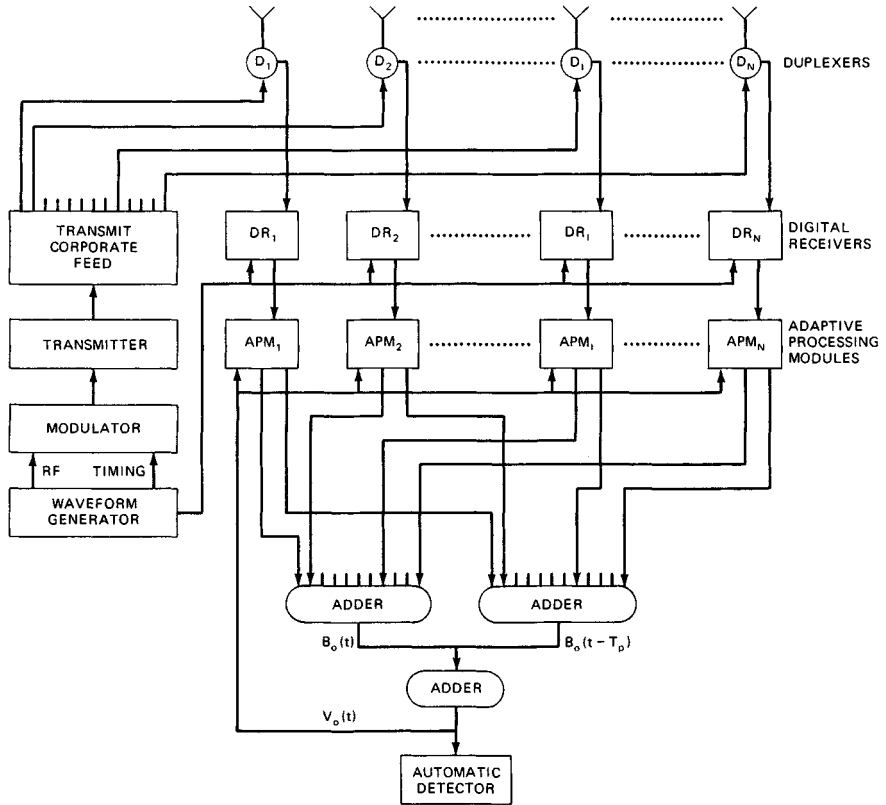
FIG. 16.23   Block diagram of a digital space-time adaptive array radar.

for space-time adaptive array processing. All components are digital processing blocks that can be implemented in various combinations of hardware and software. The complex value of the sampled signal $V_i(t)$ is multiplied by the complex adaptive weight W_{i1} to form the ith-channel input to the adder forming the undelayed antenna beam. The value is also routed to a buffer for storage. The previously stored value $V_i(t - T_p)$ is multiplied by the delayed channel weight W_{i2} to form the ith-channel input to the adder forming the delayed beam.
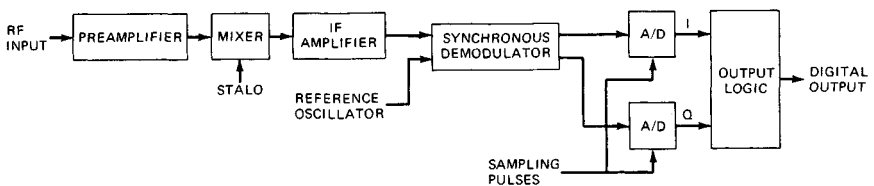
FIG. 16.24   Block diagram of a digital receiver.
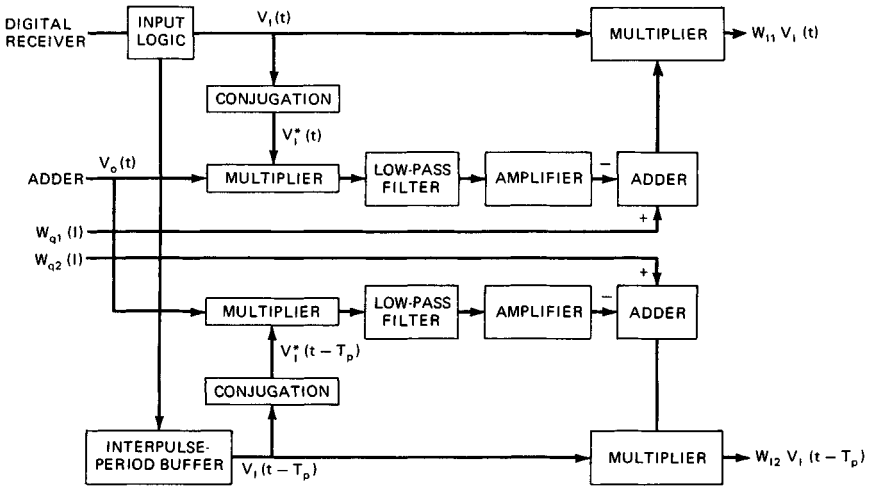
FIG. 16.25 Block diagram of an adaptive processing module.

The weights W_{i1} and W_{i2} are determined by the Howells-Applebaum algorithm.[17] The correlation between the input signal $V_i(t)$ and the output signal $V_o(t)$ is determined by multiplying $V_o(t)$ by the complex conjugate of the input signal $V^*_i(t)$ and averaging the resultant by passing it through a low-pass digital filter. This correlation is amplified and subtracted from the appropriate quiescent weight W_{q1i} or W_{q2i} to obtain the slowly varying weight W_{i1} or W_{i2}. The quiescent weight is the product of the antenna illumination factor for the ith element that will yield the desired antenna pattern and the MTI weight for the delayed or undelayed pulse. This would include the phase component required to steer the beam in a given direction. The closed-loop action will drive the weight so that the average correlation of the output V_o and each input V_i approximates the value of the quiescent weight.

$$(V^*_i V_o) \approx W_{q1} \qquad \text{for} \qquad i = 1 \text{ to } 2N$$

$$(V^*_i \Sigma W_j V_j) \approx W_{q1}$$

$$\Sigma (V^*_i V_j) W_j \approx W_{q1} \qquad\qquad (16.23)$$

where ( ) indicates the time average. If we define the N values of each delayed variable by extending the subscript range over $N + 1$ to K, where $K = 2N$, and define $m_{ij} = (V^*_i V_j)$, we have the set of equations

$$m_{11}W_1 + W_{12}W_2 + \cdots + m_{1K}W_K = W_{q1}$$

$$M_{21}W_1 + m_{22}W_2 + \cdots + m_{2K}W_K = W_{q2}$$

$$\begin{matrix} \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \end{matrix}$$

$$m_{K1}W_1 + m_{K2}W_2 + \cdots + m_{KK}W_K = W_{qK}$$

which can be expressed in matrix notation as

$$MW = W_q \qquad (16.24)$$

This set of equations can be solved for the steady-state set of weights $W1$ to W_K, which can be expressed in matrix notation as the familiar equation

$$W = M^{-1}W_q \qquad (16.25)$$

These weights have been shown[17,18] to be the optimum set which maximizes the signal-to-interference ratio. Because of the smoothing required to keep the weights from jittering,[18] the weights adapt to their steady-state values in a time determined by the clutter power and the allowable steady-state variation in the weights. Other algorithms[19] can speed up the adaptation rate, but a more complex mechanization is required.

This process results in a delayed beam and an undelayed beam, whose phase centers are offset to compensate for platform motion. If jamming is present in the sidelobes or on the shoulder of the main beam, nulls will be formed in each jamming direction. If mixed clutter and jamming are present, the weights will adjust to maximize the signal-to-total-interference ratio. The signal is defined as a short-duration pulse return from the direction determined by the quiescent weights and at a doppler frequency corresponding to half of the PRF. Independent quiescent weights for each subchannel could be utilized to optimize the doppler response for another frequency.

Performance Capability of Space-Time Adaptive Arrays. The performance attainable from the space-time array is limited by the aircraft speed, the array alignment with respect to the aircraft ground track, and the system accuracies. A performance analysis for a 16-element, two-pulse space-time array is presented in Figs. 16.26 and 16.27. The antenna elements were spaced at a half wavelength and assumed to be omnidirectional. The clutter model was assumed to be homogeneous. The improvement factor is based on the fully adapted weights. The clutter-to-noise ratio was selected to limit performance to 92 dB.

The improvement factor for adaptive arrays is usually defined as the ratio of the signal-to-interference-power ratio at the output of the processor to the signal-to-interference ratio at the input of the processor. The signal is specified as coming from the direction and at the doppler frequency specified by the quiescent weights. In MTI systems, it is customary to define the improvement factor as the average response over the doppler interval. For a single-delay processor, the MTI improvement factor is 3 dB less than the adapted improvement factor shown in Figs. 16.28 and 16.27.

Figure 16.28 shows the improvement factor for an array pointing along the ground track of the aircraft as a function of motion expressed in terms of wavelengths per interpulse period. The dashed line shows the unadapted single-delay improvement factor for an antenna with a Dolph-Chebyshev aperture illumination that provides a 28 dB uniform peak sidelobe level. The solid line shows the adapted improvement factor to be 92 dB for a stationary antenna, reducing to 89 dB for 4 wavelengths per interpulse-period platform motion. Figure 16.27 shows
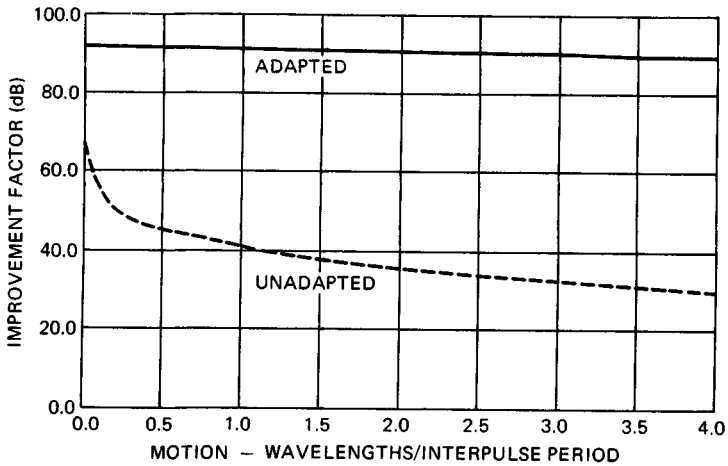
FIG. 16.26 Adapted and unadapted improvement factor as a function of normalized antenna motion per interpulse period; 16-element (half-wavelength spacing), two-pulse space-time adaptive processor; antenna array aligned to the ground track.
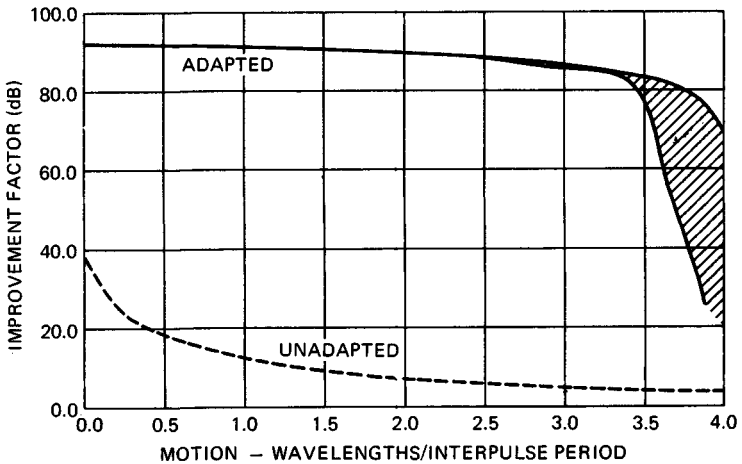
FIG. 16.27 Adapted and unadapted improvement factor as a function of normalized antenna motion per interpulse period; 16-element (half-wavelength spacing), two-pulse space-time adaptive processor; antenna array aligned to perpendicular to the ground track.

the performance when the antenna is pointing abeam. This is the standard DPCA case. The performance holds to 1.5 wavelengths per interpulse-period platform motion, then decreases slightly, and drops off sharply above 3.5 wavelengths per interpulse period. The shaded region is where the improvement factor varies within the limits. A peak occurs when the platform motion is a multiple of a quarter wavelength. The clutter-to-noise-ratio limitation results in cusping not being
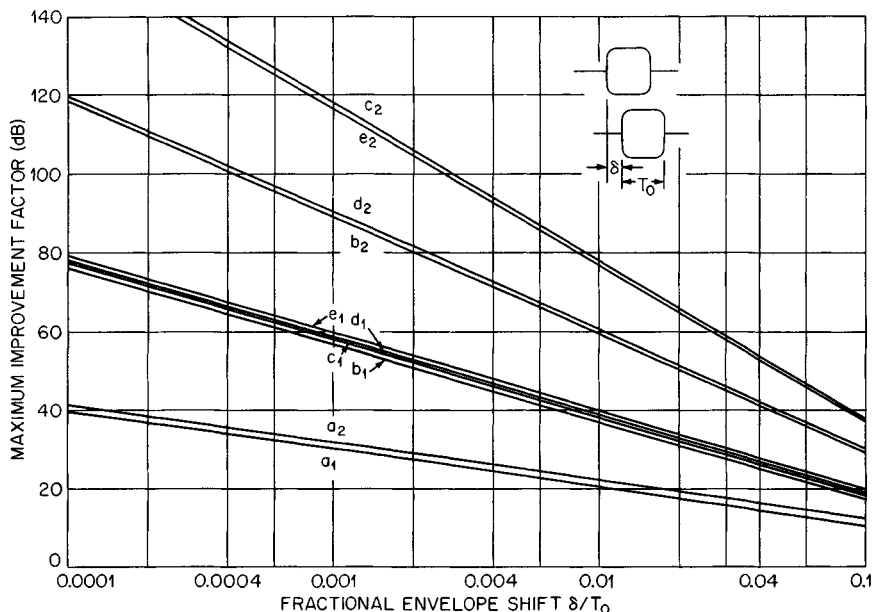
FIG. 16.28 Maximum improvement factor as a function of a fractional envelope shift for typical pulse envelope shapes. Subscript 1 indicates single delay; 2, double delay. Pulse shape a = rectangular; b = cosine; c = cosine-squared; d = triangular; e = gaussian.

visible at speeds less the 3.5 wavelengths per interpulse period. The performance at other angles is between these two cases, and peaking does not occur.

16.9 LIMITATION OF IMPROVEMENT FACTOR DUE TO PULSE ENVELOPE SHIFT

The doppler frequency that arises because of the radial component of aircraft motion results in an incremental phase shift between successive radar pulses.

The envelope of the radar pulse is also delayed a corresponding amount. The TACCAR circuit usually compensates for the phase delay at IF by changing the phase of a CW reference oscillator. Hence the envelope of a single pulse is unaffected. The mismatch in the envelope delay time between successive pulses results in a residual signal, sometimes called *ranging noise*.

Figure 16.28 shows the effect of this residual on the performance of a single-delay canceler. The idealized rectangular pulse envelope gives a pessimistic picture of this effect. Most conventional representations of pulse shape give about the same performance, 20 dB per decade.

Figure 16.28 also shows the effect of the residual on a double cancellation system. In the case of double-delay cancellation the rectangular-pulse case shows only 1.8 dB improvement over the single-delay case. However, more realistic pulse representations show substantial improvement. The triangular and cosine representations have a rolloff of about 30 dB per decade; the smoother cosine-squared and gaussian representations roll off at 40 dB per decade.

16.10   EFFECT OF MULTIPLE SPECTRA

An airborne search-radar system may be operated at an altitude so that the radar horizon is approximately at the maximum range of interest. This results in sea or ground clutter being present at all ranges of interest. Other clutter sources such as rain and chaff may coexist with the surface clutter. In most instances these sources are moving at a speed determined by the mean wind aloft and have a mean doppler frequency significantly different from that of the surface clutter. If the MTI filter is tracking the surface clutter, the spectra of the sources with a different mean doppler frequency lie in the passband of the MTI filter. A 20-kn differential in an S-band system corresponds to 200 Hz, which would be at an optimum response in a 400-PRF system. A single-delay secondary canceler can be cascaded with either a single-delay or a double-delay primary canceler. The primary canceler tracks the mean surface velocity and rejects surface clutter. The single-delay canceler tracks the secondary source and rejects it. Since the pass and rejection bands of the two cancelers overlap, the MTI improvement factor for each clutter source is a function of their spectral separation.

Figure 16.29 shows the improvement factor for a double canceler which consists of two single cancelers, each tracking one of the spectra. It can be seen that, as the separation varies from 0 to one-half of the PRF, the performance degrades from that equivalent to a double canceler to the performance of a single canceler at half of the PRF.

The triple canceler has a double-delay canceler tracking the primary spectra and a single-delay canceler tracking the secondary spectra. The performance of the primary system varies from that of a triple canceler to a level less than that of a double canceler. The secondary-system performance varies from that of a triple canceler to a performance level lower than that of a single canceler.

16.11   DETECTION OF GROUND MOVING TARGETS

Vehicles and ships may have radial speeds that are significantly greater than the clutter velocity spectrum. This allows these targets to be detected. However, for an AEW system operating overland, ground traffic can saturate the tracking system. Furthermore, traffic density on major highways, target aspect changes causing strong scintillation, and shadowing by cultural features that occurs at low grazing angles can result in misassociation in the target-tracking system, causing false and runaway tracks. Runaway tracks are false tracks whose high speed causes them to rapidly move away from the true target position. They can associate with other reported positions or false alarms and thus perpetuate themselves. Since ground traffic is not of interest in the AEW case, these undesired targets are censored on the basis of highway grid maps, the small change in range during the antenna dwell, or the small velocity determined by a scan-to-scan processor or the tracking system. High-PRF pulse doppler radars use a low-frequency stopband filter to reject these velocity regions along with main-beam clutter. Low-PRF AEW radars bypass these censor circuits in the portion of the surveillance region that is over water. This allows slow-moving shipping targets to be detected.

Air-to-surface search radars, as well as airborne battlefield surveillance radars, are designed to maximize detection of slowly moving targets. Higher-
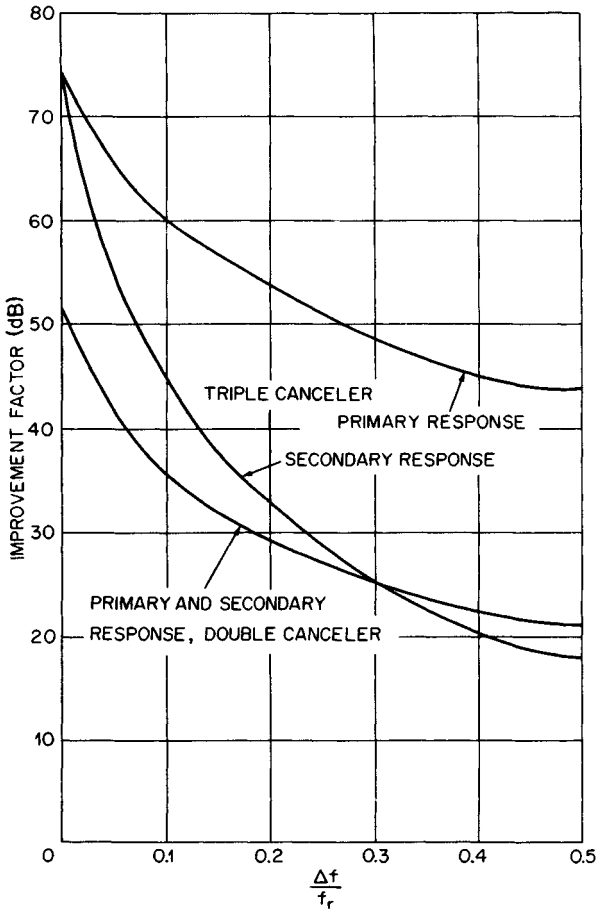
FIG. 16.29   MTI improvement factor for a double-notch canceler tracking two spectra as a function of the normalized spectra separation $\Delta f/f_r$. Normalized spectral width $\sigma_c/f_r = 0.01$.

frequency bands (X or K) are chosen to maximize the doppler shift. The PRF is chosen to optimize detection over the expected doppler frequency region of these targets. Since a strong clutter background is usually present, battlefield surveillance radars can effectively utilize noncoherent MTI techniques. However, the clutter spectrum is convolved with the target spectrum, which broadens the resultant target spectrum, thus widening the blind-speed zone and reducing doppler resolution. When the target phase coincides with the clutter phase, the targets are suppressed. In regions that are shadowed by hills or mountains, the targets are not detected.

Side-looking radars can produce a large number of pulses, thus increasing radar sensitivity. If a coherent radar is used, improved sensitivity and resolution can be obtained by using doppler filter banks or digital fast Fourier transform (FFT) processing. If the platform motion compared with the aperture length is sufficiently large, platform-motion compensation will be required.

Ship detection can be improved by rapidly scanning the antenna so that sea clutter is decorrelated and surface-target returns are integrated or leave a pattern of returns indicating their track. In some cases, frequency agility can also be utilized to decorrelate clutter and integrate ship target returns. Scan-to-scan video cancellation can be utilized for detecting moving targets overland if their scan-to-scan motion is of the order of the radar pulse width.

REFERENCES

1. Emerson, R. C.: Some Pulsed Doppler MTI and AMTI Techniques, *Rand Corporation Rept. R-274, DDC Doc. AD 65881*, Mar. 1, 1954. (Reprinted in Ref. 6.)

2. George, T. S.: Fluctuations of Ground Clutter Return in Airborne Radar Equipment, *Proc. IEE (London)*, vol. 99, pt. IV, pp. 92–99, April 1952.

3. Dickey, F. R., Jr.: Theoretical Performance of Airborne Moving Target Indicators, *IRE Trans.*, vol. PGAE-8, pp. 12–23, June 1953.

4. Berkowitz, R. S. (ed.): "Modern Radar: Analysis, Evaluation and System Design," John Wiley & Sons, New York, 1966.

5. Barton, D. K.: "Radar Systems Analysis," Prentice-Hall, Englewood Cliffs, N.J., 1964.

6. Schlerer, D. C. (ed.): "MTI Radar," Artech House, Inc., Norwood, Mass., 1978.

7. Dickey, F. R., Jr., and M. M. Santa: Final Report on Anticlutter Techniques, *General Electric Company Rept. R65EMH37*, Mar. 1, 1953.

8. Anderson, D. B.: A Microwave Technique to Reduce Platform Motion and Scanning Noise in Airborne Moving Target Radar, *IRE WESCON Conv. Rec.*, vol. 2, pt. 1, pp. 202–211, 1958.

9. "Final Engineering Report on Displaced Phase Center Antenna," vol. 1, Mar. 26, 1956; vols. 2 and 3, Apr. 18, 1957, General Electric Company, Schenectady, N.Y.

10. Urkowitz, H.: The Effect of Antenna Patterns on Performance of Dual Antenna Radar Moving Target Indicators, *IEEE Trans.*, vol. ANE-11, pp. 218–223, December 1964.

11. Tsandoulis, G. N.: Tolerance Control in an Array Antenna, *Microwave J.*, pp. 24–35, October 1977.

12. Shroeder, K. G.: Beam Patterns for Phase Monopulse Arrays, *Microwaves*, pp. 18–27, March 1963.

13. Grissetti, R. S., M. M. Santa, and G. M. Kirkpatrick: Effect of Internal Fluctuations and Scanning on Clutter Attenuation in MTI Radar, *IRE Trans.*, vol. ANE-2, pp. 37–41, March 1955.

14. Andrews, G. A.: Airborne Radar Motion Compensation Techniques: Optimum Array Correction Patterns, *Naval Res. Lab. Rept.* 7977, Mar. 16, 1976.

15. Lopez, A. R., and W. W. Ganz: CPCT Antennas for AMTI Radar, vol. 2: Theoretical Study, *Air Force Avionics Lab. Rept. WL1630.22, AD 51858*, June 1970. (Not readily available.)

16. Brennan, L. E., J. D. Mallett, and I. S. Reed: Adaptive Arrays in Airborne MTI Radar, *IEEE Trans.*, vol. AP-24, pp. 607–615, September 1976.

17. Applebaum, S. P.: Adaptive Arrays, *IEEE Trans.*, vol. AP-24, pp. 585–598, September 1976.

18. Brennan, L. E., E. L. Pugh, and I. S. Reed: Control Loop Noise in Adaptive Array Antennas, *IEEE Trans.*, vol. AES-7, March 1971.

19. Monzingo, R. A., and T. W. Miller: "Introduction to Adaptive Arrays," John Wiley & Sons, New York, 1980.

CHAPTER 17
PULSE DOPPLER RADAR

William H. Long
David H. Mooney
William A. Skillman
Westinghouse Electric Corporation

17.1 CHARACTERISTICS AND APPLICATIONS

Nomenclature. For the purpose of this chapter, the term *pulse doppler* (PD) will be used for radars to which the following apply:

1. They utilize coherent transmission and reception; that is, each transmitted pulse and the receiver local oscillator are synchronized to a free-running, highly stable oscillator.

2. They use a sufficiently high pulse repetition frequency (PRF) to be ambiguous in range.

3. They employ coherent processing to reject main-beam clutter, enhance target detection, and aid in target discrimination or classification.

Applications. PD is applied principally to radar systems requiring the detection of moving targets in a severe clutter environment. Table 17.1 lists typical applications [1-10] and requirements. This chapter will deal principally with airborne applications, although the basic principles can also be applied to the ground-based case.

PRFs. Pulse doppler radars are generally divided into two broad PRF categories: medium and high PRF.[11] In a medium-PRF radar[12-14] the target and clutter ranges and velocities of interest are usually ambiguous, while in a high-PRF radar[15] the range is ambiguous but the velocity is unambiguous (or has at most a single velocity ambiguity as discussed later).

A low-PRF radar, commonly called a moving-target indicator (MTI),[16] is one in which the ranges of interest are unambiguous while the velocities are usually ambiguous. MTI radars are generally not categorized as pulse doppler radars, although the principles of operation are similar. A comparison of MTI and pulse doppler radars is shown in Table 17.2.

TABLE 17.1   Pulse Doppler Applications and Requirements

Radar application	Requirements
Airborne or spaceborne surveillance	Long detection range; accurate range data
Airborne interceptor or fire control	Medium detection range; accurate range, velocity data
Ground-based surveillance	Medium detection range; accurate range data
Battlefield surveillance (slow-moving target detection)	Medium detection range; accurate range, velocity data
Missile seeker	May not need true range information
Ground-based weapon control	Short range; accurate range, velocity data
Meteorological	High velocity and range data resolution
Missile warning	Short detection range; very low false-alarm rate

TABLE 17.2   Comparison of MTI and Pulse Doppler (PD) Radars

	Advantages	Disadvantages
MTI—low PRF	Can sort clutter from targets on basis of range. No range ghosts. Front-end STC suppresses sidelobe detections and reduces dynamic range requirements.	Low doppler visibility due to multiple blind speeds. Poor slow-moving target rejection. Cannot measure radial target velocity.
PD—medium PRF	Good performance at all target aspects. Good slow-moving target rejection. Measures radial velocity. Less range eclipsing than in high PRF.	Range ghosts. Sidelobe clutter limits performance. High stability requirements due to range folding.
PD—high PRF	Can be sidelobe clutter–free for some target aspects. Single doppler blind zone at zero velocity. Good slow-moving target rejection. Measures radial velocity. Velocity-only detection can improve detection range.	Sidelobe clutter limits performance. Range eclipsing. Range ghosts. High stability requirements due to range folding.

Pulse Doppler Spectrum.   The transmitted spectrum of a pulse doppler radar consists of discrete lines at the carrier frequency f_0 and at sideband frequencies $f_0 \pm if_R$, where f_R is the PRF and i is an integer. The envelope of the spectrum is determined by the pulse shape. For the rectangular pulses usually employed, a $(\sin x)/x$ spectrum is obtained.

The received spectrum from a stationary target has lines that are doppler-shifted proportionally to the line of sight, or radial velocity, between the radar platform and the target. The two-way doppler shift is given by $f_d = (2V_R/\lambda) \cos \psi_0$, where λ is the radar wavelength, V_R is the radar platform velocity, and ψ_0 is
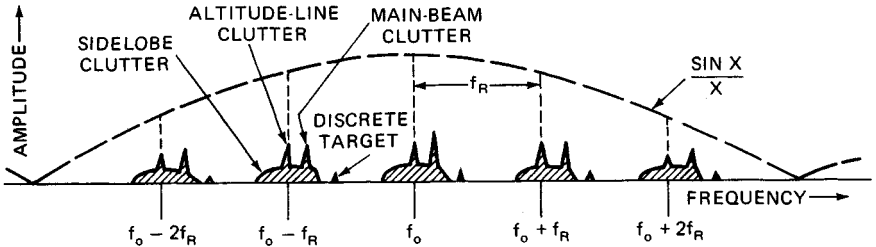
FIG. 17.1 Clutter and target frequency spectrum from a horizontally moving platform.

the angle between the velocity vector and the line of sight to the target. Illustrated in Fig. 17.1 is the received pulsed spectrum with returns from continuous clutter, such as the ground or clouds, and from discrete targets, such as aircraft, automobiles, tanks, etc.

Figure 17.2 shows the unfolded spectrum (i.e., no spectral foldover from adjacent PRF lines) in the case of horizontal motion of the radar platform, with a velocity V_R. The clutter-free region is defined as that portion of the spectrum in which no ground clutter can exist. (A clutter-free region usually does not exist with medium PRFs.) The sidelobe clutter region, $4V_R/\lambda$ in width, contains ground clutter power from the sidelobes of the antenna, although the clutter power may be below the noise level in part of the region. The main-beam region, located at $f_0 + (2V_R/\lambda)$ cos ψ_0, contains the strong return from the main beam of the antenna striking the ground at a scan angle of ψ_0, measured from the velocity vector. Rain and chaff clutter may also be large when the main beam illuminates a rain or chaff cloud. Motion due to winds may displace and/or spread the return in frequency.

Altitude-line clutter, which is due to ground clutter at near normal incidence directly below the radar platform, is at zero doppler if there is no vertical component of platform velocity. A discrete target return in the main beam is shown at $f_T = f_0 + (2V_R/\lambda)$ cos $\psi_0 + (2V_T/\lambda)$ cos ψ_T, where the target velocity is V_T, with an angle ψ_T between the target velocity vector and the radar target line of sight.
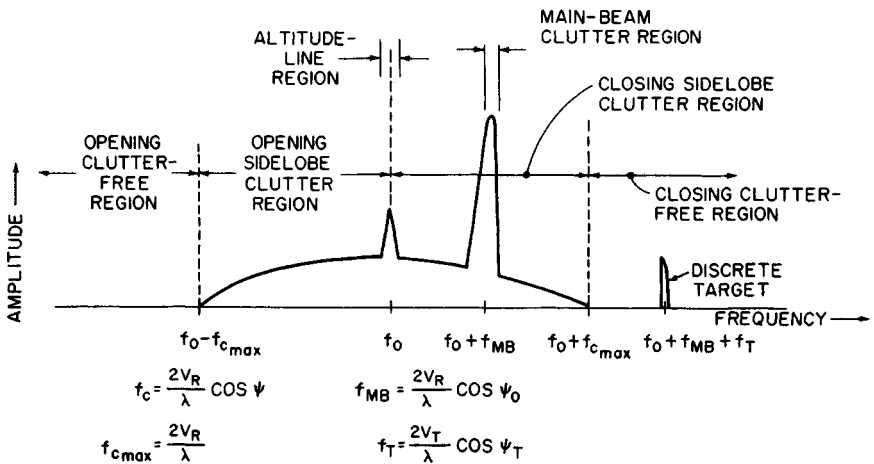
FIG. 17.2 Unfolded spectrum (with no clutter tracking).

The components of the spectrum shown in Fig. 17.2 will also vary with range as discussed later.

Figure 17.3 illustrates the various clutter doppler frequency regions as a function of the antenna azimuth and relative radar and target velocities, again for an unfolded spectrum. The ordinate is the radial, or line-of-sight, component of target velocity in units of radar platform velocity, so that the main-beam clutter region is at zero velocity and the sidelobe clutter region frequency boundaries vary sinusoidally with antenna azimuth. Thus, it shows the doppler regions in which the target can become clear of sidelobe clutter. For example, if the antenna azimuth angle is at zero, any head-on target ($V_T \cos \psi_T > 0$) is clear of sidelobe clutter, whereas if the radar is in trail behind the target ($\psi_T = 180°$ and $\psi_0 = 0°$), the target's radial velocity has to be greater than twice that of the radar to become clear of sidelobe clutter.

The sidelobe clear and clutter regions can also be expressed in terms of the aspect angle with respect to the target,[14] as shown in Fig. 17.4. Here, collision geometry is assumed in which the radar and target aircraft fly straight-line paths toward an intercept point; the look angle of the radar ψ_0 and the aspect angle of the target ψ_T are constant for a given set of radar and target speeds V_R and V_T, respectively. The center of the diagram is the target, and the angle to the radar on the circumference is the aspect angle. The aspect angle and look angles satisfy the equation $V_R \sin \psi_0 = V_T \sin \psi_T$, which is defined as a collision course. The target aspect angle is zero for a head-on condition and 180° for a tail chase. The aspect angle corresponding to the boundary between the sidelobe clutter region and the sidelobe clear region is a function of the relative radar-target velocity ratio and is shown in Fig. 17.4 for four cases. Case 1 is where the radar and target speeds are equal and the target can be seen clear of sidelobe clutter in a head-on
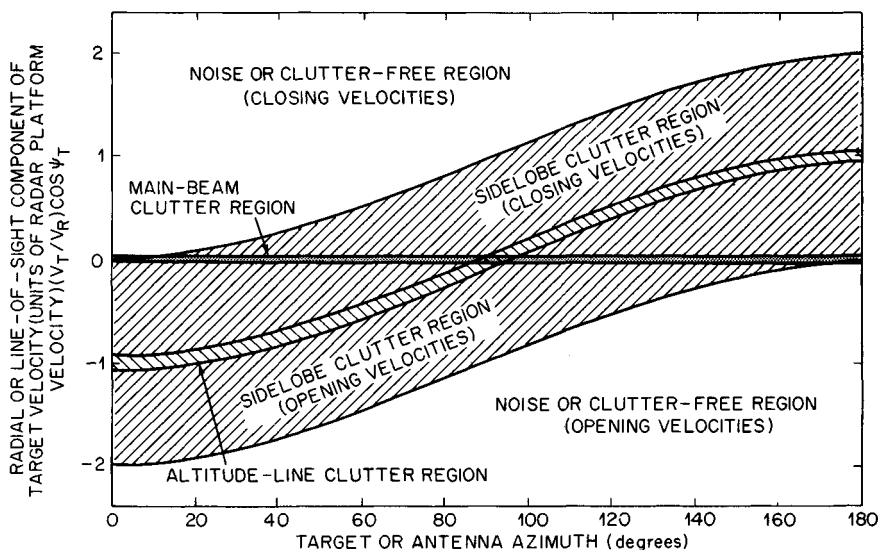
NOTE: WIDTH OF ALTITUDE-LINE AND MAIN-BEAM CLUTTER REGIONS VARIES WITH CONDITIONS; AZIMUTH IS MEASURED FROM RADAR PLATFORM VELOCITY VECTOR TO THE ANTENNA BORESIGHT OR TO THE LINE OF SIGHT TO THE TARGET; HORIZONTAL-MOTION CASE.

FIG. 17.3 Clutter and clutter-free regions as a function of target velocity and azimuth.
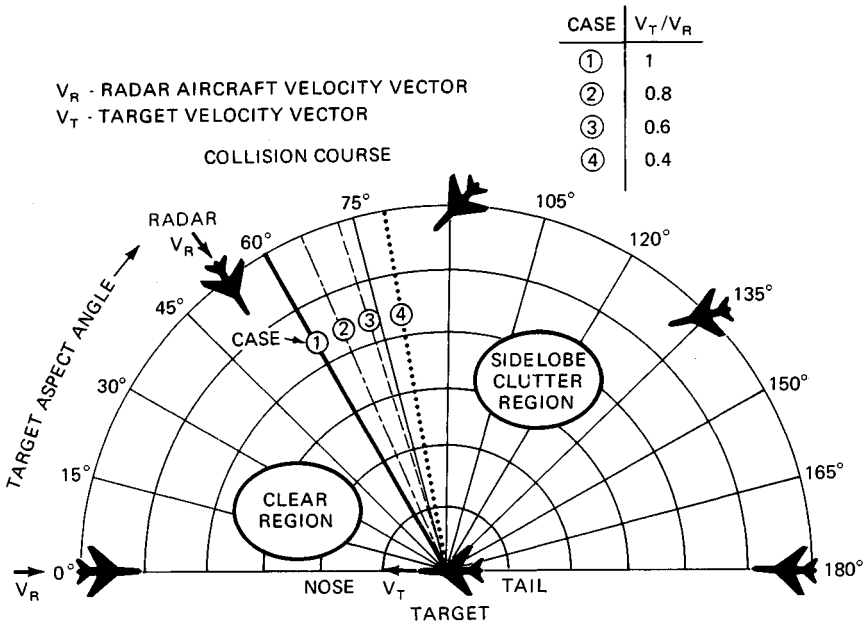
FIG. 17.4 Sidelobe clutter-clear regions versus target aspect angle.

aspect out to 60° on either side of the target's velocity vector. Similarly, Cases 2 to 4 show conditions where the target's speed is 0.8, 0.6, and 0.4 times the radar's speed, in which case the target can be seen clear of sidelobe clutter over a region of up to ±78.5° relative to the target's velocity vector. Again, these conditions are for an assumed collision course. As is evident, the aspect angle of the target clear of sidelobe clutter is always forward of the beam aspect.

Ambiguities and PRF Selection. Pulse doppler radars are generally ambiguous in either range or doppler, or both. The unambiguous range R_u is given by $c/2f_R$, where c is the speed of light and f_R is the PRF.

If the maximum target velocity to be observed is $\pm V_{T\max}$, then the minimum value of PRF, $f_{R\min}$, which is unambiguous in velocity (both magnitude and doppler sense, i.e., positive and negative), is

$$f_{R\min} = 4V_{T\max}/\lambda \tag{17.1}$$

However, some pulse doppler radars employ a PRF which is unambiguous in velocity magnitude only, i.e., $f_{R\min} = 2V_{T\max}/\lambda$, and rely on detections in multiple PRFs during the time on target to resolve the ambiguity in doppler sense. These types of radars can be considered to be in the high-PRF category if the older definition of high PRF (no velocity ambiguity) is extended to allow one velocity ambiguity, that of doppler sense. The lower PRF eases the measurement of true range while retaining the high-PRF advantage of a single blind-speed region near zero doppler.

The choice between high and medium PRF involves a number of consider-ations, such as transmitter duty cycle limit, pulse compression availability, signal-processing capability, missile illumination requirements, etc., but often de-pends on the need for all-aspect target detectability. All-aspect coverage requires good performance in tail chase, where the target doppler is in the sidelobe clutter region near the altitude line. In a high-PRF radar, the range foldover may leave little clear region in the range dimension, thus degrading target detectability. By using a lower or medium PRF, the clear region in range is increased at the ex-pense of velocity foldover for high-doppler targets that are in the clutter-free re-gion in high PRF. For example, Fig. 17.5 shows the clutter-plus-noise-to-noise ratio in range doppler coordinates for a 12-kHz PRF at an altitude of 6000 ft showing the main-beam clutter, altitude line, and sidelobe clutter. The range di-mension represents the unambiguous range interval R_u, and the frequency dimen-sion represents the PRF interval. As is evident, there is a range doppler region in which the sidelobe clutter is below thermal noise and in which good target detectability can be achieved. The main-beam clutter is filtered out.
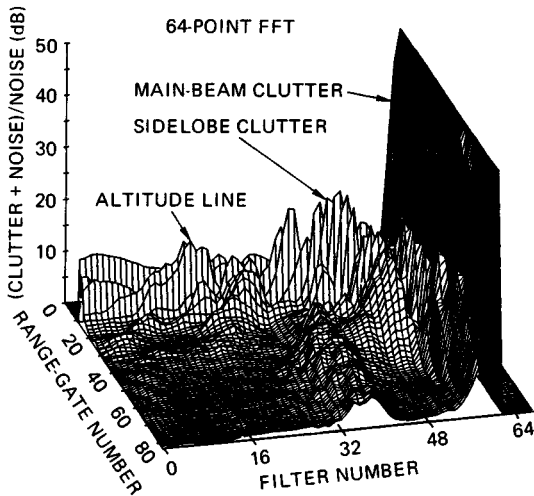
FIG. 17.5   Clutter-plus-noise-to-noise ratio in range doppler space.

Because the clutter is folded in both range and doppler with medium PRF, a number of PRFs may be required to obtain a satisfactory probability of sufficient detections to resolve the range and doppler ambiguities. The multiple PRFs move the relative location of the clear regions so that all-aspect target coverage is achieved. Since the sidelobe clutter generally covers the doppler region of inter-est, the ratio of the region with sidelobe clutter below noise relative to the total range-doppler space is a function of the radar altitude, speed, and antenna sidelobe level.

If a high-PRF waveform is used, the clear-range region disappears because the sidelobe clutter folds in range into the unambiguous range interval (assuming the target doppler is such that it still competes with the sidelobe clutter). However, in

those doppler regions free of sidelobe clutter, as shown in Figs. 17.3 and 17.4, target detectability is limited only by thermal noise, independently of radar altitude, speed, and sidelobe level. This requires system stability sidebands to be well below noise for the worst-case main-beam clutter. Thus, although medium PRF provides all-aspect target coverage, the target is potentially competing with sidelobe clutter at all aspects, whereas with high PRF a target can become clear of sidelobe clutter at aspect angles forward of the beam aspect.

Basic Configuration.   Figure 17.6 shows a representative configuration of a pulse doppler radar utilizing digital signal processing under the control of a central computer. Included are the transmitter suppression circuits, main-beam and sidelobe discrete rejection circuits, and ambiguity resolvers. The radar computer receives inputs from the on-board systems, such as the inertial unit and operator controls, and performs as a master controller for the radar. As such, it does the track loop and automatic gain control (AGC) loop filtering, antenna scan pattern generation, and clutter positioning as well as the target-processing functions (such as centroiding). In addition, the computer performs the multiple-target track functions when the radar is in a track-while-scan mode and may execute radar self-test and calibration routines. For simplicity only the search processing is shown.

Duplexer.   The duplexer in a pulse doppler radar is usually a passive device such as a circulator which effectively switches the antenna between the transmitter and receiver. Considerable power may be coupled to the receiver since typically 20 to 25 dB isolation may be expected from ferrite circulators.

Receiver-Protector (R/P).   The receiver-protector is a fast-response, high-power switch which prevents the transmitter output from the duplexer from damaging the sensitive receiver front end. Fast recovery is required to minimize desensitization in the range gates following the transmitted pulse.

RF Attenuator.   The RF attenuator is used both for suppressing transmitter leakage from the R/P into the receiver (so that the receiver is not driven into saturation, which could lengthen recovery time after the transmitter is turned off) and for controlling the input signal levels into the receiver. The received levels are kept below saturation levels, typically with a clutter AGC in search and a target AGC in single-target track, to prevent spurious signals, which degrade performance, from being generated.

Clutter Positioning.   A voltage-controlled oscillator (VCO), usually part of the stable local oscillator (stalo), is used to heterodyne main-beam clutter to zero frequency, or dc. With the clutter at dc the in-phase (I) and quadrature (Q) channel amplitude and phase-balance requirements are eased, as the images resulting from unbalance also fall near dc and can be filtered out along with the main-beam clutter.

Transmit Pulse Suppressor.   Further attenuation of transmitter leakage is provided by the transmit pulse suppressor in the receiver IF, which is a gating device.

Signal Processing.   The analog output of the receiver is downconverted to baseband (dc) via quadrature mixing. The in-phase and quadrature signals are passed through a matched filter and converted to digital words by an analog-to-digital (A/D) converter. Following the A/D is typically a delay-line clutter canceler and doppler filter bank for main-beam clutter rejection and coherent integration. The filter bank is usually realized by using the fast Fourier transform
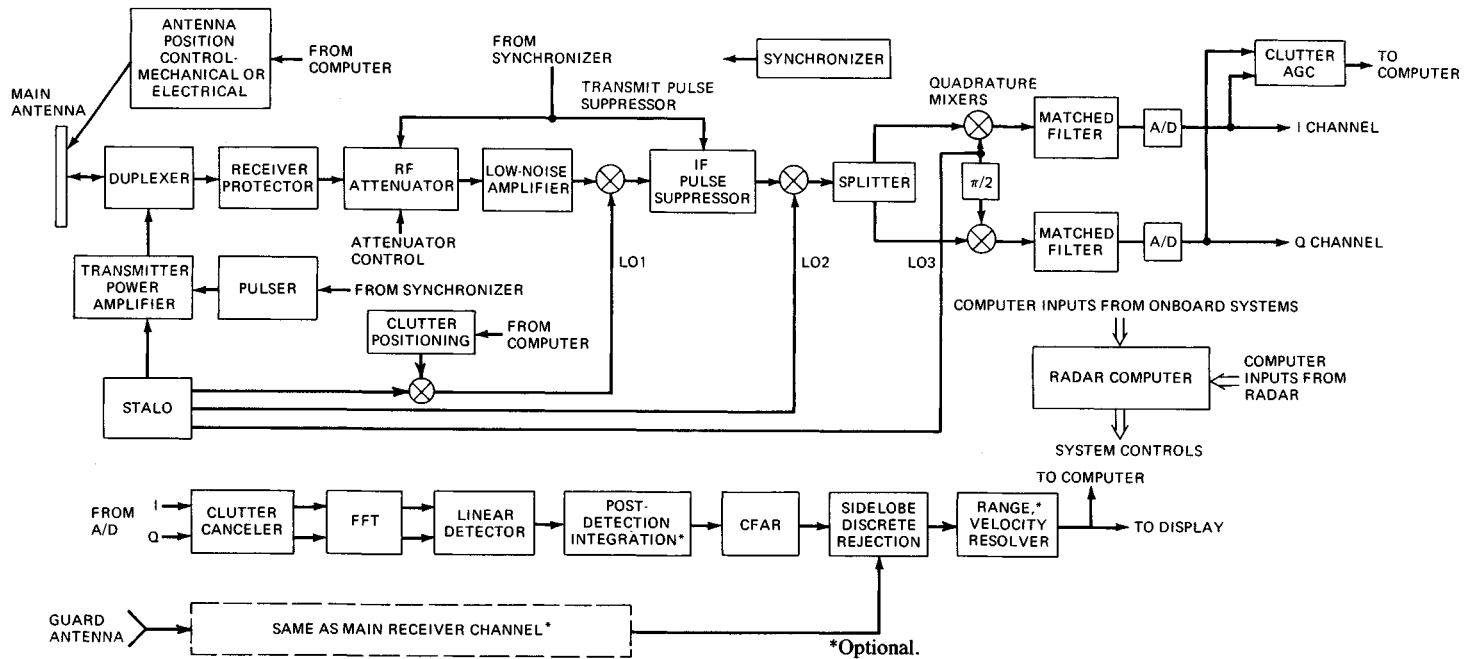
FIG. 17.6 Typical pulse doppler radar configuration.

17.8

(FFT) or by the discrete Fourier transform (DFT) for a small number of filters. Appropriate weighting is employed to reduce the filter sidelobes.

The voltage envelope at the output of the FFT is formed by using an I/Q combining approximation. Postdetection integration (PDI) may be used where each range-gate–doppler-filter output is linearly summed over several coherent looks. The PDI output is compared with a detection threshold determined by a constant-false-alarm-rate (CFAR)[17–20] process.

Following the CFAR is the sidelobe discrete rejection logic, discussed in Sec. 17.2, and the range and velocity ambiguity resolvers (if used). The final detection outputs are passed to the radar display and computer.

17.2 PULSE DOPPLER CLUTTER

General. Clutter returns from various scatterers have a strong influence on the design of a pulse doppler radar as well as an effect on the probability of detection of point targets. Clutter scatterers include terrain, both ground and water, rain, snow, and chaff. Since the antennas generally used in pulse doppler radars have a single, relatively high-gain main beam, main-beam clutter may be the largest signal handled by the radar when in a down-look condition, which is a principal reason for the use of medium- and high-PRF pulse doppler radars. The narrow beam limits the frequency extent of this clutter to a relatively small portion of the doppler spectrum. The remainder of the antenna pattern consists of sidelobes which result in sidelobe clutter. This clutter is generally much smaller than the main-beam clutter but covers much more of the frequency domain. The sidelobe clutter from the ground directly below the radar, the altitude line, is frequently large owing to a high reflection coefficient at steep grazing angles, the large geometric area, and the short range. Range performance is degraded for targets in the sidelobe clutter region wherever the clutter is near or above the receiver noise level. Multiple PRFs may be used to move the target with respect to the clutter, thus avoiding completely blind ranges or blind frequencies due to high clutter levels. This relative motion occurs owing to the range and doppler foldover. If one PRF folds sidelobe clutter and a target to the same apparent range and doppler, a sufficient change of PRF will separate them.

Ground Clutter in a Stationary Radar. When the radar is fixed with respect to the ground, both main-beam and sidelobe clutter returns occur at zero-doppler offset, the transmit frequency. The sidelobe clutter is usually small compared with main-beam clutter as long as some part of the main beam strikes the ground. The clutter can be calculated as in a pulse radar, then folded in range as a function of the PRF.

Ground Clutter in a Moving Radar. When the radar is moving with a velocity V_R, the clutter is spread over the frequency domain as illustrated in Fig. 17.2 for the special case of horizontal motion. The foldover in range and doppler is illustrated in Fig. 17.7 for a medium-PRF radar where the clutter is
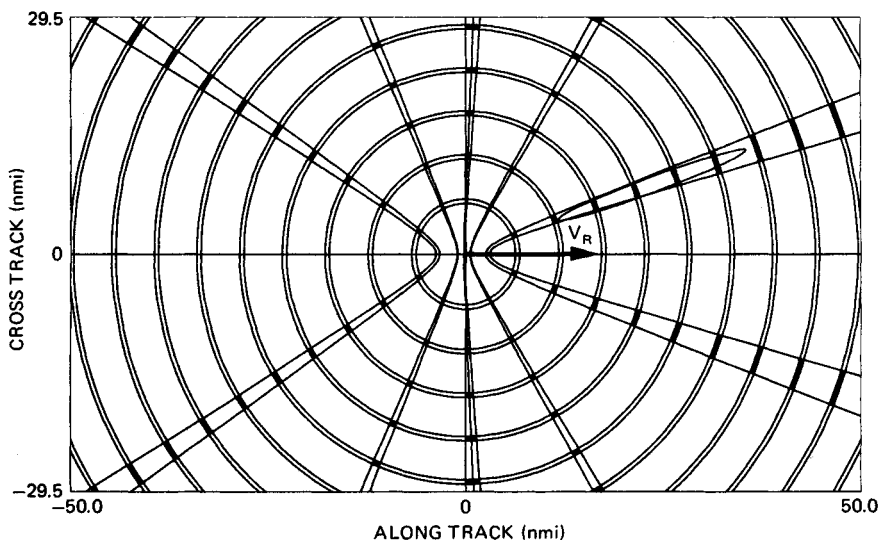
FIG. 17.7 Plan view of range-gate and doppler filter areas. Radar altitude, 10,000 ft; velocity, 1000 kn to right; dive angle, 10°; radar wavelength, 3 cm; PRF, 15 kHz; range-gate width, 6.67 μs; gate, 4; doppler filter, at 2 kHz; bandwidth, 1 kHz; beamwidth, 5° (circular); main-beam azimuth, 20°; depression angle, 5°.

ambiguous in both range and doppler. The radar platform is moving to the right at 1000 kn with a dive angle of 10°. The narrow annuli define the ground area that contributes to clutter in the selected range gate. The five narrow hyperbolic bands define the area that contributes to clutter in the selected doppler filter. The shaded intersections represent the area that contributes to the range-gate–doppler-filter cell. Each area contributes clutter power dependent on antenna gain in the direction of the area and the reflectivity of the area.

The main beam illuminates the elliptical area to the left of the ground track. Since this area lies entirely within the filter area, the main-beam clutter falls within this filter, and all other filters receive sidelobe clutter. Five range annuli are intersected by the main-beam ellipse; so the main-beam clutter in this range gate is the vector sum of the signals received from all five areas. Owing to this high degree of range foldover, all range gates will have approximately equal clutter.

If the main beam were scanned 360° in azimuth, the main-beam clutter would scan in frequency so that it would appear in the selected filter 10 times (twice for each hyperbolic band). In between, the filter would receive sidelobe clutter from all darkened intersections.

Clutter Return: General Equations. The clutter-to-noise ratio from a single clutter patch with incremental area dA at a range R is

$$C/N = \frac{P_{av} \, G_T \, G_R \, \lambda^2 \, \sigma^0 \, dA}{(4\pi)^3 \, R^4 \, L_c \, k \, T_s \, B_n} \tag{17.2}$$

where P_{av} = average transmit power

λ = operating wavelength

σ^0 = clutter backscatter coefficient

L_c = losses applicable to clutter

G_T = transmit gain in patch direction

G_R = receive gain in patch direction

k = Boltzmann's constant = 1.38054×10^{-23} W/(Hz/K)

T_s = system noise temperature, K

B_n = doppler filter bandwidth

The clutter-to-noise ratio from each radar resolution cell is the integral of Eq. (17.2) over the doppler and range extent of each of the ambiguous cell positions on the ground.[21-25] Under certain simplified conditions, the integration can be closed-form[25] while numeric integration may be used generally.

Sidelobe Clutter. The entire clutter spectrum can be calculated for each range gate by Eq. (17.2) if the antenna pattern is known in the lower hemisphere. In preliminary system design, the exact gain function may not be known, so that one useful approximation is that the sidelobe radiation is isotropic with a constant gain of G_{SL}.

Sidelobe Discretes. An inherent characteristic of airborne pulse doppler radars is that echoes from large objects on the ground (discretes), such as buildings, may be received through the antenna sidelobes and appear as though they were smaller moving targets in the main beam. This is a particularly severe problem in a medium-PRF radar, where all-aspect target performance is usually desired, as these returns compete with targets of interest. In a high-PRF radar, there is little if any range region clear of sidelobe clutter, such that the sidelobe clutter portion of the doppler spectrum is often not processed (since target detectability is severely degraded in this region). Further, in a high-PRF radar, especially at higher altitudes, the relative amplitudes of the distributed sidelobe clutter and the discrete returns are such that the discretes are not visible in the sidelobe clutter.

The apparent radar cross section (RCS), σ_{app}, of a sidelobe discrete with an RCS of σ is $\sigma_{app} = \sigma\, G_{SL}^2$, where G_{SL} is the sidelobe gain relative to the main beam. The larger-size discretes appear with a lower density than the smaller ones, and a model commonly assumed at the higher radar frequencies is as shown in Table 17.3. Thus, as a practical matter 10^6 m^2 discretes are rarely present, 10^5 m^2 sometimes, and 10^4 m^2 often.

Two mechanizations for detecting and eliminating false reports from sidelobe discretes are the guard channel and postdetection sensitivity time control (STC). These are discussed in the paragraphs which follow.

TABLE 17.3 Discrete Clutter Model

Radar cross section, m^2	Density, per mi^2
10^6	0.01
10^5	0.1
10^4	1

Guard Channel. The guard channel mechanization compares the outputs of two parallel receiving channels, one connected to the main antenna and the second to a guard antenna, to determine whether a received signal is in the main beam or the sidelobes.[26-28] The guard channel uses a broad-beam antenna that (ideally) has a pattern above the main-antenna sidelobes. A range-cell, doppler-filter by range-cell, doppler-filter comparison is made of the returns in both channels. Sidelobe returns are rejected (blanked) when they are larger in the guard receiver, and main-lobe returns are passed without blanking since they are larger in the main receiver.

A block diagram of a guard channel mechanization is shown in Fig. 17.8. After the CFAR circuits (which ideally would be identical in both channels), there are three thresholds: the main channel, guard channel, and main-to-guard-ratio threshold. The detection logic of these thresholds is also shown in Fig. 17.8.

The blanking which occurs because of the main-guard comparison affects the detectability in the main channel, the extent of which is a function of the threshold settings. The threshold settings are a tradeoff between false alarms due to sidelobe returns and detectability loss in the main channel. An example is shown in Fig. 17.9 for a nonfluctuating target, where the ordinate is the probability of detection in the final output and the abscissa is the signal-to-noise ratio (SNR) in the main channel. The quantity B^2 is the ratio of the guard channel SNR to the main channel SNR and is illustrated in Fig. 17.10. B^2 is small for a target in the main beam and large, 0 dB or so, for a target at the sidelobe peaks. In the example shown, there is a 0.5 dB detectability loss due to the guard blanking for targets in the main beam.

Ideally, the guard antenna gain pattern would exceed that of the main antenna at all angles in space (except for the main beam) to minimize detections through the sidelobes. If not, however, as illustrated in Fig. 17.10, returns through the sidelobe peaks above the guard pattern have a significant probability of detection in the main channel and would represent false detections.

Postdetection STC. A second approach to blanking sidelobe discretes is the postdetection STC,[29] the logic of which is shown in Fig. 17.11. Basically, the CFAR output data is correlated (resolved) in range 3 times. Each correlator calculates unambiguous range using M out of the N sets of detection data (e.g., three detections required out of eight PRFs). No doppler correlation is used since the doppler is ambiguous. The results of the first two correlations are used to blank all outputs which are likely to be sidelobe discretes from the final range correlator. Here, three range correlators are used in which the first, the A correlator, resolves the range ambiguities within some nominal range, say, 10 nmi, beyond which sidelobe discretes are not likely to be detected. A second correlator, the B correlator, resolves the range ambiguities out to the same range, but before a target can enter the B correlator, its amplitude is thresholded by a range-varying threshold (the STC threshold). A range-cell by range-cell comparison is made of the correlations in the A and B correlators, and if a range gate correlates in A and not in B, that gate is blanked out of the third correlator, the C correlator. The C correlator resolves the range ambiguities within the maximum range of interest.

The principle behind the postdetection STC approach is illustrated in Fig. 17.12, where the return of a target in the main beam and a large discrete target in the sidelobes is plotted versus unambiguous range (that is, after the range ambiguities have been resolved). Also shown are the normal CFAR threshold and the STC threshold versus range. As is evident, a discrete return in the sidelobes is

M	G	MGR	DET.
0	0	0	NO
0	0	1	NO
0	1	0	NO
0	1	1	NO
1	0	0	YES
1	0	1	YES
1	1	0	NO
1	1	1	YES

DETECTION LOGIC

DETECTIONS OUT

0 ⇒ NO DETECT;
1 ⇒ DETECT.

FIG. 17.8 Two-channel sidelobe blanker.
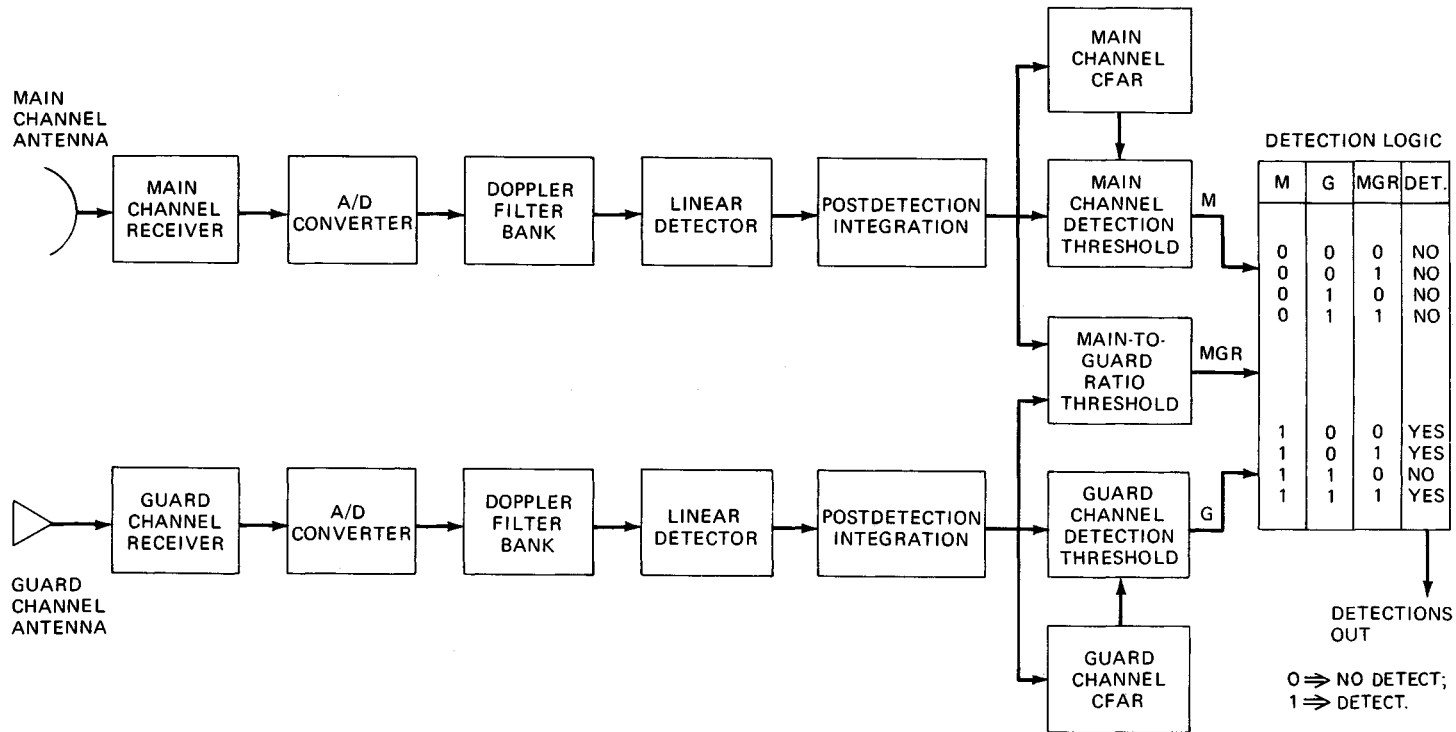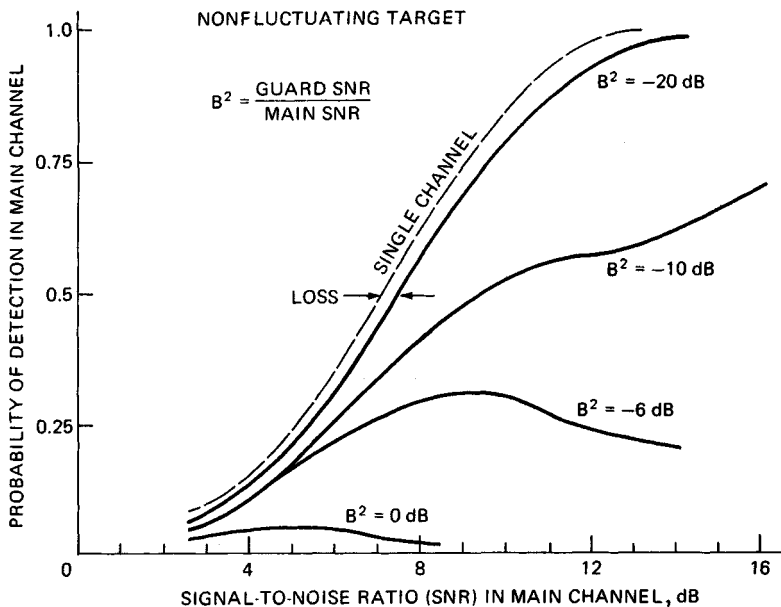
FIG. 17.9 Probability of detection versus signal-to-noise ratio with a guard channel.
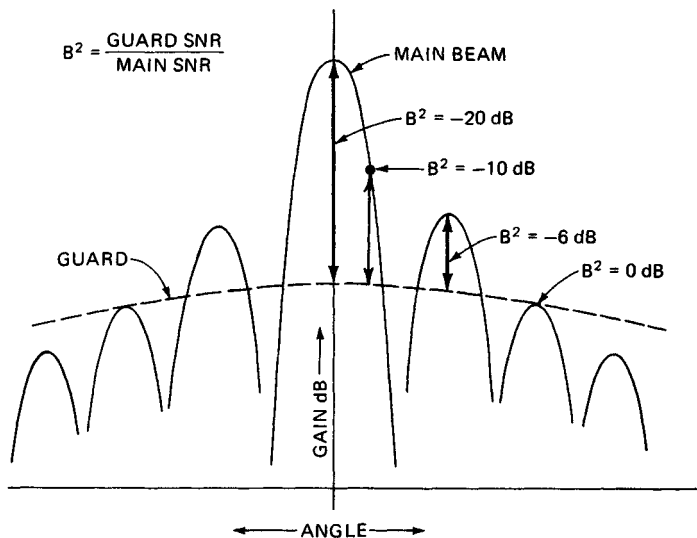
FIG. 17.10 Main and guard antenna patterns.
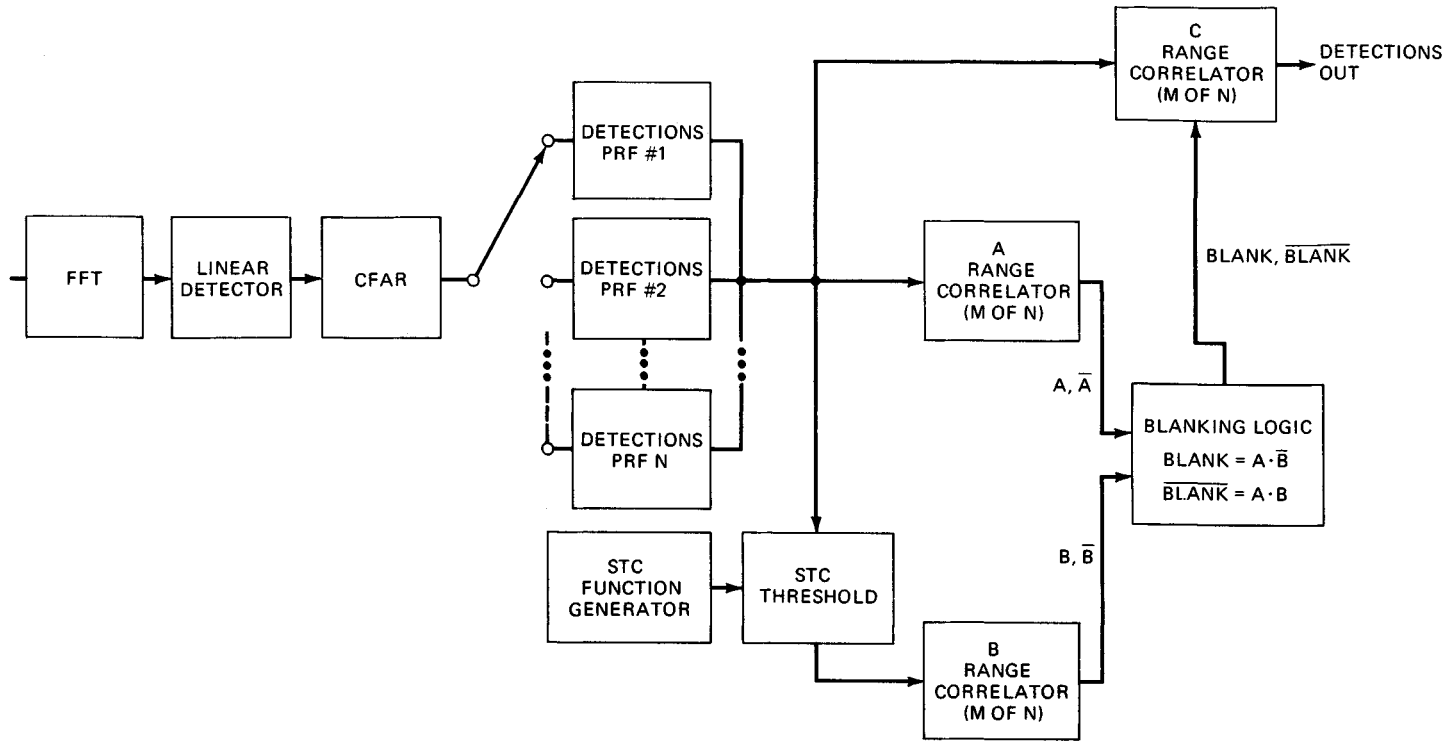
FIG. 17.11 Single-channel sidelobe blanker.
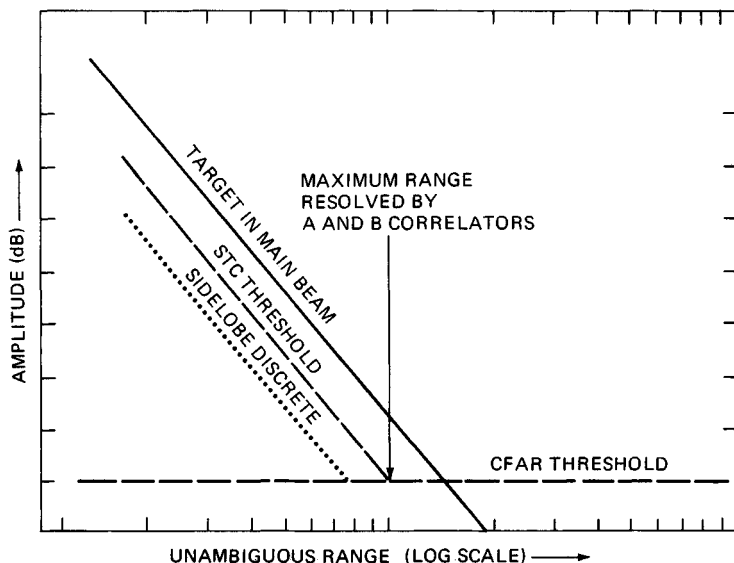
FIG. 17.12 Postdetection STC levels.

below the STC threshold, and a return in the main beam is above the threshold, such that the sidelobe discrete can be recognized and blanked from the final output but the target will not be blanked.

Main-Beam Clutter. The main-beam clutter-to-noise power can be approximated from Eq. (17.2) by substituting the intersected area for dA and summing over all intersections within the main beam.[30]

$$\frac{C}{N} = \frac{P_{av}\lambda^2\,\theta_{az}\,(c\tau/2)}{(4\pi)^3\,L_c\,K\,T_s\,B_n}\sum\frac{G_T\,G_R\,\sigma^0}{R^3\cos\alpha} \tag{17.3}$$

The summation limits are the lower and upper edges of the smaller of the transmit and receive beams and

where θ_{az} = azimuth one-half power beamwidth, rad
$\quad\quad\tau$ = compressed pulsewidth
$\quad\quad\alpha$ = grazing angle at clutter patch

The remaining terms are as defined following Eq. (17.2).

Main-Beam Clutter Filtering. In a pulse doppler radar utilizing digital signal processing, main-beam clutter is rejected by either a combination of a delay-line clutter canceler followed by a doppler filter bank or by a filter bank with low filter sidelobes. In either case, the filters around the main-beam clutter are blanked to minimize false alarms on main-beam clutter.

The choice between these options is a tradeoff of quantization noise and com-

plexity versus the filter-weighting loss. If a canceler is used, filter weighting can be relaxed over that with a filter bank alone, since the canceler reduces the dynamic-range requirements into the FFT (if the main-beam clutter is the largest signal). Without a canceler, heavier weighting is needed to reduce sidelobes to a level so that the filter response to main-beam clutter is below the thermal-noise level. This weighting increases the filter noise bandwidth and hence increases the loss in signal-to-noise ratio.

The improvement factor for a DFT filter[31] is given by

$$I(K) = \frac{\left[\sum_{n=0}^{N-1} A_n^2\right]}{\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} A_n A_m \exp\{-2[\pi(n-m)\,\sigma_c\,T]^2\}\cos[2\pi K(n-m)/N]} \tag{17.4}$$

where A_i = DFT weight, $0 \leq i \leq N-1$
  N = number of points in DFT
  σ_c = standard deviation of clutter spectrum
  K = filter number ($K = 0$ is dc filter)
  T = interpulse period

Here, the improvement factor for a filter (versus the more common definition applied to a delay-line canceler) is defined as the ratio of the total clutter power input to the filter to the clutter residue in that filter. Expressed another way, the improvement factor is the ratio of the clutter power out of a filter if it were centered over the clutter, and the clutter width reduced to zero, to the power out of the filter in actual operation.[32,33] Figure 17.13 shows the improvement factor of a 256-point, Dolph-Chebyshev weighted FFT as a function of the clutter width for various filter numbers in the filter bank.

If the main beam is pointed below the horizon and is greater than a beamwidth from 0° azimuth, the 6 dB clutter width due to platform motion Δf is

$$\Delta f = \frac{2V_R}{\lambda}\,\theta_B\,\sin\psi_0 \tag{17.5}$$

where V_R = radar ground speed
  ψ_0 = main-beam angle relative to velocity vector
  θ_B = 3 dB one-way antenna beamwidth, rad
  λ = RF wavelength

Clutter-Transient Suppression.   When the PRF is changed for multiple-PRF ranging, or the slope is changed in linear FM ranging, or the RF carrier is changed, the transient change in the clutter return may cause degradation unless it is properly handled.[34] Since the clutter is usually ambiguous in range in a pulse doppler radar, the clutter power increases at each interpulse period (IPP) as clutter return is received from longer-range ambiguities, until the horizon is reached. This phenomenon is called *space charging*. Note that although an increasing number of clutter returns are received during the charging period, the vector sum may actually decrease owing to the random phase relations of the returns from different patches.
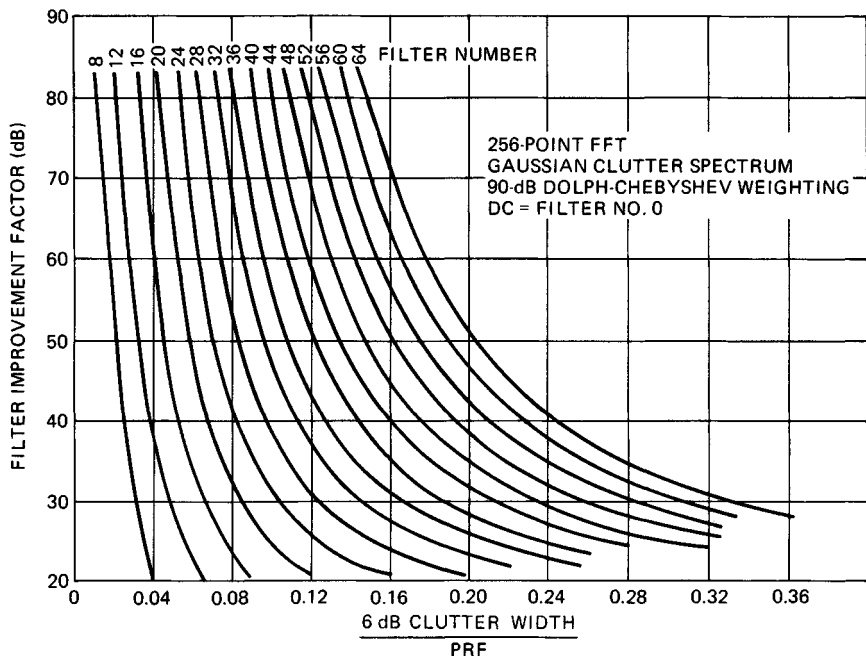
FIG. 17.13 Filter improvement factor versus clutter width.

If a clutter canceler is used, the output cannot begin to settle to steady-state value until space charging is complete. Some settling time must be allowed before signals are passed to the filter bank. Therefore, the coherent integration time available at each look is reduced from the total look time by the sum of the space charge time and the transient settling time. The canceler settling time can be eliminated by "precharging" the canceler with the steady-state input value.[35] This is done by changing the canceler gains so that all delay lines achieve their steady-state values on the first IPP of data.

If no canceler is used, signals can be passed to the filter bank after the space charge is complete, so that the coherent integration time is the total look time minus the space charge time.

Altitude-Line Clutter Filtering. The reflection from the earth directly beneath an airborne pulse radar is called altitude-line clutter. Because of specular reflection over smooth terrain, the large geometric area, and the relatively short range, this signal can be large. It lies within the sidelobe clutter region of the pulse doppler spectrum.

Because it can be much larger than diffuse sidelobe clutter and has a relatively narrow spectral width, altitude-line clutter is often removed by either a special CFAR which prevents detection of the altitude line or by a tracker-blanker which removes these reports from the final output. In the case of the tracker-blanker, a closed-loop tracker is used to position range and velocity gates around the altitude return and blank the affected range-doppler region.

17.3   TIME GATING

Time gating of the receiver permits blanking of transmitter leakage and its noise sidebands, elimination of excess receiver noise from competing with the signal, range gating for target tracking, and true range measurement, provided the ambiguity can be resolved.

Transmitted-Pulse Suppression.   One major advantage of pulse doppler over CW systems is the time blanking of transmitter leakage so that receiver sensitivity is not degraded owing to saturation effects or to noise sidebands on the transmitter.

Harmonic Frequencies.   Extreme care is required to prevent spurious signals from appearing in the system output. For example, if a 30-MHz IF receiver is being gated at a 110-kHz PRF, the 272d harmonic of the gating transient will fall at 29.92 MHz and the 273d at 30.03 MHz. Either of these harmonics may be within the doppler passband and therefore appear in the output. Although high-order harmonics of the gating transient are relatively small, they may be large compared with the signal since gating occurs early in the receiver.

Gating and Synchronization.   One solution to the gating-harmonic problem is the use of balanced gating circuits and synchronization of the IF passband and the PRF so that the PRF harmonics all fall outside the useful portion of the passband. An alternative solution is to heterodyne the clutter to a frequency that is a multiple of the PRF so that the PRF harmonics are rejected with the clutter. However, such solutions preclude a variable-PRF system other than in discrete, accurately known steps.

Although synchronization of the PRF and the IF passband is usually necessary, synchronization at RF is not usually required. The harmful harmonics are of a much higher order and therefore are much smaller. In addition, the RF gating transients are usually further reduced in amplitude by the IF gating circuit.

Transmitter Leakage.   The on-off ratio required for the overall transmitter blanking circuits is fairly large (more than can be obtained readily at RF without excessive insertion loss). Thus a combination RF and IF blanking system is usually employed. The transmitter leakage through the blanking circuits can be allowed to be as large as main-beam clutter if there is zero-doppler filtering to remove it. Alternatively, it must be a fraction of the noise power in a detection filter if there is no such filtering.

Range Gating.   Range gating eliminates excess receiver noise from competing with the signal and permits target tracking and range measurement. Range gating is very similar to transmitted-pulse suppression. In a single-channel 0.5-duty-cycle system, one pulse-suppressor circuit serves both functions. In multiple-range-gated systems the range gates can serve both functions. If one circuit serves both functions, the on-off ratio must be adequate for pulse suppression, whereas if two are used, the range gate does not need as much rejection.

17.4   RANGE-AMBIGUITY RESOLUTION

Several methods of ranging are commonly employed in high PRF, while medium PRF is usually confined to multiple discrete PRF ranging.

High-PRF Ranging. Range-ambiguity resolution in high PRF is performed by modulating the transmitted signal and observing the phase shift of the modulation on the return echo. Modulation methods include varying the PRF, either continuously or in discrete steps; varying the RF carrier, with either linear or sinusoidal FM; or some form of pulse modulation such as pulse-width modulation (PWM), pulse-position modulation (PPM), or pulse-amplitude modulation (PAM). Of these modulation techniques, PWM and PPM may have large errors because of clipping of the received modulation by eclipsing or straddling (discussed in Sec. 17.7), and PAM is difficult to mechanize in both the transmitter and the receiver. Consequently, they will not be further considered here.

Multiple Discrete PRF Ranging. Ranging by use of several (usually two or three) fixed PRFs involves sequential measurement of the ambiguous range in each PRF, followed by comparison of the measurements to eliminate ambiguities.[36,37]

Figure 17.14 illustrates the principle of multiple-PRF ranging for a two-PRF, high-PRF radar. The PRFs are chosen to have a common submultiple frequency $1/T_u$. If the transmitted-pulse trains are compared in a coincidence detector, the common submultiple frequency is obtained. Similarly, if the received gates are compared in a coincidence detector, the same submultiple frequency shifted in time by the target range delay T_r is obtained. Measuring the time delay between the two sets of coincidence pulses yields the true target range. If desired, a three-PRF system can be mechanized similarly. The advantage obtained is the increased unambiguous range achievable.
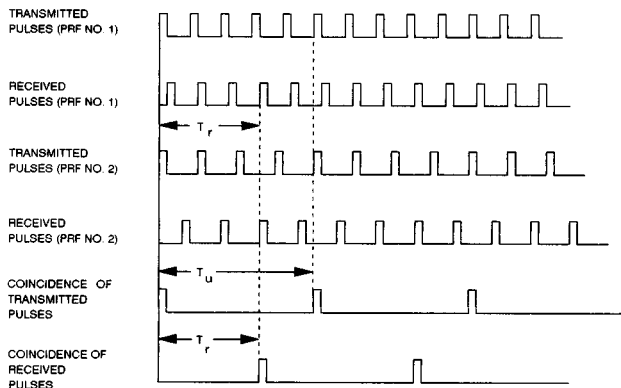
FIG. 17.14   Principle of two-PRF ranging.

In a surveillance radar a number of receiver gates are used to detect targets that may appear at any range within the interpulse period. Figure 17.15 illustrates a common method of spacing the gates for the general case where the gate spacing τ_s, the gate width τ_g, and the transmitted pulse τ_t are all unequal. Selecting $\tau_g > t_s$ reduces the range-gate straddle loss but increases the possibility of range ghosts. Selecting $\tau_t = \tau_g$ maximizes range performance.
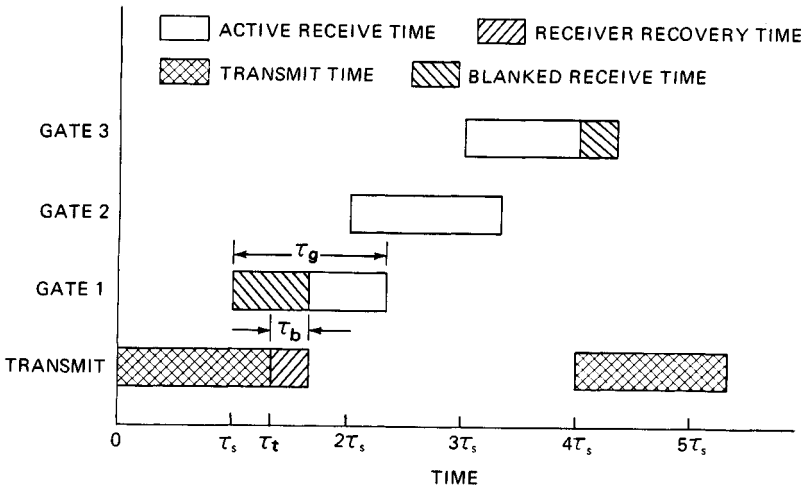
FIG. 17.15 Example of three range gates equally spaced in the interpulse period.

The PRFs are usually related by the ratios of closely spaced, relatively prime integers m_1, m_2, and m_3, as indicated in Table 17.4. Thus a three-PRF system using the seventh, eighth, and ninth submultiples of the range-gate clock frequency $f_c = 1/\tau_s$ as PRFs yields an unambiguous range of $7 \times 9 = 63$ times that of the middle PRF alone.

TABLE 17.4 Multiple-PRF Ranging Parameters

Item	Two-PRF	Three-PRF
Ranging parameters: $m_1 > m_2 > m_3$	m_1, m_2	m_1, m_2, m_3
Number of range-gated channels	$m_1 - 1$	$m_1 - 1$
PRFs		
f_{R1}	$1/m_1\tau_s$	$1/m_1\tau_s$
f_{R2} $(f_{R3} > f_{R2} > f_{R1} \geq f_{R\min})$	$1/m_2\tau_s$	$1/m_2\tau_s$
f_{R3}		$1/m_3\tau_s$
Unambiguous range ($R_{\max}$)	$m_2c/2f_{R1}$	$m_2m_3c/2f_{R1}$
Transmitter duty cycle, d	$\tau_t f_{R2}$	$\tau_t f_{R3}$
Ratio of highest and lowest PRF	m_1/m_2	m_1/m_3

NOTE: m_1, m_2, m_3 must be relatively prime integers.
τ_t = transmitted pulsewidth
τ_g = range-gate width
τ_b = blanking width due to receiver recovery
τ_s = range-gate spacing
f_c = range-gate clock = $1/\tau_s$

Figure 17.16 shows the maximum unambiguous range as a function of the minimum PRF, $f_{R\min}$, and the ranging parameter m_1, for the case where m_1, m_2, m_3 are consecutive integers. It is usually desirable to keep m_1 in the region from about 8 to 50. Thus the unambiguous range of a two-PRF system is seen to be rather limited, whereas three PRFs give much larger ranges. Some of the considerations influencing the choice of m_1 to this range are as follows:
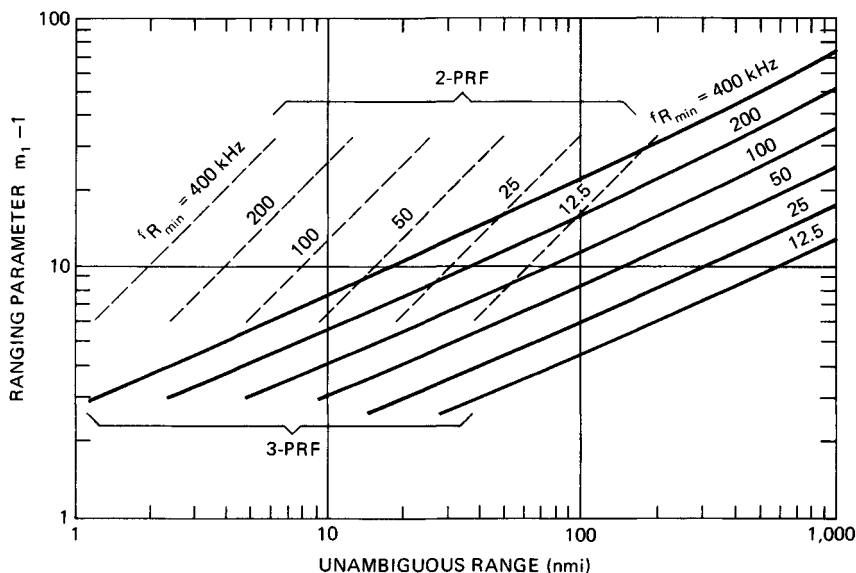
FIG. 17.16 Unambiguous range for two- and three-PRF ranging systems as a function of ranging parameters, $m_1 - 1$, and minimum value of PRF for the case where m_1, m_2, m_3 are consecutive integers and m_1 is odd.

1. To minimize hardware, m_1 should be small since a maximum of $m_1 - 1$ range gates must be processed.

2. The probability of eclipse in at least one PRF is about $3/m_1$ for a three-PRF system, and so m_1 should be at least eight or higher since range cannot be measured if any PRF is eclipsed.

3. To get a long unambiguous range, m_1 should be large.

4. For good range resolution, τ_T must be small, which requires that m_1 be large. (Target range change during the dwell limits the minimum τ_r.)

5. To minimize the transmitter duty cycle and hence the average-power variation between PRFs, m_1 should be relatively large.

The Chinese remainder theorem is one means for calculating the true range from the several ambiguous measurements in a range-while-search system.[38] This approach permits a unique direct computation of the true-range cell number R_c from the three ambiguous-range cell numbers A_1, A_2, and A_3 (or two numbers for a two-PRF system). (The cell number is the range expressed in units of the pulse width and ranges from 0 to $m_i - 1$.) The theorem for a three-PRF system is expressed by the congruence

$$R_c \quad (C_1 A_1 + C_2 A_2 + C_3 A_3) \text{ (modulo } m_1 m_2 m_3) \tag{17.6}$$

The smallest value of R_c that satisfies Eq. (17.6) is the remainder of the term within parentheses when divided by $m_1 m_2 m_3$ as many times as possible. There-

fore, $0 \leq R_c < m_1 m_2 m_3$. The constants C_1, C_2, and C_3 are related to m_1, m_2, and m_3 by the congruences

$$C_1 = b_1 m_2 m_3 \equiv 1 \text{ (modulo } m_1) \tag{17.7}$$

$$C_2 = b_2 m_1 m_3 \equiv 1 \text{ (modulo } m_2) \tag{17.8}$$

$$C_3 = b_3 m_1 m_2 \equiv 1 \text{ (modulo } m_3) \tag{17.9}$$

where b_1 is the smallest positive integer which, when multiplied by $m_2 m_3$ and divided by m_1, gives unity as the remainder (and similarly for the other b's).

Once m_1, m_2, and m_3 have been chosen, the range can be computed from Eq.(17.6) by using the C values and the ambiguous-range cell numbers (A_1, A_2, A_3) in which the target is detected. For example, if $m_1 = 7$, $m_2 = 8$, $m_3 = 9$, then $b_1 = 4$, $b_2 = 7$, $b_3 = 5$, and the range is $R_c = (288A_1 + 441A_2 + 280A_3)$ (modulo 504). If the target is in the first gate after the transmit pulse, $A_1 = A_2 = A_3 = 1$ and $R_c = (288 + 441 + 280)$ (modulo 504) $= 1$. An alternative to the Chinese remainder theorem is either a hard-wired correlator or a special-purpose computer that accepts detections from all PRFs and outputs all double or triple correlations.

Continuously Variable PRF Ranging. In a single-target tracking radar, the range ambiguity can be resolved by varying the PRF so that the target return is centered in the interpulse period. A high duty cycle, 0.333 to 0.5, may be used. Range R can then be calculated by

$$R = -\frac{\dot{R} f_R}{\dot{f}_R} \tag{17.10}$$

This method of range measurement has poor accuracy because of the errors involved in measuring the derivatives. An advantage of this technique is that the target return is never eclipsed by the transmitter pulse, thus improving tracking. It has a disadvantage, however, in that PRF harmonics can appear within the doppler band as spurious signals.

Linear-Carrier FM. Linear frequency modulation of the carrier can be used to measure range, especially in range-while-search applications. The modulation and demodulation to obtain range are the same as used in CW radar, but the transmission remains pulsed.

Assume that the dwell time is divided into two periods. In the first period, no FM is applied, and the doppler shift of the target is measured. In the second period, the transmitter frequency is varied linearly at a rate $\dot{f}$ in one direction. During the round-trip time to the target, the local oscillator has changed frequency so that the target return has a frequency shift, in addition to the doppler shift, that is proportional to range. The difference in the frequency Δf of the target return in the two periods is found, and the target range calculated from

$$R = \left| \frac{c \Delta f}{2\dot{f}} \right| \tag{17.11}$$

The problem with only two FM segments during a dwell time is that, with more than a single target in the antenna beamwidth, range ghosts result. For example, with two targets present at different dopplers, the two frequencies observed during the FM period cannot be unambiguously paired with the two frequencies observed during the no-FM period. Thus, typical high-PRF range-while-search radars use a three-segment scheme in which there are no-

FM, FM-up, and FM-down segments. The range is found by selecting returns from each of the three segments that satisfy the relations

$$f_1 < f_0 < f_2 \qquad (17.12)$$

$$f_1 + f_2 = 2f_0 \qquad (17.13)$$

where $f_0, f_1,$ and f_2 are the frequencies observed during the no-FM, FM-up, and FM-down segments, respectively. The range then is found from Eq. (17.11), where

$$\Delta f = f_2 - f_0 \qquad \text{or} \qquad (f_2 - f_1)/2 \qquad \text{or} \qquad f_0 - f_1 \qquad (17.14)$$

An example is shown in Fig. 17.17.

Target	A	B
Range, nmi	10	20
Doppler frequency, kHz	21	29
FM shift, kHz	3	6
Observed frequencies		
f_0, no FM, kHz	21	29
f_1, FM up, kHz	18	23
f_2, FM down, kHz	24	35

Possible sets which satisfy the relations shown in Eqs. (17.12) and (17.13) are:

f_1	f_0	f_2	$2f_0$	$f_1 + f_2$	Target?	Range, nmi
18	21	24	42	42	Yes	10
18	21	35	42	53	No	
18	29	35	58	53	No	
23	29	35	58	58	Yes	20

FIG. 17.17 Three-slope FM ranging example. There are two targets, A and B; $\dot{f}$ = FM slope = 24.28 MHz/s.

If more than two targets are encountered during a dwell time, ghosts again result, as only $N - 1$ simultaneously detected targets can be resolved ghost-free where N is the number of FM slopes. This is not a severe problem in practice, however, for multiple targets in a single beamwidth are usually a transient phenomenon.

The accuracy of the range measurement improves as the FM slope increases since the observed frequency differences can be more accurately measured. However, the FM slope is limited by clutter-spreading considerations since during the FM periods the clutter is smeared in frequency and can appear in frequency regions normally clear of clutter. Range accuracies on the order of 1 or 2 mi can be reasonably achieved.

Sinusoidal-Carrier FM. This method is similar to that sometimes used in CW radar but retains the pulse transmission. It is particularly useful for tracking, ei-

ther continuously or in a pause-to-range mode (discussed in Sec. 17.5). It is not suitable for range-while-search because of the relatively long time required to measure the phase shift of the sinusoidal modulation.

Medium-PRF Ranging. Multiple discrete PRF ranging, as discussed for high PRF, is also used for medium PRF except that the PRF selection criterion differs.[13] The technique of using closely spaced PRFs can be extended to medium PRF by employing three groups of three closely spaced PRFs, the groups being widely spaced to improve doppler visibility. The center PRF in each group is called the *major* PRF, and the adjacent ones the *minor* PRFs. Ranging is accomplished by requiring a detection in the major PRF and its adjacent minor PRFs and is effectively a detection criterion of exactly three detections out of three opportunities. This approach is attractive from a ghosting standpoint but suffers owing to the poor doppler visibility that results from having only three PRFs visible.

A better technique for medium PRF is to use seven or eight PRFs which cover nearly an octave in frequency and to require detections in at least three of these to declare a target report. The advantage is that doppler visibility is better than with the major-minor approach, and hence better range performance in sidelobe clutter is achieved (where some PRFs may be obscured by clutter). However, it is more susceptible to ghosting owing to the high doppler visibility. This problem is mitigated by also resolving the doppler ambiguities and using the true doppler for correlation to reject ghosts.

The basic accuracy of multiple-PRF ranging is on the order of the range-gate size (150 m/μs), but this can be improved to a fraction of the gate width by amplitude centroiding.

17.5 TARGET TRACKING

Target tracking can be performed on either a single target, using more or less conventional angle, range, and velocity tracking servo loops, or on multiple targets, using track-while-scan.

Single-Target Tracking. Angle tracking can be identical to a conventional pulse radar using monopulse, sequential lobing, or conical scan. Monopulse is more difficult to mechanize because of the problem of phase and amplitude matching of the multiple receiver channels, but the problem can be mitigated by using self-calibration routines controlled by the radar computer.

In a low-duty-cycle radar, range tracking is similar to pulse radar tracking in that split-gate tracking is used. In a high-duty-cycle radar, continuously variable PRF ranging or linear-FM ranging may be used. In a pulse doppler radar, the tracked range is usually ambiguous, so that provisions must be made to track through multiple interpulse periods and during eclipse (that is, when the target return overlaps the transmitted pulse).

Velocity (or doppler) tracking in a pulse doppler radar is carried out by forming a centroid on the target's doppler return in the filter bank. A closed-loop tracker then positions a doppler window around the tracked target such that returns which differ in doppler by more than a predetermined value are discarded by the tracker. In medium PRF, the PRF has to be adjusted to keep the doppler return away from the main-beam clutter notch as well as to avoid range eclipse.

Tracking through Eclipse.   Because of the range ambiguities in medium and high PRF, the radar must cope with the loss of target each time that it passes through eclipse. Automatic tracking systems might recycle to the search mode if eclipse is not recognized and preventive measures taken.

The multiple-PRF true-ranging system is the most positive solution. Once true range has been determined so that the range ambiguity is resolved, PRF switching eliminates eclipsing. The onset of eclipse is detected by the range tracker by noting when the range gate begins to overlap the transmitted pulse. Then, before eclipse occurs, the PRF is switched to one of the other values. One of these values is certain to be uneclipsed, owing to the synchronization and relative PRF values. Since tracking is carried out in true range, no transient occurs and eclipse-free tracking continues indefinitely.

The continuously variable PRF system also permits eclipse-free tracking, but because of the spurious-signal problem it has not received much favor.

Other ranging systems do not permit eclipse-free tracking. FM ranging is not accurate enough to predict when eclipsing is about to occur. If the range is not accurately known, there is no way to anticipate an eclipse. In this case, an after-the-fact eclipse detection is made. A target-presence-detector circuit notes the absence of a signal and assumes that this is due to an eclipse. This circuit then commands the PRF to change value in an attempt to bring the target out of eclipse.

The problems with this approach are that target scintillation can cause PRF cycling and that there is no accurate way to predict which PRF will prevent eclipse. The latter problem can be reduced, if crude FM range data is available, by selecting the PRF from groups of PRFs, each group having values appropriate for a particular region of ranges. This reduces the time required in searching for an uneclipsed PRF.

Multiple-Target Tracking.   Multiple-target tracking can be accomplished in several ways. One, track-while-scan, is to use the normal search mode with FM or multiple-PRF ranging and store the range, angle, and doppler of the reported detections in the computer. These detections are then used to form track files. The antenna scans in a normal search pattern, and a scan-to-scan correlation is made on the detections which update the track files. Although tracking accuracies are less than can be achieved in a single-target track, multiple targets can be tracked simultaneously over a large volume in space.

A second method of multiple-target tracking, pause-while-scan, particularly applicable to electronic scan antennas, is to scan in a normal search pattern, pause on each search detection, and enter a single-target track mode for a brief period. The advantage is that the resulting range, angle, and doppler measurements are more accurate than those made with a scanning antenna, but the time to search a volume in space is increased.

17.6 DYNAMIC-RANGE AND STABILITY REQUIREMENTS

Dynamic Range.   Dynamic range as discussed here is the linear region above thermal noise over which the receiver and signal processor operate before any saturation (clipping) or gain limiting occurs. If saturations occur,

spurious signals which degrade performance may be generated. For example, if main-beam clutter saturates, spurious frequencies can appear in the doppler passband normally clear of main-beam clutter and generate false-target reports. An AGC function is often employed to prevent saturations on either main-beam clutter in search or the target in single-target track mode. If saturations do occur in a range gate during an integration period, an option in a multiple-range-gated system is simply to blank detection reports from that gate.

Th most stressing dynamic-range requirement is due to main-beam clutter when searching for a small low-flying target. Here, full sensitivity must be maintained in the presence of the clutter to maximize the probability of detecting the target.

The dynamic-range requirement of a pulse doppler radar, as determined by main-beam clutter, is a function not only of the basic radar parameters such as power, antenna gain, etc., but of radar altitude above the terrain and the radar cross section (RCS) of low-flying targets. As an example, Fig. 17.18 shows the maximum clutter-to-noise ratio (C/N_{max}) which appears in the ambiguous-range interval, i.e., after range folding, for a medium-PRF radar as a function of radar altitude and the range of the main-beam center. Note that the quantity plotted is the rms value of the clutter-to-noise ratio. A pencil-beam antenna pattern is as-
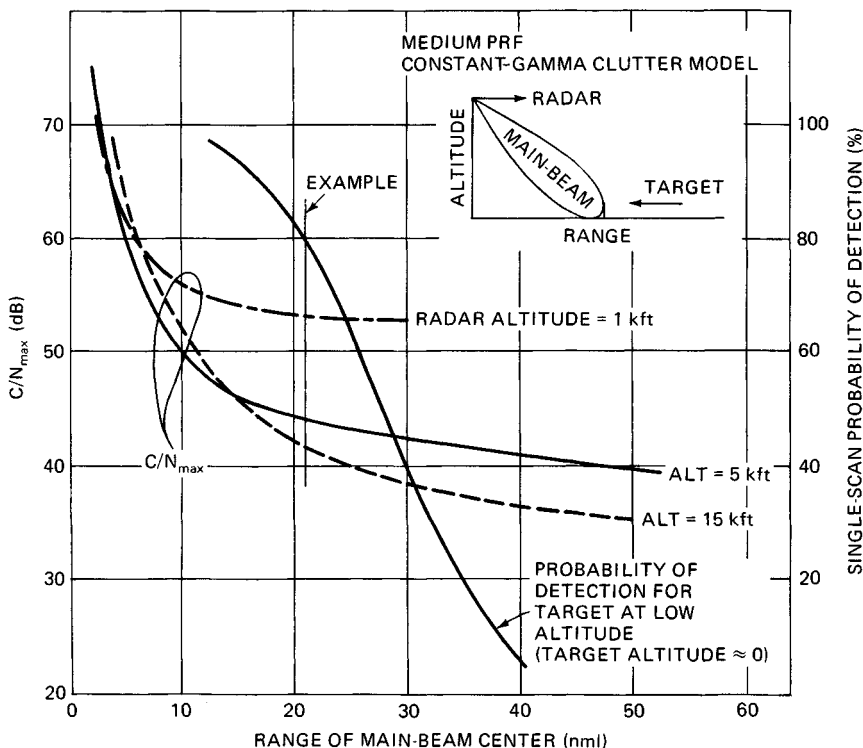
FIG. 17.18 Dynamic-range example.

sumed. At the longer ranges (small look-down angles), the clutter decreases with increasing altitude since range folding is less severe owing to less of the main beam intersecting the ground. At the shorter ranges, clutter increases with altitude since the clutter patch size on the ground increases. While Fig. 17.18 is for a medium-PRF radar, similar curves result for a high-PRF radar.

Also shown in Fig. 17.18 is the single-scan probability of detection P_d versus range for a given RCS target in a receiver with unlimited dynamic range. If it is desired to have the low-flying target reach at least, say, an 80 percent P_d before any gain limiting occurs, the dynamic-range requirement due to main-beam clutter is 53 dB at 1000 ft, 44 dB at 5000 ft, and 41 dB at 15,000 ft for this example. As is evident, the higher the desired probability of detection or the lower the radar altitude, the more dynamic range that is required. Further, if the specified target RCS is reduced, the dynamic-range requirement for the same desired P_d increases as the P_d-versus-range curve in Fig. 17.18 shifts to the left.

In a PD radar using digital signal processing, the dynamic range is most often limited by the A/D converters. The maximum signal level relative to thermal noise that can be processed linearly is related to the number of amplitude bits in the A/D by

$$\frac{S_{max}}{N} = 20 \log \left(\frac{2^{NAD} - 1}{\text{noise}} \right) \qquad (17.15)$$

where S_{max}/N = maximum input level relative to noise, dB
$\quad$ NAD = number of amplitude bits in the A/D
$\quad$ noise = rms thermal-noise level at the A/D, quanta

From the relationships described above and assuming that the A/D limits the dynamic range, the A/D size can now be determined. An additional factor, that of a margin to allow for main-beam clutter fluctuations above the rms value, also needs to be considered. Since main-beam clutter time fluctuation statistics are highly dependent on the type of clutter being observed, such as sea clutter or clutter from an urban area, and are generally unknown, a value of 10 to 12 dB above the rms value is often assumed for the maximum level. Thus, the required number of amplitude bits in the A/D converter as determined by the main-beam clutter is

$$NAD \geq CEIL \left[\frac{(C/N)\text{max (dB)} + \text{fluc margin (dB)} + 20 \log (\text{noise})(q)}{6} \right] \qquad (17.16)$$

where CEIL is the next larger integer.

For the example cited in Fig. 17.18 where the maximum C/N is 53 dB at a 1000-ft altitude, and with a fluctuation margin of 10 dB and thermal noise at 1.414 quanta, the A/D requires at least 11 amplitude bits (plus a sign bit).

Stability Requirements. To achieve the theoretical clutter rejection and target detection and tracking performance of a pulse doppler system, the reference frequencies, timing signals, and signal-processing circuitry must be adequately stable.[39-42] In most cases, the major concern is with short-term stability rather than long-term drift. Long-term stability mainly affects velocity or range accuracy or spurious signals (due to PRF harmonics) but is relatively

easy to make adequate. Short-term stability refers to variations within the round-trip radar echo time or during the signal integration time. The most severe stability requirements relate to the generation of spurious modulation sidebands on the main-beam clutter, which can appear as targets to the target detection circuitry. Thus, the ratio of main-beam clutter to system noise measured at the receiver output (C/N), including the fluctuation margin as discussed above, is the predominant parameter that determines stability requirements. However, at low spurious modulation frequencies, other constraints may become limiting.

Types of Spurious Modulation. The various spurious modulations that can appear on the received signal (clutter or target return) include both carrier and pulse modulation.

Carrier modulation can be amplitude modulation (AM), common FM, or independent FM. Common FM refers to identical modulation on both the transmitted signal and the receiver local-oscillator signal; independent FM appears on only one or the other or in the receiver following the first mixer.

Pulse modulation can be common pulse-position modulation (PPM), independent pulse-width modulation (PWM), or pulse droop. In these cases, common modulation refers to in-phase modulation on both the transmitter pulse and the receiver range gate or transmitter blanking gate. Independent modulation occurs if only one of these pulses is affected. Common PWM does not usually occur. However, if it does, it has requirements similar to those of common PPM. Also, pulse-amplitude modulation (PAM) can occur but is usually negligible when the other requirements are satisfied.

Sinusoidal Modulations. Any of these types of modulation may be caused by a sinusoidal disturbance, such as power supply ripple, line-frequency pickup, or sinusoidal vibration. Discrete sidebands at the modulation frequency and possibly higher harmonics will be introduced on the clutter and target signals. Since the pulse doppler receiver is basically a spectral analyzer, the radar requirements are most readily defined in terms of the allowable level of these modulation sidebands.

The predominant effect of these sidebands depends on the sideband frequency. For sideband frequencies greater than f_{min}, the sidebands on clutter signals fall outside the clutter rejection filter, where f_{min} is the minimum frequency separation of a detection filter from the edge of the main-beam clutter. This spread clutter must be kept below receiver noise; otherwise, it either will be detected as a target or will desensitize the receiver at these frequencies.

For sideband frequencies between $B_n/2$ and f_{min}, the concern is with either a target SNR loss or generation of false targets, where B_n is the receiver predetection-filter bandwidth. The SNR loss results from the sideband energy falling outside the detection filter. The false-target effect is due to modulation sidebands on a strong target appearing to the detection filters as weak targets.

For sideband frequencies below the reciprocal of the postdetection integration time T_1, sidebands per se cease to be of concern. However, the instantaneous-signal-frequency excursion should not be greater than the predetection bandwidth during the integration time, or a SNR loss results. Also, the signal amplitude should not be modulated significantly from one integration period to the next, or the sensitivity could change.

Sidebands in the immediate vicinity of the antenna lobing frequency or range track jitter frequency must be small enough to prevent excessive tracking error or noise.

TABLE 17.5 Allowable Deviation of Pulse or Carrier Sinusoidal Modulation as a Function of Modulating Frequency

Modulation frequency	Criterion	Factor Q	PAM-δ AM-M	Carrier modulation		Pulse modulation		
				Common FM δ_F	Independent FM δ_F	Independent PWM δ_W	Common PPM δ_P	Independent PPM δ_P
0 to $1/T_I$	Signal constant for T_I	0.1	$2Q$	. . . .	. . . .	$2\tau d_{min}Q$	$\dfrac{\tau d_{min}Q}{\pi f_m T_c}$	$2\tau d_{min}Q$
		$B_n/4\pi f_m T_I$	. . . .	$Q/\pi f_{mTc}$	$2Q$			
$1/T_I$ to $1/\pi T_c$	Target *SNR* loss	$(0.1i!)^{1/i}$	No require-ment	$Q/\pi T_c$	$2f_m Q$	No requirement (for natural sampling)		
	Clutter spreading	$[j!\sqrt{K_s/(C/N)}]^{1/j}$		$Q/\pi T_c$	$2f_m Q$			
$f_L - f_{bw}$ to $f_L + f_{bw}$	Tracking error	$K_M \sigma_E/\sqrt{2}(\theta \ or \ \tau)$	$2Q$	No require-ment	No require-ment	$2\tau d_{min}Q$	$\tau d_{min}Q$	$2\tau d_{min}Q$
$1/\pi T_c$ to $B_n/2$	Target *SNR* loss	. . . .	No require-ment	$f_m(0.1i!)^{1/i}$	$2f_m(0.1i!)^{1/i}$	No requirement (for natural sampling)		
	Clutter spreading	$[j!\sqrt{K_s/(C/N)}]^{1/j}$		$f_m Q$	$2f_m Q$			
$B_n/2$ to f_{min}	Target *SNR* loss	. . . .	No require-ment	$f_m(0.1i!)^{1/i}$	$2f_m(0.1i!)^{1/i}$	No requirement (for natural sampling)		
	Clutter spreading	$[j!\sqrt{K_{s/(C/N)}}]^{1/j}$		$f_m Q$	$2f_m Q$			

$5B_n$ to f_{min}	False targets	$\sqrt{K_s/(\text{max}SNR)}$	$2Q$	f_mQ	$2f_mQ$	$2\tau Q$	τQ	$2\tau Q$
f_{min} and higher	Clutter spreading	$\sqrt{K_s/(C/N)}$	$2Q$	f_mQ	$2f_mQ$	$2\tau Q$	τQ	$2\tau Q$

M = peak fractional carrier amplitude modulation

δ_F = peak carrier frequency deviation, Hz

f_m = modulating frequency, Hz

T_c = two-way delay for main-beam clutter, s

δ_W = peak pulse-width deviation, s

d_{min} = minimum target signal duty cycle when partly eclipsed (say, $\tau f_R/5$)

δ_p = peak pulse-position deviation, s

K_s = safety factor of modulation power sidebands relative to system noise ($K_s \leq 1$)

f_L = tracking subcarrier frequency, Hz

f_{bw} = tracking-loop noise bandwidth, Hz

K_M = modulation sensitivity (fractional modulation/beamwidth or pulse-width error)

σ_E = rms tracking error due to modulation sidebands

θ = antenna beamwidth

max SNR = maximum target power SNR that does not cause automatic gain control or limiting

i = integer part of $[(B_n/2) + f_m]/f_m$

j = integer part of $(f_{min} + f_m)/f_m$

δ_Δ = peak fractional pulse-amplitude modulation

Applying all these constraints to the various types of modulations expected gives the sinusoidal modulation allowances of Table 17.5 where 0.1 dB tolerable loss was assumed in *SNR*. The various modulation-frequency regions (not sideband frequencies) corresponding to the predominant sideband effects (i.e., the criterion column in Table 17.5) are listed. A factor Q for each region is indicated; it is related to the maximum allowable deviation indicated in other columns. For example, for independent pulse-width modulation, for a modulation frequency greater than f_{min}, the maximum allowable deviation is

$$\delta_w = 2\tau Q = 2\tau \left[\frac{K_s}{(C/N)\text{max}} \right]^{1/2} \qquad (17.17)$$

for acceptable clutter spreading. The safety factor K_s assures that the clutter sidebands will be buried in receiver noise.

Each tabulated value assumes that only one source of modulation is present. If multiple modulations are expected, an appropriate reduction factor must be provided so that the composite sidebands do not exceed the tolerable value.

In the receiver following the main-beam clutter filter, the clutter-spreading requirements are not germane. The *SNR* and false-target considerations then become the limiting factors at high-sideband frequencies as well as low-sideband frequencies.

Narrowband Noise. Although the requirements in Table 17.5 relate to a single-frequency sinusoidal modulation, they can be interpreted for narrowband-noise modulation. For this interpretation, the listed modulation values represent $\sqrt{2}$ times the allowable rms noise modulation referenced to a bandwidth of B_n. This interpretation is fairly good for modulating frequencies much greater than $B_n/2$, but for lower frequencies it is only a rough guide.

PPM and PWM. The pulse-position and pulse-width values are based on the assumption of natural sampling,[43] as is normally the case. In natural sampling, the deviation of the pulse edge is determined by the amplitude of the modulating signal at the time of occurrence of the pulse edge. Because of the range delay of target clutter signals and the range gating, PPM can be converted to PWM, which generates much larger sidebands. Most clutter signals are partially eclipsed or partially outside the range gate; therefore, the position of one edge of the pulse in the receiver is determined by the transmitted pulse, and the other edge is determined by the range gate or transmitter blanking pulse. Thus the original PPM can be converted into PWM as a function of modulating frequency and time delay to the clutter.

The deviation requirements in Table 17.5 were derived for the worst case, where the two common modulating signals are 180° out of phase or one edge of the received clutter pulse is completely eclipsed in the independent case.

Frequency Modulation. Common FM permits relaxed modulation requirements at low frequencies but 6 dB more severe requirements at high frequencies compared with independent FM.[44] This effect is caused by the range delay of the clutter. Since the transmitter and the receiver local oscillator are frequency-modulated in synchronism, the deviation of the IF difference signal is dependent on the range delay.

Droop. Although it does not cause new spectral sidebands and therefore was not included in Table 17.5, pulse droop on the transmitter modulator pulse is also of interest. This is because of the high phase-modulation sensitivity of traveling-wave tubes and klystrons typically used as RF amplifiers. A linear droop of the modulating voltage will serrodyne the RF signal, shifting the peak of the spectral

envelope relative to the RF carrier frequency and reducing the useful signal power when passed through a filter matched to the pulse width. For small droop this loss is given by

$$\text{Fractional power loss} = \frac{1}{3}\left[\frac{\pi}{3.6}K_\phi\frac{\Delta V}{V}\right]^2 \qquad (17.18)$$

where K_ϕ = transmitter phase sensitivity (degrees phase change/percent voltage)
$\Delta V/V$ = fractional voltage droop on modulator pulse

Pulse-to-Pulse Random Modulation. In addition to the sinusoidal or narrowband-noise modulation, pulse-to-pulse random modulation may also be present. The predominant effect is clutter spreading noise into the detection filter. With the same notation as for Table 17.5, the factor Q is $[K_s f_R/(C/N)B_n]^{1/2}$, and the rms allowable fractional AM is equal to Q, as is the rms phase modulation (given in radians). The rms PPM or single-edge PWM allowable modulation (given in seconds) is τQ.

17.7 RANGE PERFORMANCE

Chapter 2 discusses the general radar range equation and the calculation of detection probability. This section extends those concepts to pulse doppler radars and includes a discussion of system losses and false-alarm probability. Generalized detection curves, which include multilook detection criteria, are presented.

Range Equation. In the doppler region where the signal does not fall in clutter, performance is limited only by system noise. The signal-to-noise ratio in the detection filter prior to postdetection integration for a target at range R is given by

$$SNR = \left(\frac{R_o}{R}\right)^4 \qquad (17.19)$$

$$R_o = \left(\frac{P_A\,G_T\,G_R\,\lambda^2\,\sigma_T}{(4\pi)^3 k\,T_s\,B_n\,L}\right)^{1/4} \qquad (17.20)$$

where R_o = range at which $S/N = 1$
σ_T = target radar cross section
L = losses applicable to the target

The remaining terms are as defined following Eq. (17.2).

System Losses. Some of the losses inherent in, but not necessarily unique to, pulse doppler radars which employ digital signal processing are discussed below.

Quantization Noise Loss. This loss is due to the noise added by the A/D conversion process and to truncation due to finite word lengths in the signal-processing circuits which follow.[45]

CFAR Loss. This is caused by an imperfect estimate of the detection threshold compared with the ideal threshold. The fluctuation in the estimate necessitates that the mean threshold be set higher than the ideal, hence a loss.

Doppler Filter Straddle Loss. This loss is due to a target not always being in the center of a doppler filter. It is computed by assuming a uniformly distributed target doppler over one filter spacing and is a function of the FFT sidelobe weighting.

Amplitude-Weighting Loss. This loss results from the increased noise bandwidth of the doppler filters that occurs because the filter sidelobe weighting. It can also be accounted for by an increase of the doppler filter noise bandwidth instead of as a separate loss.

Pulse Compression Mismatch Loss. This is caused by the intentional mismatching of the pulse compression filter to reduce the time (range) sidelobes.

Guard Blanking Loss. This is the detectability loss in the main channel caused by spurious blanking from the guard channel. (See Fig. 17.9.)

Eclipsing and Range-Gate Straddle Loss. Because of eclipsing, the value of R_o, given by Eq. (17.20), may fall anywhere between zero and a maximum value, depending on the exact location of the target return in the interpulse period. When the PRF is high, so that many range ambiguities occur, the target range delay may be considered to be random from scan to scan, with a uniform distribution over the interpulse period. An approximate measure of performance in this case is found by first computing a detection curve averaged over target ambiguous ranges from zero to the range corresponding to the interpulse period. The loss is equal to the increase in signal-to-noise ratio required to obtain the same probability of detection with eclipsing or straddle as in the case when the transmit pulse is received by a matched gate with no straddle. Since the detection curve changes shape, the loss depends on the probability of detection selected. A less accurate approximation compares the average signal-to-noise ratio over the interpulse period with the signal-to-noise ratio of the matched case. In the case of M contiguous range gates of width τ that occupy the entire interpulse period except for the transmitted pulse also of width τ, the average eclipsing and straddle loss on a signal-to-noise-ratio basis is

$$\text{Eclipse and straddle loss} = \frac{Y}{3(M + 1)} \qquad \tau_t = \tau_g \qquad (17.21)$$

where $Y_1 = (1 - R)(2 + R)$ $M = 1$

$Y = (1 - R)(1 - R + 2X) + 2 + 1.75(M - 2)$ $M > 1, R \geqslant 0.618$

$Y = (1 - R)(1 + R + Z) + (Z - R)[Z(Z + X)]$
       $+ (1 - Z)[Z(Z + 1) + 1] + 1 + 1.75(M - 2)$ $M > 1, R < 0.618$

$Z = 1/(1 + X)$

$X = \sqrt{1 - R}$

$R = \tau_b/\tau$

τ_b = width of first gate blanking

τ = width of transmitted pulse τ_t and receiver gate τ_g

M = number of contiguous gates

The loss is plotted in Fig. 17.19. Signals cannot be received during the time τ_b following each transmit pulse owing to slow shutoff of the transmitter or recovery of the duplexer and/or receiver-protector.
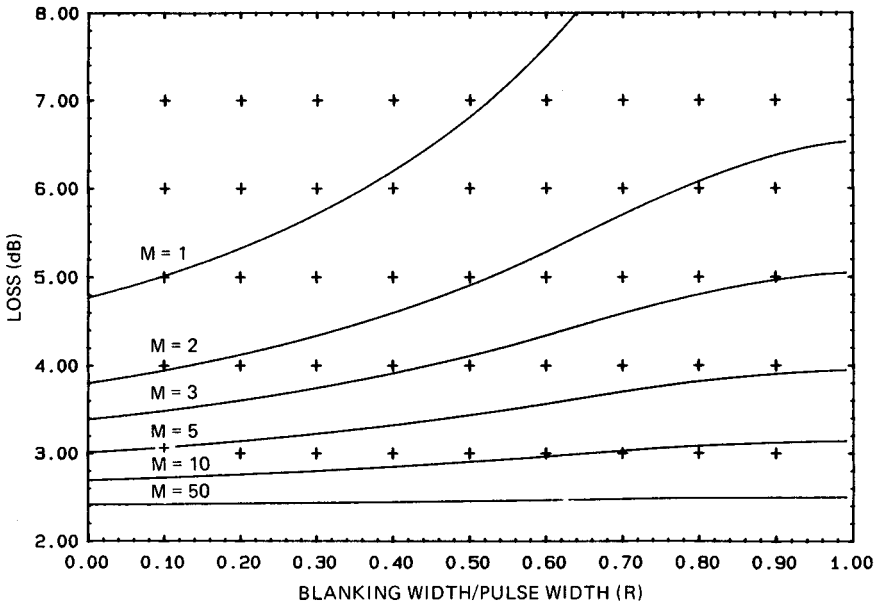
FIG. 17.19 Eclipsing and straddle loss for M contiguous range gates and equal transmitted pulse and range-gate widths as a function of the number of receive gates and blanking width.

Although Eq. (17.21) assumes contiguous range gates, the loss factor can be reduced by the use of overlapping gates at the expense of extra hardware and possibly more range ghosts.

Probability of False Alarm. PD radars often employ a multilook detection criterion to resolve range ambiguities such that during the time on target (dwell time) several PRFs are transmitted in successive looks and a threshold detection in more than one look is required for the radar to output a target report. For the case where a doppler filter bank in each range gate is used for coherent integration, possibly followed by a postdetection integrator, the probability of false alarm P_{FA} in each range gate–doppler filter required to obtain a given false report time T_{FR} is given approximately by

$$P_{FA} = \frac{1}{N_F}\left[\frac{0.693\ T_d}{\binom{n}{m} N_g\ T_{FR}}\right]^{1/m} \tag{17.22}$$

where N_F = number of independent doppler filters visible in the doppler passband (number of unblanked filters/FFT weighting factor)

  n = number of looks in a dwell time

  m = number of detections required for a target report (For example, 3 detections out of 8 PRFs is $m = 3$ and $n = 8$.)

  T_d = total dwell time of the multiple PRFs including postdetection integration (if any) and any dead time

  $\binom{n}{m}$ = binomial coefficient $n!/[m!(n-m)!]$

N_g = number of range gates in the output unambiguous-range interval (display range/range-gate size)

T_{FR} = false-report time [per Marcum's definition where the probability is 0.5 that at least one false report will occur in the false-report time (Ref. 46)]

Equation (17.22) is for the case where no doppler correlation is required for a target report. In the case where both range and doppler correlation are used, the required P_{FA} is

$$P_{FA} = \left[\frac{0.693 \, T_d}{\binom{n}{m} N_{fu} \, N_g \, T_{FR} \, W^{m-1}} \right]^{1/m} \tag{17.23}$$

where N_{fu} = number of independent doppler filters in the unambiguous doppler region and W = width (in filters) of the correlation window applied to detections following initial detection.

Probability of Detection. The probability-of-detection curves presented in Chap. 2 have been extended to include multilook detection criteria and are presented in a generalized fashion after Ref. 47; that is, they are presented in terms of gR/R_o, where the g factor is a function of the number of pulses integrated noncoherently and the probability of false alarm. For coherent integration $N = 1$ and g becomes

$$g = [- \ell n(P_{FA})]^{1/4} \tag{17.24}$$
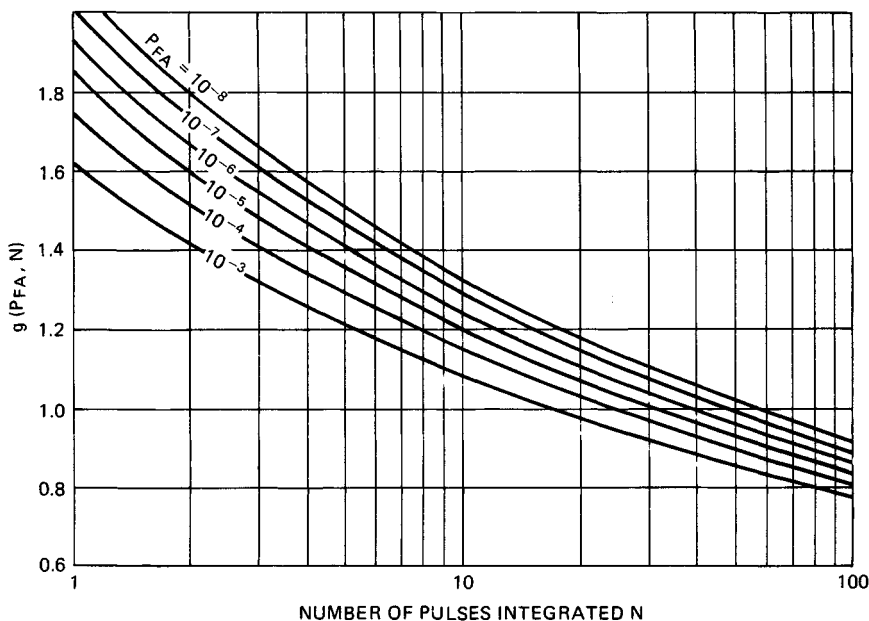
FIG. 17.20 Swerling's g factor as a function of P_{FA} and N (after Ref. 47).

The g factor is plotted in Fig. 17.20 for various false-alarm probabilities and number of pulses integrated.

The generalized results are based on the realization that Marcum's curves[46] are very similar over a wide range of parameters; one can reasonably use a single universal Marcum curve as shown in Fig. 17.21. This is patterned after the universal curve presented in Ref. 47, except that here it is for a nonfluctuating target. It is accurate to within 1 dB over the entire range (and closer over most of the range) of integration samples N from 1 to 100, probability of false-alarm values P_{FA} from 10^{-3} to 10^{-8}, and probability from 1 to 99 percent. To use Fig. 17.21, a value of $g(P_{FA},N)$ is first found from Fig. 17.20; this can then be used to convert gR/R_o values to R/R_o or signal-to-noise-ratio values.
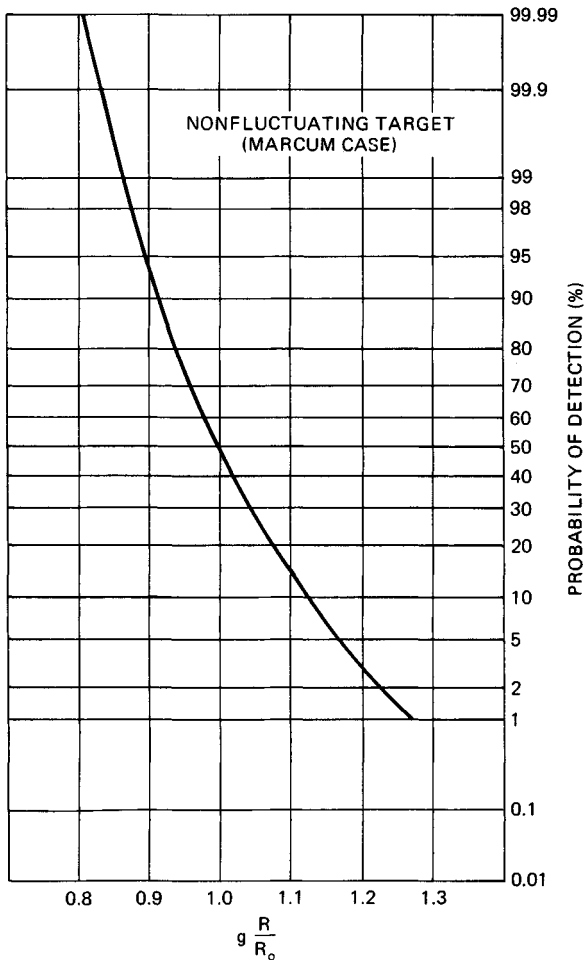
FIG. 17.21   Universal Marcum curve.

Generalized Curves. By using the universal Marcum curve, several detection cases have been generated and are shown in Fig. 17.22. These are all for a Swerling Case 1 target in which the target amplitude fluctuates independently from scan to scan but is constant within the dwell time. No losses other than the fluctuation loss have been included in these curves, so that any losses such as range-gate straddle and eclipsing can be accounted for in the computation of R_o.

The Swerling Case 1 single-scan detection curves can be closely approximated by

$$P_d = P_{\text{FA}}^{\left(\frac{1}{a+b\,SNR}\right)} \qquad (17.25)$$

where P_d = single-scan detection probability
P_{FA} = probability of false alarm
SNR = signal-to-noise ratio = $(R_o/R)^4$
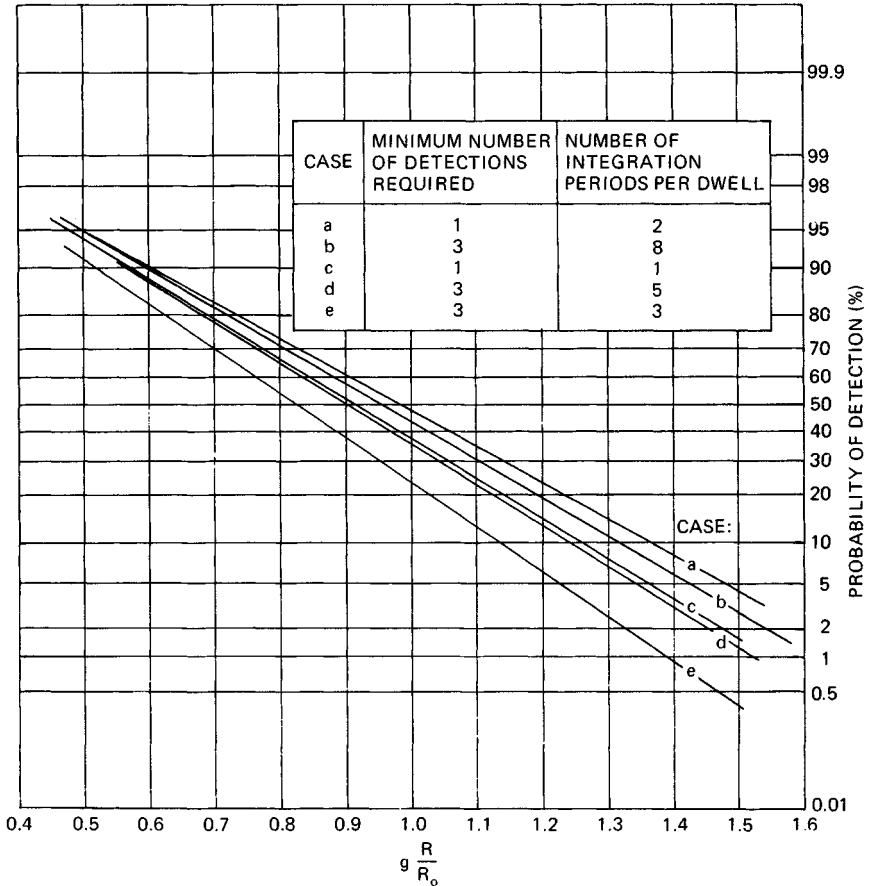
FIG. 17.22 Generalized single-scan probability of detection for a scan-to-scan, Swerling Case 1, Rayleigh fluctuating target.

CASE	MINIMUM NUMBER OF DETECTIONS REQUIRED	NUMBER OF INTEGRATION PERIODS PER DWELL
a	1	2
b	3	8
c	1	1
d	3	5
e	3	3
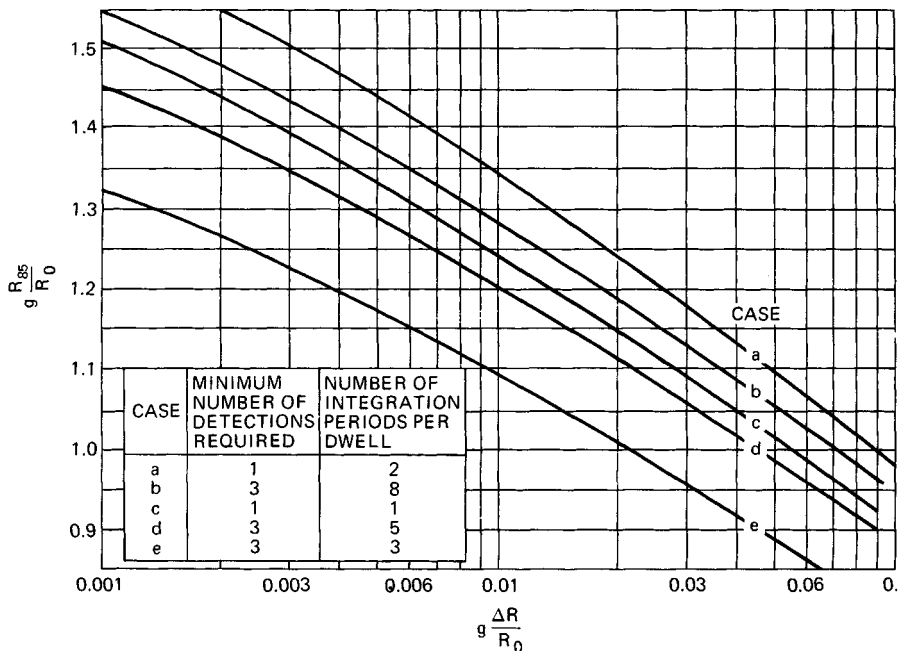
FIG. 17.23 Generalized 85 percent cumulative probability of detection for scan-to-scan, Swerling Case 1, Rayleigh fluctuating target.

The constants a and b can be found by substitution, using two pairs of P_d and gR/R_o values from Fig. 17.22, converting gR/R_o to SNR, and solving the resultant simultaneous equations.

The cumulative probability of detection (probability of detecting the target at least once in k scans), Pc_k, is defined as

$$Pc_k = 1 - \prod_{i=1}^{k}[1 - P_d(i)] \qquad (17.26)$$

where $P_d(i)$ is the probability of detection on the ith scan. The cumulation may occur over a variable number of scans, such as when it begins at a range where $P_d(i)$ is approximately zero, or over a defined number of scans, where a 1-out-of-N acquisition criterion must be satisfied. The single-scan probability-of-detection curves shown in Fig. 17.22 have been used to compute the 85 percent cumulative probability of detection for the variable-scan case, shown in Fig. 17.23. ΔR is the change in range between successive scans for a fixed-velocity, radially moving target.

Clutter-Limited Case. The foregoing discussion assumed that the target fell in the noise-limited (i.e., clutter-free) part of the doppler band. If the target falls in the sidelobe clutter region, the range performance will be degraded, since the total power (system noise plus clutter) with which the target must compete is increased. The foregoing discussion can be applied to the sidelobe clutter region, however, by interpreting R_o as the range where the signal is equal to sidelobe clutter plus system

noise.[48–50] The CFAR loss may also be higher owing to the increased variability of the threshold when the clutter varies over the target detection region.

REFERENCES

1. Skolnik, M. I.: Fifty Years of Radar, *Proc. IEEE*, vol. 73, pp. 182–197, February 1985.

2. Perkins, L. C., H. B. Smith, and D. H. Mooney: The Development of Airborne Pulse Doppler Radar, *IEEE Trans.*, vol. AES-20, pp. 290–303, May 1984.

3. Clarke, J., D. E. N. Davies, and M. F. Radford: Review of United Kingdom Radar, *IEEE Trans.*, vol. AES-20, pp. 506–520, September 1984.

4. Moaveni, M. K.: Corrections to "Radio Interference in Helicopter-Borne Pulse Doppler Radars," *IEEE Trans.*, vol. AES-14, p. 688, July 1978.

5. Moaveni, M. K.: Radio Interference in Helicopter-Borne Pulse Doppler Radars, *IEEE Trans.*, vol. AES-14, pp. 319–328, March 1978.

6. Ringel, M. B., D. H. Mooney, and W. H. Long: F-16 Pulse Doppler Radar (AN/APG-66) Performance, *IEEE Trans.*, vol. AES-19, pp. 147–158, January 1983.

7. Skillman, W. A.: Utilization of the E-3A Radar in Europe, *Proc. Mil. Electron. Def. Expo '78*, Wiesbaden, Germany, 1978.

8. Skillman, W. A.: Microwave Technology, Key to AWACS Success, *Proc. IEEE MTT/ S*, Boston, 1983.

9. Clarke, J.: Airborne Early Warning Radar, *Proc. IEEE*, vol. 73, pp. 312–324, February 1985.

10. Doviak, R. J., D. S. Zrnic, and D. S. Sirmans: Doppler Weather Radar, *Proc. IEEE*, vol. 67, no. 11, pp. 1522–1553, 1979.

11. Stimson, G. W.: "Introduction to Airborne Radar," Hughes Aircraft Company, El Segundo, Calif., 1983, pt. 7.

12. Hovanessian, S. A.: Medium PRF Performance Analysis, *IEEE Trans.*, vol. AES-18, pp. 286–296, May 1982.

13. Aronoff, E., and N. M. Greenblatt: Medium PRF Radar Design and Performance, *20th Tri-Service Radar Symp.*, 1974. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. IV-7, pp. 261–276.

14. Long, W. H., and K. A. Harriger: Medium PRF for the AN/APG-66 Radar, *Proc. IEEE*, vol. 73, pp. 301–311, February 1985.

15. Goetz, L. P., and J. D. Albright: Airborne Pulse Doppler Radar, *IRE Trans.*, vol. MIL-5, pp. 116–126, April 1961. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. IV-3, pp. 215–225.

16. Skolnik, M. I.: "Introduction to Radar Systems," 2d ed., McGraw-Hill Book Company, New York, 1984, chap. 4.

17. Finn, H. M., and R. S. Johnson: Adaptive Detection Mode with Threshold Control as a Function of Spatially Sampled Clutter-Level Estimates, *RCA Rev.*, pp. 414–464, September 1968.

18. Steenson, B. O.: Detection Performance of a Mean-Level Threshold, *IEEE Trans.*, vol. AES-4, pp. 529–534, July 1968.

19. Rohling, H.: Radar CFAR Thresholding in Clutter and Multiple Target Situations, *IEEE Trans.*, vol. AES-19, pp. 608–621, July 1983.

20. Hansen, V. G.: Constant False Alarm Rate Processing in Search Radars, *Proc. IEEE Int. Radar Conf.*, pp. 325–332, London, 1973.

21. Farrell, J., and R. Taylor: Doppler Radar Clutter, *IEEE Trans.*, vol. ANE-11, pp.

162–172, September 1964. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. VI-2, pp. 351–361.

22. Helgostam, L., and B. Ronnerstam: Ground Clutter Calculation for Airborne Doppler Radar, *IEEE Trans.*, vol. MIL-9, pp. 294–297, July–October 1965.

23. Friedlander, A. L., and L. J. Greenstein: A Generalized Clutter Computation Procedure for Airborne Pulse Doppler Radars, *IEEE Trans.*, vol. AES-6, pp. 51–61, January 1970. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. VI-3, pp. 363–374.

24. Ringel, M. B.: An Advanced Computer Calculation of Ground Clutter in an Airborne Pulse Doppler Radar, *NAECON '77 Rec.*, pp. 921–928. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. VI-4, pp. 375–382.

25. Jao, J. K., and W. B. Goggins: Efficient, Closed-Form Computation of Airborne Pulse Doppler Clutter, *Proc. IEEE Int. Radar Conf.*, pp. 17–22, Washington, 1985.

26. Harvey, D. H., and T. L. Wood: Designs for Sidelobe Blanking Systems, *Proc. IEEE Int. Radar Conf.*, pp. 410–416, Washington, 1980.

27. Maisel, L.: Performance of Sidelobe Blanking Systems, *IEEE Trans.*, vol. AES-4, pp. 174–180, March 1968.

28. Finn, H. M., R. S. Johnson, and P. Z. Peebles: Fluctuating Target Detection in Clutter Using Sidelobe Blanking Logic, *IEEE Trans.*, vol. AES-7, pp. 147–159, May 1971.

29. Mooney, D. H.: Post Detection STC in a Medium PRF Pulse Doppler Radar, U.S. Patent 690,754, May 27, 1976.

30. Skillman, W. A.: "SIGCLUT: Surface and Volumetric Clutter-to-Noise, Jammer and Target Signal-to-Noise Radar Calculation Software and User's Manual," Artech House, Norwood, Mass., 1987.

31. Ziemer, R. E., and J. A. Ziegler: MTI Improvement Factors for Weighted DFTs, *IEEE Trans.*, vol. AES-16, pp. 393–397, May 1980.

32. Skillman, W. A.: "Radar Calculations Using the TI-59 Programmable Calculator," Artech House, Norwood, Mass., 1983, p. 308.

33. Skillman, W. A.: "Radar Calculations Using Personal Computers," Artech House, Norwood, Mass., 1984.

34. Ward, H. R.: Doppler Processor Rejection of Ambiguous Clutter, *IEEE Trans.*, vol. AES-11, July 1975. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. IV-11, pp. 299–301.

35. Fletcher, R. H., Jr., and D. W. Burlage: An Initialization Technique for Improved MTI Performance in Phased Array Radar, *Proc. IEEE*, vol. 60, pp. 1551–1552, December 1972.

36. Skillman, W. A., and D. H. Mooney: Multiple High-PRF Ranging, *Proc. IRE Conf. Mil. Electron.*, pp. 37–40, 1961. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. IV-1, pp. 205–213.

37. Hovanessian, S. A.: An Algorithm for Calculation of Range in Multiple PRF Radar, *IEEE Trans.*, vol. AES-12, pp. 287–289, March 1976.

38. Ore, O.: "Number Theory and Its History," McGraw-Hill Book Company, New York, 1948, pp. 246–249.

39. Goetz, L. P., and W. A. Skillman: Master Oscillator Requirements for Coherent Radar Sets, *IEEE-NASA Symp. Short Term Frequency Stability*, NASA-SP-80, November 1964.

40. Raven, R. S.: Requirements for Master Oscillators for Coherent Radar, *Proc. IEEE*, vol. 54, pp. 237–243, February 1966. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. V-1, pp. 317–323.

41. Gray, M., F. Hutchinson, D. Ridgely, F. Fruge, and D. Cooke: Stability Measurement

Problems and Techniques for Operational Airborne Pulse Doppler Radar, *IEEE Trans.*, vol. AES-5, pp. 632–637, July 1969.

42. Acker, A. E.: Eliminating Transmitted Clutter in Doppler Radar Systems, *Microwave J.*, vol. 18, pp. 47–50, November 1975. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. V-3, pp. 331–336.

43. Black, H. S.: "Modulation Theory," D. Van Nostrand Company, Princeton, N.J., 1953, p. 265.

44. Barton, D. K.: "Radar Systems Analysis," Prentice-Hall, Englewood Cliffs, N.J., 1964, p. 206.

45. Ziemer, R. E., T. Lewis, and L. Guthrie: Degradation Analysis of Pulse Doppler Radars Due to Signal Processing, *NAECON 1977 Rec.*, pp. 938–945. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. IV-12, pp. 303–312.

46. Marcum, J. I.: A Statistical Theory of Target Detection by Pulsed Radar, *IRE Trans.*, vol. IT-6, pp. 59–267, April 1960.

47. Swerling, P.: Probability of Detection for Fluctuating Targets, *IRE Trans.*, vol. IT-6, pp. 269–308, April 1960.

48. Mooney, D., and G. Ralston: Performance in Clutter of Airborne Pulse MTI, CW Doppler and Pulse Doppler Radar, *IRE Conv. Rec.*, vol. 9, pt. 5, pp. 55–62, 1961. Reprinted in Barton, D. K.: "CW and Doppler Radars," vol. 7, Artech House, Norwood, Mass., 1978, sec. VI-1, pp. 343–350.

49. Ringel, M. B.: Detection Range Analysis of an Airborne Medium PRF Radar, *IEEE NAECON Rec.*, Dayton, Ohio, pp. 358–362, 1981.

50. Holbourn, P. E., and A. M. Kinghorn: Performance Analysis of Airborne Pulse Doppler Radar, *Proc. IEEE Int. Radar Conf.*, pp. 12–16, Washington, 1985.

CHAPTER 18
TRACKING RADAR

Dean D. Howard
Locus, Inc., a subsidiary of Kaman Corp.

18.1 INTRODUCTION

A typical tracking radar has a pencil beam to receive echoes from a single target and track the target in angle, range, and/or doppler. Its resolution cell—defined by its antenna beamwidth, transmitter pulse length, and/or doppler bandwidth—is usually small compared with that of a search radar and is used to exclude undesired echoes or signals from other targets, clutter, and countermeasures. Electronic beam-scanning phased array tracking radars may track multiple targets by sequentially dwelling upon and measuring each target while excluding other echo or signal sources.

Because of its narrow beamwidth, typically from a fraction of 1° to 1 or 2°, a tracking radar usually depends upon information from a search radar or other source of target location to acquire the target, i.e., to place its beam on or in the vicinity of the target before initiating a track. Scanning of the beam within a limited angle sector may be needed to fully acquire the target within its beam and center the range-tracking gates on the echo pulse prior to locking on the target or closing the tracking loops.

The primary output of a tracking radar is the target location determined from the pointing angles of the beam and position of its range-tracking gates. The angle location is the data obtained from synchros or encoders on the antenna tracking axes shafts (or data from a beam-positioning computer of an electronic-scan phased array radar). In some cases, tracking lag is measured by converting tracking-lag-error voltages from the tracking loops to units of angle. This data is used to add to or subtract from the angle shaft position data for real-time correction of tracking lag.

There are a large variety of tracking-radar systems, including some that achieve simultaneously both surveillance and tracking functions. A widely used type of tracking radar and the one to be discussed in detail in this chapter is a ground-based system consisting of a pencil-beam antenna mounted on a rotatable platform which is caused by motor drive of its azimuth and elevation position to follow a target (Fig. 18.1). Errors in pointing direction are determined by sensing the angle of arrival of the echo wavefront and corrected by positioning the antenna to keep the target centered in the beam.

The principal applications of tracking radar are weapon control and missile-

FIG. 18.1 AN/FPQ-6 C-band monopulse precision tracking radar installation at the NASA Wallops Island Station, Va. It has a 29-ft-diameter antenna and a specified angle precision of 0.05 mrad rms.

range instrumentation. In both applications a high degree of precision and an accurate prediction of the future position of the target are generally required. The earliest use of tracking radar was in gunfire control. The azimuth angle, the elevation angle, and the range to the target were measured, and from the rate of change of these parameters the velocity vector of the target was computed and its future position predicted. This information was used to point the gun in the proper direction and to set the fuzing time. The tracking radar performs a similar role in providing guidance information and steering commands for missiles.

In missile-range instrumentation, the tracking-radar output is used to measure the trajectory of the missile and to predict future position. Tracking radar which computes the impact point of a missile continuously during flight is also important for range safety. Missile-range instrumentation radars are normally used with a beacon to provide a point-source target with high signal-to-noise ratio. Some of these systems achieve a precision of the order of 0.1 mil in angle and a range accuracy of 5 yd.

This chapter describes the conical-scan, sequential-lobing, and monopulse (both phase comparison and amplitude comparison) tracking-radar techniques, with the main emphasis on the amplitude-comparison monopulse radar.

18.2 SCANNING AND LOBING

The first technique used for angle tracking of targets by radar was to sense the target location with respect to the antenna axis by rapidly switching the antenna beam from one side of the antenna axis to the other, as in Fig. 18.2. The original tracking radars of this type, such as the SCR-268, used an array of radiating elements which could be switched in phase to provide two beam positions for the lobing operation. The radar operator observed an oscilloscope that displayed side by side the video returns from the two beam positions. When the target was on axis, the two pulses were of equal amplitude (Fig. 18.2a); when the target moved off axis, the two pulses became unequal (Fig. 18.2b). The radar operator, observing the existence of an error and its direction, could position the antenna to regain a balance between the two beam positions. This provided a manual tracking loop.

This lobing technique was extended to continuous rotation of the beam around the target (conical scan) as in Fig. 18.3.[1] Angle-error-detection circuitry is provided to generate error voltage outputs proportional to the tracking error and with a phase or polarity to indicate the direction of errors. The error signal actuates a servosystem to drive the antenna in the proper direction to reduce the error to zero.

Continuous beam scanning is accomplished by mechanically moving the feed of an antenna since the antenna beam will move off axis as the feed is moved off the focal point. The feed is typically moved in a circular path around the focal point, causing a corresponding movement of the antenna beam in a circular path around the target. A typical block diagram is shown in Fig. 18.4.[2] A range-tracking system is included which automatically follows the target in range, with range gates that turn on the radar receiver only during the time when the echo is expected from the target under track. Range gating excludes undesired targets and noise. The system also includes an automatic gain control (AGC) necessary to maintain constant angle sensitivity (volts of error-detector output per degree of
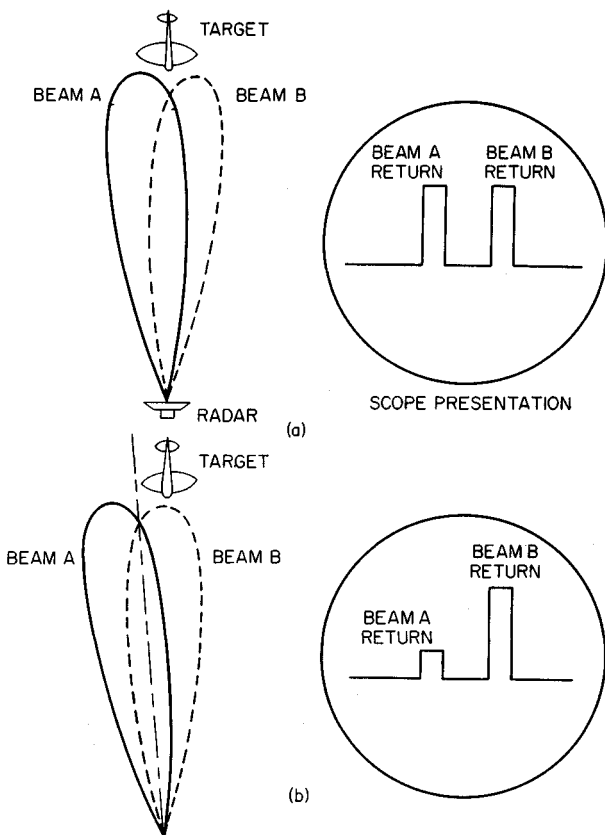
FIG. 18.2 Angle error sensing in one coordinate by switching the antenna beam position from one side of the target to the other. (*a*) Target located on the antenna axis. (*b*) Target at one side of the antenna axis.

error) independent of the amplitude of the echo signal. This provides the constant gain in the angle-tracking loops necessary for stable angle tracking.

The feed scan motion may be either a rotation or a nutation. A rotating feed turns as it moves with circular motion, causing the polarization to rotate. A nutating feed does not rotate the plane of polarization during the scan; it has a motion like moving one's hand in a circular path.

The radar video output contains the angle-tracking-error information in the envelope of the pulses, as shown in Fig. 18.5. The percentage modulation is proportional to the angle-tracking error, and the phase of the envelope function relative to the beam-scanning position contains direction information. Angle-tracking-error detection (error demodulation) is accomplished by a pair of phase detectors using a reference input from the scan motor. The phase detectors perform essentially as dot-product devices with sine-wave reference signals at the frequency of scan and of proper phases to obtain elevation error from one and
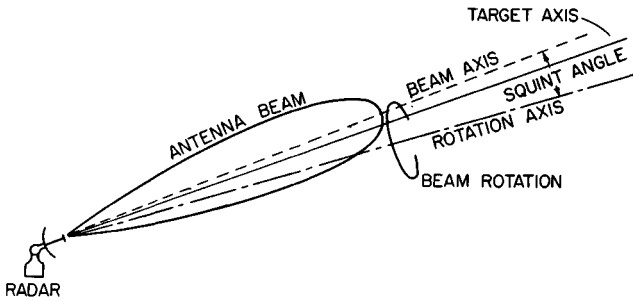
FIG. 18.3 Conical-scan tracking.
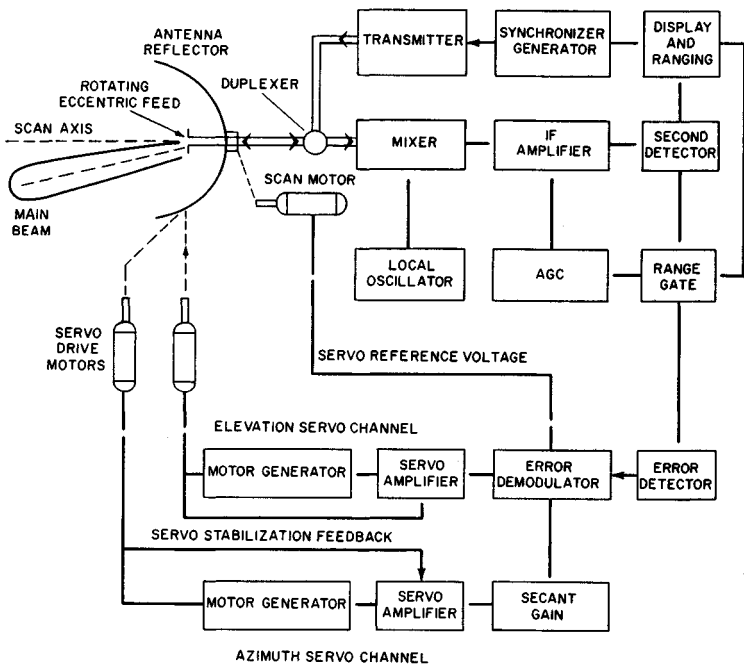
FIG. 18.4 Block diagram of a conical-scan radar.

azimuth error from the other. For example, the top scan position may be chosen as zero phase for a cosine function of the scan frequency. This provides a positive voltage output proportional to the angle error when the target is above the antenna axis. The reference signal to the second phase detector is generated with a 90° phase relation to the original reference. This provides a similar error voltage proportional to the azimuth-angle error and with polarity corresponding to the direction of error.

A secant correction (Fig. 18.4) is necessary in any conventional elevation-over-azimuth tracking radar where the elevation drive system rotates when the
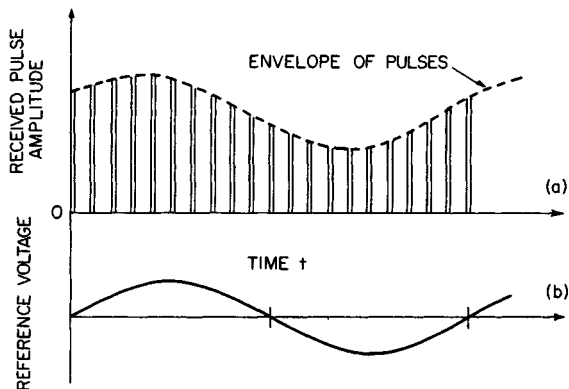
FIG. 18.5 (*a*) Angle error information contained in the envelope of the received pulses in a conical-scan radar. (*b*) Reference signal derived from the drive of the conical-scan feed.

antenna changes azimuth. A target flying a passing course by the radar will, at its closest point to the radar, cause the azimuth servo to drive faster at high elevation angles than at low elevation angles. In the extreme, when the target passes directly overhead, the azimuth drive would have to flip the antenna 180° at the instant when the target crosses overhead. This effect requires the azimuth-tracking loop gain to change approximately as the secant of elevation angle in order to maintain essentially constant overall azimuth loop gain. In practical tracking radars with the conventional elevation-over-azimuth mount the elevation angle is typically limited by this effect to a maximum of 85° since the servo bandwidth required for higher elevation angles exceeds practical limits.

A major parameter in a conical-scan radar is the size of the circle to be scanned relative to the beamwidth. Figure 18.6 shows a circle representing the 3 dB contour of the beam at one position of its scan. The half-power beamwidth is θ_B. The dashed circle represents the path described by the center of the beam as it is scanned. The radius of the dashed circle is β, the offset angle. The compromise that must be made in choosing β is between the loss of signal or antenna gain L_k (crossover loss) and the increase in angle sensitivity k_s of the angle-sensing circuits.

High angle-error sensitivity is desired to obtain higher voltage from the angle error detectors for a given true angle error relative to undesired voltages in the receiver output. The undesired receiver output includes angle errors caused by receiver thermal noise. For a given signal-to-noise ratio (*SNR*) the thermal-noise effects are inversely proportional to angle error sensitivity. Unfortunately, increasing β to increase k_s also increases the loss L_k, which reduces the *SNR*.

The relative values of k_s and L_k depend upon whether the target provides a beacon response that removes the transmitting modulation from the received signal or whether two-way skin tracking is performed. Two-way tracking gives a greater depth of modulation, or angle sensitivity, for a given β but doubles the loss in decibels. Figure 18.7 shows the loss of antenna gain and angle sensitivity as a function of β for the two cases.[2] The rms error caused by receiver thermal noise is inversely proportional to k_s and proportional to $\sqrt{L_k}$ (L_k expressed as power loss). The peaks of the dotted curves labeled $k_s/\sqrt{L_k}$ indicate the optimum offset angle β for minimizing receiver thermal-noise effects on angle tracking. However, the range-tracking system of the radar is affected by L_k only, and a β
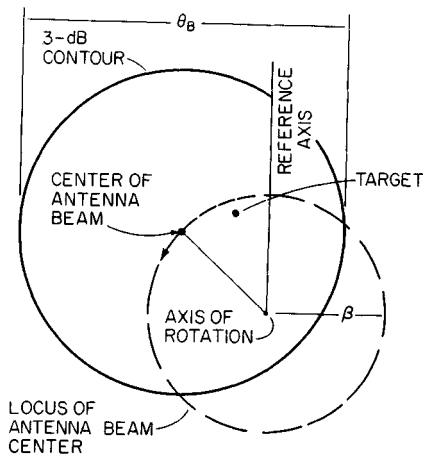
FIG. 18.6 Conical-scan-radar antenna beam 3 dB contour (solid circle) and path of rotation (dashed circle) of the beam center.
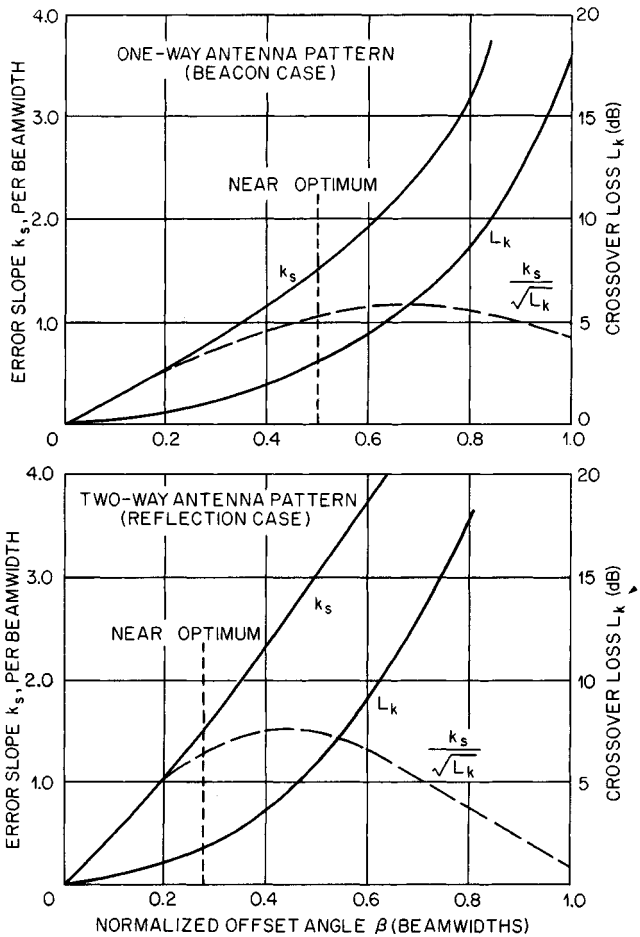
FIG. 18.7 Error slope k_s and crossover loss L_k.

of zero maximizes range-tracking performance. Therefore, values of β indicated by the vertical dashed lines are chosen smaller than optimum for angle tracking as a compromise between angle- and range-tracking performance.

The tracking-error information in beam-scanning tracking radars is a time fluctuation of the echo signal amplitude. Other sources of echo-signal-amplitude fluctuation such as target scintillation (Sec. 18.8) can cause false indications of tracking error. The undesired fluctuations that cause difficulty occur at about the same rate as the scan rate. Since target scintillation energy of aircraft is concentrated in the lower frequency range below approximately 100 Hz (particularly the troublesome propeller modulation), it is desirable to increase the scan rate as high as possible. The maximum practical rate is one-fourth of the pulse repetition frequency (PRF) so that four pulses provide a complete scan with one each up, down, right, and left. The maximum PRF and, consequently, maximum scan rate are limited by the maximum range of targets to be tracked by the radar. At a PRF of 1000 Hz the unambiguous range extends to about 80 nmi (at this range the echo is returning at the time when the radar is ready for its next transmission). The radar can track beyond this range by using nth-time-around tracking as described in Sec. 18.5.

High scan rates are difficult to achieve with mechanical scanning devices in a large antenna, and a variety of techniques to scan electronically have been used. In a small antenna such as that used in missile-homing heads, the dish rather than the feed may be tilted and rotated at high revolutions per minute (r/min) to achieve high scan rates. Scan rates of hundreds of r/min are frequently used—in some instances, as high as 2400 r/min, as in the AN/APN-58 target seeker. In the target seeker application the PRF can be high since the target is at short range. The coming of the jet aircraft caused additional problems for lobing systems because jet turbines cause significant modulations at high frequencies in regions near the maximum practical mechanical or electronic lobing rates. A further problem in scanning and lobing systems is a limitation on long-range tracking. At long ranges the time required for the radar signal to travel to the target and back becomes a significant portion of a scan cycle. For example, at a 100-Hz scan rate and the target at 460 mi, a signal transmitted on an up lobe will return as an echo when the antenna is looking on a down lobe, canceling the effect of the scan and the angle-error-sensing capability. In applications where this effect is significant, compensation can be provided if the range to the target is measured.

18.3 MONOPULSE (SIMULTANEOUS LOBING)

The susceptibility of scanning and lobing techniques to echo amplitude fluctuations was the major reason for developing a tracking radar that provides simultaneously all the necessary lobes for angle-error sensing. The output from the lobes may be compared simultaneously on a single pulse, eliminating any effect of time change of the echo amplitude. The technique was initially called *simultaneous lobing*, which was descriptive of the original designs. Later the term *monopulse* was used, referring to the ability to obtain complete angle error information on a single pulse. It has become the commonly used name for this tracking technique.

The original monopulse trackers suffered in antenna efficiency and complexity of microwave components since waveguide signal-combining circuitry was a relatively

new art. These problems were overcome, and monopulse radar with off-the-shelf components can readily outperform scanning and lobing systems. The monopulse technique also has an inherent capability for high-precision angle measurement because its feed structure is rigidly mounted with no moving parts. This has made possible the development of pencil-beam tracking radars that meet missile-range instrumentation-radar requirements of 0.003° angle-tracking precision.

This chapter is devoted to tracking radar, but monopulse is used in other systems including homing devices, direction finders, and some search radars. However, most of the basic principles and limitations of monopulse apply for all applications. A more general coverage is found in Refs. 3 and 4.

Amplitude-Comparison Monopulse.   A method for visualizing the operation of an amplitude-comparison monopulse receiver is to consider the echo signal at the focal plane of an antenna.[5] The echo is focused to a "spot" having a cross-section shape approximately of the form $J_1(X)/X$ for circular apertures, where $J_1(X)$ is the first-order Bessel function. The spot is centered in the focal plane when the target is on the antenna axis and moves off center when the target moves off axis. The antenna feed is located at the focal point to receive maximum energy from a target on axis.

An amplitude-comparison monopulse feed is designed to sense any lateral displacement of the spot from the center of the focal plane. A monopulse feed using the four-horn square, for example, would be centered at the focal point. It provides a symmetry so that when the spot is centered equal energy falls on each of the four horns. However, if the target moves off axis, causing the spot to shift, there is an unbalance of energy in the horns. The radar senses the target displacement by comparing the amplitude of the echo signal excited in each of the horns. This is accomplished by use of microwave hybrids to subtract outputs of pairs of horns, providing a sensitive device that gives signal output when there is an unbalance caused by the target being off axis. The RF circuitry for a conventional four-horn square (Fig. 18.8) subtracts the output of the left pair from the output of the right pair to sense any unbalance in the azimuth direction. It also subtracts the output of the top pair from the output of the bottom pair to sense any unbalance in the elevation direction.

The Fig. 18.8 comparator is the circuitry which performs the addition and subtraction of the feedhorn outputs to obtain the monopulse sum and difference signals. It is illustrated with hybrid-T or magic-T waveguide devices. These are four-port devices which, in basic form, have the inputs and outputs located at right angles to each other. However, the magic T's have been developed in convenient "folded" configurations for very compact comparator packages. The performance of these and other similar four-port devices is described in Ref. 3, Chap. 4.

The subtractor outputs are called difference signals, which are zero when the target is on axis, increasing in amplitude with increasing displacement of the target from the antenna axis. The difference signals also change 180° in phase from one side of center to the other. The sum of all four horn outputs provides a reference signal to allow angle-tracking sensitivity (volts per degree error) even though the target echo signal varies over a large dynamic range. AGC is necessary to keep the gain of the angle-tracking loops constant for stable automatic angle tracking.

Figure 18.9 is a block diagram of a typical monopulse radar. The sum signal, elevation difference signal, and azimuth difference signal are each converted to intermediate frequency (IF), using a common local oscillator to maintain relative phase at IF. The IF sum-signal output is detected and provides the video input to
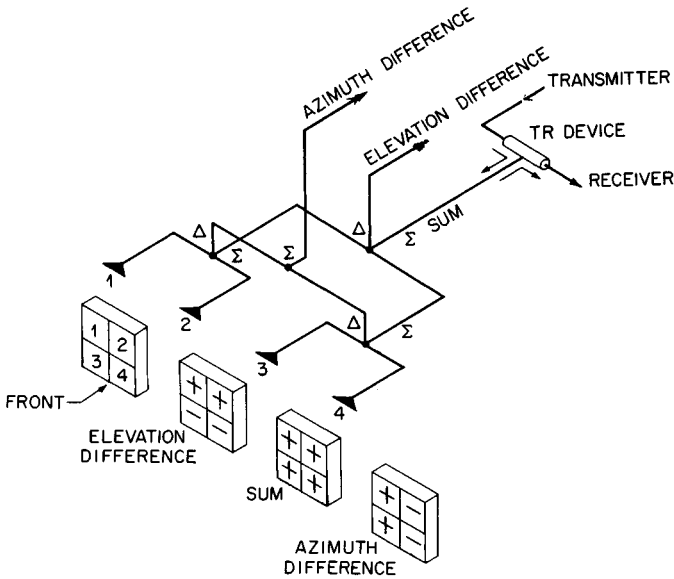
FIG. 18.8 Microwave-comparator circuitry used with a four-horn monopulse feed.
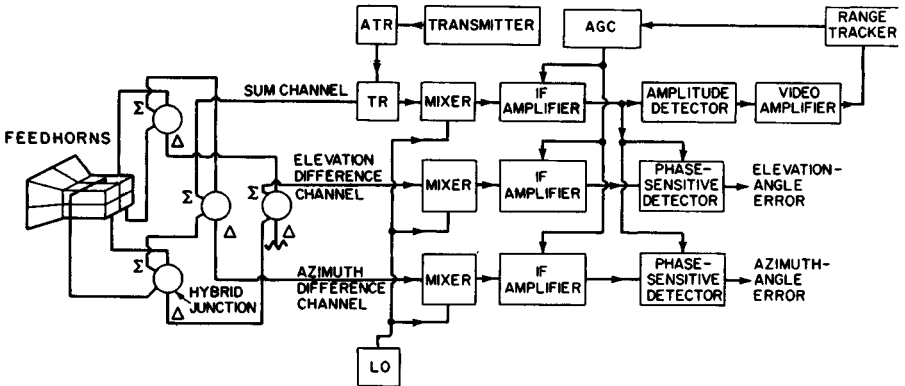
FIG. 18.9 Block diagram of a conventional monopulse tracking radar.

the range tracker. The range tracker determines the time of arrival of the desired target echo and provides gate pulses which turn on portions of the radar receiver only during the brief period when the desired target echo is expected. The gated video is used to generate the dc voltage proportional to the magnitude of the Σ signal or $|\Sigma|$ for the AGC of all three IF amplifier channels. The AGC maintains constant angle-tracking sensitivity (volts per degree error) even though the target echo signal varies over a large dynamic range by controlling gain or dividing by $|\Sigma|$. AGC is necessary to keep the gain of the angle-tracking loops constant for stable automatic angle tracking. Some monopulse systems, such as the two-

channel monopulse, can provide instantaneous AGC or normalizing as described later in this section.

The sum signal at the IF output also provides a reference signal to phase detectors which derive angle-tracking-error voltages from the difference signal. The phase detectors are essentially a dot-product device producing the output voltage

$$e = \frac{|\Sigma|\,|\Delta|}{|\Sigma|\,|\Sigma|} \cos \theta \qquad \text{or} \qquad e = \frac{\Delta}{|\Sigma|} \cos \theta$$

where e = angle-error-detector output voltage

$|\Sigma|$ = magnitude of sum signal

$|\Delta|$ = magnitude of difference signal

θ = phase angle between sum and difference signals

The dot-product error detector is only one of a wide variety of monopulse angle error detectors described in Ref. 3, Chap. 7.

Normally, θ is either 0° or 180° when the radar is properly adjusted, and the only purpose of the phase-sensitive characteristic of the detector is to provide a plus or minus polarity corresponding to $\theta = 0°$ and $\theta = 180°$, respectively, giving direction sense to the angle-error-detector output.

In a pulsed tracking radar the angle-error-detector output is bipolar video; that is, it is a video pulse with an amplitude proportional to the angle error and whose polarity (positive or negative) corresponds to the direction of the error. This video is typically processed by a boxcar circuit which charges a capacitor to the peak video-pulse voltage and holds the charge until the next pulse, at which time the capacitor is discharged and recharged to the new pulse level. With moderate low-pass filtering, this gives a dc error voltage output employed by the servo amplifiers to correct the antenna position.

The three-channel amplitude-comparison monopulse tracking radar is the most commonly used monopulse system. The three signals may sometimes be combined in other ways to allow use of a two-channel or even a single-channel IF system as described later in this section.

Monopulse-Antenna Feed Techniques.   Monopulse-radar feeds may have any of a large variety of configurations. For two-angle tracking such as azimuth and elevation, the feeds may include three or more apertures.[6] Single apertures are also employed by using higher-order waveguide modes to extract angle-error-sensing difference signals. There are many tradeoffs in feed design because optimum sum and difference signals, low sidelobe levels, omnipolarization capability, and simplicity cannot all be fully satisfied simultaneously. The term *simplicity* refers not only to cost saving but also to the use of noncomplex circuitry which is necessary to provide a broadband system with good boresight stability to meet precision-tracking requirements. (Boresight is the electrical axis of the antenna or the angular location of a signal source within the antenna beam at which the angle-error-detector outputs go through zero.)

Some of the typical monopulse feeds are described to show the basic relations involved in optimizing the various performance factors and how the more important factors can be optimized by a feed configuration but at the price of lower performance in other areas. Many new techniques have been added since the original four-horn square feed in order to provide good or excellent performance in all desired feed characteristics in a well-designed monopulse radar.

The original four-horn square monopulse feed is inefficient since the optimum feed size in the plane of angle measurement for the difference signals is approx-
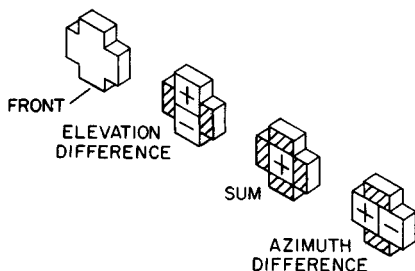
FIG. 18.10 Approximately ideal feed-aperture E-field distribution for sum and difference signals.

imately twice the optimum size for the sum signal.[7] Consequently, an intermediate size is typically used with a significant compromise for both sum and difference signals. The optimum four-horn square feed, which is subject to this compromise, is described in Ref. 3 as based on minimizing the angle error caused by receiver thermal noise. However, if sidelobes are a prime consideration, a somewhat different feed size may be desired.

The limitation of the four-horn square feed is that the sum- and difference-signal E fields cannot be controlled independently. If independent control could be provided, the *ideal* would be approximately as described in Fig. 18.10 with twice the dimension for the difference signals in the plane of error sensing than that for the sum signal.[7]

A technique used by the MIT Lincoln Laboratory to approach the ideal was the 12-horn feed (Fig. 18.11). The overall feed, as illustrated, is divided into small parts and the microwave circuitry selects the portions necessary for the sum and difference signals to approach the ideal. One disadvantage is that this feed requires a very complex microwave circuit. Also, the divided four-horn portions of the feed are each four element arrays which generate large feed sidelobes in the H plane because of the double-peak E field. Another consideration is that the 12-horn feed is not practical for focal-point-fed parabolas or reflectarrays because of its size. A focal-point feed is usually small to produce a broad pattern and must be compact to avoid blockage of the antenna aperture. In some cases the small size required is below waveguide cutoff, and dielectric loading becomes necessary to avoid cutoff.

A more practical approach to monopulse-antenna feed design uses higher-order waveguide modes rather than multiple horns for independent control of sum- and difference-signal E fields. This allows much greater simplicity and flexibility. A triple-mode two-horn feed used by RCA[7,8] retracts the E-plane septa to allow both the TE_{10} and TE_{30} modes to be excited and propagate in the double-width septumless region as illustrated in Fig. 18.12. At the septum the double-humped E field is represented by the combined TE_{10} and TE_{30} modes subtracting at the center and adding at the TE_{30}-mode outer peaks. However, since the two modes propagate at different velocities, a point is reached farther down the double-width guide where the two modes add in the center and subtract at the outer humps of the TE_{30} mode. The result is a sum-signal E field concentrated toward the center of the feed aperture.

This shaping of the sum-signal E field is accomplished independently of the difference-signal E field. The difference signal is two TE_{10}-mode signals arriving at the septum of Fig. 18.12 out of phase. At the septum it becomes the TE_{20} mode, which propagates to the horn aperture and uses the full width of the horn as desired. The TE_{20} mode has zero E field in the center of the waveguide where the septum is located and is unaffected by the septum.

The AN/FPS-16 radar feed used two retracted septum horns illustrated in Fig. 18.13. The TE_{20}-mode signals are added for the H-plane difference signal, the combined TE_{10} and TE_{30} modes are added for the sum signal, and they are subtracted for the E-plane difference signal. Since this is a focal-point feed, it is small in size (wavelengths) and RF currents tend to flow around the top and bottom
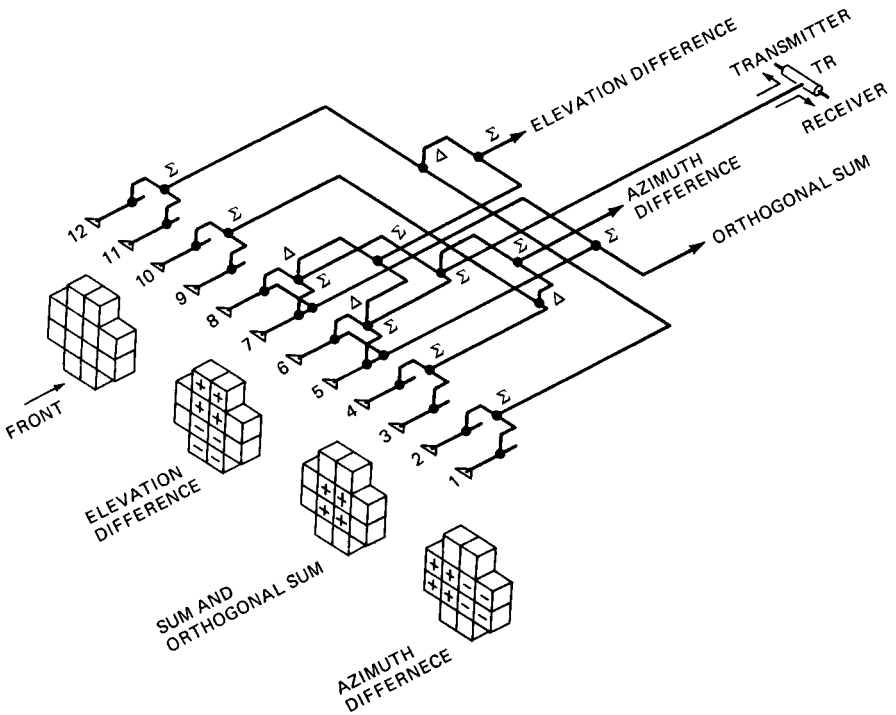
FIG. 18.11 Twelve-horn feed.

edges at the E-field peaks in the middle of the horns. This results in the need for the top and bottom matching stubs seen in Fig. 18.13.

A further step in feed development is the four-horn triple-mode feed illustrated in Fig. 18.14.[7] This feed uses the same approach as described above but with the addition of a top and bottom horn. This allows the E-plane difference signal to couple to all four horns and uses the full height of the feed. The sum signal uses only the center two horns to limit its E field in the E plane as desired for the ideal field shaping. The use of smaller top and bottom horns is a simpler method of concentrating the E field toward the center of the feed, where the full horn width is not needed.

The feeds described thus far are for linear-polarization operation. When circular polarization is needed in a paraboloid-type antenna, square or circular cross-section horn throats are used. The vertical and horizontal components from each horn are separated and comparators provided for each polarization. The sum and difference signals from the comparators are combined with 90° relative phase to obtain circular polarization. Use of the previously described feeds for circular polarization would require the waveguide circuitry to be prohibitively complex. Consequently, a five-horn feed is used as illustrated in Fig. 18.15.

The five-horn feed is selected because of the simplicity of the comparator which requires only two magic (or hybrid) T's for each polarization. The sum and difference signals are provided for the two linear-polarization components and, in an AN/FPQ-6 radar, are combined in a waveguide switch for selecting
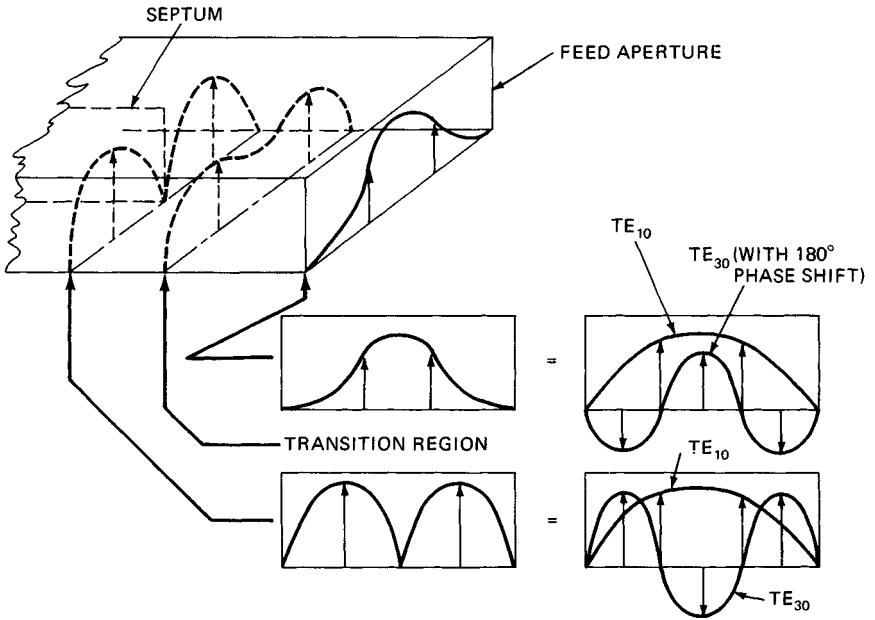
FIG. 18.12 Use of retracted septum to shape the sum-signal E field.
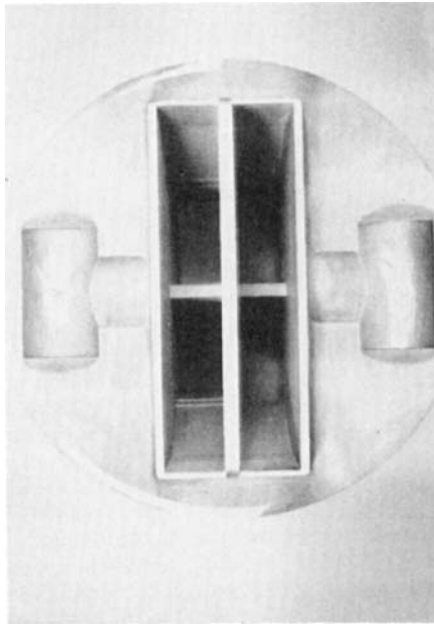
FIG. 18.13 AN/FPS-16 feed, front view. (*From S. M. Sherman, Ref. 3.*)
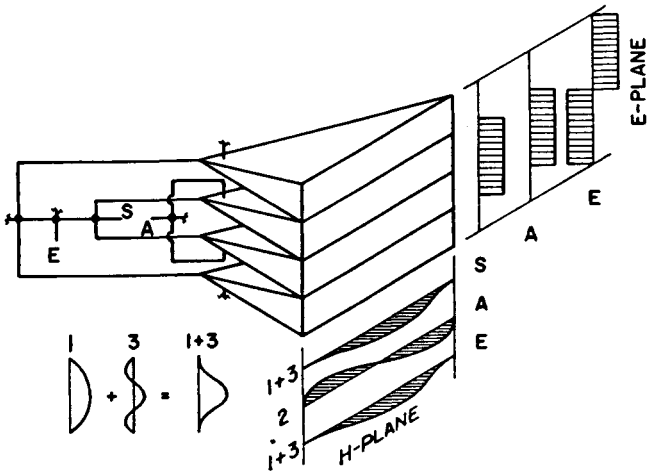
FIG. 18.14 Four-horn triple-mode feed. (*From P. W. Hannan, Ref. 7; copyright 1961, IEEE.*)
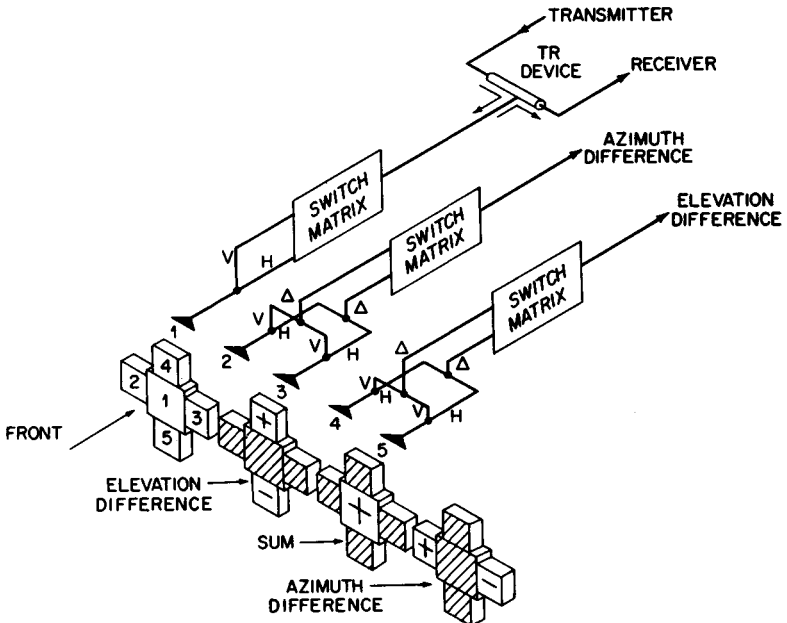
FIG. 18.15 Five-horn feed with coupling to both linear-polarization components, which are combined by the switch matrix to select horizontal, vertical, or circular polarization.

polarization. The switch selects either the vertical or the horizontal input component or combines them with a 90° relative phase for circular polarization. This feed does not provide optimum sum- and difference-signal E fields because the sum horn occupies space desired for the difference signals. Generally an undersized sum-signal horn is used as a compromise. However, the five-horn feed is a

practical choice between complexity and efficiency. It has been used in several instrumentation radars including the AN/FPQ-6, AN/FPQ-10, AN/TPQ-18, and AN/MPS-36[9,10] and in the AN/TPQ-27 tactical precision-tracking radar.[11]

The multimode feed techniques can be expanded to other higher-order modes for error sensing and E-field shaping.[12,13,14] The difference signals are contained in unsymmetrical modes such as the TE_{20} mode for H-plane error sensing and combined TE_{11} and TM_{11} modes for E-plane error sensing. These modes provide the difference signals, and no comparators, as shown in Fig. 18.8, are used.[12] Generally, mode-coupling devices can give good performance in separating the symmetrical and unsymmetrical modes without significant cross-coupling problems.

Multiband monopulse feed configurations are practical and in use in several systems. A simple example is a combined X-band and K_a-band monopulse paraboloid antenna radar. Separate conventional feeds are used for each band, with the K_a-band feed as a Cassegrain feed and the X-band feed at the focal point.[15] The Cassegrain subdish is a hyperbolic-shaped grid of wires reflective to parallel polarization and transparent to orthogonal polarization. It is oriented to be transparent to the X-band focal-point feed behind it and reflective to the orthogonally polarized K_a-band feed extending from the vertex of the paraboloid.

Monopulse feed horns at different microwave frequencies can also be combined with horns interlaced. The multiband feed clusters will sacrifice efficiency but can satisfy multiband requirements in a single antenna.

AGC (Automatic Gain Control). To maintain a stable closed-loop servosystem for angle tracking, the radar must maintain essentially constant loop gain independent of target size and range. The problem is that monopulse difference signals from the antenna are proportional to both the angle displacement of the target from the antenna axis and the echo signal amplitude. For a given tracking error, the error voltage would change with echo amplitude and cause a corresponding change in loop gain.

AGC is used to remove the angle-error-detector-output dependence on echo amplitude and retain constant tracking loop gain. A typical AGC technique is illustrated in Fig. 18.16 for a one-angle coordinate tracking system. The AGC system detects the peak voltage of the sum signal and provides a negative dc voltage proportional to the peak signal voltage. The negative voltage is fed to the IF amplifier stages, where it is used to decrease gain as the signal increases. A high gain in the AGC loop is equivalent to dividing the IF output by a factor proportional to its amplitude.
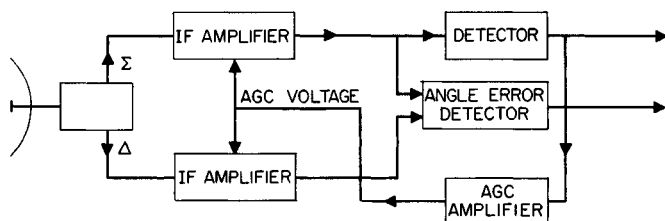
FIG. 18.16   AGC in monopulse tracking.

In a three-channel monopulse radar, all three channels are controlled by the AGC voltage, which effectively performs a division by the magnitude of the sum

signal or echo amplitude. Conventional AGC with a control voltage is band-limited by filters, and the gain is essentially constant during the pulse repetition interval. Also, the AGC of the sum channel normalizes the sum echo pulse amplitude to similarly maintain a stable range-tracking servo loop.

The angle-error detector, assumed to be a product detector, has an output

$$|e| = k \, \frac{\Delta \, \Sigma}{|\Sigma||\Sigma|} \cos \theta$$

where $|e|$ is the magnitude of the angle error voltage. Phases are adjusted to provide 0 or 180° on a point-source target. The resultant is

$$|e| = \pm \, k \, \frac{\Delta}{|\Sigma|}$$

Complex targets can cause other phase relations as a part of the angle scintillation phenomenon.[3] The above error voltage proportional to the ratio of the difference signal divided by the sum signal is the desired angle-error-detector output, giving a constant angle error sensitivity.[3]

With limited AGC bandwidth, some rapid signal fluctuations modulate $|e|$, but the long-time-average angle sensitivity is constant. These fluctuations are largely from rapid changes in target reflectivity, $\sigma(t)$, that is, from target amplitude scintillation. The random modulation of $|e|$ causes an additional angle noise component that affects the choice of AGC bandwidth.

At very low signal-to-noise ratios ($SNR < 4$ dB) the AGC voltage is limited to a minimum value by the noise level. Therefore, as the signal decreases into the noise, the IF gain remains constant and the resultant angle sensitivity decreases. Consequently, the effect of thermal noise on tracking performance differs from the linear relations, which are accurate to within about 1 dB for an SNR of 4 dB or greater. Reference 2 discusses means for calculating thermal-noise effects for very low SNR conditions.

The AGC performance in conical-scan radars provides a similar constant angle error sensitivity. One major limitation in conical-scan radars is that the AGC bandwidth must be sufficiently lower than the scan frequency to prevent the AGC from removing the modulation containing the angle error information. The very low SNR effects on conical scan differ from the effects on monopulse, as discussed in detail in Ref. 2.

Phase-Comparison Monopulse.   A second monopulse technique is the use of multiple antennas with overlapping (nonsquinted) beams pointed at the target. Interpolating target angles within the beam is accomplished, as shown in Fig. 18.17, by comparing the phase of the signals from the antennas (for simplicity a single-coordinate tracker is described). If the target were on the antenna boresight axis, the outputs of each individual aperture would be in phase. As the target moves off axis in either direction, there is a change in relative phase. The amplitudes of the signals in each aperture are the same so that the output of the angle error phase detector is determined by the relative phase only. The phase-detector circuit is adjusted with a 90° phase shift in one channel to give zero output when the target is on axis and an output increasing with increasing angular displacement of the target with a polarity corresponding to the direction of error.[3,4]

Typical flat-face corporate-fed phased arrays compare the output of halves of
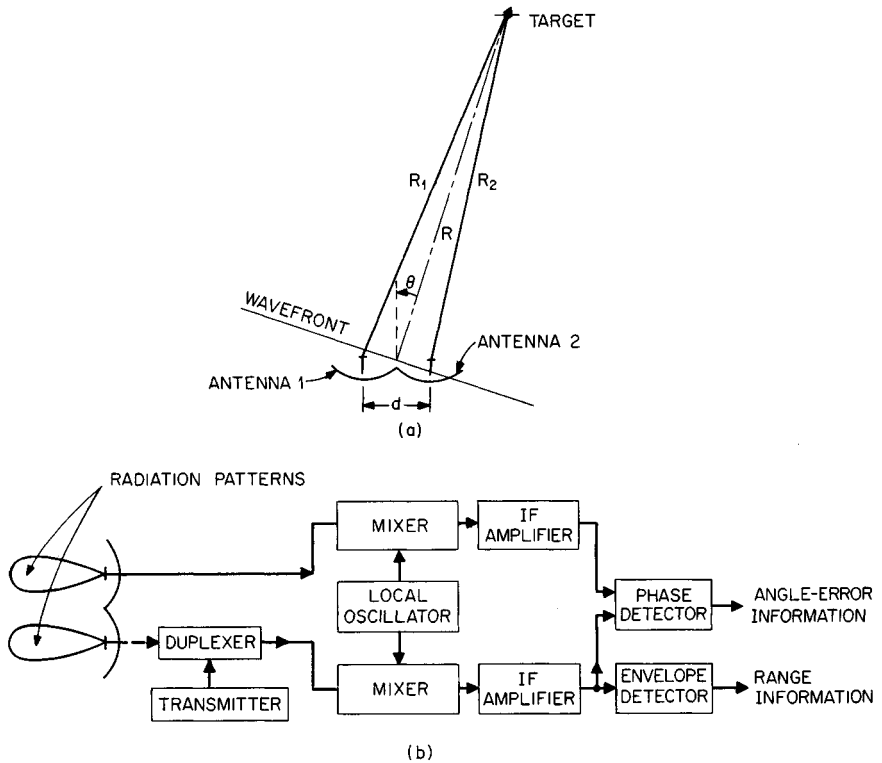
FIG. 18.17 (a) Wavefront phase relationships in a phase comparison monopulse radar. (b) Block diagram of a phase comparison monopulse radar (one angle coordinate).

the aperture and fall into the class of phase-comparison monopulse. However, the basic performance of amplitude- and phase-comparison monopulse is essentially the same.[3]

Figure 18.17 shows the antenna and receiver for one angular-coordinate tracking by phase-comparison monopulse. Any phase shifts occurring in the mixer and IF amplifier stages cause a shift in the boresight of the system. The disadvantages of phase-comparison monopulse compared with amplitude-comparison monopulse are the relative difficulty in maintaining a highly stable boresight and the difficulty in providing the desired antenna illumination taper for both sum and difference signals. The longer paths from the antenna outputs to the comparator circuitry make the phase-comparison system more susceptible to boresight change due to mechanical loading or sag, differential heating, etc.

A technique giving greater boresight stability combines the two antenna outputs at RF with passive circuitry to yield sum and difference signals, as shown in Fig. 18.18. These signals may then be processed as in a conventional amplitude-comparison monopulse receiver. The system shown in Fig. 18.18 would provide a relatively good difference-channel taper, having smoothly tapered E fields on each antenna. However, a sum-signal excitation with the two antennas provides a two-hump in-phase E-field distribution which causes high sidelobes since it looks like a two-element array. This problem may be reduced by allowing some

aperture overlap but at the price of loss of angle sensitivity and antenna gain.

Monopulse Tracking with Phased Arrays. In general, phased-array tracking radars fall in either the amplitude- or the phase-comparison class, depending on the feed technique. Feedthrough lens arrays acting as an RF lens and reflectarrays acting like a parabolic reflector may use any of the described multihorn or multimode feeds, and the same general factors in optimizing a feed apply. Monopulse angle-error sensing can be accomplished in a corporate-feed array by using the two halves of the array (the top and bottom halves can also be used for elevation) as a phase-comparison tracker. Array antennas which use the two halves of the aperture for phase-comparison angle error sensing generally provide a good taper for the sum pattern, but the difference-signal E field across the array reaches a peak toward the center with a sudden 180° phase change. This sharp discontinuity at a maximum amplitude point causes undesired high sidelobes. Techniques such as use of separate feeds can provide the desired shaping of the difference-signal E-field distribution.

FIG. 18.18 (a) RF phase-comparison monopulse system with sum and difference outputs. (b) Vector diagram of the sum and difference signals.

The monopulse electronic-scan phased array is used in instrumentation radar to meet requirements for simultaneously tracking multiple targets.[16–18] An example is the Multiobject Tracking Radar (MOTR), AN/TPQ-39, built for the White Sands Missile Range for high-precision tracking. The MOTR provides a 60° cone of electronic pulse-to-pulse beam coverage plus mechanical pedestal movement to cover the hemisphere. High accuracy and efficiency can be maintained over the ± 30° electronic-scan coverage, and this coverage moves mechanically as needed to optimally move the electronic-scan coverage along with the target configuration to be tracked.[14,19]

One- and Two-Channel Monopulse. Monopulse radars may be constructed with fewer than the conventional three IF channels. This is accomplished by combining the sum and difference signals by some means so that they may be individually retrieved at the output. These techniques provide some advantages in AGC or other processing techniques but at the cost of *SNR* loss or of cross coupling between azimuth and elevation information.

A single-channel monopulse system called SCAMP (single-channel monopulse processor)[20] provides the desired constant angle error sensitivity by normalizing the difference signals with the sum signal in a single IF channel, as shown in Fig. 18.19. The signals are each converted from RF to different IF frequencies by separate local oscillators (LOs) of different frequencies for each signal. They are amplified in a single IF amplifier of sufficient bandwidth for all three signals at different frequencies. At the IF output the signals are hard-limited and separated by three narrowband filters. The signals are then converted to the same frequency
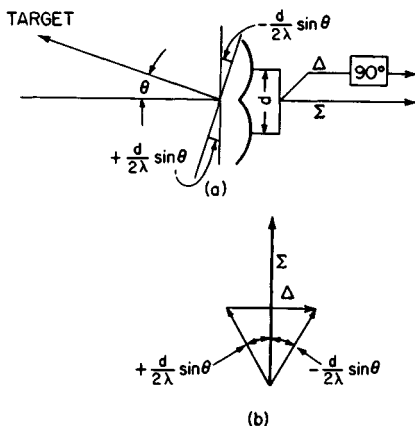
by beating two of the signals with the frequency difference between their LOs and the LO of the third signal. The angle-error voltage is then determined by either a conventional phase detector or simply an amplitude detector.[14] The effect of AGC action and normalizing is performed by hard limiting which causes a weak-signal suppression of the difference signal similar to the effects of hard limiting on a weak signal in noise.[20]
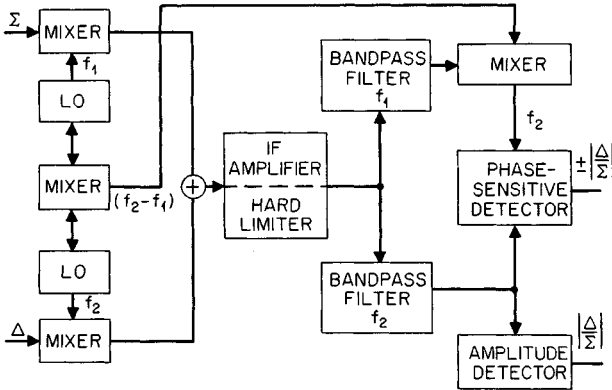
FIG. 18.19 Block diagram of SCAMP, a single-channel monopulse tracking system, demonstrating angle tracking in one angle coordinate. (The system is capable of tracking in both coordinates.) (*From W. L. Rubin and S. K. Kamen, Ref. 20.*)

The single-channel monopulse provides in effect an instantaneous AGC. Performance in the presence of thermal noise is about equal to that of the three-channel monopulse. However, the limiting process generates a significant cross-coupling problem,[21] causing a portion of the azimuth-error signal to appear in the elevation-angle-error-detector output and elevation error to appear in the azimuth-angle-error-detector output. Depending on the receiver configuration and the choice of IF frequencies, the cross modulation could cause serious errors and allow vulnerability to jamming. Reference 21 describes how the sum- and difference-channel bands in the wideband IF can be arranged to minimize cross coupling and suggests use of narrow banding to separate each signal before limiting in order to reduce jamming effects.

A two-channel monopulse receiver[22] may also be used by combining the sum and difference signals at RF, as shown in Fig. 18.20. The microwave resolver is a mechanically rotated RF coupling loop in circular waveguide. The azimuth and elevation difference signals are excited in this guide with E-field polarization oriented at 90°. The energy into the coupler contains both difference signals coupled as the cosine and sine of the angular position of the coupler, $\omega_s t$, where ω_s is the angular rate of rotation. The hybrid adds the combined difference signals Δ to the sum signal Σ. The $\Sigma + \Delta$ and $\Sigma - \Delta$ outputs each look like the output of a conical-scan tracker except that the modulation function differs by 180°. In case of failure of one channel, the radar can be operated as a scan-on-receive-only conical-scan radar with essentially the same performance as a conical-scan radar. The advantage of two channels with opposite-sense angle-error information on one with respect to the other is that signal fluctuations in the received signal are
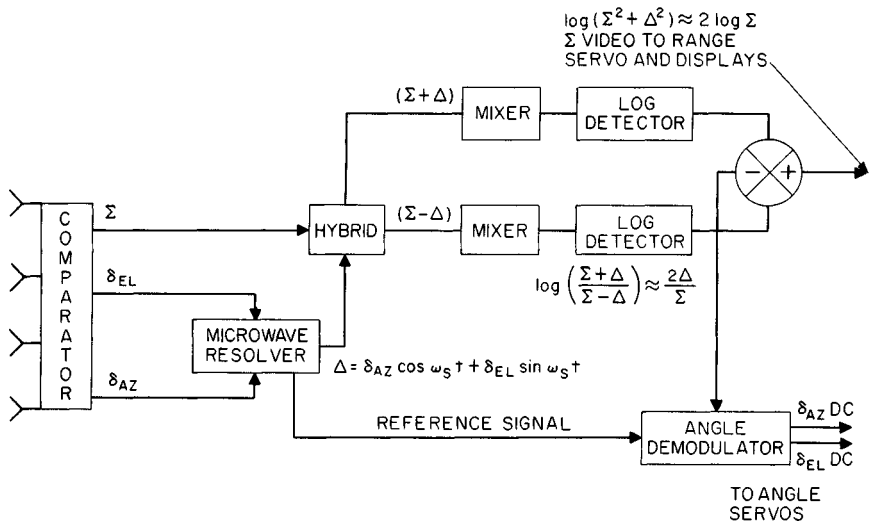
FIG. 18.20 Block diagram of a two-channel monopulse radar system. (*From R. S. Noblit, Ref. 22.*)

canceled in the postdetection subtractor at the IF output which retrieves the angle-error information. The log IF performs essentially as instantaneous AGC, giving the desired constant angle-error sensitivity of the difference signal normalized by the sum signal. The detected Δ information is a bipolar video where the error information is contained in the sinusoidal envelope. This signal is separated into its two components, azimuth- and elevation-error information, by an angle demodulation. The demodulator, using a reference from the drive on the rotating coupler, extracts the sine and cosine components from Δ to give the azimuth- and elevation-error signals. The two-channel monopulse technique is used in the AN/SPG-55 tracking radar and the AN/FPQ-10 missile-range instrumentation radar. The modulation caused by the microwave resolver is of concern in instrumentation radar applications because it adds spectral components in the signal which complicate the possible addition of pulse doppler tracking capability to the radar.

This system provides instantaneous AGC operation with only two IF channels and operation with reduced performance in case of failure of either channel. However, there is a loss of 3 dB *SNR* at the receiver inputs although this loss is partly regained by coherent addition of the Σ-signal information. The design of the microwave resolver must minimize loss through the device, and high precision is required to minimize cross coupling between azimuth and elevation channels. The resolver performance is improved by the use of ferrite switching devices to replace the mechanical rotating coupler.

Conopulse. Conopulse (also called scan with compensation) is a radar tracking technique that is a combination of monopulse and conical scan.[23,24] A pair of antenna beams is squinted in opposite directions from the antenna axis and rotated like a conical-scan-radar beam scan. Since they exist simultaneously, monopulse information can be obtained from the pair of beams. The plane in which monopulse information is measured rotates. Consequently,

elevation and azimuth information is sequential and must be separated for use in each tracking coordinate. Conopulse provides the monopulse advantage of avoiding errors caused by amplitude scintillation, and it requires only two receivers. However, it has the disadvantage over conventional monopulse radar of lower angle data rate and the mechanical complexity of providing and coupling to a pair of rotating antenna feedhorns.

18.4 SERVOSYSTEMS FOR TRACKING RADAR

The servosystem of a tracking radar is the portion of the radar that receives as its input the tracking-error voltage and performs the task of moving the antenna beam in a direction that will reduce to zero the alignment error between the antenna axis and the target. For two-angle tracking with a mechanical-type antenna there are typically separate axes of rotation for azimuth and elevation and separate servosystems to move the antenna about each axis. A conventional servosystem is composed of amplifiers, filters, and a motor that moves the antenna in a direction to maintain the antenna axis on the target. Range tracking is accomplished by a similar function to maintain range gates centered on the received-echo pulses. This may be accomplished by analog techniques or by digital-counter registers that retain numbers corresponding to target range to provide a closed tracking loop digitally rather than mechanically.

Servosystems may contain hydraulic-drive motors, conventional electric motors geared down to drive the antenna, or direct-drive electric motors where the antenna mechanical axis shaft is part of the armature and the motor field is built into the supporting case. The direct drive is heavier for a given horsepower but eliminates gear backlash. The conventional motors may be provided in a duplicate drive with a small residual opposing torque to reduce backlash. Amplifier gain and filter characteristics as well as motor torque and inertia determine the velocity and acceleration capability or the ability to follow higher-order motions of the target.

It is desired that the antenna beam follow the center of the target as closely as possible, which implies that the servosystem should be capable of moving the antenna quickly. The combined velocity and acceleration characteristics of a servosystem can be described by the frequency response of the tracking loop, which is essentially a low-pass filter characteristic. Increasing the bandwidth increases the quickness of the servosystem and its ability to follow closely a strong, steady signal. However, a typical target causes scintillation of the echo signal, giving erroneous error-detector outputs, and at long range the echo is weak, allowing receiver noise to cause additional random fluctuations in the error-detector output. Consequently, a wide servo bandwidth which reduces lag errors allows the noise to cause erroneous motions of the tracking system. Therefore, for best overall performance it is necessary to limit the servo bandwidth to the minimum necessary to maintain a reasonably small tracking-lag error. There is an optimum bandwidth that minimizes the rms of the total erroneous outputs including both tracking lag and random noise, depending upon the target, its trajectory, and other radar parameters.

The optimum bandwidth for angle tracking is range-dependent. A target with typical velocity at long range has low angle rates and a low SNR, and a narrower servo passband will follow the target with reasonably small tracking lag while minimizing the response to receiver thermal noise. At close range the signal is

strong, overriding receiver noise, but target angle scintillation errors proportional to the angular span of the target are large. A wider servo bandwidth is needed at close range to keep tracking lag within reasonable values, but it must not be wider than necessary or target scintillation errors become excessive.

The low-pass closed-loop characteristic of a servosystem is unity at zero frequency, typically remaining near this value up to a frequency near the low-pass cutoff, where it may peak up to higher gain, as shown in Fig. 18.21. The peaking is an indication of system instability but is allowed to be as high as tolerable, typically up to about 3 dB above unity gain to obtain maximum bandwidth for a given servomotor drive system. System A of Fig. 18.21 is a case of excessive peaking of about 8 dB. The effect of the peaking is observed by applying a step error input to the servosystem. The peaking of the low-pass characteristic results in an overshoot when the antenna axis moves to align with the target. High peaking causes a large overshoot and a return toward the target with an additional overshoot. In the extreme, as in system A of Fig. 18.21, the antenna zeros in on the target with a damped oscillation. An optimum system compromise between speed of response and overshoot, as in system B, allows the antenna to make a small overshoot with a reasonably rapid exponential movement back to the target. This corresponds to about 1.4-dB peaking of the closed-loop low-pass characteristic.
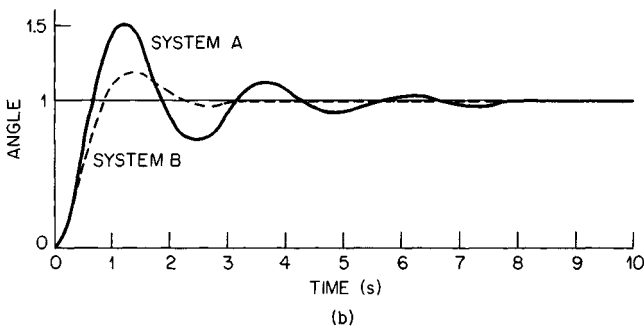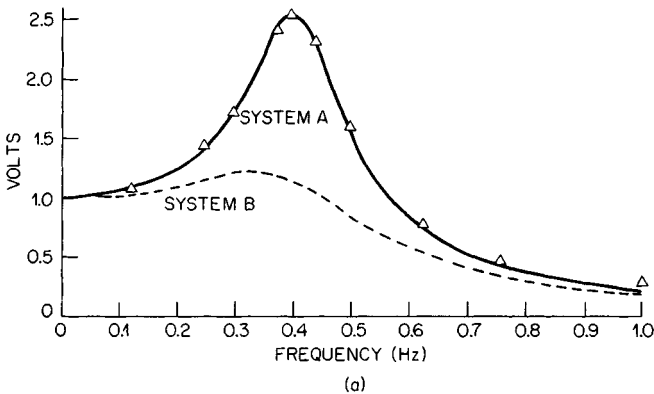
FIG. 18.21 (*a*) Closed-loop frequency-response characteristics of two servosystems. (*b*) Their corresponding time response to a step input.

To maximize the servo closed-loop bandwidth for a given motor drive system, tachometer feedback must be used. The tachometer provides a negative feedback voltage proportional to the servomotor speed. As the servomotor rotates the antenna toward a target, the error voltage decreases, dropping to zero when the antenna axis arrives at the target. However, the tachometer feedback provides a retarding motor torque opposing the system inertia that causes the overshoot, thus reducing the overshoot.

Resonances of the antenna and servosystem structure (the structure foundation is one of the most critical items) must be kept well above the bandwidth of the servosystem; otherwise the system can oscillate at the resonant frequency. A factor of at least 10 is desirable for the ratio of system resonant frequency to servo bandwidth. The high resonant frequency is difficult to obtain with a large antenna, such as the AN/FPQ-6 radar with a 29-ft dish, because of the large mass of the system. The ratio was pushed to a very minimum of about 3 to obtain a servo bandwidth of about 3.5 Hz. A smaller radar with a 12-ft dish, for example, can provide a servo bandwidth up to 7 or 8 Hz with conventional design.

A convenient method for calculating tracking error for a given target trajectory and servosystem is the use of the equation[25]

$$e(t) = \frac{\dot{\theta}(t)}{K_v} + \frac{\ddot{\theta}(t)}{K_a} - \frac{\dddot{\theta}(t)}{K_j} - \cdots$$

where $e(t)$ = tracking error as function of time
   $\dot{\theta}(t)$ = angular velocity of target relative to radar versus time
   $\ddot{\theta}(t)$ = angular acceleration
   $\dddot{\theta}(t)$ = next higher order of angular motion (descriptively called *jerk*)
   K_v = velocity constant of servosystem
   K_a = acceleration constant
   K_j = jerk constant of servosystem

A servosystem could be specified by selecting values for the three constants and, with a known target trajectory $\theta(t)$, the lag error calculated by using the
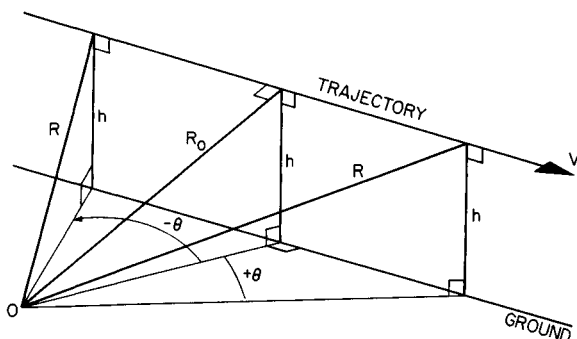
FIG. 18.22 Target trajectory on a passing course. (*From A. S. Locke, Ref. 25.*)

above equation. An example trajectory is shown in Fig. 18.22 for an aircraft flying a straight course past the radar with a minimum distance R_0 at altitude h. The figure shows that minimum range and maximum elevation angle occur when the radar is looking normal to the target path. The azimuth angle θ, with minimum range as the $0°$ azimuth reference point, starts at about $-90°$ and ends at about $+90°$, with its maximum rate of change $\dot{\theta}(t)$ at minimum range. For target parameters of a velocity of 500 kn, crossover range of 1000 yd, and altitude of 1500 ft, the derivatives of azimuth angle are $\dot{\theta}_{max} = 18.6°/s$, $\ddot{\theta}_{max} = 4°/s^2$, and $\dddot{\theta}_{max} = 4.2°/s^3$. Choosing example servosystem constants $K_v = 100$, $K_a = 111$, and $K_j = 1111$ results in a tracking-lag function

$$e(t) = \frac{\dot{\theta}(t)}{100} + \frac{\ddot{\theta}(t)}{111} - \frac{\dddot{\theta}(t)}{1111} \cdots$$

By substituting values for the derivatives, a time plot of tracking lag may be determined as shown in Fig. 18.23.
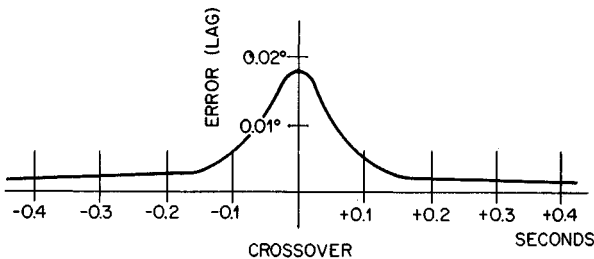
FIG. 18.23 Azimuth-tracking error for the passing-course target. (*From A. S. Locke, Ref. 25.*)

The range and elevation lag errors may be similarly determined by calculation of derivatives of $R(t)$ and $\phi(t)$, respectively, and use of the constants for these tracking systems. Elevation constants are similar to those for the azimuth system. The range tracker may be an inertialess electronic system with a double integration in the tracking loop. This is called a Type II system[25] with a $K_v = \infty$, causing velocity lag to be zero. The remaining significant lag components are the acceleration and jerk lags.

Electronically steerable arrays provide a means for inertialess angle tracking. However, because of this capability the system can track multiple targets by rapidly switching from one to another rather than continuously tracking a single target. The tracker simply places its beam at the location where the target is expected, corrects for the pointing error by converting error voltages (with a known angle error sensitivity) to units of angle, and moves to the next target. The system determines where the target was and, from calculations of target velocity and acceleration, predicts where it should be the next time the beam looks at the target. The lag error in this case is dependent on many factors, including the accuracy of the value of angle sensitivity used to convert error voltages to angular error, the size of the previous tracking error, and the time interval between looks.

CHAPTER 19
RADAR GUIDANCE OF MISSILES*

Alex Ivanov
Missile Systems Division
Raytheon Company

19.1 INTRODUCTION

Radar guided missiles represent one of the most widely used applications of the radar art, yet one about which much less has been published in the open literature than about other, more "conventional" radars. There is no nonmilitary use of this part of radar technology, and much of the detailed data is still classified. However, drawing solely on the unclassified published data permits at least a tutorial overview of radar guidance to be presented in this chapter.

Guided missiles can be characterized in several ways,[1-4] based on their mission, type of guidance, sensing wavelength, source of guidance energy, etc. The discussion here will narrow down to the particular radar homing types which form the vast majority of operational systems.

Based on their use, missile systems can be categorized as surface-to-surface, air-to-surface, surface-to-air, and air-to-air. The types of guidance are inertial, map-following, command, beam-riding, and homing. Types other than inertial can use the broadest range of the electromagnetic spectrum, from radio frequencies (RF) through infrared (IR) to the visible spectrum and beyond, to perform the guidance function.

Within these general categories, the surface-to-surface types [especially the intercontinental ballistic missile (ICBM) and the shorter-range ballistic] are usually inertially guided and fall outside the scope of this discussion. The primary exception is the antiship missile, which uses radar guidance and may be surface- (as well as air-) launched.[4,5] The main users of radar guidance are the air defense systems—surface-to-air or air-to-air. These also can employ IR or laser radars, but we shall restrict our discussion to the microwave radar category. Air-to-surface systems for use against ships, armored vehicles, or hard fixed targets

---

such as bridges, can use the visible spectrum (TV), IR, laser, or radar. Only the last-named category will be discussed.

Whether they are used against airborne or surface targets, guided missiles are intended to achieve a much higher accuracy than conventional artillery, which relies on open-loop prediction (even when optical or radar target tracking is employed for fire control). To achieve the required accuracy, a guidance system utilizes automatic closed-loop control by continuously sensing errors in the missile-to-target intercept geometry and translating them into corrective missile maneuvers designed to reduce miss distance to zero, although in practice a finite miss distance usually results.

Radar has been extensively used for command, beam-riding, and homing guidance. The simplest form of guidance is the beam rider. The target is tracked by a tracking radar (or, in early systems, by an operator using an optical sight with a radar slaved to it) which keeps the beam always pointed at the target. The missile itself does not perceive the target but detects its own position relative to the tracking beam. By keeping itself centered in the beam, it attempts, like the radar beam it rides, to pass through the target. In command guidance the target and the missile are tracked by separate radars (or by separate beams of a phased array radar). The missile itself does not perceive the target. Measured target and missile states are fed to a computer which calculates the missile trajectory required for intercept and develops the guidance commands which are continuously transmitted to the missile. In both these systems accuracy is inversely proportional to range from the radar, since a fixed angular error at the radar becomes an increasing linear error at increasing ranges.

Homing provides the highest accuracy at the cost of complexity of the missile-borne hardware. Whereas the beam rider and command systems require only a single receiver in the missile, to sense the beam or receive commands, the homing system perceives the target with its own radar (called the *seeker*), extracts tracking data from the received signal, and computes its own steering commands. As it closes on the target, a fixed angular error at the missile results in a decreasing linear error, providing the higher accuracy characteristic of homing guidance.

Homing systems can be further categorized on the basis of the source of the sensed radar energy into passive, semiactive, and active. Passive homing uses energy originating from the target (i.e., jamming or radar transmissions). An active homing system is a self-contained radar which transmits its own radar energy at the target and tracks the target-reflected energy. The semiactive system includes an external radar which illuminates the target while the missile receives and tracks the target-reflected energy to extract guidance information.

The waveforms used vary from noncoherent pulse (used in some early systems) to continuous wave (CW) and coherent pulse doppler (PD). The most widely used operational systems have, over the years, employed CW semiactive homing. Since the active systems differ only by virtue of the presence of the illuminator-transmitter on board the missile, a discussion of the semiactive system can be easily extended to cover the active type as well. Similarly, passive homing can be considered a subset of the semiactive.

Because of antenna size constraints, operating frequencies have generally been in the C, X, or K_u bands. The increased availability of components at higher frequencies has permitted operation at K_a and millimeter-wave frequencies in later-generation systems.

The nature of missile systems and the environment in which they are developed result in evolutionary changes rather than revolutionary innovations. Thus to understand today's systems we must understand how they became what they

are. This chapter will, therefore, begin with a discussion of the CW semiactive system and trace the evolution of the CW seeker through several generations. Once these concepts have been explained, they will be extended to the active and passive systems. Functional operations (i.e., acquisition, tracking) and characteristics of subsystems will then be discussed.

19.2 OVERVIEW OF SEMIACTIVE CW SYSTEMS[6]

The basic semiactive system is conceptually illustrated in Fig. 19.1. The illuminator maintains the target within its radar beam throughout the engagement. The missile receives the target-reflected illumination in its front antenna and a sample of the directly received illumination (often through sidelobes of the illuminator antenna) in its rearward-looking reference (rear) antenna. The front and rear signals are coherently detected against each other, resulting in a spectrum which contains the doppler-shifted target signal at a frequency roughly proportional to closing velocity. A narrowband frequency tracker searches the spectrum, locks onto the target return, and extracts guidance information from it. The use of CW provides the capability of discriminating against clutter on the basis of doppler frequency and thus allows low-altitude operation.

Doppler Frequency Relationships. The geometry for the doppler frequency relationships in a general semiactive case is shown in Fig. 19.2. The doppler shift is a function of the transmitted frequency f_0 and radial velocity V_R [the component of velocity along the line of sight (LOS) from the source to the observer—either a receiver or a reflector].

$$f_{\text{dop}} = (f_0/c)V_R$$

where c = velocity of light.

In the geometry of Fig. 19.2, the *rear*, or reference, doppler is a function of the radial velocity of the missile with respect to the illuminator. The *front*, or target, doppler frequency, depends on the radial velocities of the illuminator, missile, and target. The resulting spectrum, when the front signal is coherently detected
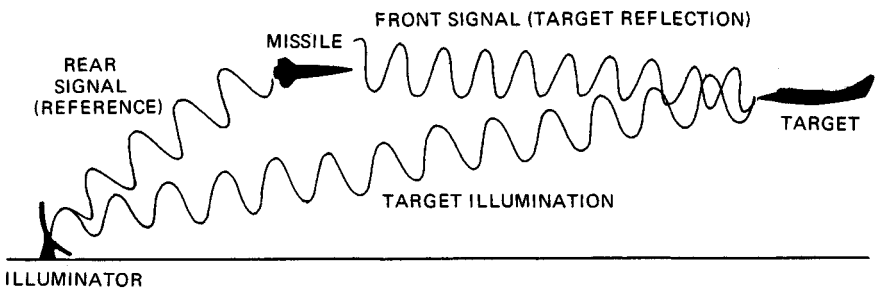
FIG. 19.1 Semiactive homing employs an illuminator to illuminate the target and provide a reference to the missile, which compares the reference with the reflected-target illumination to extract guidance information.

(mixed) against the rear, is the difference of the two.

$$f_{\text{rear}} = -\frac{f_0}{c}V_M \cos \theta + \frac{f_0}{c}V_I \cos A$$

$$f_{\text{front}} = \frac{f_0}{c}(V_T \cos \phi + V_T \cos \beta + V_M \cos \alpha + V_I \cos B)$$

$$f_d = f_{\text{front}} - f_{\text{rear}}$$

$$= \frac{f_0}{c}(V_M \cos \theta + V_M \cos \alpha + V_T \cos \phi + V_T \cos \beta + V_I \cos B - V_I \cos A)$$

For a stationary illuminator $V_I = 0$. Closing velocity is $V_C = V_M \cos \alpha + V_T \cos \beta$. For the head-on case, all angles become zero and $V_C = V_M + V_T$, with the result that $f_d = (f_0/c)2V_C$. The constant of proportionality f_0/c at X band (approximately 10 GHz) is 10 Hz/(ft/s), and hence the rule of thumb is that target doppler is 20 Hz for each foot per second of closing velocity at X band. Scaling is a convenient way to handle other frequency bands. Where more exact doppler frequencies must be known, the exact transmitted frequency f_0 should be used.

It is important to note that, in addition to the target, there exist large interfering signals within the spectrum of interest: clutter and feedthrough (spillover or leakage of the rear signal into the front receiver through backlobes of the front antenna). Because the frequency of the feedthrough is the same as the rear signal, for a system in which the front and rear signals are mixed directly (baseband conversion), feedthrough would occur at dc (zero frequency), with the approaching and receding spectra folded around it. As will be shown later, it is usually desirable to unfold the spectrum, to separate the approach and recede portions, so that feedthrough will occur at some arbitrary offset frequency. Figure 19.3 illustrates the latter case.

The clutter doppler can be calculated by using the same equations used for the target. Let the reflecting clutter patch be the target, with a velocity $V_T = 0$. Use the appropriate angles which relate the missile velocity vector to the missile-to-clutter patch LOS. Not only main-lobe but also sidelobe clutter must be considered.

The spectrum of Fig. 19.3a shows the case of a ground-to-air missile. For small look angles (α and θ in Fig. 19.2), the main-lobe clutter (MLC) occurs at a frequency corresponding to a velocity of approximately $2V_M$. Sidelobe clutter extends all the way from $2V_M$ to zero-doppler velocity (feedthrough) as the angle between the missile velocity vector and the reflecting clutter patch varies from 0° (head on) to 180° (backlobe clutter).

The air-to-air case of Fig. 19.3b differs in that the clutter spectrum extends below feedthrough due to the airborne illuminator's backlobe, which can produce a return from a clutter patch behind the aircraft, i.e., an angle of 180° with respect to the illuminator velocity vector.
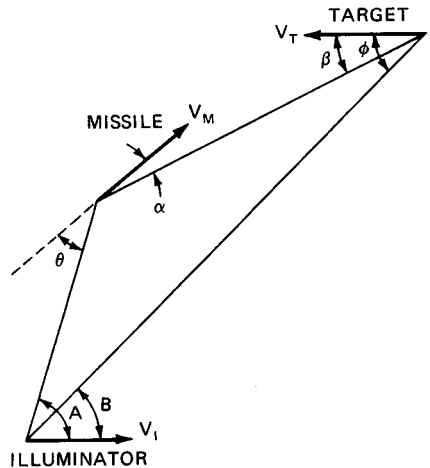
FIG. 19.2 Semiactive geometry. The radial velocity components of the three system elements contribute to the doppler shift of the received signal.

It should also be clarified what is meant by *approaching* and *receding* targets within the missile doppler spectrum. A target approaching the missile will yield a signal at a frequency above that of MLC (which corresponds to missile velocity). For the X-band case, let $V_M = 2000$ ft/s and $V_T = 500$ ft/s in level flight. MLC will then be at roughly 40 kHz and the target at 50 kHz. If the target were flying away from the missile at the same 500 ft/s velocity, its doppler frequency would be 30 kHz, or 10 kHz below MLC. However, the missile is still closing on this target at 1500 ft/s; so it is in the *approach* part of the spectrum, above feedthrough, even though it is an outbound, or receding, target. (Note that if the outbound target were faster than the missile, its doppler would be below feedthrough, but of course the missile would never catch up with it. Thus, it is a
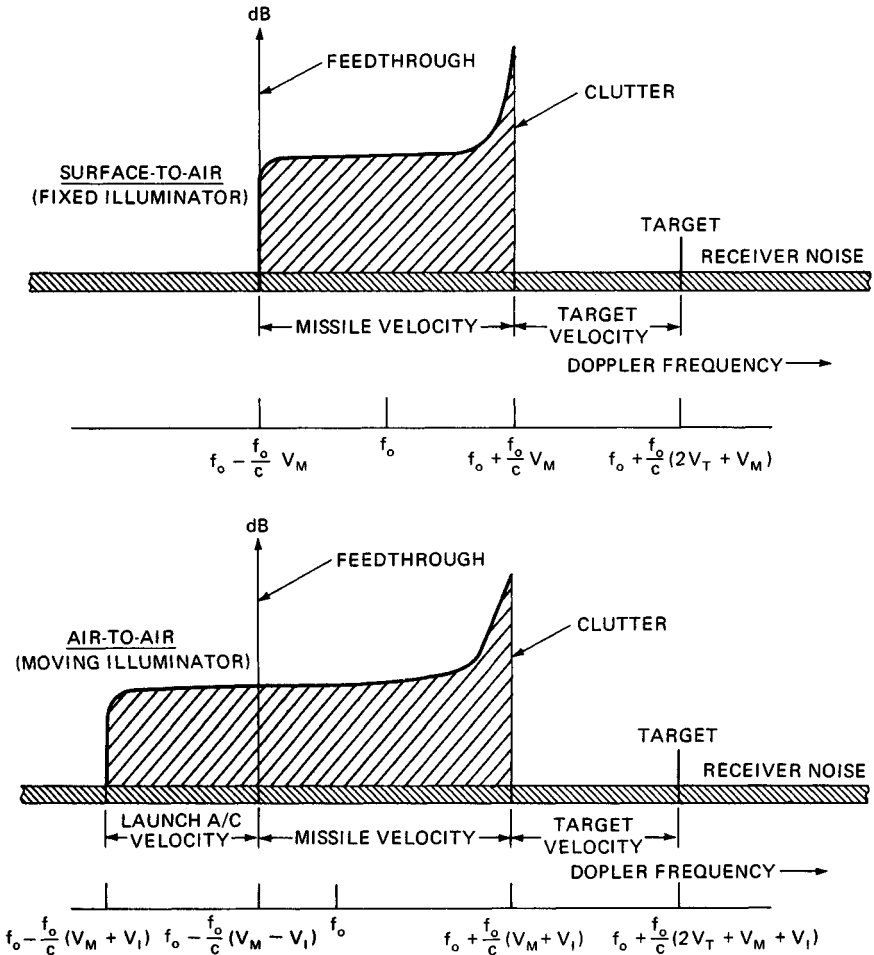
FIG. 19.3 Signal spectra for semiactive homing indicate the clutter and feedthrough (spillover) with which the target signal must compete. Both fixed and moving illuminator cases are shown. The frequencies shown are for maximum clutter extent; i.e., all angles shown in Fig. 19.2 are zero and all velocity vectors colinear.

meaningless case.) This is different for a ground-based radar, where the outbound target is in the *receding* part of the spectrum, below feedthrough. The significance of the above discussion is that while approaching targets are in a clutter-free region of the spectrum, the receding target lies in the sidelobe clutter and must compete directly with it: if the clutter in the detection cell exceeds the target, the target cannot be detected. This is of primary concern in look-down tail-chase air-to-air engagements.

One additional important factor must be noted. Although the target signal can be discriminated from the feedthrough and clutter on the basis of frequency (except for the receding target in sidelobe clutter), this is only true for spectrally pure signals. Noise on the transmitted signal and on any conversion oscillators within the missile will be spread throughout the doppler spectrum by the feedthrough and clutter and mask the target signal if the noise is not adequately controlled. Noise reduction is thus one of the key technologies required for good radar seeker performance.

Clutter and Feedthrough Considerations.[2,5,7] The presence of clutter and feedthrough is one of the primary limitations on the performance achievable in a seeker and has been one of the main design drivers in the evolution of radar guided missiles. There are three main problems which must be addressed in connection with these large interfering signals. The first is the need to prevent lock on the clutter signal and its harmonics. Clutter in some geometries may be spectrally very narrow, resembling a target signal. Preventing lock is generally accomplished by limiting the portion of the doppler spectrum which is searched during the acquisition process, to exclude the clutter frequency. However, because clutter varies in frequency during missile flight, avoiding it can be a relatively complex problem, especially for slow-radial-velocity targets (small frequency separation from clutter). Feedthrough, on the other hand, is fixed in frequency and can thus be significantly attenuated with fixed filters and be easily avoided during search.

The second problem is often termed the *subclutter visibility* (SCV) or *subfeedthrough visibility* (SFV) problem. In essence, this refers to the maximum ratio of clutter (or feedthrough) to signal with which the system can operate. In its simplest form, this can be related to the dynamic range of the seeker receiver (i.e., its range of linear operation). One must consider not only possible suppression of the target signal by the clutter or feedthrough but also potential cross-modulation or intermodulation effects. As will be shown, gain normalization (automatic gain control, or AGC) is a key concern in achieving the required SCV.

The third problem is also related to SCV (and SFV) and is, as noted earlier, concerned with the spectral purity of the transmitter and the local oscillator. The spectrum of Fig. 19.3 will be broadened by noise, so that noise sidebands of feedthrough and clutter will appear at the target doppler frequency. In view of the magnitude of feedthrough and clutter, very low noise is required to prevent performance degradation (masking of the target). Maximum feedthrough levels can typically range from 80 to 100 dB above the target signal, while main-lobe clutter can be 40 to 80 dB greater than the target. However, the frequency separation between target and clutter is much smaller than between target and feedthrough; so, depending on the specific design and conditions, clutter may establish the more stringent noise requirement. Also, the effects of feedthrough noise can be reduced through cancellation in the missile receiver (see Sec. 19.4). Since amplitude-modulation (AM) noise in sources is generally well below frequency-modulation (FM) noise (20 dB is typical), the noise reduction techniques concentrate on FM noise.

Guidance Fundamentals.[7-12]   A detailed discussion of guidance is beyond the scope of this chapter. However, to understand the effect that the radar seeker's ability to measure the target's radar observables has on the performance of the missile (miss distance), a brief overview of guidance principles is presented here.

Virtually all homing systems employ some form of proportional navigation (PN), although modern control theory (such as Kalman filtering) has been used extensively to optimize performance of later-generation systems. The important fact to note is that PN can be accomplished with angle-only measurements and can thus become a fallback mode even if range or doppler (range rate) information—required for advanced guidance techniques—is unobtainable.

Proportional navigation is based on the fact that if two objects are closing on each other, they will collide if the LOS does not rotate in inertial space, as illustrated in Fig. 19.4. Any rotation of the LOS (i.e., an LOS rate) is indicative of a deviation from the collision course which must be corrected by a missile maneuver. In PN, the rate of rotation of the LOS is measured, and a lateral acceleration of the missile is commanded according to the equation

$$n_L = N'V_c \dot{\lambda}$$

where n_L = lateral acceleration
N' = effective navigation ratio (constant, selected as discussed below)
V_c = closing velocity
$\dot{\lambda}$ = rate of change of the line of sight

The lateral acceleration ideally should be normal to the LOS; in practice the deflection of the missile control surfaces will result in acceleration normal to the missile velocity vector. The closing velocity can be estimated or, in the case of a doppler radar, measured (the target doppler is an approximate measure of V_c, as noted above). The LOS rate $\dot{\lambda}$ is measured by the seeker—this is the seeker's primary function. The value of N' is chosen to optimize performance in the face of initial errors or disturbances which would increase miss distance: heading error, target maneuver, system biases, and noise. For example, increasing values of N' cause early correction of collision course errors, reserving the missile's maneuver capability near intercept for countering target maneuvers and noise. Too high a value of N', however, results in too great a sensitivity to noise inputs, especially glint, which increases with decreasing range. In practice, N' values in the range of 3 to 5 are normally chosen.

The missile does not respond instantaneously to an LOS rate; rather, a finite response time, made up of several components, governs the process. This equivalent time constant, referred to as the guidance time constant τ_g, is a key parameter affecting miss distance.

Several time lags in series combine to produce τ_g. These are the track-loop time constant, the noise-filter time constant, and the autopilot-airframe response time. The antenna track-loop time constant can be eliminated as a contributor in certain configurations (LOS or LOS rate reconstruction[12]). The airframe aerodynamic response will vary with missile speed and altitude, and the autopilot must compensate for this variation. The final value of τ_g is a compromise between the rapid desired speed of response to counter target maneuvers and a long desired smoothing time to minimize glint. Moreover, the variations in τ_g brought about by *parasitic feedback* effects, such as radome aberration and imperfect antenna stabilization, must be controlled to avoid guidance loop instability.

A practical rule of thumb for a properly designed system is that a homing time
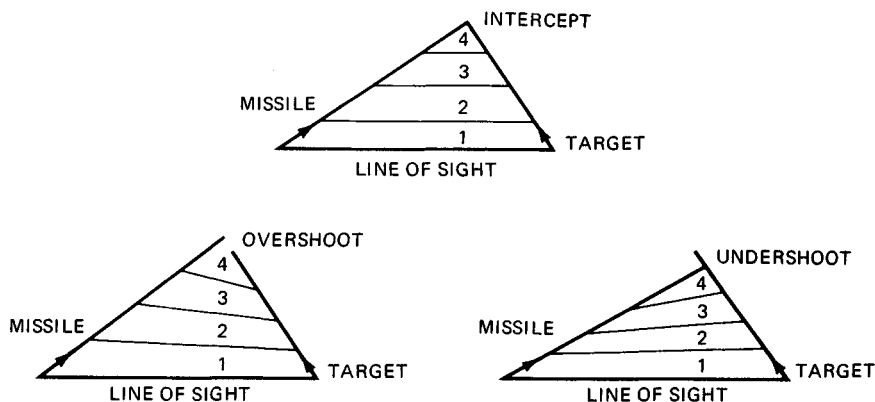
FIG. 19.4 Line-of-sight (LOS) motion of intercept. The line-of-sight rate is constant when missile and target are on an intercept (collision) trajectory.

of 10 τ_g will reduce miss distance to the asymptotically achievable value. This will therefore establish minimum range capability as well as set the requirement on the terminal guidance mode of a multimode missile.

The seeker's primary function is to generate an estimate of the inertial LOS rate. To accomplish this, it must track the target in angle and stabilize the antenna LOS against missile body motions, which could be erroneously interpreted as target motions. It is the accuracy of the resulting LOS rate estimate that will determine how well the missile performs.

The fundamental limit on achievable accuracy is the target's own angle noise (glint, scintillation, and depolarization). Other noise contributors must be minimized by proper design (i.e., maximize signal-to-noise ratio to minimize range-dependent noise, reduce the range-independent noise—servo and other instrumentation noise). Also, the correct angle measurement scale factor must be maintained over the full range of signal levels and over all look angles.

Finally, the effect of the radome must be considered. Because of aerodynamic considerations, the radome enclosing the gimballed antenna will be pointed rather than a hemisphere. Thus, at different gimbal (look) angles the radar signal will pass through a different portion of the radome, and the apparent LOS to the target will change with gimbal angle because of refraction (aberration). This results from different path lengths through the dielectric material (different curvature) as well as local differences in thickness or dielectric constant. A constant error would present no difficulty, since the tracking and guidance loops are driving the boresight error (LOS rate) to zero. It is the variations of the radome error with gimbal angle—radome error slope—which cause the problem by creating a feedback path.

Since the missile responds to a target LOS rate by maneuvering, the missile body orientation with respect to the observed LOS will change as a result of the maneuver. Thus the space-stabilized antenna, while maintaining track of the target, will move with respect to the radome, and the resulting change in the refraction angle will cause an apparent additional LOS rate, closing the feedback loop. The feedback can be either regenerative or degenerative, depending on the sign of the radome error slope (the direction of the radome error).

This phenomenon must be viewed in the context of the closed antenna tracking loop. Since for a constant LOS rate the residual boresight error is a constant, any radome error which tends to increase the apparent boresight error constitutes regenerative feedback. A radome error which makes the boresight error smaller is degenerative. To a first order, positive slopes (degenerative feedback) lower the guidance gain and lengthen the guidance time constant, making for a more sluggish response, while negative slopes (regenerative feedback) raise the guidance gain and shorten the guidance time constant to the point where missile instability could occur. The guidance design must avoid such an instability.

Target Illumination.[6,13] Target illumination for a CW semiactive missile system can be provided by a CW tracking radar, a CW transmitter slaved to another tracking radar, or a pulse or pulse doppler tracking radar at another frequency with the CW illumination injected into the antenna system from a separate CW transmitter.

The most capable of these configurations is the CW tracking illuminator. It is generally a two-dish radar because sufficient receiver-transmitter isolation cannot usually be achieved in a single-dish system. The CW tracking illuminator, since it uses the same radar signal to track the target as the missile utilizes for homing, sees essentially the same view of the target environment and can track targets at the same low altitudes as the missile seeker. The receiver portion of such an illuminator is very much like the seeker described in the following sections. The main differences stem from the much higher feedthrough levels in which the illuminator receiver operates and from the previously mentioned doppler spectrum differences (i.e., outbound targets are below feedthrough).

Alternatively, the illuminator can be the transmit-only portion of a radar slaved to a tracking radar—a mechanically scanned track-while-search (TWS) radar or a phased array which simultaneously maintains multiple target tracks with its electronically steered agile beam. In the third approach, where space constraints preclude use of separate antennas, such as in a fighter aircraft, a conventional pulse or PD radar tracks the target and the CW illumination is injected into the transmission port of the antenna from a separate transmitter.

Traditionally, the illuminator must continuously illuminate the target throughout the engagement. A system is therefore limited in its simultaneous-engagement capability by the number of available illuminators. A given illuminator must remain dedicated to its assigned target until the missile has achieved intercept; only then can it be reassigned to another. One of the primary reasons for active seekers is to remove this system firepower limitation, since each missile provides its own illumination. Another approach to avoiding the one-illuminator–one-target constraint is to use sampled data and time-share one illuminator (phased array or TWS) among several missiles.

19.3 SYSTEM EVOLUTION

Radar guided missiles have evolved through several generations since the first developments began in the closing years of World War II. The threat and the available technology have evolved over the years, and the systems have followed

suit. As new requirements have been generated in response to more severe threats, new approaches using new technology have been developed. This section will attempt to trace some of these evolutionary developments.

Basic Semiactive Seeker.[2,6,7]   The block diagram of Fig. 19.5 is representative of the earliest systems developed in the late 1940s and early 1950s. The simplest implementation of a CW missile seeker, it consists of a rear receiver, a front receiver, a signal processor (speedgate), and a tracking loop to control the gimballed front antenna. The missile also contains an autopilot to guide it and stabilize the airframe, a fuze to detonate the warhead at the optimum time, and a source of electrical and (in most missiles) hydraulic power.

The purpose of the rear receiver is to provide a coherent reference for detection of the front (target) signal. The rear signal, after conversion to IF, closes the automatic frequency control (AFC) loop around the microwave local oscillator (LO) and acts as the reference for the IF coherent detector. The target signal, received in the front antenna, is heterodyned to IF and amplified in a relatively wideband amplifier (typically 1 MHz or wider). It is then converted to baseband by mixing it with the rear signal in the balanced mixer (coherent detector). The doppler signal (now at baseband, with feedthrough at dc) is amplified in the video (doppler) amplifier, which has a bandwidth equal to the total range of possible doppler frequencies. It is then mixed with the speedgate LO, which is controlled by an AFC loop to keep the desired signal centered in the narrow speedgate (sometimes called the velocity gate or doppler tracker). Typical bandwidths range from 500 Hz to 2 kHz.[5]

Target acquisition is accomplished by sweeping the frequency of the speedgate LO over the designated portion of the doppler bandwidth. In essence, this sweeps the spectrum past the narrow frequency window of the speedgate. When a signal exceeds the detection threshold, the search is stopped and the signal is examined to verify that it is a coherent target rather than a false alarm due to noise. A valid target is then tracked in frequency, and guidance commands are extracted from it.

The front antenna conically scans the received beam. The resulting amplitude modulation of the received signal is recovered in the speedgate and resolved into the two orthogonal pitch and yaw gimbal axes. These pitch and yaw error signals are used to close the antenna tracking loop and to guide the missile.

The guidance error signal must be normalized (a constant scale factor of volts per degree off boresight is required) over the full dynamic range of target signal amplitudes (range, target size) in the presence of large feedthrough and clutter signals. Therefore, AGC in the receiver is necessary. Since the IF amplifier signal includes both the feedthrough and the clutter while the video amplifier includes the clutter, the specific AGC implementation must consider the degree to which these large interfering signals shall be allowed to control the gain for the target signal while preventing saturation on the interference. To maintain linear receiver operation over the large dynamic range is a major design challenge.

Unambiguous (Offset Video) Receiver.   The basic receiver described above folds the spectrum around feedthrough, which occurs at dc. Although in the moving missile this does not produce an ambiguity, in the tracking illuminator the inbound targets must be distinguished from the outbound. This *unfolding* of the spectrum was initially achieved by use of a quadrature receiver (Chap. 14).
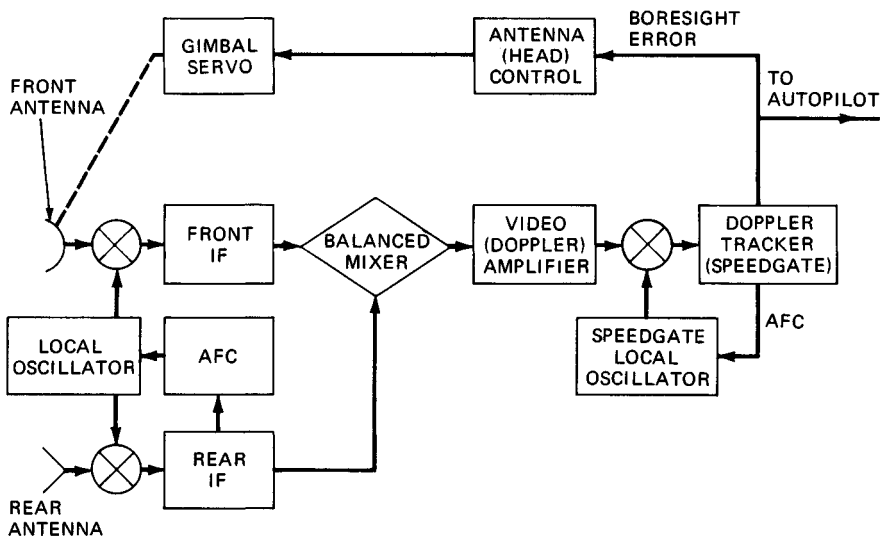
FIG. 19.5   This semiactive-seeker block diagram of a baseband conversion system is represen-
tative of the early-generation systems in which the rear (reference) and front signals are mixed
directly to extract the doppler-shifted target signal.

There are two additional drawbacks to the original configuration. Folding
the spectrum around feedthrough folds the receiver noise as well, resulting in
a 3 dB higher noise level (hence a 3 dB loss in sensitivity).[5] The other problem
stems from the fact that main-lobe clutter is the dominant signal in the doppler
spectrum. Clutter harmonics can be misclassified as targets and must there-
fore be avoided, thus limiting the usable range of target dopplers. For exam-
ple, consider a clutter-to-signal ratio of 60 dB. A mere 0.1 percent second har-
monic distortion would yield a clutter harmonic of the same magnitude as the
target. If a missile velocity of 2000 ft/s is assumed (40 kHz doppler at X band),
the harmonic would occur at 80 kHz, and the usable doppler spectrum, which
the speedgate would be able to search, could extend no further than 80 kHz (in
practice a safety margin of a few kilohertz would have to be maintained at
both ends of the search region, further limiting achievable performance). This
is illustrated in Fig. 19.6a.

However, by introducing a frequency offset before the coherent detector,[5] the
resulting spectrum will be as shown in Fig. 19.6b. This can be accomplished by
offsetting either the signal or the reference channel. Figure 19.7 shows the offset
reference configuration. Clutter harmonic distortion, noise foldover, and (for the
case of the illuminator) approach-recede ambiguity are eliminated. However,
feedthrough rejection now requires a complex notch filter at the relatively high
offset frequency rather than a simple high-pass filter. Also, clutter still controls
the gain normalization in the doppler amplifier.

For each frequency conversion, spurious higher-order mixer products must be
considered and kept out of the target spectrum. As additional conversions are
added, this task becomes increasingly difficult and seeker complexity grows. Ex-
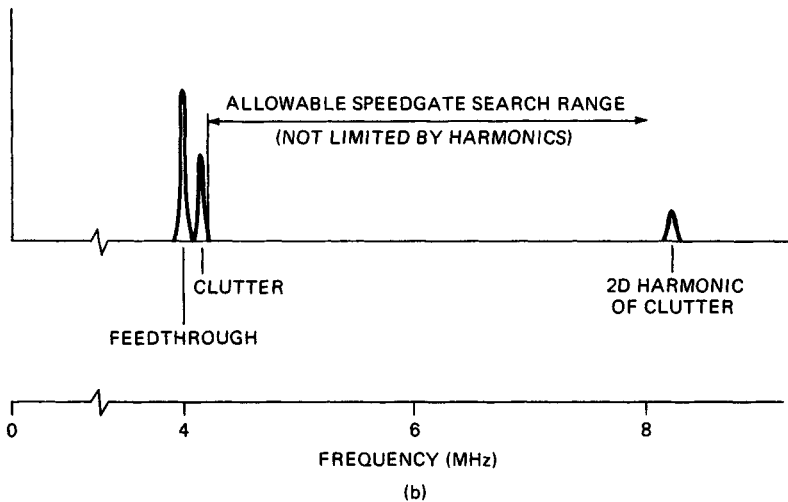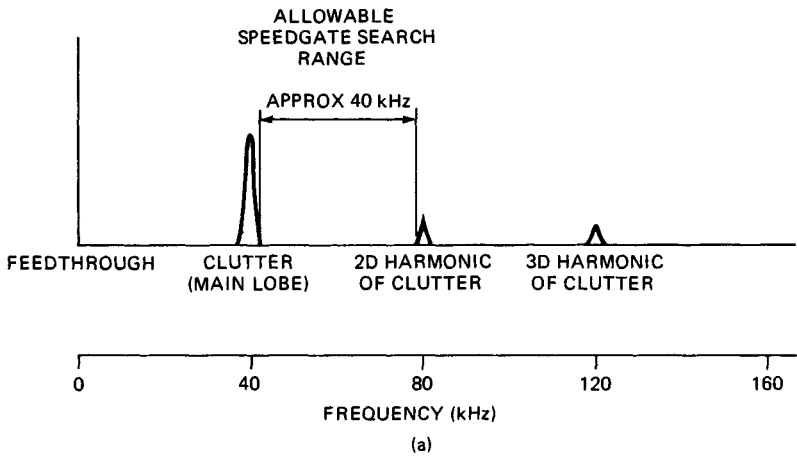
FIG. 19.6   Target spectra of the baseband (folded or ambiguous) receiver (*a*) and the offset video (unambiguous) receiver (*b*) indicate the limitation which clutter harmonics impose on the achievable range of target velocities which can be handled.

tending the speedgate's frequency coverage to cope with faster targets and attempts to eliminate—or at least attenuate—clutter required additional conversions, which, even with the introduction of solid-state circuitry to replace vacuum tubes, resulted in prohibitive increases in size, complexity, and reduced reliability.[2,6]

Inverse Receiver.[2,6,14]   The major breakthrough was the introduction of the *inverse receiver*, which gets its name from the fact that the bandwidth "funnel" of the conventional receiver (wide IF, narrower doppler amplifier, final narrowband speedgate) is inverted, with the final narrow banding (speedgating) placed right after the first conversion from microwave to IF. The crit-
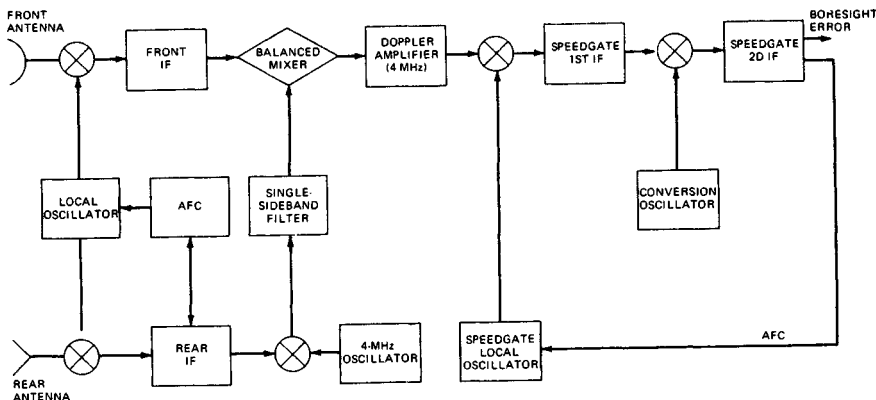
FIG. 19.7 Offset video receiver block diagram. It provides an unambiguous (unfolded) doppler spectrum by offsetting the rear reference before it is mixed with the front signal.

ical components necessary for the inverse receiver are highly selective filters at IF frequencies and low-noise tunable microwave sources.

The simplified block diagram of an inverse receiver is shown in Fig. 19.8. In the conventional receiver, the target signal must compete with feedthrough, clutter, and jamming until the final stages, with the dynamic-range requirements of the receiver and its AGC loops dictated by these large undesired signals. The inverse receiver, on the other hand, excludes them virtually at the input. The narrowband filter (usually a quartz crystal type), constituting the speedgate bandwidth, is placed in the IF after only a nominal amount of fixed preamplifier gain, sufficient to establish noise figure. One additional conversion is used in the receiver to avoid the problem of too much gain at one frequency. In the resulting two-conversion system, complexity is significantly reduced and unwanted signals
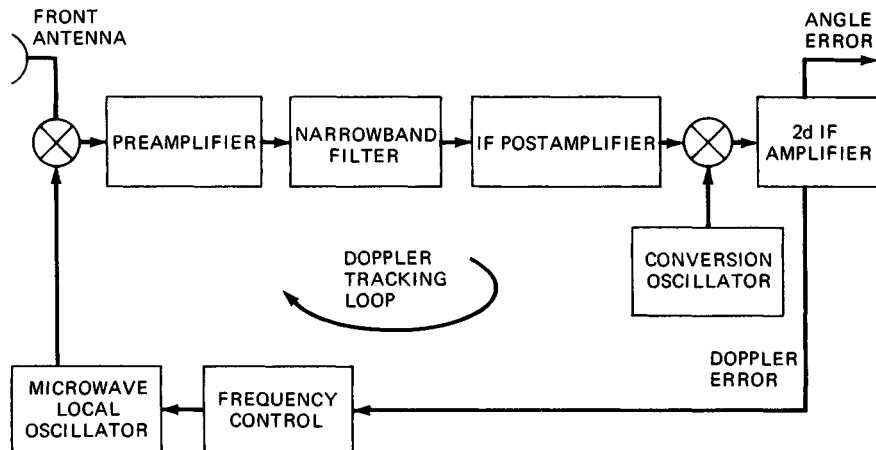
FIG. 19.8 Inverse-receiver block diagram. The narrow banding is placed very early in the receiver, inverting the bandwidth "funnel" of the conventional receiver and excluding interference from subsequent stages of the seeker.

are rejected very early in the signal path, thus reducing dynamic-range requirements and avoiding most possible sources of distortion.

The doppler tracking loop is closed through the microwave LO, which must, therefore, be tunable over the doppler frequency range of interest. This LO essentially fulfills the role of the speedgate LO in the conventional receiver of Fig. 19.5. The inverse receiver can be thought of as a double-conversion speedgate with the speedgate AFC loop closed around the microwave LO and the input to the speedgate being the microwave output of the seeker front antenna.

The IF spectrum at the mixer output will have the same form as Fig. 19.3. Sweeping the LO moves the spectrum past the narrowband filter to accomplish acquisition as in the conventional speedgate. Doppler tracking is similarly accomplished by controlling the LO frequency to keep the target in the narrow filter. The angle error signals required for guidance are extracted after the second IF amplifier. A single AGC loop (not shown), required to cope with only the target signal variations, is used to normalize the angle error signals.

Angle Tracking: Conical Scan to Monopulse. This subsection assumes that the reader is familiar with the conical-scan and monopulse angle-tracking concepts described in Chap. 18.

Conical scan requires only a single channel and extracts the angle information which is contained in the amplitude and phase of the scan amplitude modulation by simple envelope detection. Conventional monopulse normally requires three complete channels, which must track in gain and phase to maintain the proper relationship between the sum and difference channel signals (the angle information is contained in the difference/sum ratio).[15] The complexity of monopulse, however, provides well-known performance advantages over conical scan.

Conical-scan processing requires that both the amplitude and the phase of the AM be preserved (at least one cycle of the scan is needed to make an angle measurement). The AGC which is required for gain normalization must therefore be slow enough not only to prevent it from following the scan AM envelope but to avoid any phase shift of the envelope[15] (since this would cause cross coupling between channels; i.e., a pitch error would couple into the yaw plane, and vice versa). Thus any externally generated amplitude fluctuations (propeller modulation, target fading noise, or jamming) at or near the scan frequency will be detected along with the target BSE and will result in noise or false data. In particular, this makes conical-scan systems susceptible to AM jamming at the scan frequency (the *spin frequency jammer*).[5]

The monopulse system extracts the angular information instantaneously by comparing the difference and sum channel signals. The gain normalization can therefore be made instantaneous (fast or instantaneous AGC), and the external amplitude variations, since they affect sum and difference channels by the same relative amount, are never detected as erroneous guidance signals.

The early systems all used conical scan for angle tracking because of its simplicity. The limited available volume and discrete-component tube technology of the period mandated a single-channel approach despite the performance limitations of conical scan. The inverse receiver permitted the performance of monopulse to be achieved with the single-channel simplicity of conical scan.[2,6,14]

Three identical mixers, preamplifiers and crystal filters, first process the three monopulse signals. Immediately after the narrowband filters, however, the difference channels are multiplexed with the sum channel at a moderate frequency (several kilohertz, much higher than the filter bandwidth). Interference at the multiplexing frequency is, therefore, prevented from passing through the filters. The modulated difference channels are combined with the sum signal, and the

composite signal is processed in a single channel (just as a conical-scan signal). The AGC, required for gain normalization, is made faster than the bandwidth of the filters and thus acts as an instantaneous AGC. Its dynamic range has to cope only with target signal variations. The normalized monopulse error signals are then demultiplexed and used for closing the guidance loops. Frequency (AM or FM) or time multiplexing can be employed. If the frequency multiplexing is phased so as to produce AM sidebands, the processing is identical to that of conical scan.

Pulse Doppler (PD) Operation.[2–6] Semiactive systems using other than CW illumination have been employed. Some early systems employed noncoherent pulse waveforms, but they are not suitable for operation in clutter (except for very large target cross sections). Coherent PD systems, however, can approach the performance of CW.

The motivation for the use of PD in the seeker was to simplify the illuminator in air-to-air systems. For early-generation airborne radars, which employed a noncoherent pulse waveform, CW injection was the only practical solution. With the advent of coherent PD radars, an alternative way to achieve virtually CW operation without the penalty of the additional transmitter became available. This was to select a high-PRF (pulse repetition frequency), high-duty-cycle (30 to 50 percent) waveform and to use only the central line of the PD spectrum, both in the radar and in the seeker. This has sometimes been called *interrupted* CW (ICW).[3]

A high PRF is defined as one which is unambiguous in doppler. Thus when the receiver selects the central line, the spectrum is identical to the CW case. The radar receiver must be protected during transmission (duplexing and/or gating). In addition, the receiver may or may not use a range gate. If only the central-line power of the PD spectrum is used (no range gate), the resulting loss must be accepted. Use of a range gate matched to the pulse avoids this loss. In either case, the rest of the receiver and signal processing is the same as for a CW system.

The seeker implementation follows this same pattern. However, range gating in the seeker is generally not used with a high-duty-cycle system. The loss resulting from use of only the central line is essentially the duty cycle d_t. For a peak transmitted power P_t, the average power in the central line is $P_t(d_t)^2$, compared with the average power of the transmitted waveform of $P_t(d_t)$.

A low duty cycle (less than 10 percent) can also be used, but for this case the central-line power loss becomes prohibitive. Low-duty-cycle systems, therefore, must use range gating to optimize performance. In addition to retaining doppler resolution capability the range-gated system provides range resolution.

Active Seekers.[4,16] Active seekers can provide increased firepower as well as fire-and-forget (or launch-and-leave) operation. Thus, they have found application in both the air defense and the surface-target attack roles.

An active seeker is functionally the same as a semiactive seeker, with the exception that it carries along its own illuminator. Besides adding the transmitter, the other main difference in the active seeker configuration is elimination of the rear receiver, with the reference generated by offsetting the transmitter excitation (or drive) signal, as shown in Fig. 19.9.

Active seekers, since they use a single antenna both to transmit and to receive, cannot use CW because of the very limited isolation achievable. Noncoherent pulse or coherent PD waveforms have been employed, and either the central-line processing or the range-gated approach can be used for coherent operation.
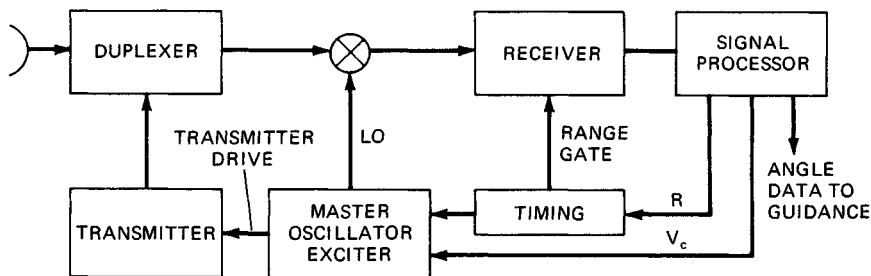
FIG. 19.9 The active seeker block diagram differs from the semiactive in that target illumination is provided by a self-contained transmitter.

Surface Targets.[4,5,17] Noncoherent pulse waveforms have been widely used in active seekers designed for attacking large-cross-section surface targets. For example, in antiship applications the slow target speed prevents effective doppler resolution from clutter, but the large target reflectivity provides an effective discriminant since the target return exceeds sea clutter by several orders of magnitude (large signal-to-clutter ratio). Even in antitank applications, such *contrast* discrimination of the target can be achieved if the size of the competing clutter patch can be reduced by the use of narrow-beamwidth antennas and narrow range gates. These noncoherent systems utilize low-duty-cycle short pulses or highly coded waveforms to achieve narrow range resolution. The resolution cell is determined by the range-gate duration in the range dimension and by antenna beamwidth in the cross-range (azimuth) dimension. The resulting surface clutter return, even for rough seas and fairly severe ground reflections, will contribute much less energy than the target echo even when the target fills only a small portion of the resolution cell. Thus the angle information derived will be primarily from the target, because of its large contrast with respect to the clutter background, and accurate homing can be achieved. It should be noted that radar cross sections of ships can be several thousand square meters, while those of tanks typically range from 25 to 125 m^2.[17]

For severe clutter and reduced target cross sections, coherent processing may be required. Although stationary or very slow-moving targets cannot be discriminated from clutter on the basis of doppler frequency, use of *doppler beam sharpening* or synthetic aperture techniques can reduce the effective size of the resolution cell and hence increase the signal-to-clutter ratio in the cell containing the target.[18] Angle tracking can thus take place to provide the required data for guidance.

Air Targets.[4,16] Although some early systems were designed to use noncoherent pulse waveforms, they were not suitable for low-altitude (high-clutter) operation against small-cross-section aircraft targets. Therefore, the typical air defense active missile employs some form of PD transmission and coherent processing to resolve and track the target in doppler (velocity) and sometimes also in range.

The active PD seeker can be thought of as a miniature airborne fire control radar. The waveform selection tradeoffs, especially the clutter-waveform interactions, are the same in the active seeker as in the airborne intercept (AI) radar. The unique problems of clutter ambiguities, eclipsing, range determination, etc., are the same as described in Chap. 17 and will not be repeated here. A key point is that the active seeker is a monostatic radar, whereas the semiactive system is bistatic. The doppler frequency relationships can be simply derived from the ge-

ometry of Fig. 19.2 by colocating the illuminator and missile and making the illuminator and missile velocity vectors coincident. The doppler spectrum will be like that of Fig. 19.3b.

Although it is highly desirable to select a high PRF (HPRF), which is unambiguous in doppler, it may be necessary in some system applications to use a medium PRF (MPRF) and operate with both range and velocity ambiguities (which must be resolved).[19] The tail-chase look-down air-to-air scenario is a key example. Since the target return must directly compete with the sidelobe clutter in the doppler resolution cell, it may be necessary to reduce the absolute amount of clutter contained in the cell. A positive signal-to-clutter ratio (S/C) is necessary to permit target visibility. One way to achieve this is to range-gate and thus reduce the size of the clutter patches, the return from which is accepted in the receiver. Reducing the PRF reduces the number of intervening range-ambiguous clutter patches which fold into the target doppler cell, further improving S/C.

To maximize range performance (not clutter-limited) average transmitter power must be the maximum practically achievable. Within the constraints of a tactical missile—small size, limited weight—this will tend to drive the design to higher-duty-cycle lower-peak-power systems. This is quite compatible with HPRF, where high-duty-cycle central-line processing has generally been used. If clutter is the limiting factor rather than receiver thermal noise, lower average power is acceptable—consistent with MPRF. The difficulty arises if the same system must achieve both long-range (noise-limited, approaching target) and tail-chase (clutter-limited) performance. Transmitter hardware constraints make it difficult to vary the waveform at will over a wide range of PRF and pulse width. Thus if an MPRF system is employed, the tendency will be toward long pulses (to keep average power high without increasing peak power). Therefore, to achieve good range resolution may require some form of pulse compression.

System Implementation. Active seekers have used both conical scan and monopulse angle tracking, and the receivers have evolved from the conventional to the inverse configuration, just as with the semiactive.

Because of the limitation on achievable antenna size as well as transmitter power, the range performance of an active seeker will be considerably less than for a comparable size of semiactive seeker operating with a high-power large-antenna illuminator.[5] Thus active systems are used in short-range homing-all-the-way applications or as the terminal guidance mode of a multimode long-range system. For example, a midcourse mode employing inertial or command guidance can be used to bring the missile within the terminal guidance range (typically the last 10 guidance time constants or a few kilometers from intercept). The target coordinates in angle (antenna pointing) and range and/or velocity (doppler), provided by prelaunch data or by command updates during flight, initialize the seeker. The target uncertainty is searched by the seeker, and when the target is acquired, the missile transitions into the terminal phase of flight. Seeker operation then proceeds as for the semiactive case until target intercept.

Passive Seekers. Three passive operating modes have been employed for missile guidance: antiradiation homing (ARH), home-on-jam (HOJ), and radiometric. ARH operation is used in missiles for attacking hostile radars, usually in an air-to-surface application (although air-to-air and surface-to-surface systems are also potential configurations). HOJ is an essential adjunct for semiactive and active systems to counter noise jamming.[4] The radiometric homing mode has been employed as a terminal guidance mode in millimeter-wave antitank missiles.[17]

Antiradiation Homing.[20]   ARH systems differ from the active or semiactive air defense or the ship or ground attack systems in that they are very wideband (octave bandwidth or wider). This need for wideband operation is the main driver in seeker design. The receiver configuration is very similar to the emitter location and identification systems often called ESM (electronic support measures).[21] However, the size and weight constraints of missile-borne hardware restrict usable approaches.

The parameters available to an ARH sensor include frequency, PRF, pulse width, angle (direction of arrival), and signal amplitude. Various combinations of these can be used to discriminate and select a specific emitter from among the multiplicity of signals present (estimates of 10^6 pulses per second have been cited as a "high-density" environment[21]). Signals must be initially sorted on the basis of frequency and then the pulse trains *deinterleaved* to select a particular emitter. Most radars will be detected primarily through their sidelobes, requiring reasonable sensitivity, but the dynamic range must be able to cope with main-lobe signals as well.

Broadband antennas can be gimballed or body-fixed. Since directional information must be determined on each received pulse, some form of monopulse antenna is required. Because the seeker may encounter any incident polarization, the antenna should have a uniform response to all senses of linear polarization.[20]

Four types of antenna systems are possible: an amplitude monopulse with four squinted beams, a three-channel phase-amplitude monopulse using four elements or apertures, an interferometer, and the two-channel polar monopulse using a four-arm dual-mode spiral.[22] The first three are conventional configurations except that each of the antenna elements is a broadband device such as a log-periodic type, conical log spiral, or cavity-backed planar spiral. The dual-mode spiral has been the preferred choice, since a single aperture generates all the direction-finding information (and thus makes full use of the limited available space), requires only two receiver channels, has excellent polarization characteristics, and is frequency-independent.[20]

The four arms of the spiral are fed by a mode-forming network to form a sum (Σ) and a delta (Δ) mode (hence the name *dual-mode*). The directional information is contained in the relative amplitude and phase of the Σ and Δ channels. The Δ/Σ ratio represents the magnitude of the BSE (the angle off axis in a cone of rotation about the boresight), while the relative phase indicates the direction on the cone of rotation. This polar information is then converted into the more conventional pitch and yaw coordinates.[23]

A variety of receiver types can be used to analyze the signal spectrum: wideband crystal video, instantaneous-frequency measurement (IFM), channelized, scanning superheterodyne, compressive (microscan), or Bragg cell (acoustooptic).[21,24] Size and weight limitations dictate that a single-channel approach be used in a seeker. The contradictory requirements of wide instantaneous bandwidth for rapid acquisition and narrow bandwidth for high sensitivity can be achieved by using switchable bandwidths, such as shown in the typical block diagram of Fig. 19.10. This also includes a compressive receiver which is ideal for CW signals. In this approach, the local oscillator is swept rapidly (*chirped*) to impress linear FM on the signal. A matched compressive delay-line filter then compresses the signal, producing a short pulse, the time of occurrence being indicative of RF frequency.[25,26]

Home-on-Jam.[3,4,27]   The HOJ mode is an essential part of a semiactive or active seeker. The use of wideband noise represents the earliest brute-force active jamming technique which can mask the desired target reflection. The jamming,
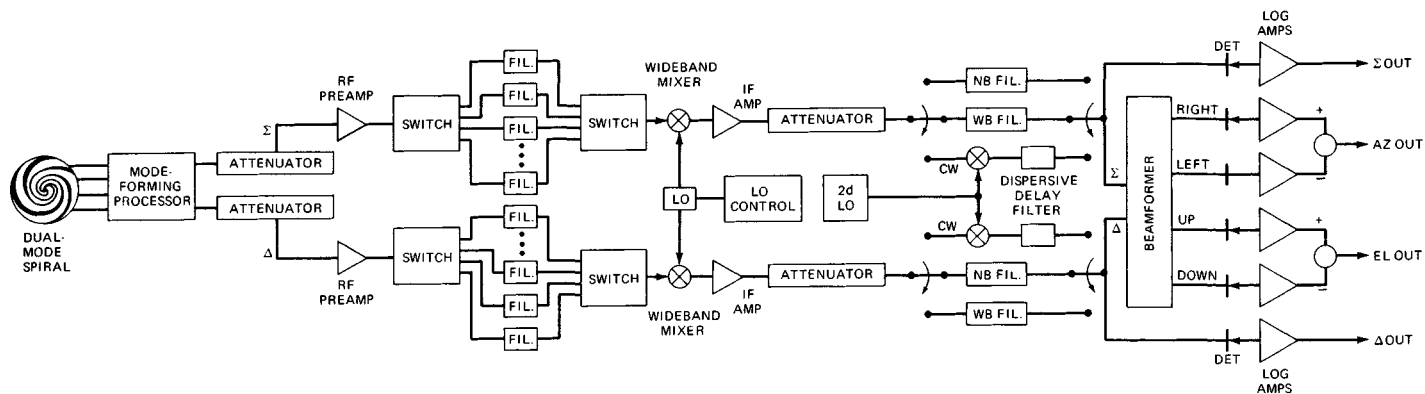
FIG. 19.10 Antiradiation homing (ARH) receiver. A typical configuration is shown, including a dual-mode spiral antenna, high first IF, switchable bandwidths, and compressive (microscan) CW processing. (*From Ref. 20.*)

however, is a powerful point source of radar energy which can provide more than adequate angle information for homing. All that is required is a means to allow the seeker angle track circuits to process the noise energy. When the jamming is such that tracking of the target *skin* return is not possible, the seeker switches to passive tracking of the received jamming energy. If the jammer-to-signal ratio (J/S) decreases to the point that skin track is again possible, this is given preference over HOJ. Also, provision must be made to allow switching between HOJ and skin if the jamming is intermittent. The criterion in all cases is that the mode which provides the better quality of guidance information should be given precedence.

Radiometric Homing. This mode utilizes the natural thermal radiation from targets for guidance. The very sensitive receiver detects the difference in radiation between the target and the ambient background. Use of this technique in millimeter-wave seekers against surface targets provides a terminal mode with significantly lower glint than the active or the semiactive radar mode.

Other System Configurations. Variations of the above seeker types as well as multimode combinations have been studied and in some cases implemented to take advantage of new and emerging technologies.

Sampled-Data Operation.[2,4,6] To overcome the limitation of tying up an illuminator for the duration of a semiactive engagement, a single radar can be time-shared among several missiles. This generally implies a phased array radar, although mechanically scanned track-while-scan (TWS) radars can provide this option in some cases.

The advent of phased array radars permitted a single transmitter to illuminate many targets by sequentially stepping its agile beam from one target to the next. The illumination was no longer continuous, and the missile would thus have to operate in a sampled-data mode, extracting information during the time that its target was illuminated (dwell time) and then holding the information until the next sample. The illumination waveform could be CW during the dwell time (interrupted or keyed CW), or PD could be employed with or without pulse compression.

For sampled-data operation, the primary difference is in the doppler acquisition scheme. Since target illumination occurs only in short bursts, the use of a sweeping gate for acquisition would result in excessively long acquisition times. The doppler uncertainty region must, therefore, be examined simultaneously by a bank of contiguous doppler filters. The illumination burst must be shaped or the received signal time-gated to prevent the spreading of clutter through the target doppler spectrum owing to the pulsed nature of the transmission. Finally, sample-and-hold circuits must be added in AFC, AGC, and angle track loops.

Either the conventional or the inverse receiver can be made to operate in the sampled-data mode. Sampled data can be used all the way to intercept or as a midcourse mode for an active terminal seeker. Lower data rates are allowable in midcourse than in terminal, providing an additional degree of freedom in the system design.

Retransmission Guidance.[13,28-31] Retransmission guidance, also known as TVM (target-via-missile or track-via-missile), was initially conceived as a simplification of missile-borne hardware, placing all the processing on the ground and making the seeker a simple repeater. In practice the repeater is limited in gain by the usual transmit/receive isolation (ring-around problem); so additional complexity must be added. At the same time, use of more complex pulse compression waveforms could be more easily accommodated by not requiring the sophisticated processing hardware to be missile-borne.

TVM is essentially a variation of semiactive homing. The target-reflected illumination is received in the missile, but instead of being processed on board it is retransmitted to the illuminating radar. Here the complex waveform is processed, guidance information extracted, and steering commands transmitted to the missile as in a command guidance system.

Multimode Systems.[2,4,16]   The early CW semiactive systems were generally designed to home all the way from launch to intercept. In later-generation, more sophisticated systems, homing generally lasts for only the last few seconds of flight (typically 10 guidance time constants). In these systems, a midcourse phase (inertial, beam rider, or command) is employed to get the missile to an appropriate point on its trajectory, where it acquires the target (using prelaunch or in-flight commanded designation data) and enters the terminal (homing) phase of its flight. This is more efficient from the standpoint of both missile trajectory and radar power. The missile can fly out to longer ranges by a commanded or inertial up-and-over trajectory, spending less time in the denser air at low altitude. The radar power needed for illumination (semiactive or active) is sized by the terminal phase of flight, a fraction of the total intercept range. Midcourse commands impose much less severe demands on radar power since this is a one-way transmission path.

Combinations of semiactive or active radar with IR or ARH modes and the trend for operation at higher frequencies offer a large number of potential multimode seeker configurations.

19.4 SYSTEM FUNCTIONAL OPERATION

There are a number of necessary functions all of which must be successfully accomplished to permit a lethal intercept of the target by a guided missile. These begin with initial target detection and decision to engage and include missile launch, proper operation of the propulsion, guidance, and control systems through the flight, and fuzing and detonation of the warhead at intercept. We shall now consider the radar functions of target acquisition and tracking which provide the intelligence for guidance. Emphasis is on semiactive or active coherent operation unless noted otherwise.

Reference-Channel Operation.[2,5,6,32]   Within the context of a coherent system, the seeker must have available as a reference a precise replica of the illuminating signal. In semiactive systems this has generally been provided by the rear (reference) receiver (although an alternative *on-board reference* approach is also possible). In active systems the reference is derived directly from the transmitter-exciter.

The reference must be spectrally pure (low-noise), and its frequency must accurately represent the illumination frequency to allow the target echo to fall within the bandwidth of the receiver. These requirements are relatively easy to meet in an active system because the same microwave source provides the reference and the transmitter exciter (drive) signal. In semiactive systems, particularly in the early-generation systems in which the transmitters were not crystal-controlled, providing a coherent reference posed a significant challenge.

The early-generation illuminator transmitters generally employed magnetrons or klystrons as power sources which, while possessing good short-term stability and low near-carrier FM noise, lacked the setability and long-term drift charac-

CHAPTER 20
HEIGHT FINDING AND 3D RADAR

David J. Murrow
General Electric Company

20.1 HEIGHT FINDING RADARS AND TECHNIQUES

Early Radar Techniques for Height Finding. Early radar techniques employed to find target height were classified according to whether or not the earth's surface was used in the measurement. The practice of using the earth's surface for height finding was quite common in early radar because antenna and transmitter technologies were limited to lower radio frequencies and broad elevation beams. The first United States operational shipborne radar, later designated CXAM and developed in 1939 by the U.S. Naval Research Laboratory (NRL), used the range of first detection of a target to estimate its height, based on a knowledge of the shape of the pattern near the horizon due to the primary multipath null. Later a refinement was made as the target traversed the higher-elevation multipath nulls or "fades." This technique, illustrated in Fig. 20.1a, was extensively employed on early shipborne radars, where advantage could be taken of the highly reflective nature of the sea surface. Of course, the technique was limited in performance by such uncontrollable factors as sea state, atmospheric refraction, target radar cross section, and target maneuvers.[1,2]

Reflections from the earth's surface were also used by other early contemporary ground-based radars, such as the British Chain Home (CH) series, which was employed in World War II for the defense of Britain. This radar was a pulsed high-frequency (HF) radar which made height measurements by comparing amplitudes of the (multipath-lobed) main beams of a pair of vertically mounted receiving antennas. Conceptualized in Fig. 20.1b, the technique was also utilized in early United States radars, notably, the Canadian-built United States radar SCR-588, and the United States–built SCR-527, both based on the British Type 7 radar design.[3]

One of the earliest and perhaps most direct form of radar height finding was to mechanically direct and hold a narrow-elevation-beam antenna pointed toward the target. The elevation angle of the target corresponds to the elevation readout on the antenna mount. In early radar systems employing this technique, an op-
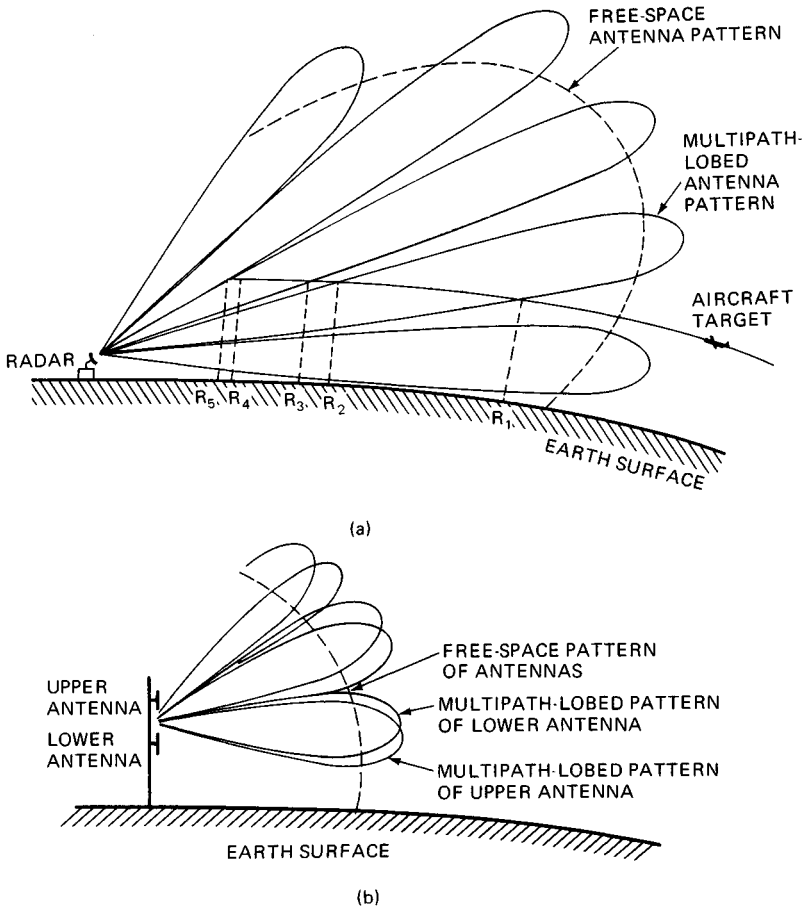
FIG. 20.1 Early radar height finding techniques. (a) Method of multipath nulls. (b) Amplitude comparison using multipath lobes.

erator would keep the antenna boresighted on the target with a handwheel while monitoring the target return strength. It was quickly learned that maximizing the signal strength of a target echo in a beam was not sensitive enough to provide the desired accuracies, and so alternative techniques were ultimately developed for this purpose. One of the first of these, called *lobe switching*, was first demonstrated in 1937 on a prototype of what later became the U.S. Army Signal Corps SCR-268 radar.[4] This radar was designed for directing antiaircraft gunfire and was the first production radar to use lobe-switching techniques to center the antenna on the target. Two separate identical beams, one above and one below the antenna boresight, are formed at the antenna on receive. By switching between the two beams and keeping the observed amplitudes equal, the SCR-268 elevation operator could keep the antenna boresighted on the target accurately.

If a dish antenna, which generates a narrow pencil-type beam in azimuth and elevation, is mechanically boresighted and trained at or in the vicinity of a target,

allowing determination of its azimuth and elevation, the technique is called *searchlighting*. The searchlight technique was successfully employed on the British CMH radar and on the widely deployed United States SCR-584[5] as well as on the United States SCR-615 and the U.S. Navy SM radar. All these radars were S-band dish antenna radar systems. Some of these dish antennas employed conical scanning of a single beam to provide the elevation error signal required to accurately center the beam on the target. The accuracy of such a technique is very good but obviously is limited to one target at a time. Conical scanning and lobe switching are special cases of a general technique for developing off-boresight error signals called *sequential lobing*. The fundamental accuracy and limitations of the sequential-lobing technique are presented in Sec. 20.3. The searchlighting technique was the forerunner of modern-day tracking radars discussed in Chap. 18, many of which now employ monopulse techniques to develop off-boresight error signals. Obviously, techniques which require the antenna to be boresighted on the target are limited in simultaneous surveillance and height finding capability. Typically they make a measurement on a single target at a time and usually also require a designation at least in range and azimuth by an accompanying search radar. The concept of searchlighting and lobe switching is illustrated in Fig. 20.1c.

A widely used early radar dedicated to finding the height of a target in augmentation of a 2D surveillance set was the nodding antenna.* In this type of radar a horizontal fan beam, with a narrow elevation beamwidth, is mechanically scanned in elevation by rocking or "nodding" the entire antenna structure (Fig. 20.1d). As the radar beam traverses the target continuously transmitting pulses, the main-lobe target echoes that return are displayed to an operator by means of a range-height-indicator (RHI) type of display. This allowed the operator to precisely and directly estimate the target height of the target by a process termed *beam splitting*, referring to the process of estimating the center of the displayed target video. Although some nodding-antenna height finders had a slow azimuth rotation search mode, most relied on designations of azimuth from an operator. The operator would observe a detection by the 2D surveillance radar and then command a height determination by the height finder. The height finder would then slew to the commanded azimuth and obtain a height and range measurement. This method of operation was relatively slow and limited in multiple-azimuth target-tracking capability compared with 3D radars. These drawbacks seriously limited the continued use of the manual nodding-antenna height finder in military applications.

Several nodding-antenna height finders, notably the British Type 13 and the widely deployed United States AN/TPS-10, appeared in the mid- to late 1940s, when higher-frequency technology began to emerge.[6] The AN/TPS-10 X-band nodding-antenna height finder radar series was subsequently replaced by the AN/FPS-6, an S-band nodding-antenna radar also designed for the U.S. Army.[7] The AN/MPS-14 was a mobile version of the radar, and the AN/FPS-89 was an improved fixed-site version. The elevation beamwidth of the AN/FPS-6 was 0.9°, its azimuth beamwidth was 3.2°, and the entire antenna nodded at a rate of 20 to 30 nods per minute. The radar could scan in azimuth at a rate of 45° per second. It transmitted 2-μs pulses with pulse repetition frequencies (PRFs) from 300 to 400 Hz, and operated with a peak power of 4.5 MW.

---

*Nodding-antenna height finders have also been used for raid counting.

(c)

(d)
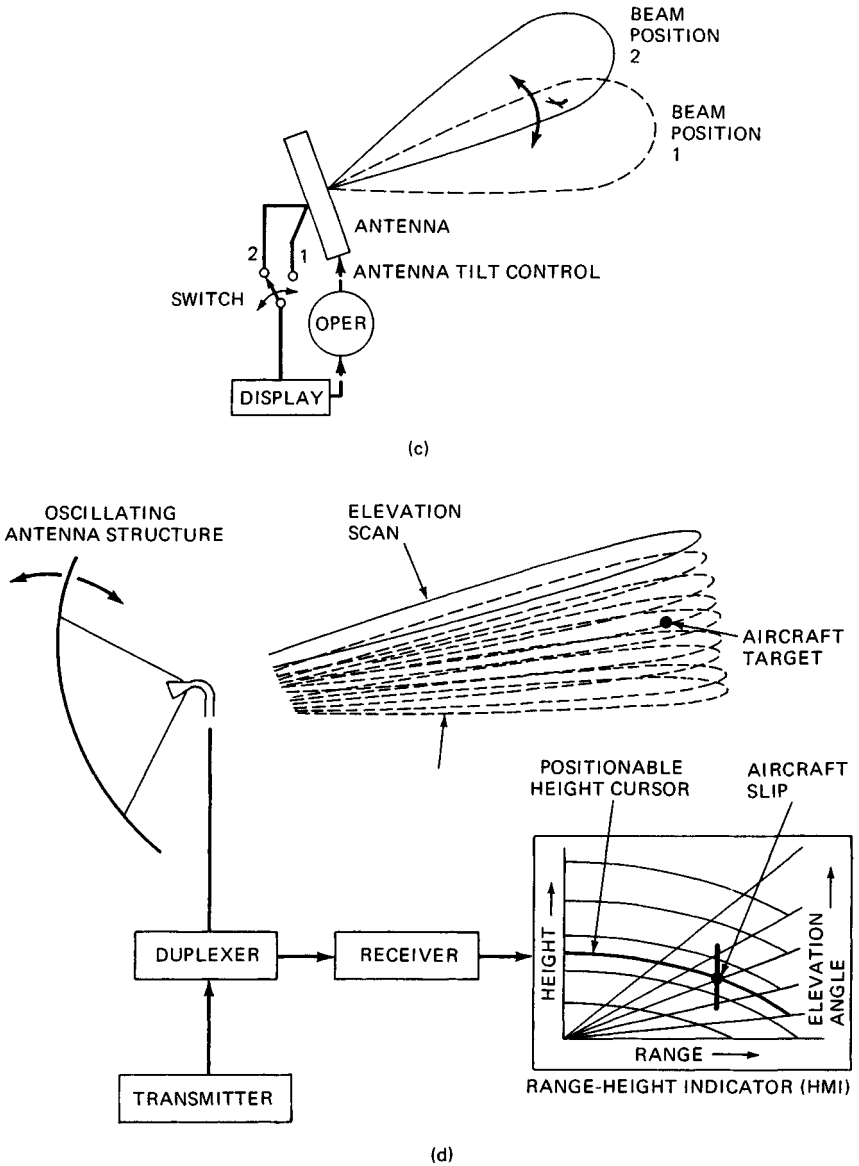
FIG. 20.1 *(Continued)* Early radar height finding techniques. (*c*) Searchlighting with lobe switching. (*d*) Nodding antenna.

The data rates of later versions of the nodding-antenna radar have been considerably improved over their predecessors. For example, the S600 series C-band nodding-antenna height finder is computer-controlled and -managed for maximum data rate, enabling it to obtain up to 22 height measurements per minute.[8]
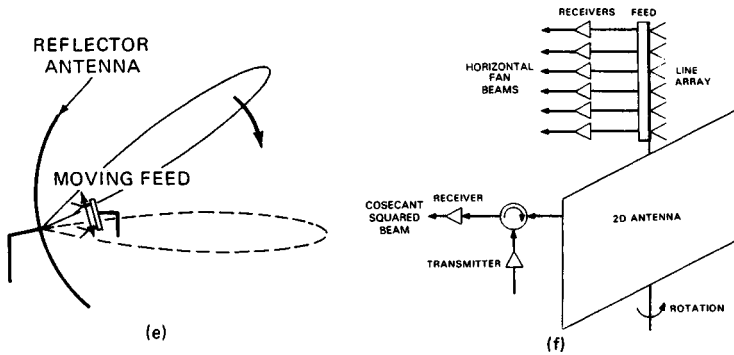
FIG. 20.1 (*Continued*) Early radar height finding techniques. (*e*) Electromechanical beam scanning. (*f*) 2D antenna with vertical line array.

Of course, it is possible to rapidly scan a horizontal fan beam in elevation by electromechanical means (Fig. 20.1*e*) instead of by mechanically rocking the entire antenna structure. Many nodding-beam-type height finders with ingenious means of beam scanning have been successfully deployed over the years. Two notable examples are the World War II SCI radar and the AN/SPS-8 shipborne radar, both of which used a Robinson-type electromechanical feed to rapidly scan the beam in elevation.

The development of higher-frequency microwave technology facilitated electrically larger apertures and correspondingly narrower beams, all in convenient physical sizes. Accompanying this evolution was a series of inventions of rapid electromechanical scanners based on geometric optic principles and developed for surveillance radar applications. These include the Robinson, delta *a* (Eagle), organ-pipe, Foster, Lewis, and Schwartschild scanners, along with a variety of polarization-switching mirror scanners. These scanners all utilized the motion of the feed structure of the antenna to control the incidence of illumination on the aperture, thereby scanning the beam. The reader may refer to a number of excellent sources for a detailed treatment of the method of operation of electromechanical scanners.[6,9,10] In principle, a relatively inexpensive volumetric 3D radar could be created by using an electromechanical feed and scanning a narrow pencil-type beam in elevation while rotating the antenna in azimuth. In practice this approach has not been employed because of the lack of waveform flexibility versus elevation angle imposed by the constant-rate scanning of the electromechanical feed.

During World War II, the British developed a very-high-frequency (VHF) phased array antenna height finder called the Variable Elevation Beam (VEB) radar. This 200-MHz radar utilized mechanically adjustable phase shifters to control the relative phase of nine groups of eight dipoles on a 240-ft mast. The resulting elevation beamwidth was approximately 1° in width and was phase-scanned over an elevation interval slightly more than 6° in extent.

A technique which has seen limited service for radar height finding is a vertical receive-only line array mounted on a conventional 2D surveillance radar as shown in Fig. 20.1*f*. The line array is processed to form a stack of receive horizontal fan beams, each of which is relatively narrow in elevation. Since the (narrow-azimuth) transmit beam is generated by the 2D radar antenna, the resulting stack of two-way beams is narrow in both azimuth and elevation. The stack of

receive beams may be formed in a number of ways from the line array. One technique is the Butler matrix, an RF feed analog of the discrete Fourier transform. A second technique was designed for an experimental version of the FAA AHSR-1 S-band air traffic control.[11] The technique augmented the 2D surveillance aperture with a vertical receive-only line array of elements. Each beam of a vertical stack of horizontal fan beams was generated from the line array by combining energy coupled out of waveguide runs at the appropriate length from the element to produce a linear-phase gradient element to element. This produced a set of uniformly illuminated beams, each time-delay-steered to the desired elevation.

One of the early radars combining 2D surveillance with height finding was the AN/CPS-6B, which utilized the V-beam principle. The V-beam radar consisted of a primary and a secondary antenna aperture mounted on the same rotating shaft. The primary aperture operated as a conventional 2D radar, generating a vertical cosecant-squared fan beam which provided detections and the range and azimuth coordinates of targets in the surveillance volume. The secondary antenna aperture was similar to the primary aperture except that it rotated about an axis normal to the aperture. This produced a second fan beam which was tilted from the vertical plane. The two beams might be powered by the same or separate transmitters, but each beam had its own receiver. The tilted beam provided a second set of detections to the radar operator as the antenna rotated. The azimuth separation of the center of the two sets of detections corresponding to a single target, correlated by the operator using range, was directly proportional to the height of the target, to within flat-earth and normal propogation approximations. The concept of the V-beam radar is illustrated in Fig. 20.1g.

The V-beam radar has been referred to as a 3D radar by some authors.[3,7,10] Technically, however, it should not be classified as a true 3D radar because it lacks resolution in the elevation dimension. This shortcoming limits its use to low-density aircraft situations where it is unlikely to encounter two aircraft appearing at the same range and azimuth but at different heights.

The Japanese have developed a radar based on phase interferometry to find height in air traffic control applications,[12] but it also does not have resolution in the elevation dimension. The concept employs a set of four horizontal line arrays vertically displaced in a staggered fashion about a conventional 2D reflector-type antenna. The principle utilized by phase interferometry is that the phase difference between offset antennas is proportional to the sine of the angle of arrival of a received target echo as sketched in Fig. 20.1h.

Radars such as the VEB, the V-beam, the vertical line array plus 2D, the crossed-line array, and the phase interferometer plus 2D, which obtain simultaneous tricoordinate (range, azimuth, and elevation or height) measurements on a target but do not have significant resolution in the elevation dimension compared with their elevation coverage, might appropriately be termed 2½D radars.

Height Finding Techniques in 3D Radars.   There are many types of radars that provide 3D information by simultaneously measuring the three basic position coordinates of a target (range, azimuth, and elevation). In this handbook, however, the convention is followed in which a *3D radar* is taken to be a surveillance radar whose antenna mechanically rotates in azimuth (to measure range and azimuth) and which obtains the elevation-angle measurement either by scanning one or more beams in elevation or by using contiguous, fixed-elevation beams.

Military interest in 3D radar stems from its ability to determine the height of a noncooperating target, along with its range and azimuth. Because of its better angu-

(g)

(h)
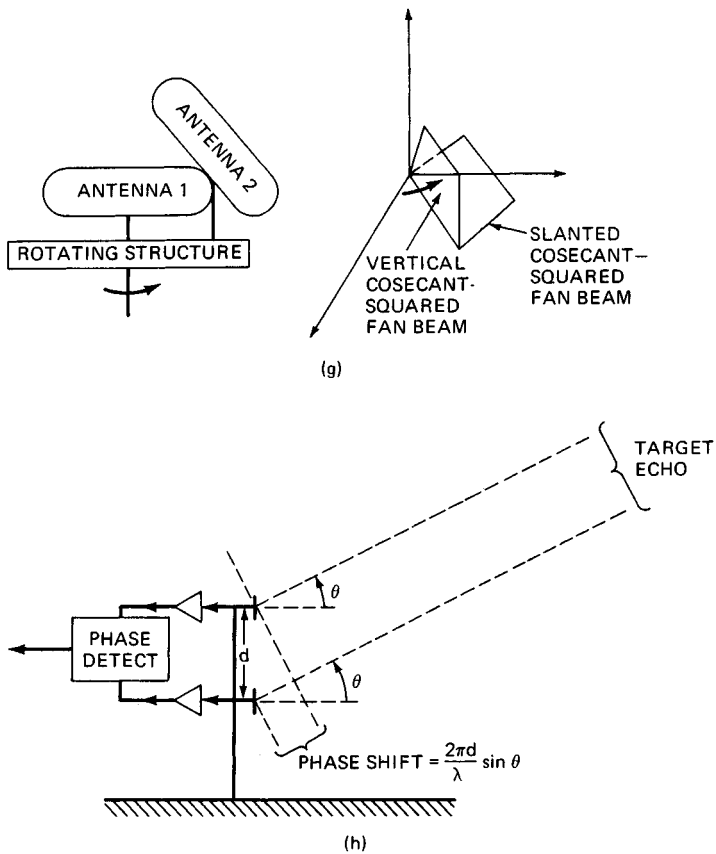
FIG. 20.1 *(Continued)* Early radar height finding techniques. *(g)* V-beam radar. *(h)* Phase interferometry.

lar resolution the 3D radar provides a higher-gain antenna and, arguably, a greater resistance to jamming and other forms of electronic countermeasures (ECM) than a combination of 2D and dedicated height finder. The counterargument points out that the 2D and height finder may be implemented in two separate frequency bands, forcing the jammer to spread out its energy, thereby diluting it.

Rotating 3D radars can be implemented as stacked-beam radars, frequency-scanned radars, phase-scanned radars, electromechanically scanned radars, and digital beamforming radars, according to how the elevation beams are formed and/or scanned in elevation.

Stacked-Beam Radars. Stacked-beam radars employ a vertical *stack* of simultaneously formed receive beams in elevation which are mechanically rotated in azimuth in order to perform search and tricoordinate target position estimation. The target is illuminated by a single transmit beam which is broad enough to cover the receive beam main lobes containing the target. Elevation-angle estimation may be accomplished in such a radar by an amplitude comparison technique, by which the amplitudes of the return at the target range in two or more adjacent

simultaneous elevation receive beams are compared. The technique is thus a special case of the general technique of simultaneous lobing. Target height is determined via computer table lookup, using range and elevation entries. The sensitivity of this approach depends on the relative spacings of the beams in the stack, the aperture illumination used to form the beams, and the elevation angle of arrival of the target relative to the beam boresight placements, along with other equipment-related error sources.

It should be noted that the overall height finding performance of a stacked-beam radar is greatly influenced by the extent to which the designer includes anticlutter moving-target indication (MTI) and/or doppler processing in the beam stack, even in clear weather conditions. This is especially important for the lowest beam in the stack, as its main lobe intercepts the earth's surface, admitting surface clutter returns. However, it may also be important for all the beams in the stack, depending on the severity of the clutter. This is so because the elevation patterns of a stacked-beam radar are primarily one-way patterns, being dominated by the receiver elevation pattern. Thus, the elevation sidelobes protecting the radar upper beams from ground or sea clutter are the (one-way) elevation receive sidelobes. This is true because the transmit beam main lobe must be broad enough in elevation to cover all receive beams. This is different from the scanning pencil-beam radar, in which the product of the transmit and receive elevation sidelobes protects the radar upper beams from surface clutter. In benign surface clutter applications, it is economical to implement the stacked-beam radar without MTI or doppler processing in the beam stack, reserving this processing for a single cosecant-squared receive detection beam.

The AN/TPS-43 is an example of a widely deployed operational stacked-beam radar. Deployed in the 1970s, it is a transportable ground-based S-band radar which has been extensively used for air surveillance in the U.S. Air Force Tactical Air Command System (TACS). The radar employs a multiple-horn feed illuminating a reflector-type antenna rotating at 6 r/min to generate a stack of six receive beams in elevation. The original version of the radar utilized a linear-beam Twystron tube to generate approximately 4 MW of RF peak power in a 6.5-μs simple pulse. The radar is instrumented to a range of 240 nmi and operated with six PRFs averaging 250 Hz. The receive beams in the stack are 1.1 in azimuth and variable in elevation beamwidth in such a way that the six span the 20° of total elevation coverage. Subsequent versions of the radar provided pulse compression and improved MTI waveforms and processing.[13] The AN/TPS-75 is an upgraded version of this radar with a planar array low-sidelobe antenna.

Another example of a stacked-beam radar is the S713 Martello (Fig. 20.2a), an L-band transportable radar with an eight-beam stack. The Martello S713 radar employs IF processing to form the receive beam stack in elevation. In operation, a cosecant-squared transmit beam is formed and eight narrow beams are formed and processed on receive. A ninth receive beam, cosecant-squared in shape, is used for surveillance and detection. Azimuth and range are determined as in a conventional 2D radar. Height finding is accomplished by interpolating the received signal strengths in adjacent elevation beams of the stack to determine the target elevation angle. The array is 10.6 m high by 6.1 m wide and consists of 60 center-fed rows of 32 radiating elements, each equipped with 60 receivers to downconvert received RF to IF. The azimuth beam is 2.8° wide. The tube transmitter generates 3 MW of RF power at the peak of a 10-μs pulse and an average RF power of 8 kW. The radar is instrumented to 256 nmi and up to 30° in elevation and 100 kft in height. The antenna rotates at 6 r/min. A height accuracy of
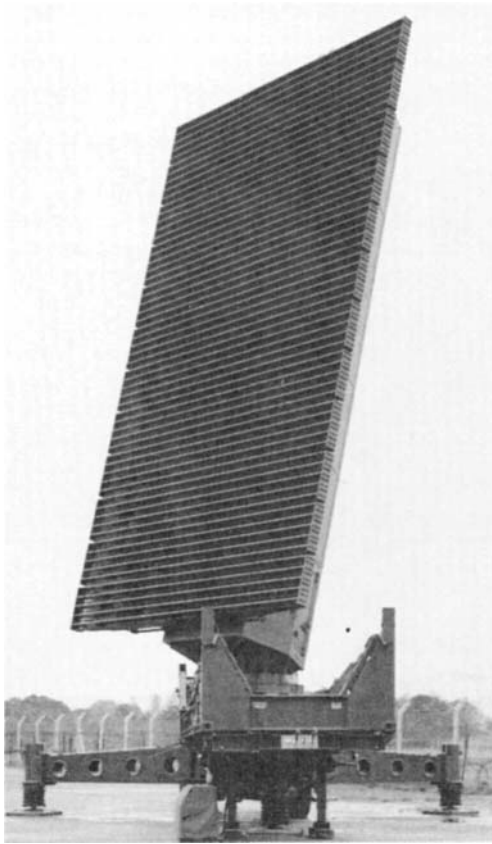
(a)

FIG. 20.2 Exemplar 3D radars. (*a*) S713 Martello stacked-beam 3D radar (*Courtesy Marconi Company*).

1000 ft on a small fighter aircraft at 100 nmi is claimed by the radar manufacturer. A solid-state transmitter version of the radar, the S273 with a shorter and wider array, is also available. This version offers a six- or eight-beam stack, with a 1.4° azimuth beamwidth but with wider beams in elevation covering to 20° total elevation. The solid-state transmitter consists of up to 40 modules generating 132 kW of total RF power at the peak of a 150-μs pulse and up to 5 kW of average power. The height accuracy claimed for the radar by its manufacturer is 1700 ft on a small fighter at 100 nmi.[14]

A radar which is a hybrid mix of stacked beams and phase steering is the RAT-31S, an S-band radar which phase-steers a stack of four beams in elevation to cover the surveillance volume. The radar employs monopulse to determine target height. It rotates a 13.2-ft-square array at 5 to 10 r/min while generating a stack of three receive beams covering 21° in elevation. The array is divided into three vertical sections. Each section of the array then generates its own beam, which is phase-steered over a designated section of the elevation coverage.[15]

Scanning Pencil-Beam Radars.   Another method of achieving 3D volumetric coverage suitable for high-air-traffic situations is the scanning pencil-beam radar. The most common radars in this class obtain high volumetric coverage by employing an antenna feed technique which electronically scans a narrow pencil-type beam through the elevation coverage as it rotates in azimuth, producing an azimuth-elevation scanning pattern similar to that of a TV raster scan.

For the air volume surveillance mission, electronic scanning provides flexibility and performance not readily available in electromechanical scanners. These advantages include (1) shorter volume surveillance frame times; (2) highly flexible computer-programmable waveform versus elevation time and energy management; (3) electronic compensation for moving platforms and mobile applications such as ground vehicles, ships, and aircraft; and (4) wide and flexible elevation coverage, including highly agile beam placement in elevation, programmable elevation coverage versus range and azimuth, good beam shape preservation over wide coverage, and flexible, precise control of beam placement versus azimuth, which is especially critical for low-elevation-beam performance.

Frequency Scanned Radars.   One of the earlier 3D radar techniques that has found application for the air surveillance mission is frequency scanning. Frequency-scanned arrays utilize the frequency-dependent phase characteristics of a length of transmission line, usually waveguide, to scan a pencil beam.[16] The waveguide is folded into a serpentine configuration on the side (or sides) or back of the array to provide output taps at the locations of the closely spaced antenna elements. A controlled change of transmit/receive RF frequency produces a different phase gradient across the aperture, electronically steering the beam to the desired elevation angle. Frequency scanning may be accomplished from pulse to pulse by changing the transmitter and receiver frequency sequentially from one pulse to the next, or "within" the pulse, by transmitting a chirp linear-frequency-modulated (LFM) pulse or sequence of contiguous subpulses each stepped in frequency, and by processing each of a stack of receive beams in elevation each at one of the subpulse frequencies.[17] The AN/SPS-39 S-band shipborne radar used a parabolic-cylinder reflector antenna fed by a line source to produce the change in phase with frequency necessary to scan its beam electronically in elevation. Upgraded with a planar array, this radar evolved into the AN/SPS-52[18] (Fig. 20.2b). The within-pulse approach was employed on the AR3D S-band surveillance radar. It transmits LFM pulses and extracts the target height via frequency discrimination in the receiver.[19]

The U.S. Marine Corps AN/TPS-32, the U.S. Navy shipborne AN/SPS-48, and the Series 320 radars are all examples of S-band 3D surveillance radars consisting of a small stack of frequency-scanned beams which are then step-scanned as a group to cover the elevation surveillance volume.[20,21]

The use of frequency-scanning beams in elevation as a height finding technique is a form of the general technique of sequential lobing, in which amplitudes from adjacent sequentially formed beams are compared to estimate the target elevation angle. The elevation-angle accuracy achievable in this class of radars is not as good as that of stacked-beam or phase-scanned monopulse radars, e.g., radars employing simultaneous lobing. There are several fundamental reasons for this. One is that because different frequencies are required in order to steer the beam, amplitude fluctuations in the target return are induced. These tend to dilute the quality of target angle information available in the multiple-beam target returns. The effect can be compensated by averaging out the target fluctuation effects by the use of noncoherent integration of multiple-frequency diversity subpulses in each beam. However, the diversity subpulse frequency separations
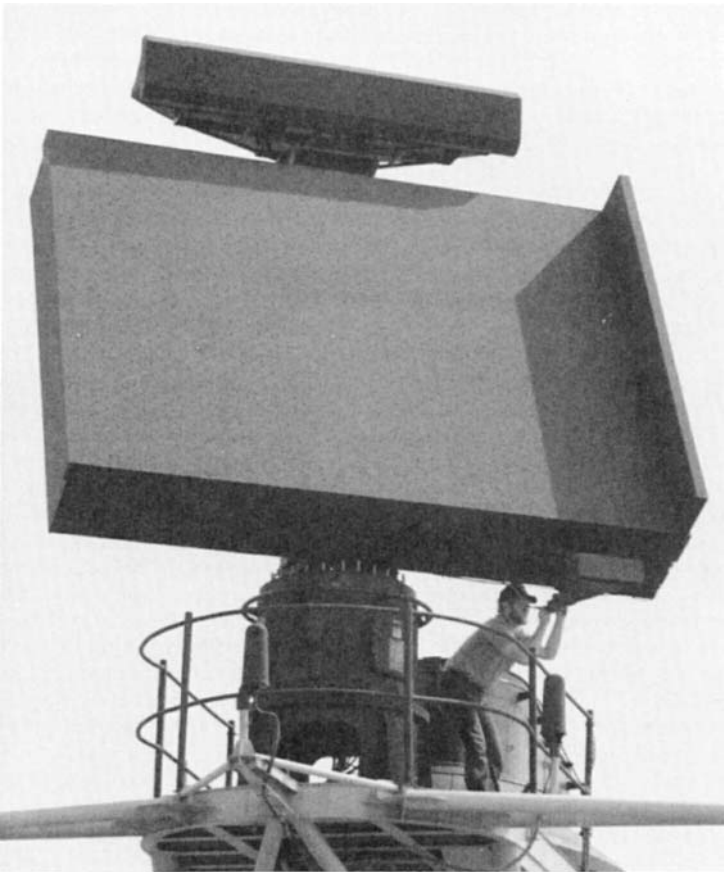
(b)

FIG. 20.2 *(Continued)* Exemplar 3D radars. *(b)* AN/SPS-52C shipboard frequency-scanned 3D radar *(Courtesy Hughes Aircraft Company)*.

must be enough to induce target amplitude fluctuations while not causing too much beam steering—a difficult tradeoff in some applications. Sequential-lobing angle estimation techniques are also vulnerable to time-varying or amplitude-modulation jammers such as blinkers. The fact that RF frequencies correspond one to one with elevation angle in the frequency-scanned radar constrains it in the use of frequency agility for electronic counter-countermeasures (ECCM) purposes. It also tends to limit its flexibility in waveform time and energy management. The electronically steered phased array provides considerable relief to the designer from these limitations.

Phased Array Radars.   Scanning or steering of a pencil-type narrow beam in elevation can be accomplished by means of electronically controlled phase shifters placed at the row feed outputs of an array antenna. This approach is the most flexible of the various 3D radar height finding techniques, allowing full use of the frequency band for purposes beyond beam scanning and allowing for com-

plete independence of waveform and beam position. Height finding techniques which can be used with the phase-scanned array include a variety of coherent simultaneous-lobing (monopulse-multipulse* and phase-interferometry) techniques, as well as amplitude comparison sequential-lobing techniques. The phased array radar is becoming more commonplace in the present-day military marketplace, owing to an ever-escalating target and threat environment and dynamics.

The AN/TPS-59 L-band radar is an example of a long-range transportable 3D tactical radar with phase scanning to steer the beam in elevation. Developed for the U.S. Marine Corps, it is unique among air surveillance radars in that it was the first to employ an all-solid-state transmitter. The solid-state transmitter of this radar is distributed over the antenna aperture in the form of individual row transmitter units. The total transmit power is combined only in space, in the far-field collimated beam. The planar array antenna consists of 54 rows of horizontal stripline linear arrays. Each of the 54 rows contains its own solid-state modular transceiver consisting of a 1-kW nominal RF peak-power solid-state transmitter, integral power supply, low-noise receiver, phase shifter, duplexer, and logic control, all mounted on the antenna. The feed structure of the 15-ft by 30-ft planar array generates a full two-axis monopulse beam set on receive, consisting of a sum and two delta beams. An additional column feed provides a special low-angle height finding capability for the lowest angle beam positions. The feed generates a pair of squinted sum-type beams carefully placed in elevation and processed as a monopulse pair. The technique minimizes the effects of multipath. Its fundamental accuracy performance is considered in Sec. 20.3. The 1.6° by 3.2° monopulse beam set is electronically phased-scanned from -1 to 19°.[22–24] Fixed-site variants of this radar are the AN/FPS-117 SEEK IGLOO radar (Fig. 20.2c) and the GE-592 radar, both of which are distributed aperture solid-state and similar to the AN/TPS-59 but which employ a 24- by 24-ft array antenna with 44 rows and additional digital signal processing. The square-aperture array of the GE-592/FPS-117 radar generates a 2.2° azimuth by 2° elevation two-axis monopulse beam set.[25–28]

The HADR, deployed mainly in Europe for North Atlantic Treaty Organization (NATO) applications, is a ground-based 3D S-band phased array radar which also uses phase scanning in elevation and mechanical rotation in azimuth. The radar's 12.5-ft by 16-ft planar array rotates at 5 r/min while phase-scanning a single pencil-type 1.45° azimuth by 1.9° elevation beam through 12 long-range search beam positions and 4 MTI beam positions in elevation. The instrumented coverage of the radar is 250 nmi in range, 20° in elevation, 360 degrees° in azimuth, and 120 kft in height. Target height is estimated by using sequential lobing between contiguous beams in elevation.[29]

A significant example of a long-range airborne 3D surveillance radar is the AWACS (Airborne Warning and Control System) AN/APY-1 S-band radar used on the E-3A aircraft. Because of the limited vertical aperture extent of the AN/APY-1 radar, the elevation beam is relatively broad. Consequently the height accuracies achieved by the radar do not compare well with those of its surface-based counterparts.

---

*The term *multipulse* is used by the author to refer to a target angle estimation technique discussed in Sec. 20.3 for coherently combining received monopulse sum- and delta-channel (I,Q) target echo samples from multiple-pulse transmissions. This technique is to be distinguished from the combining of individual monopulse angle measurements from each of the multiple-pulse transmissions.
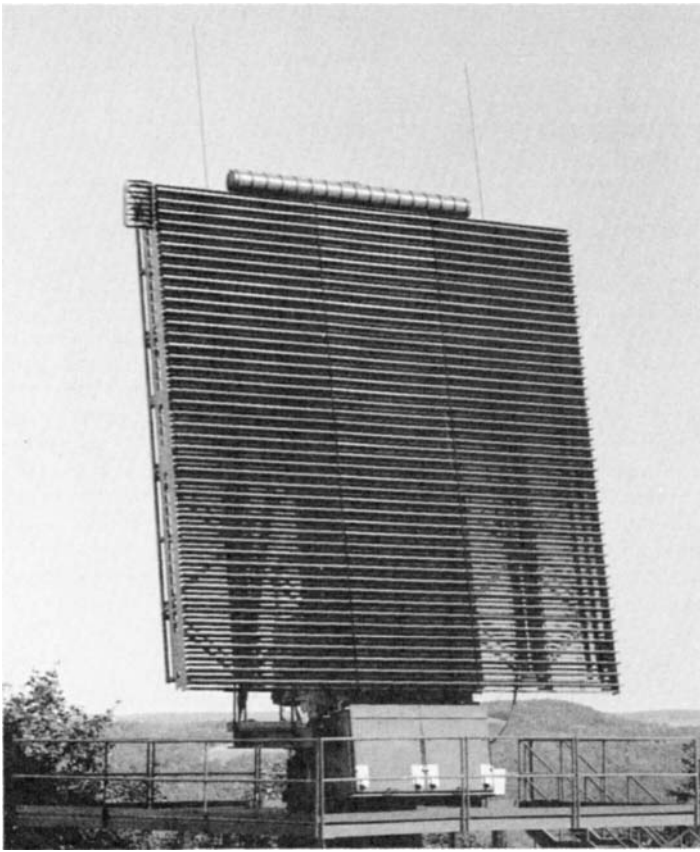
(c)

FIG. 20.2 *(Continued)* Exemplar 3D radars. (*c*) AN/FPS-117 fixed-site solid-state phase-scanned 3D radar (*Courtesy General Electric Company*).

Digital Beamforming Radar.   A technology with considerable attractiveness for radar is digital beamforming. As a technique for finding target height, digital beamforming involves placing a receiver on each element of a vertical array of elements, or rows of elements. By digitally weighting and linearly combining the analog-to-digital (A/D) converted receiver outputs, a stack of receive beams or a single scanning receive beam in elevation can be generated. In this form, a digital beamforming radar is a special case of a stacked-beam radar, implemented in a technology that offers several advantages over conventional stacked-beam technology. The major advantage offered by digital beamforming technology is that of full adaptive control of the beam patterns for ECCM purposes. The major challenge faced by digital beamforming radar designers relative to height finding is to develop techniques to preserve monopulse ratios in the presence of adaptive array cancellation of jamming, including those in the main lobe. Monopulse beam pairs or stacked beams are easily generated digitally, but the accuracy of height

finding depends on a precise, unambiguous knowledge of the relative patterns of the (adapted) height finding beams.

20.2 DERIVATION OF HEIGHT FROM RADAR MEASUREMENTS

Height in radar is always a derived rather than a measured quantity. This is true because a radar can only measure range and angle of arrival of target returns. Surface-based radars derive the height of a target from the range (time) of the echo return and elevation coordinate measurements. A radar on a ship, aircraft, or space satellite may be required to convert tricoordinate measurements relative to the antenna to an inertial reference system as part of the height calculation. The accurate calculation of height from radar measurements must provide for such effects as the location and orientation of the radar antenna in the desired reference coordinate system, the curvature of the earth, the refractive properties of the atmosphere, and the reflective nature of the earth's surface. Furthermore, if the target height is to be referenced to the local terrain, then the height of that possibly irregular terrain below the target must also be taken into account. The effects of some of the systematic internal equipment errors can also be partially offset by incorporating internal calibration measurements into the range and angle estimation algorithms.

Flat-Earth Approximation.   For very-short-range targets, a sufficiently good estimate of target height is given by the flat-earth approximation:

$$h_T = h_a + R_T \sin \theta_T \qquad (20.1)$$

where h_a is the radar antenna height, R_T is the measured target range, θ_T is the measured or estimated target elevation angle, and h_T is the estimated target height.

Spherical Earth: Parabolic Approximation.   A somewhat better approximation to target height which models the earth's curvature as parabolic in range can be derived by reference to Fig. 20.3b. For a radar located near the surface of the earth, it can be readily shown from the law of cosines that, to a first approximation,

$$h_T = h_a + R_T \sin \theta_T + R_T^2/2R_0 \qquad (20.2)$$

where r_0 is the radius of the earth and the other parameters are as defined above.

The height calculated with the above curved-earth algorithm exceeds that calculated by using the flat-earth algorithm, increasing quadratically with measured target range. The discrepancy reaches a value of about 88 ft for a measured target range of 10 nmi.

Spherical Earth: Exact Geometry.   Again with reference to Fig. 20.3b, the exact target height can be calculated as follows:

$$h_T = [(R_0 + h_a)^2 + R_T^2 + 2(R_0 + h_a)R_T \sin \theta_T]^{1/2} - R_0 \qquad (20.3)$$
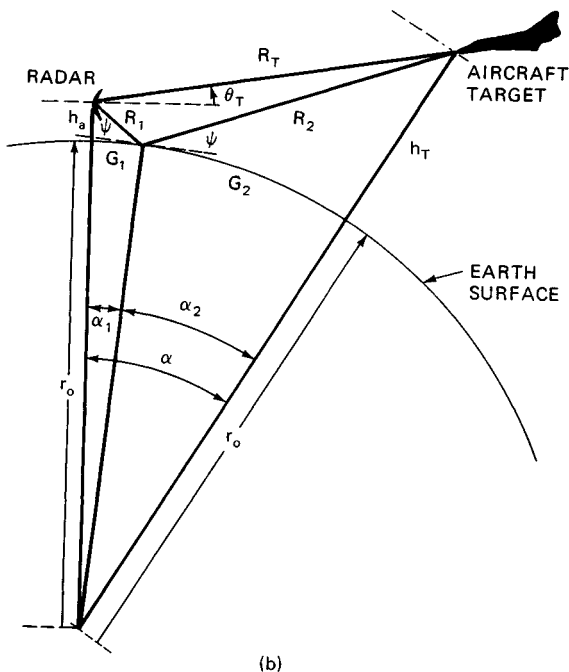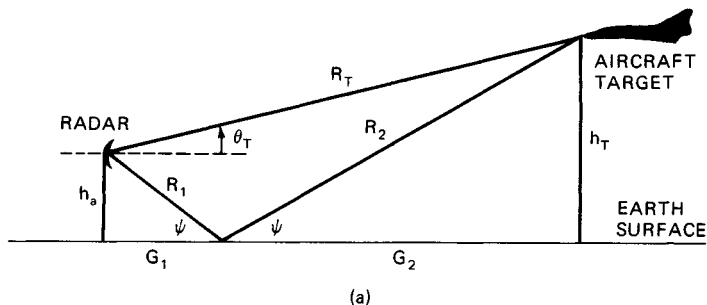
FIG. 20.3 Geometric considerations. (*a*) Flat-earth geometry. (*b*) Spherical-earth geometry.

Corrections for Atmospheric Refraction.* To further improve the accuracy of height computation, refraction of the radar beam along the ray path to the target must be taken into account. In free space, radio waves travel in straight lines. In the earth's atmosphere, however, electromagnetic waves are generally bent or refracted downward. The bending or refracting of radar waves in the atmosphere is caused by the variation with altitude of the index of refraction, which is defined as the ratio of the velocity of propagation in free space to the

---

*The data tables, figures, and portions of the following discussion on corrections for atmospheric refraction were extracted from and follow closely the original text of Burt Brown's Chap. 22, "Radar Handbook," 1st ed., edited by Merrill Skolnik.[10]

(c)

(d)
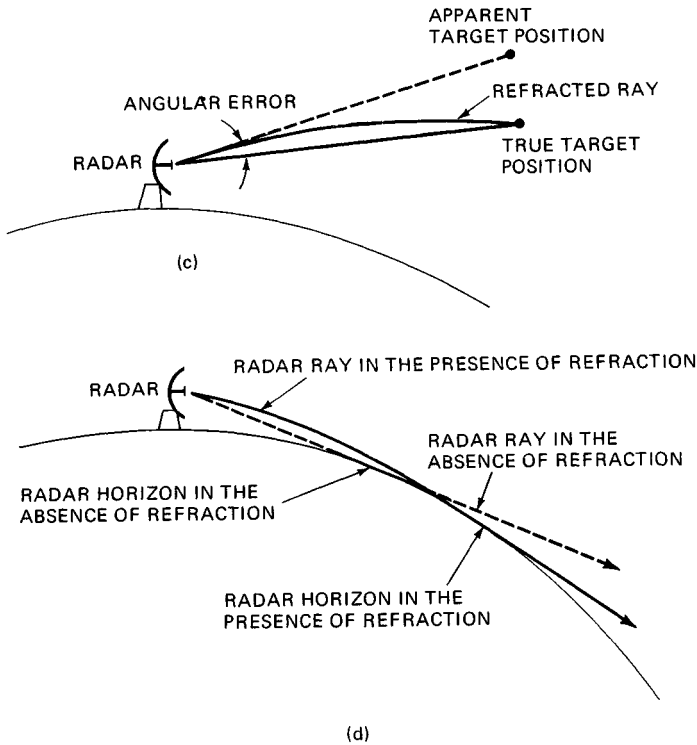
FIG. 20.3 *(Continued) (c)* Effect of refraction on radar horizon. *(d)* Angular error due to refraction.

velocity in the medium in question. One effect of refraction is to extend the radar distance to the horizon, as suggested in Fig. 20.3c. Another effect is the introduction of errors in the radar measurement of elevation angle. In the tropospheric portion of the atmosphere, the index of refraction n is a function of such meteorological variables as temperature, pressure, and water vapor and can be represented by[30]

$$(n - 1) \times 10^6 = N = \frac{77.6p}{T} + \frac{3.73 \times 10^5 e}{T^2} \tag{20.4}$$

where T = air temperature, K; p = barometric pressure, in millibars; and e = partial pressure of water vapor, in millibars. The parameter N is a scaled index of refraction termed *refractivity*.

   Since the barometric pressure p and the water vapor content e decrease rapidly with height, the index of refraction normally decreases with increasing altitude. In a standard atmosphere, the index decreases at a rate of about 4.5×10^{-8} per meter of altitude. A typical value of the index of refraction at the surface

of the earth is of the order of 1.0003. The Cosmic Ray Physics Laboratory (CRPL) standard atmosphere has been defined as one having an index of refraction of 1.000313 (or 313 N units for the refractivity N) and having an exponential decrease of refractive index with altitude.

$$N = N_s \exp (- ah) \qquad (20.5)$$

where $N_s = 313$ N units is the surface refractivity and $a = 0.04385$ per kft when h is in thousands of feet.

The classic method of accounting for atmospheric refraction in radar height computations is to replace the actual earth radius R_0 ($= 3440$ nmi) by an equivalent earth of radius $R_e = kR_0$ and to replace the actual atmosphere by a homogeneous atmosphere in which electromagnetic waves travel in straight lines rather than curved lines (Sec. 2.6). It may be shown by Snell's law in spherical geometry that the value of the factor k by which the earth's radius must be multiplied in order to plot the ray paths as straight lines is

$$k = \frac{1}{1 + R_0 \, (dn/dh)} \qquad (20.6)$$

where dn/dh is the rate of change of refractive index n with height. The vertical gradient of the refractive index dn/dh is normally negative. If, contrary to the CRPL standard atmosphere assumption, it is assumed that this gradient is constant with height, the value of k is 4/3. The use of the 4/3 effective earth's radius to account for the refraction of radio waves has been widely adopted in radio communications, propagation work, and radar.[31] The height calculated by using a 4/3 effective earth's radius is less than that calculated by using the actual earth radius, the difference increasing quadratically with measured target range, attaining a value of about 22 ft at 10 nmi.

The distance d to a horizon from a radar at height h_a may be shown from simple geometry to be approximately

$$d = \sqrt{2kR_0h_a} \qquad (20.7)$$

where h_a is assumed to be small compared with R_0. For $k = 4/3$, the above expression reduces to a particularly convenient relationship if d and h_a are measured in nautical miles and feet respectively:

$$d(\text{nmi}) = 1.23\sqrt{h_a(\text{ft})} \qquad (20.8)$$

Refined computations of the angular deviations introduced when electromagnetic waves traverse a medium other than free space are discussed elsewhere.[32,33] Estimates of the height error for a target located at an altitude of 100,000 ft based on CRPL Reference Refractivity Atmosphere—1958, are contained in Table 20.1. It is noted that the magnitude of the height error is directly related to the surface refractivity and that above approximately 40° elevation angle the height error is independent of the surface refractivity. The height error is given in Table 20.2 as a function of slant range.

TABLE 20.1 Estimates of Height Error at an Altitude of 100,000 Ft Based on the CRPL Reference Refractivity Atmosphere—1958

Elevation angle,°	Height error, kft		
	Surface refractivity, $N_0 = 280$	Surface refractivity, $N_0 = 315$	Surface refractivity, $N_0 = 370$
1	9.14	11.12	14.73
2	5.65	6.75	8.63
4	2.63	3.08	3.82
6	1.44	1.68	2.06
8	0.89	1.03	1.26
10	0.60	0.69	0.84
15	0.28	0.32	0.39
20	0.16	0.18	0.22
40	0.04	0.04	0.05
70	0.01	0.02	0.02

TABLE 20.2 Estimate of Height Error at Slant Ranges of 100, 200, and 300 nmi Based on the CRPL Reference Refractivity Atmosphere—1958

Elevation angle,°	Height error, kft					
	Slant range, 100 nmi		Slant range, 200 nmi		Slant range, 300 nmi*	
	$N_0 = 280$	$N_0 = 370$	$N_0 = 280$	$N_0 = 370$	$N_0 = 280$	$N_0 = 370$
1	1.61	2.68	5.08	8.07	9.14	14.73
2	1.38	2.32	4.20	6.34		
4	1.12	1.73				
6	0.93	1.36				
8	0.78	1.13				

*Approximate slant range.

TABLE 20.3 Comparison of Heights Based on 4/3-Earth's-Radius Principle with Heights Based on the Exponential Model*

Elevation angle,*	Slant range, 100 nmi			Slant range, 200 nmi			Slant range, 300 nmi		
	h_{exp}	$h^{4/3}$	Δh†	h_{exp}	$h^{4/3}$	Δh	h_{exp}	$h^{4/3}$	Δh
0.0	6.9	6.8	0.1	28.0	26.8	1.2	65.1	60.2	4.9
0.5	12.3	12.1	0.2	39.4	37.4	2.0	82.9	75.9	7.0
1.0	17.8	17.5	0.3	50.6	48.0	2.6			
2.0	28.6	28.1	0.5	72.8	69.1	3.7			
4.0	50.0	49.2	0.8						

*After Bauer, Mason, and Wilson.[34,35]
†$\Delta h = h_{exp} - h_{4/3}$; all heights in kilofeet.

A comparison of the heights based on a 4/3-earth's-radius principle with the heights based upon the exponential model is illustrated in Table 20.3. The data shows that, for a given elevation angle, the difference in height computation increases with slant range and that, for a given range, the height difference increases with elevation angle.

Compensation for Surface Refractivity Variation. For extremely accurate height calculations at long ranges, it is possible to correct for variations in surface refractivity in otherwise normal atmospheric refraction conditions. Such a technique is used, for example, in the General Electric series of solid-state 3D radars. The approach is to use offline ray tracing with an exponential model for atmospheric refraction. The resulting heights are then pretabulated as a function of elevation angle, range, and surface refractivity along with the partial derivatives of the height function with respect to the three above variables. These calculations are then stored in the radar computer database. In normal radar operation, the surface refractivity is measured periodically at the radar site, where it is used in conjunction with measured target elevation θ_T and range R_T to perform online table lookup of the tabulated height-refractivity parameters. The final height is computed by means of the interpolation

$$h_T = h_T(R_k, \theta_k, N_k) + \frac{\partial h_T}{\partial R_k}(R_T - R_k) + \frac{\partial h_T}{\partial \theta_k}(\theta_T - \theta_k) + \frac{\partial h_T}{\partial N_k}(N - N_k) \tag{20.9}$$

where R_k, θ_k, and N_k are the closest stored values of range, elevation angle, and surface refractivity to the measured values.

Practical Corrections

Terrain Height Adjustments. If the height of the target above local terrain is to be obtained, the height relative to mean sea level must be corrected by the height of the terrain below the target. This involves calculation of ground range from target slant range and elevation angle and computer lookup of terrain height versus ground range and azimuth.

Platform Location, Orientation, and Stabilization. The calculation of target height with a radar on a moving platform, such as a ship, aircraft, or satellite (all of which are subject to uncertainties in location and orientation) is somewhat more complicated. Coordinate conversion of measured target range, azimuth, and elevation is necessary to determine the target height. Platform location and orientation must be sensed, and perhaps stabilized, and provided to the radar computer. Some of these quantities are also required for platform navigation and therefore may be available from the navigation gyros.

20.3 HEIGHT ACCURACY PERFORMANCE LIMITATIONS

The accuracy of the measurement of target height with a radar is conveniently expressed in terms of the root-mean-square error (rmse), i.e., the square root of the expected value of the square of the difference between the estimated target height and the actual target height. Because height is a derived quantity from the basic radar measurements of range and elevation angle, height accuracy can be expressed in terms of the rms errors associated with those measurements, as suggested in Table 20.4. The remainder of this section is devoted to analysis of the errors involved in the basic radar measurements, primarily elevation angle.

All radar measurements are in error because of the contamination of the received-signal echo with thermal noise. The common assumption about the nature of thermal noise, well justified by practical experience, is that it is a narrowband zero-mean gaussian random process. A particular pair of samples consisting of in-phase (I) and quadrature (Q) components can be properly

TABLE 20.4 Relationship of Target Height Error to Radar Range and Elevation Angle Measurement Errors*

Flat earth:
$$\sigma_h = (\sigma_R^2 \sin^2 \theta + R^2 \sigma_\theta^2 \cos^2 \theta)^{1/2}$$
Spherical earth: parabolic approximation:
$$\sigma_h = [\sigma_R^2 (R/R_e + \sin \theta)^2 + R^2 \sigma_\theta^2 \cos^2 \theta]^{1/2}$$
Spherical earth: exact geometry:
$$\sigma_h = \{[\sigma_R^2 (R^2 + (R_e + h_a)^2 \sin^2 \theta) + (R_e + h_a)^2 R^2 \sigma_\theta^2 \cos^2 \theta]/(R_e + h)^2\}^{1/2}$$
R_e = effective earth radius = kR_0
h_a = antenna height above earth surface
R = radar-measured target range
θ = radar-measured target elevation angle
h = radar-measured target height
σ_R = rmse of radar range measurement
σ_θ = rmse of radar elevation-angle measurement
σ_h = rmse of target height estimate

*Colocated and exactly known platform-antenna location and orientation and small measurement errors relative to target coordinate values are assumed.

viewed, therefore, as a single complex zero-mean gaussian random variable.* The rmse associated with the accuracy performance of a particular radar angle measurement technique, as limited solely by the presence of thermal noise on the technique, is termed herein the *fundamental accuracy* of the technique.

The fundamental accuracy of two general categories of elevation-angle estimation techniques is presented in this section: the sequential-lobing technique and the simultaneous-lobing technique. Other practical effects influencing the height accuracy of a radar system include: beam-pointing errors, pattern errors, channel mismatch errors, calibration, platform orientation and gyration, stabilization, compensation, ECM-ECCM and clutter errors, multipath, target fluctuations and thresholding effects, and multiple hit and channel combining.

Fundamental Accuracy of Sequential Lobing. Sequential lobing is a technique used in radar for estimating the angle of arrival of electromagnatic radiation incident on an antenna by comparing the amplitudes of the received echoes in two or more sequentially formed or selected antenna beams. The technique is used for height finding by time—sequentially scanning a beam in elevation while transmitting and receiving pulses in each beam position. The pulse amplitudes in each beam position are envelope-detected and stored for use in a comparison with those from the other beams. The simplest form of radar sequential lobing compares the envelope-detected returns from a single pulse in each of two adjacent beams. The ratio of the detected pulse amplitude in one beam position to that in the other forms the basis for a table lookup or readout of target elevation angle. In early radar the readout was a calibrated dial or display. In 3D radars, computers provide elaborate lookup tables to relate the ratios to the target elevation angle.

The envelope-detected target returns in each of the beam positions are corrupted by thermal noise even in the most ideal of circumstances. This noise is due

*The term *complex* here refers to $I + jQ$. The complex sample has zero mean because of its random uniformly distributed phase on the interval $(0, 2\pi)$.

to thermally generated electronic noise in the radar receivers and to noiselike electromagnetic emissions from the sky and ground entering the antenna.

For a nonfluctuating target, a good approximation for the fundamental accuracy of the sequential-lobing technique for a large signal-to-noise ratio* is

$$\text{rmse} = \frac{1}{|\dot{f}|}\left(\frac{1 + f^2}{2x}\right)^{1/2} \tag{20.10}$$

where, if θ is the target angle and $\hat{\theta}$ is its estimate, the rmse is defined by $[E(\hat{\theta} - \theta)^2]^{1/2}$. The various factors in the above rmse are defined as

$f = f(\theta) = G_2(\theta)/G_1(\theta)$ = ratio of two-way elevation beam power patterns

$\dot{f} = df/d\theta$; gives rmse in radians or milliradians (mrad)

$\dot{f} = df/d(\sin\theta)$; gives rmse in sines or millisines (msine)

$G_2(\theta)$ = two-way elevation beam power pattern in beam position 2

$G_1(\theta)$ = two-way elevation beam power pattern in beam position 1

x = signal-to-noise ratio in beam position 1

Note that this result may be put into the familiar form

$$\text{rmse} = \frac{1}{K\sqrt{2x_0}} \tag{20.11}$$

where $K = \dfrac{|\dot{f}|\,|g_1|}{(1 + f^2)^{1/2}}$ \hfill (20.12)

g_1 = two-way normalized voltage pattern in beam position 1

$= G_1^{1/2}/G_{10}^{1/2}$; G_{10} = boresight antenna power gain in beam 1

and, x_0 = boresight signal-to-noise ratio

The constant K in this form is a measure of the sensitivity of the angle estimation technique, in that the larger K is, the smaller are the rms errors and the better is the angle accuracy. A single value of the sensitivity (at beam peak or at crossover) is often used to compare techniques. However, care must be taken in this practice, because K is a strong function of the antenna beamwidth and of the target angle of arrival relative to the position of the beams in angle space. As expressed above, K includes the signal-to-noise-ratio dependence on the target angle through the two-way pattern of the beam. An alternative formulation could leave this pattern factor in the signal-to-noise ratio and omit it from the definition of K.

The fundamental accuracy of the sequential-lobing technique for a scanning beam generated by a uniformly illuminated aperture on transmit and receive is presented in Fig. 20.4a. Performance is presented in terms of a normalized ver-

---

*The signal-to-noise-ratio definition used here is E/N_0, where E is the received pulse energy and $N_0/2$ is the spectral power density of the interfering thermal noise.
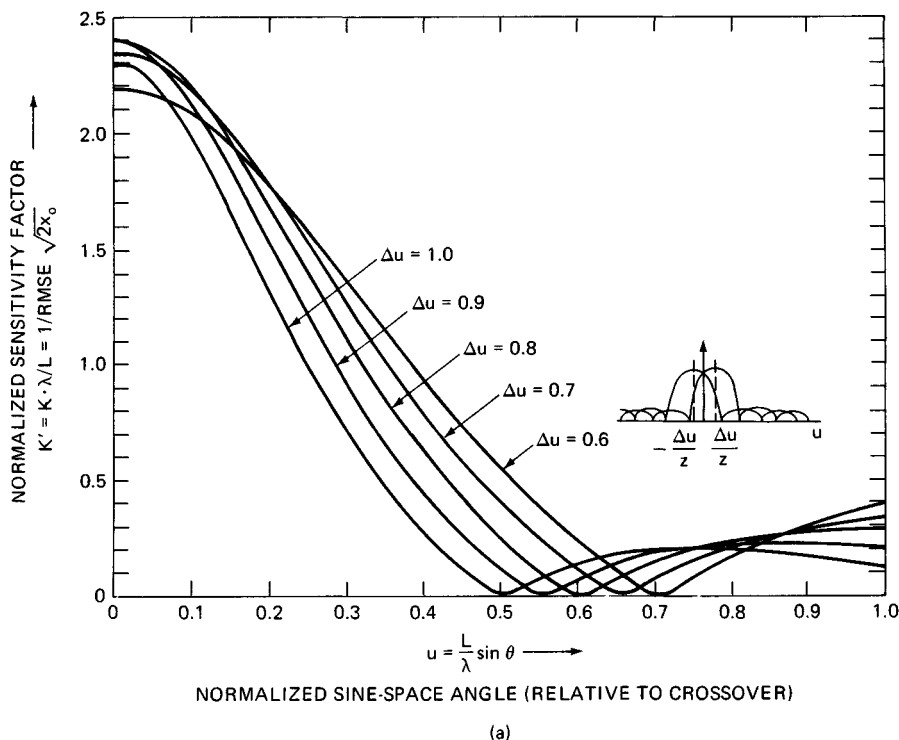
FIG. 20.4 Fundamental accuracy. (*a*) Two-beam sequential lobing: uniform (sin $\pi u/\pi u$) sum beams transmit and receive; separation = Δu; nonfluctuating target; N = one pulse per beam.

sion of the sensitivity factor $k = K\lambda/L$ versus a similarly normalized sine-space elevation angle of arrival, $u = L/\lambda \sin (\theta)$. The elevation angle θ is referenced to the crossover point halfway between the beam peaks. This normalization removes the aperture dimension L and transmit wavelength λ from the performance of the technique. The figure shows that the sensitivity factor peaks at crossover and is symmetrical about the crossover angle. The value of the normalized sensitivity factor at crossover depends on the two-way beam shape–aperture illumination and the step size between the beams. For a uniformly illuminated aperture whose beam is stepped by $\Delta u = 1$, between single-pulse transmission/reception, the normalized sensitivity factor associated with the two-beam sequential-lobing technique against a nonfluctuating target attains a value of approximately 2.15 at the crossover angle. Thus, for example, if the aperture height is 24 ft (7.3 m) and the radar is L band ($\lambda = 0.23$m), the appropriate normalization factor is $L/\lambda = 31.75$, so that the actual sequential-lobing sensitivity factor for a target at crossover is 68.25 V/(V · sine*), or 0.06825 V/(V · msine*). If the boresight

---

*A *sine* or a *millisine* is a unit of measure of the sine of an angle. For example, an angle of 0.7 rad (700 mrad) corresponds to sin (0.7)   0.64422 sine = 644.22 msines.

signal-to-noise ratio is 100 (20 dB), the rmse is easily calculated as rmse = 1.04 msines. Furthermore, if the beams have been electronically steered so that their crossover elevation angle is, say, 30° away from the antenna broadside, the angular accuracy can be calculated as rmse = 1.04/cos (30°) = 1.2 mrad.

If the beams are too closely spaced, sensitivity suffers because there isn't sufficient difference in the received echo strength to measure the angle of arrival accurately. On the other hand, if they are separated too far, there isn't sufficient signal-to-noise ratio in one of the beams for accuracy. It follows that there is a beam step size that optimizes accuracy by maximizing sensitivity and minimizing errors. This is clearly illustrated in Fig. 20.4a, which shows a maximization of the sensitivity factor for a beam spacing between $\Delta u = 0.8$ and $\Delta u = 0.9$. It can be shown analytically (for gaussian beam shapes) that the optimum spacing between beams is $\Delta u = 0.85$. It is also significant to note that the valid range of angular (u-space) coverage for the sequential-lobing technique using uniform illuminated sum beams is approximately $2 - \Delta u$, where Δu is the beam step size between transmission in u space. Within this region, the target is in the main lobe of each of the two beams, and its angle is uniquely and unambiguously determined by the ratio of the echo strengths in the two beams. Outside this region the target is in the sidelobes of at least one of the beams, and its angle of incidence cannot be unambiguously estimated with the technique. In practice, sidelobe responses to a target are eliminated by the use of sidelobe blanking. At the optimum beam spacing of $\Delta u = 0.85$, the valid angular coverage is $2 - 0.85 = 1.15$, or ± 0.575 about the crossover angle. Coverage can be increased from this value by decreased step size or by aperture weighting, but only at the expense of sensitivity.

It is possible to utilize more than two beam positions in the sequential-lobing estimation algorithm. In such an approach, performance is improved when the beam scans a small amount in angle between transmissions. As the beam steps past the target in elevation, it is possible to display intensity, or target echo strength, on a display or other indicator at the angular locations of the beam. A centroidal interpolation of the angular locations as weighted by the receive pulse amplitudes may be used to extract the target angle estimate. The mathematically equivalent estimation process is

$$\hat{\theta} = f^{-1} \frac{\displaystyle\sum_{k=1}^{N} |r_k| \theta_k}{\displaystyle\sum_{k=1}^{N} |r_k|} \tag{20.13}$$

where r_k = complex $(I_k + jQ_k)$ sample at receiver–pulse matched-fiber output
  θ_k = boresight elevation of beam position k
  N = number of beam positions used in algorithm

and $f(\theta) = \displaystyle\sum_{k=1}^{N} \theta_k \, g_k \, \bigg/ \sum_{k=1}^{N} g_k$ $g_k = g_k(\theta) = G_k^{1/2}\,(\theta)/G_0^{1/2}$

where G_0 = boresight gain of beam. The fundamental accuracy of this multiple-beam position version of the sequential-lobing estimation algorithm is

$$rmse = \left[\frac{\displaystyle\sum_{k=1}^{N}(\theta_k - f)^2}{2\dot{f}^2 x_0 \left(\displaystyle\sum_{k=1}^{N} g_k\right)^2} \right]^{1/2} \tag{20.14}$$

This performance can be calculated at an arbitrary elevation angle between the beams θ, or it can be averaged over an elevation angle in a root-sum-square (rss) sense as

$$\overline{rmse} = \left(\frac{1}{\theta_s} \int_0^{\theta_s} mse(\theta)\, d\theta \right)^{1/2} \tag{20.15}$$

where θ_s is the angular separation between the beam positions and $mse(\theta)$ is the mean square error.

Figure 20.4b displays this rss average accuracy for a nonfluctuating target versus signal-to-noise ratio for various beam spacings of a gaussian-shaped scanning beam. In general, the fundamental accuracy performance of the sequential-lobing technique is a function of the beam shapes and separations, the number of pulses
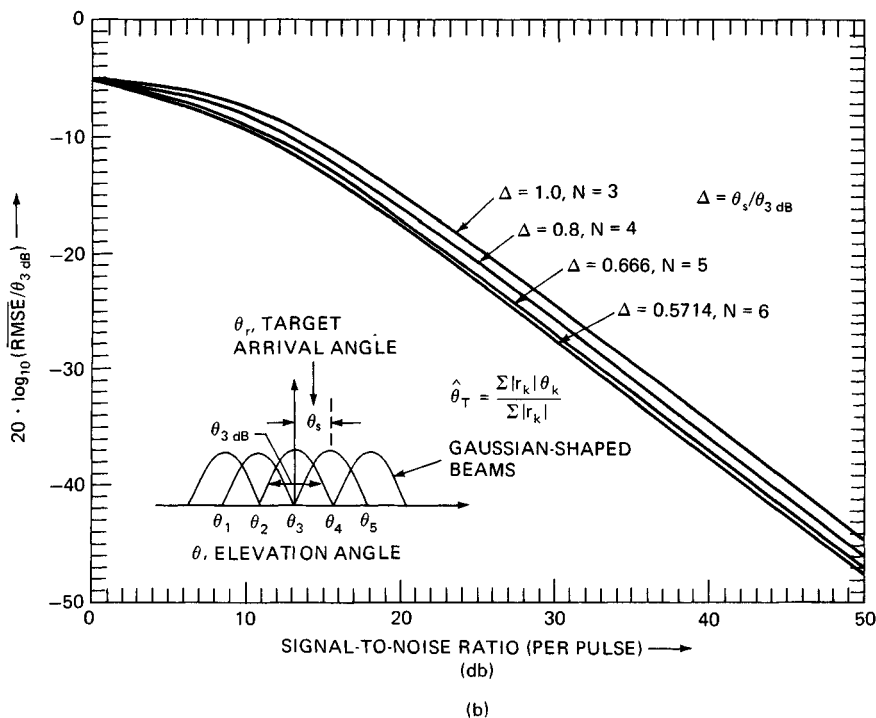
FIG. 20.4 (Continued) Fundamental accuracy. (b) Multiple-beam sequential lobing: nonfluctuating target.

noncoherently integrated in each beam, the type and amount of target echo pulse-to-pulse fluctuation within a beam, and of the prevailing beam-to-beam correlation in target fluctuation. For N pulse returns noncoherently integrated within each beam on a nonfluctuating target, the asymptotic rmse accuracy performance of two-beam sequential lobing is given by

$$\text{rmse} = \left(\frac{1 + f^2}{2\dot{f}^2 Nx}\right)^{1/2} \qquad (20.16)$$

where f = ratio of two-way elevation power patterns as before
$\quad\;\; N$ = number of pulses noncoherently integrated
$\quad\;\; x$ = per-pulse SNR

Since this form is the same as for a single pulse except for the factor of N in the denominator, numerical results may be obtained from Fig. 20.4a with appropriate scaling.

Fundamental Accuracy of Simultaneous Lobing. In the simultaneous-lobing method of angle estimation, two or more radar receive beams are simultaneously formed by the antenna and processed in parallel receive channels. A single transmit beam covers the angular region to be simultaneously processed on receive. Stacked beams, monopulse, and phase interferometry are all examples of the use of simultaneous lobing for target elevation angle estimation. While very different in implementation for a radar system, the fundamental accuracies of these techniques are all analyzed in a similar fashion, with approximate results that can be placed in the same form as Eq. (20.11). Because the receive beams in this technique are formed and processed simultaneously, the relative phase of the return between receive channels can, if desired, be used to aid the angle extraction accuracy. If it is used, the process is termed *phase-coherent* or simply *coherent*, and a close match in phase between receive channels must be maintained.

Monopulse. In general, the term *monopulse* refers to a radar technique to estimate the angle of arrival of a target echo resulting from a single-pulse transmission by using the amplitude and/or phase samples of the echo in a pair of simultaneously formed receive beams (Sec. 18.3). Historically the term has been associated with the simultaneous generation and processing of a *sum* receive beam and a *difference*, or *delta*, receive beam. These beams are so named because of the early and still common method used to form them, i.e., by adding and subtracting, respectively, the two halves of the antenna aperture. While this method is a relatively inexpensive way to produce a sum-difference beam pair, it is not necessarily the best way from a performance standpoint. Furthermore, it is unnecessarily constraining in many phased array applications, especially where the feeds account for a small fraction of the cost of the total radar. Typically a sum beam may be designed for good detectability and sidelobes. The delta beam is then optimized for accuracy performance, perhaps with other constraints. The defining characteristic of a sum beam is that it has approximately even symmetry about the beam boresight, while a delta beam has approximately odd symmetry about the same boresight. Without loss of generality, the delta beam may be assumed to be adjusted or calibrated to be in phase with the sum beam, in the sense that the ratio of the two patterns is real and odd about the beam boresight versus angle of arrival.

Monopulse techniques are classified according to the manner in which the incident radiation is sensed, i.e., according to antenna and beamforming techniques, and independently according to how the various beams and channels are subsequently processed and combined to produce a target angle estimate.[36,37] Amplitude comparison monopulse and phase comparison monopulse are categories of antenna-beamforming *sensing* techniques. In amplitude comparison monopulse, the antenna-beamformer generates a pair of sum and difference beams which, without loss of generality, may be assumed to be in phase, in the sense that their ratio is real. In phase comparison monopulse, two or more antennas or sets of radiating-receiving elements, physically separated in the elevation dimension, are used to generate two beams which have ideally identical patterns except for a phase difference which depends on the angle of incidence of the received target echo. Each of these techniques may be converted to the other, either in concept through mathematical sums and differences or physically through the use of passive RF hybrid combining devices. The fundamental accuracy performance of a phase comparison monopulse system is identical to that of an amplitude comparison monopulse system converted by this method, and vice versa. Therefore, the fundamental accuracy performance is addressed here from the conceptual viewpoint of amplitude comparison monopulse.

There are a variety of ways to implement monopulse processing on a sum-difference beam pair, depicted functionally in Fig. 20.5, some of which have a substantial impact on the fundamental monopulse accuracy performance. In each of these implementations, returns from a single transmission are received in simultaneously formed sum and difference beams and processed coherently. In the full-vector monopulse of Fig. 20.5a, two complex (I, Q) samples are fully utilized to calculate a complex monopulse ratio statistic. This calculated statistic, the *measured* monopulse ratio, provides the basis for a computer table lookup of the target angle of arrival relative to the null in the delta beam. The computer lookup function is simply a tabulated version of the assumed monopulse ratio consisting of the assumed delta beam antenna pattern to that of the assumed sum beam versus angle off-beam boresight. The tabulated monopulse ratio is inverted in the lookup process by entering the table with the measured monopulse ratio and finding the corresponding off-boresight angle. The full-vector monopulse processing in Fig. 20.5b differs somewhat from that in Fig. 20.5a, in that after low-noise amplification to establish the system noise figure, an RF quadrature hybrid device is used to combine the delta and sum beam signals 90° out of phase, i.e., as $\Sigma + j\Delta$. The purpose of this combining in the difference channel is to bring the signal strength in the difference channel to approximately the same amplitude at that in the sum channel. This causes unavoidable receiver nonlinearities to have nearly the same effect in the two channels, resulting in less degradation in accuracy performance attributable to receive-string nonlinearities. In the absence of nonlinearities, the two techniques in Fig. 20.5a and b are mathematically identical because

$$\text{Im}\left(\frac{\Sigma + j\Delta}{\Sigma}\right) = \text{Im}\left(1 + j\frac{\Delta}{\Sigma}\right) = \text{Im}\left|\frac{\Delta}{\Sigma}\right|\cos\phi \qquad (20.17)$$

Hence, they both provide the fundamental accuracy performance of full-vector monopulse processing, given by

$$\text{rmse} = \frac{\|w_\Delta - fw_\Sigma\|}{|\dot{f}|(2x)^{1/2}} \qquad (20.18)$$
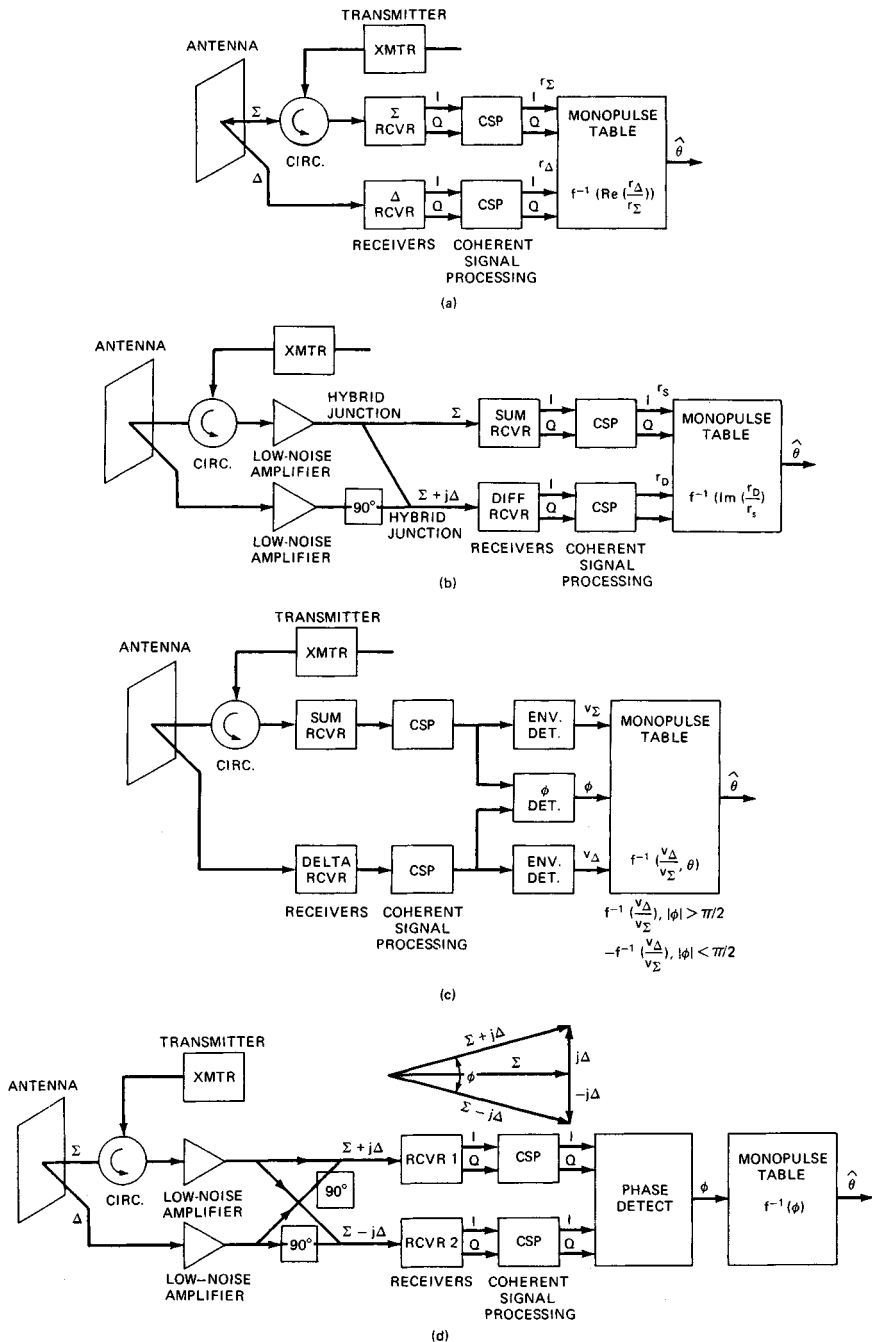
FIG. 20.5 Functional monopulse processing implementations. (*a*) Full-vector monopulse processing. (*b*) Full-vector monopulse with prehybrid combining. (*c*) Amplitude-only monopulse processing. (*d*) Phase-only monopulse processing.

where $f = f(\theta) = \Delta(\theta)/\Sigma(\theta)$, x = signal-to-noise ratio in the sum beam, and

Continuous aperture	Discrete aperture array
$w_\Delta = w_\Delta(x)$ aperture illumination	$w_\Delta = (w_{\Delta n})$ $n = 1, N$ vectors of
$w_\Sigma = w_\Sigma(x)$ functions	$w_\Sigma = (w_{\Sigma n})$ array weights
$\|w_\Delta\| = \left(\int_{-\infty}^{\infty} |w_\Delta(x)|^2 dx \right)^{1/2}$	$\|w_\Delta\| = \left(\sum_{1}^{N} |w_{\Delta n}|^2 \right)^{1/2}$
$\|w_\Sigma\| = \left(\int_{-\infty}^{\infty} |w_\Sigma(x)|^2 dx \right)^{1/2}$	$\|w_\Sigma\| = \left(\sum_{1}^{N} |w_{\Sigma n}|^2 \right)^{1/2}$
$\Delta(\theta) = \int_{-\infty}^{\infty} w_\Delta(x) \exp\left(j2\pi x (\sin\theta)\right) dx / \|w_\Delta\|$	$\Delta(\theta) = \sum_{1}^{N} w_{\Delta n} \exp\left[j2\pi x_n(\sin\theta)\right] / \|w_\Delta\|$
$\Sigma(\theta) = \int_{-\infty}^{\infty} w_\Sigma(x) \exp\left(j2\pi x(\sin\theta)\right) dx / \|w_\Sigma\|$	$\Sigma(\theta) = \sum_{1}^{N} w_{\Sigma n} \exp\left[j2\pi x_n(\sin\theta)\right] / \|w_\Sigma\|$

Various authors have defined the *monopulse sensitivity factor* in different ways.[38] For the purposes of this chapter, the monopulse sensitivity factor is defined as the constant of proportionality required in the denominator of the rmse to convert the square root of twice the boresight signal-to-noise ratio in the beam to the rmse. Defined in this manner, the monopulse sensitivity factor has the desirable property of containing all target elevation angle-of-arrival information.

The monopulse sensitivity factor for full-vector monopulse is

$$K = \frac{|\dot{f}| |g_\Sigma| |g_T|}{\|w_0 - f w_\Sigma\|} \qquad (20.19)$$

where $g_\Sigma = g_\Sigma(\theta) = \Sigma(\theta)/\Sigma_0$ = sum-beam voltage pattern normalized to unity gain

$g_T = g_T(\theta) = G_T(\theta)/G_{T0}$ = transmit-beam voltage pattern normalized to unity gain

For orthogonal aperture illumination functions, where $\Sigma_{k=1}^{N} w_{\Sigma k} w^*_{\Delta k} = 0$ (usually the case in practice), this equation reduces to

$$K = \frac{|\dot{f}| |g_\Sigma| |g_T|}{(1 + f^2)^{1/2}} \qquad (20.20)$$

This performance is presented graphically for several cases of interest in Fig. 20.6. A normalized sensitivity factor $k = K\lambda/L$ is plotted versus the u-space elevation angle of arrival $u = L/\lambda \sin\theta$, with θ referenced to the boresight of the sum and delta beams. The monopulse sensitivity factor peaks and is symmetrical about the boresight angle. The two curves in Fig. 20.6a correspond to uniformly illuminated sum beams. In one case, the delta beam is formed by subtracting the upper and lower halves of a uniformly illuminated aperture. For this delta beam illumination function, the normalized boresight monopulse sensitivity is $\pi/2 \approx 1.57$. The sensitivity degrades off boresight, despite an increasing monopulse ratio slope, owing to a decreasing signal-to-noise ratio in the two channels and to

(a)
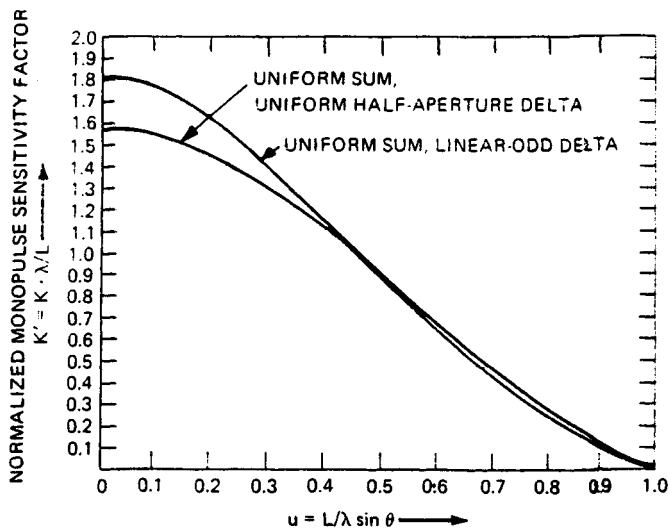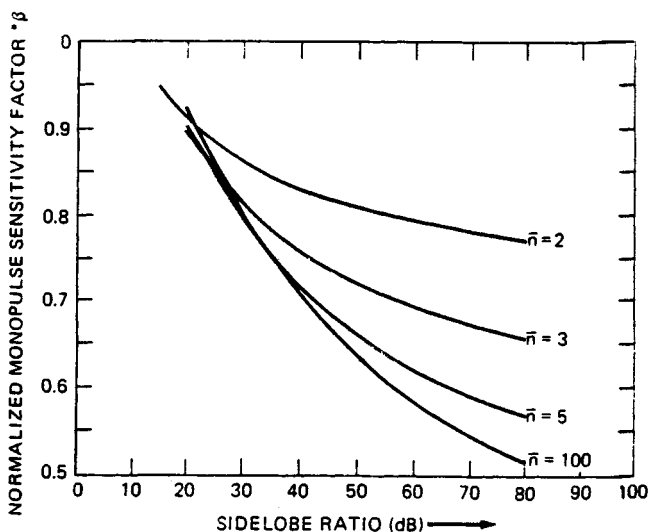
$^*\beta = \dfrac{RMSE_0}{RMSE}$; = accuracy with Taylor Σ, Bayliss Δ; $RMSE_0$ = accuracy with uniform Σ, linear-odd Δ.

(b)

FIG. 20.6 Fundamental accuracy of monopulse. (a) Vector monopulse accuracy: two common monopulse aperture illumination pairs. (b) Boresight vector monopulse sensitivities of Taylor sum and Bayliss delta aperture illuminations.

an increasing absolute value of the monopulse ratio $f(\theta)$. It can be shown that the boresight sensitivity of full-vector monopulse with a uniform sum beam can be maximized at a value of approximately 1.8 by employing a linear-odd aperture illumination function to generate the delta beam. The second curve in Fig. 20.6a illustrates performance for a linear-odd delta beam aperture illumination. The actual monopulse sensitivity factor can be calculated from the normalized sensitivity once the aperture height and RF wavelength have been specified. As an example, if $L/\lambda = 31.75$, the boresight monopulse sensitivity factor corresponding to the linear-odd delta beam illumination function is 0.05715 V/(V·msine). With a 20 dB signal-to-noise-ratio target return, this corresponds to a fundamental accuracy of 1.24 msines. For a uniform beam, the range of valid u-space angle coverage is approximately 2.0, corresponding to the sum-beam main-lobe null-to-null width. This is a principal advantage of monopulse because it allows reasonable spacings of the monopulse beams for coverage of large surveillance volumes. Coverage is increased with aperture weighting at the expense of monopulse sensitivity and fundamental accuracy. The effect on the boresight monopulse sensitivity of Taylor aperture weighting for the sum beam and Bayliss aperture weighting for the delta beam is illustrated in Fig. 20.6b. The sensitivity presented there is normalized by the sensitivity of the uniform-sum, linear-odd delta case, and is plotted for various values of the two parameters used to specify Taylor and Bayliss weighting, $\bar{n}$, and sidelobe ratio (SLR). It should be noted that not all combinations of $\bar{n}$ and SLR depicted in the figure constitute good aperture illumination design choices.

It is sometimes convenient and/or economical to perform coherent signal processing at RF or IF, by analog techniques, and then to carry out envelope and phase detection in the two channels. In amplitude-only monopulse the purpose of phase detection is solely to tell on which half of the beam the target return is incident. The angle off boresight is then determined via table lookup of the ratio of the envelope-detected signal strengths. The primary disadvantage of this approach is a degradation of accuracy at and near boresight relative to full-vector monopulse. It also provides less flexibility in coherent signal processing since it is analog instead of digital.

The fundamental accuracy performance of amplitude-only monopulse processing is degraded at boresight by the probability of incorrect phase detection, i.e., the probability of deciding that the target is below boresight when it is actually above, or vice versa. This probability is 0.5 at beam boresight, which results in boresight fundamental accuracy which is a factor of 2 worse than that of full vector monopulse. At off-boresight angles, the phase detection error probability depends on the signal-to-noise ratio. At angles far from the beam boresight, the signal-to-noise ratio diminishes, causing the error probability again to approach 0.5. A minimum-error probability–maximum-accuracy condition is reached for intermediate angles.

The last monopulse implementation illustrated (Fig. 20.5d) is termed *phase-only monopulse*. This processing is to be distinguished from the technique of phase interferometry, which has also been called by some authors[36,37] *phase-comparison monopulse*. In Fig. 20.5d, RF or IF hybrids are used to combine the sum and delta channels in quadrature, i.e., with a 90° phase shift. An accurate phase detector then detects the phase difference between the two channels. The underlying principle is that this phase difference will be in one-to-one correspondence with the delta-to-sum ratio, as illustrated in the vector diagram accompanying Fig. 20.5d. In phase-only monopulse, off-boresight accuracy is sacrificed to

gain the benefit of identical amplitude signals in the two receiver-processor channels. If desired, the signals in the two channels may be hard-limited without affecting the fundamental accuracy of the phase-only monopulse processing. In principle, phase-only monopulse can be used to alleviate stringent receiver-processor dynamic-range requirements. However, other aspects of performance may suffer and should be examined carefully in the tradeoff process. Another advantage of phase-only monopulse, relative to vector and amplitude-only, is that the need for precise amplitude matching channel to channel is reduced.

Phase-only monopulse processing does not utilize the full target angle-of-arrival information available in the two beams. For this reason, its fundamental accuracy performance suffers. The fundamental accuracy of phase-only monopulse is identical to that of vector monopulse at boresight but degrades more rapidly off boresight. Full vector monopulse, using all the available information in the target returns, shows superior sensitivity at all target incidence angles. A uniformly illuminated aperture and beam and a uniform half-aperture difference beam are used for comparison of the three implementations.

In a radar which employs vector monopulse processing for height finding, it is possible to coherently precombine the returns from multiple pulses or subpulses in the simultaneous beams to form a single estimate of the target elevation angle of arrival, as suggested in Fig. 20.7. In this approach, the returns in the delta and sum channels are coherently cross-correlated pulse to pulse, and then the real part of the cross-correlation sum is normalized by a term determined by noncoherent integration in the sum channel to form the measured monopulse ratio. The same noncoherent sum used to normalize the measured monopulse ratio may be also used in the target decision logic for detection thresholding.

The rmse for multiple-pulse coherent monopulse differs from that of a single-pulse monopulse only by the square root of the number of pulses in the denominator. The results of Fig. 20.6, appropriately scaled, are applicable.

Stacked Beams. Stacked beams are another example of simultaneous lobing for target elevation-angle estimation. The processing of a pair of beams in the stack consists of an amplitude comparison table lookup. Its fundamental accuracy can also be placed in the form of Eq. (20.11).

In the stacked-beam radar, the transmit beam must be designed to cover all the beams within the stack and is therefore relatively wide in elevation beamwidth compared with that of a receive beam in the stack. A good approximation is that it is isotropic in elevation and thus is not a factor in the fundamental accuracy performance.
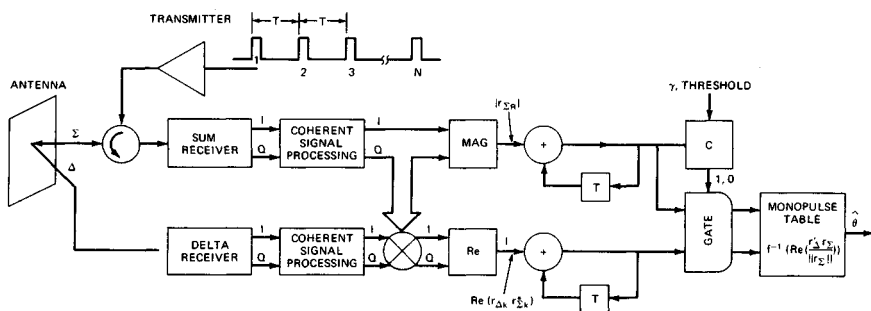
FIG. 20.7 Multiple-pulse vector monopulse processing.

The fundamental accuracy performance of a pair of uniformly illuminated beams in a stack is presented in Fig. 20.8 in terms of a normalized sensitivity factor $k = K\lambda/L$ versus normalized sine-space angle-of-arrival $u = L/\lambda \sin \theta$. The elevation u-space angle of arrival u of target energy is referenced to the crossover point halfway between the beams. Various beam separations in u space are illustrated. The sensitivity of the technique peaks at the crossover angle and is symmetrical about that angle, attaining a value at crossover which depends on the separation between the beams. A maximum crossover sensitivity of 1.95 is achieved for a u-space beam separation of 1.2. Coverage in u space provided by the uniform stacked-beam pair is approximately given by $2-\Delta u$, where Δu is the u-space beam separation corresponding to a target in the main lobes of both beams. Outside this region, the target is in the sidelobes of one of the beams. In a stacked-beam radar, detections are made in a special cosecant-squared type of surveillance beam; so this condition is not sensed in the detection process. Thus, in order to eliminate the possibility of ambiguities, uniformly illuminated beams should be stacked at $\Delta u \leq 1$. The coverage of each beam pair may be increased by aperture weighting. In this case the beams may be stacked at greater separations but will possess reduced crossover sensitivity. The normalized crossover sensitivity associated with a pair of uniformly illuminated sum beams spaced at $\Delta u = 1$ is approximately 1.8. This corresponds to a fundamental accuracy of approximately 1.24 msines for a 24-ft aperture height at L band with a 20 dB target boresight signal-to-noise ratio.
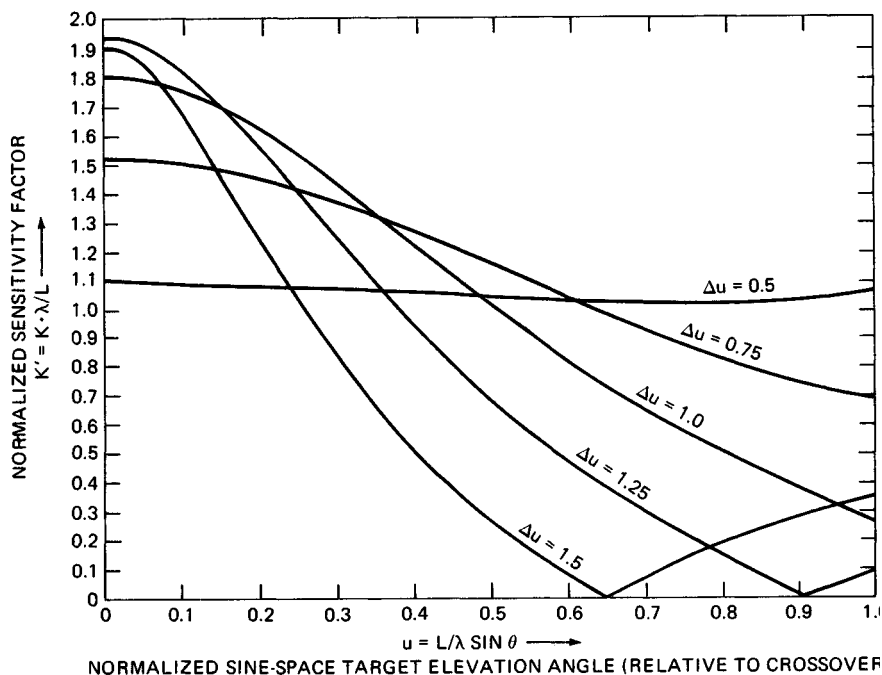
FIG. 20.8 Fundamental accuracy of stacked beams.
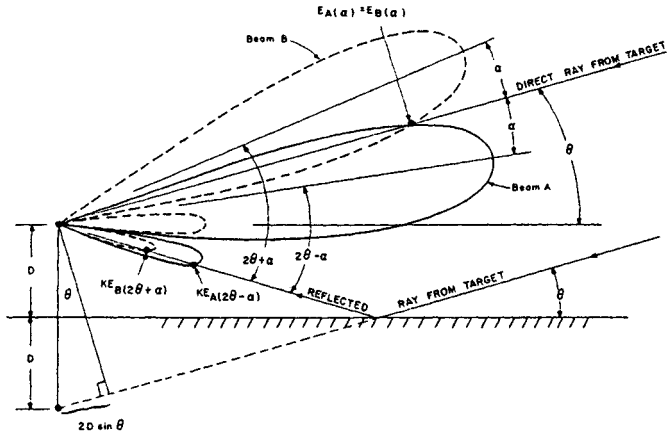
FIG. 20.9 Geometry for the analysis of elevation errors due to ground reflections in a simultaneous amplitude comparison radar.

Elevation Error Due to Surface Reflections.[*]   One of the fundamental factors limiting the height accuracy in all height finding techniques that depend on elevation-angle measurements is the elevation-angle accuracy degradation due to the multipath from surface reflections. Such surface reflections vectorially combine with the direct-path signals entering the antenna to produce amplitude and phase variations which ordinarily cannot be separated from the direct-path signals. In general, the magnitude of such elevation-angle errors is such that, at low elevation angles where an appreciable portion of the antenna beam is directed into the ground, the elevation-angle errors are prohibitively large. Therefore, as a general rule, pencil-beam height finding radars and elevation-tracking radars cannot be expected to produce reliable elevation-angle data when their beams are pointed within about one beamwidth (−3 dB beamwidth) above the ground. At larger elevation angles, the magnitude of the elevation errors is a direct function of the ground-reflected relative field strength received in the respective negative-angle elevation sidelobe (i.e., the product of the relative sidelobe level and the ground-reflection coefficient).

Radar systems that employ a simultaneous amplitude comparison technique for target elevation-angle determination derive the elevation angle inside the radar beamwidth by measuring the ratio of simultaneous signal returns on two squinted received beams after having illuminated the target in some manner. The resulting elevation-angle data is independent of the manner in which the target is illuminated by the radar and is dependent only on the squinted *receiving*-antenna patterns.

In analyzing the elevation-angle errors due to ground reflections, we shall consider the case where the boresight crossover of a pair of squinted receiving beams (A and B) is oriented exactly on the target at elevation angle θ (a condition of zero error in the absence of ground reflections). See Fig. 20.9.

With pattern functions of beams A and B assumed to be identical and with the centerlines of beams A and B oriented at elevation angles of $\theta - \alpha$ and $\theta + \alpha$, respectively, the net received field strength at the feed points of beams A and B (relative to the peak of each beam) is then

---

[*]The material in this subsection was originally written by Burt Brown and appeared in Sec. 22.3 of the first edition of the handbook.

$$E_A = E_{A(\alpha)} + KE_{A(2\theta-\alpha)}e^{-j(\phi+2D\sin\theta)}$$

$$E_B = E_{B(\alpha)} + KE_{B(2\theta+\alpha)}e^{-j(\phi+2D\sin\theta)} \quad (20.21)$$

where K = amplitude of reflection coefficient

ϕ = phase of reflection coefficient

$E_{A(\alpha)} = E_{B(\alpha)}$ = relative received field strength of beams A and B from signal arriving along direct path at angle α from peaks of beams A and B

$E_{A(2\theta-\alpha)}$ = relative received field strength of beam A from reflected path at $2\theta - \alpha$ from peak of beam A

$E_{B(2\theta+\alpha)}$ = relative received field strength of beam B from reflected path at $2\theta + \alpha$ from peak of beam B

The magnitude of the off-boresight elevation-angle error due to ground reflections is a function of the ratio of the magnitudes of E_A and E_B, or

$$\text{Elevation error} = f\,\frac{|E_A|}{|E_B|} = f\,\frac{|E_{A(\alpha)} + KE_{A(2\theta-\alpha)}e^{-j(\phi+2D\sin\theta)}|}{|E_{B(\alpha)} + KE_{B(2\theta+\alpha)}e^{-j(\phi+2D\sin\theta)}|} \quad (20.22)$$

When $KE_{A(2\theta-\alpha)}$ and $KE_{B(2\theta+\alpha)}$ are small compared with $E_{A(\alpha)}$, the maximum value of $f(|E_A|/|E_B|)$ is equal to

$$\frac{E_{A(\alpha)} + KE_{A(2\theta-\alpha)}}{E_{B(\alpha)} - KE_{B(2\theta+\alpha)}} \quad (20.23)$$

To illustrate these effects, a specific example is cited where the amplitude comparison beams A and B are assumed to have the following characteristics:

Antenna aperature (α) = 25.5λ

Receiving beam pattern function $\dfrac{\sin\,[\pi(\alpha/\lambda)\sin\theta]}{\pi(\alpha/\lambda)\sin\theta} = \dfrac{\sin\,(25.5\pi\sin\theta)}{25.5\pi\sin\theta}$

Beamwidth (at-3 dB points) = 2.0°

Squint angle (α) = 1.125°

Antenna height above ground = 50λ

The ground-reflection coefficient $Ke^{j\phi}$ is assumed to be $1.0e^{j\pi}$, which corresponds to horizontal polarization over an infinite conducting plane.

With these values, Fig. 20.10 shows a plot of the resultant relative field strengths in beams A and B as a function of the elevation boresight pointing angle θ. Note that in these curves, for each value of elevation angle θ, the antenna boresight is assumed to be directed exactly on the target. Thus, in the *absence* of ground reflections, the net field strengths in A and B would have been *equal* for all values of θ.

The corresponding off-boresight errors for the amplitude curves of Fig. 20.10 are shown in Fig. 20.11. In viewing the characteristics of the off-boresight error curves of Fig. 20.11, the important features to be noted are as follows:
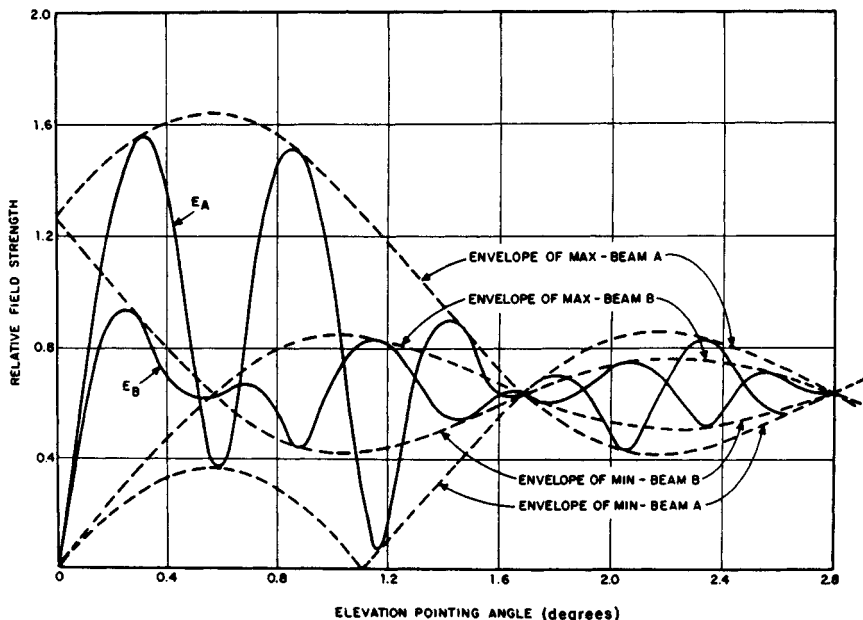
FIG. 20.10 Resultant relative field strength received in beams *A* and *B* resulting from ground reflections as a function of the elevation pointing angle of the beam crossover. Beamwidth is 2°; antenna height, 50λ; total squint angle, 2.25°; and reflection coefficient, 1.0.

1. The shape and amplitude of the *envelope of maximum errors* (dash curves) are dependent only on the antenna-beam pattern function in the elevation plane and the ground-reflection coefficient and are independent of the antenna electrical height above the ground.

2. The configuration of the error curve lying inside the envelope of maximum errors (i.e., the positions and spacing of the peak errors) is dictated chiefly by the antenna electrical height above the ground (in wavelengths) and, to a minor extent, by the phase angle of the antenna sidelobes that receive the reflected rays.

A similar analysis can be made of the errors due to surface reflections for a simultaneous phase comparison system, as found in Sec. 22.3 of the first edition of this handbook.

At high elevation pointing angles, where the errors are due only to surface reflections from the negative elevation-angle sidelobes, the errors contributed by various sidelobe levels in simultaneous amplitude comparison and phase comparison radars may be summarized approximately as follows: 0.2 to 0.3 beamwidth, when the surface-reflected sidelobe is 10 dB down from the peak; 0.07 to 0.10 beamwidth for −20 dB sidelobes; 0.025 to 0.035 for −30 dB sidelobes; 0.008 to 0.011 for −40 dB sidelobes; and about 0.003 beamwidth for −50 dB sidelobes illuminating the surface.

Low-Angle Squinted-Beam Height Finding. One height finding technique which has proved practical and effective against surface multipath is the so-called low-angle squinted-sum-beam height finding technique employed in the
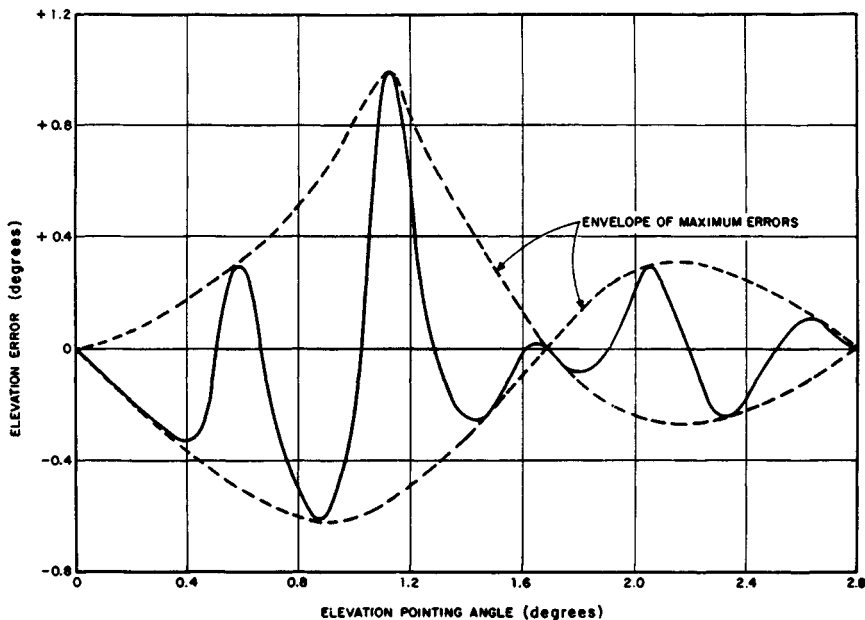
FIG. 20.11 Elevation errors due to ground reflection as a function of pointing angle for amplitude comparison system and uniform aperture distribution. Other conditions are as in Fig. 20.10.

TPS-59/GE-592/FPS-117 solid-state radar series.[27] The problem with conventional sum-delta monopulse in a surface multipath environment is that the delta-beam peak response is in the direction of the indirect-path reflection. The low-angle technique avoids this problem by using a pair of squinted beams on receive, as illustrated in Fig. 20.12. The lower beam, unweighted so as to generate as narrow a beam as permitted by the array aperture, is placed in elevation so that the indirect path is attenuated by the lower side of the beam. The upper beam, weighted to produce low sidelobes, is placed a degree or so above the lower beam so that the indirect-path echo is rejected by the sidelobes of the upper beam while the direct-path echo is received at high gain. This approach tends to minimize the amount of indirect-path energy in the two beams while maintaining coverage on the horizon.

The performance of the low-angle squinted-sum-beam technique also differs from that of a conventional stacked-beam pair, for two reasons. First, the two receive beams in the low-angle technique are not formed from identical aperture illuminations. Second, the transmit beam in the low-angle technique is narrow, reducing indirect multipath returns and to a lesser degree off-boresight signal-to-noise ratio.

The performance of a version of the low-angle squinted-sum-beam technique relative to that of a conventional monopulse is compared in Fig. 20.13 for a surface reflection coefficient of -1, approximately that of a smooth sea. The antenna in each case is situated at a height of twice the vertical aperture dimension above a flat earth. The monopulse case consists of a uniformly illuminated sum beam and a half-aperture uniform difference beam electronically phase-steered to $u_0 = 0.5$, processed, and implemented with full vector processing. The low-angle
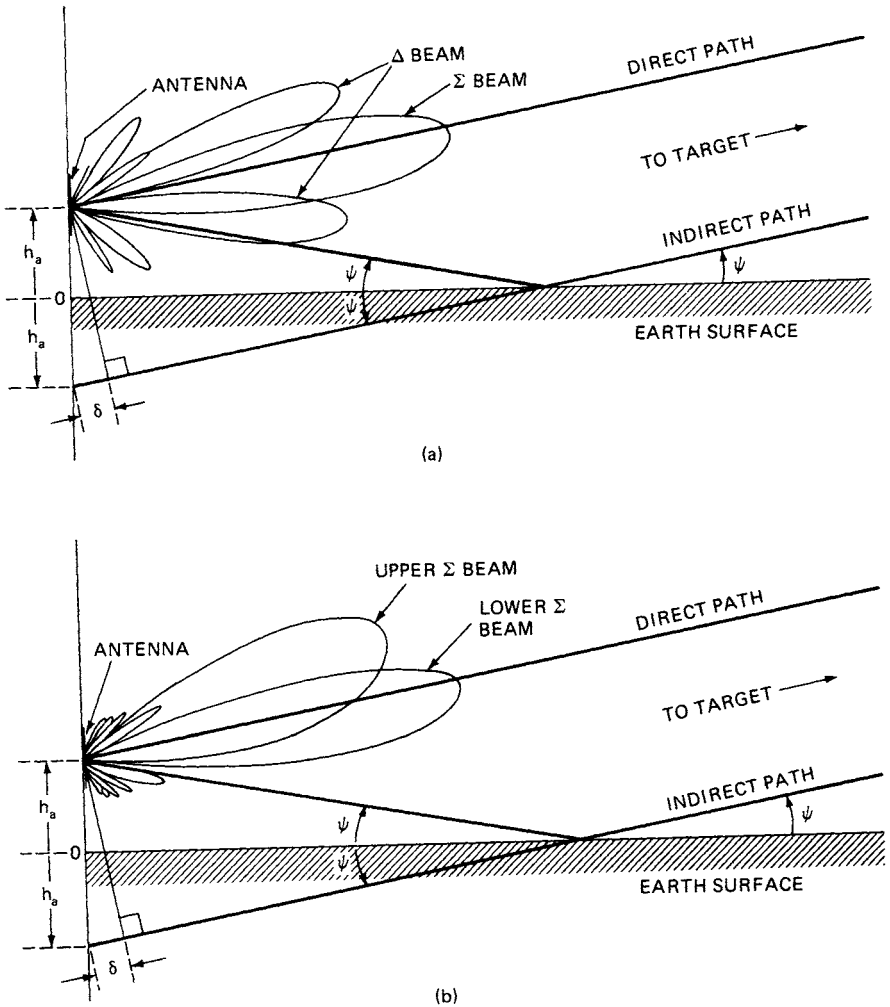
FIG. 20.12 Multipath impact on monopulse; δ = path difference. (*a*) Conventional sum-delta monopulse. (*b*) Squinted-sum-beam low-angle technique.

technique examined consists of a uniformly illuminated transmit/receive lower beam, accompanied by a weighted aperture upper beam. The lower beam is electronically phase-steered to $u_1 = 0.5$ while the upper beam is electronically phase-steered an additional $\delta u = 1.0$ (i.e., to $u_2 = 1.5$). This places the multipath largely in the sidelobe of the upper beams. In the monopulse case, the multipath introduces severe bias errors into the elevation-angle estimate with peaks on the order of 0.4 in u space. For a radar with an aperture of $L/\lambda = 32$, this corresponds to an rmse of 12.5 msines. The multipath bias errors dominate the total accuracy performance of the monopulse technique and, contrary to the behavior of the thermal errors, are not suppressed by high target signal-to-noise ratios. By
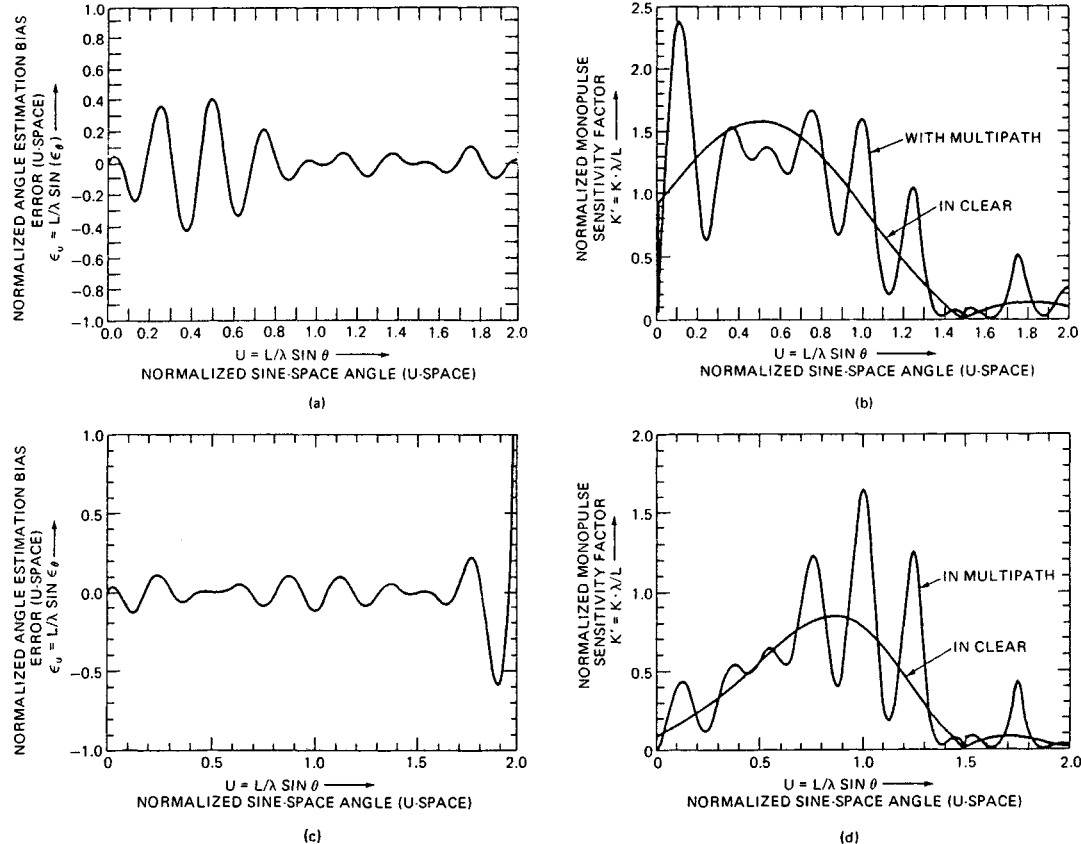
FIG. 20.13 Accuracy comparison in multipath. (*a*) Sum-delta monopulse bias error. Antenna height = 2 × aperture height; monopulse beams boresighted at $u_0 = 0.5$; reflection coefficient = − 1. (*b*) Sum-delta monopulse sensitivity factor. Antenna height = 2 × aperture height; monopulse beams boresighted at $u_0 = 0.5$; reflection coefficient = − 1. (*c*) Squinted-sum low-angle bias error. Antenna height = 2 × aperture height; lower beam boresighted at $u_0 = 0.5$; upper beam boresighted at $u_1 = 1.5$; reflection coefficient = − 1. (*d*) Squinted-sum low-angle sensitivity factor. Antenna height = 2 × aperture height;

contrast, the bias errors introduced by the multipath in the low-angle squinted-sum-beam technique are kept to peaks on the order of 0.15 in u space, corresponding to an rmse of approximately 4.7 msines for a radar of the same aperture. Further reduction might be possible with an optimization of the beam placement and aperture illumination functions.

REFERENCES

1. Skolnik, M. I.: Fifty Years of Radar, *Proc. IEEE*, vol. 73, pp. 182–197, February 1985.

2. Guerlac, H. E.: "Radar in World War II," Tomash Publishers, American Institute of Physics, Los Angeles, 1987.

3. Ridenour, L. N.: "Radar System Engineering," MIT Radiation Laboratory Series, vol. 1, McGraw-Hill Book Company, New York, 1947.

4. The SCR-268 Radar, *Electronics*, vol. 18, pp. 100–109, September 1945.

5. The SCR-584 Radar, *Electronics*, vol. 18, pp. 104–109, November 1945.

6. Schneider, E. G.: Radar, *Proc. IRE*, vol. 34, pp. 528–578, August 1946.

7. Brookner, E.: "Radar Technology," Artech House, Norwood, Mass., 1980, pp. 5–59.

8. Sutherland, J. W.: Marconi S600 Series of Radars, *Interavia*, vol. 23, pp. 73–75, January 1968.

9. Skolnik, M. I.: "Introduction to Radar Systems," 2d ed., McGraw-Hill Book Company, New York, 1980.

10. Brown, B. P.: Radar Height Finding, chap. 22 of Skolnik, M. I. (ed.): "Radar Handbook," 1st ed., McGraw-Hill Book Company, New York, 1970.

11. Simpson, T. J.: The Air Height Surveillance Radar and Use of Its Height Data in a Semi-Automatic Air Traffic Control System, *IRE Int. Conv. Rec.*, vol. 8., pt. 8, pp. 113–123, 1960.

12. Watanabe, M., T. Tamana, and N. Yamauchi: A Japanese 3-D Radar for Air Traffic Control, *Electronics*, p. 68, June 21, 1971.

13. AN/TPS-43E Tactical Radar System, brochure, Westinghouse Corporation.

14. The Martello High Power 3-D Radar System, brochure, Marconi Company.

15. RAT-31S 3D Surveillance Radar, brochure, Selenia Radar and Missile Systems Division, Rome.

16. Hammer, I. W.: Frequency-Scanned Arrays, chap. 13 of Skolnik, M. I. (ed.): "Radar Handbook," 1st ed., McGraw-Hill Book Company, New York, 1970.

17. Milne, K.: The Combination of Pulse Compression with Frequency Scanning for Three Dimensional Radars, *Radio Electron. Eng.*, vol. 28, pp. 89–106, August 1964.

18. Polmar, N.: "Ships and Aircraft of The U.S. Fleet," 14th ed., Naval Institute Press, Annapolis, Md., 1987, chap. 29, Electronic Systems.

19. AR-3D Mobile Air Defense Radar System, brochure, Plessey.

20. Pretty, R. T. (ed.): "Jane's Weapon Systems, 1981–1982," pp. 449–596.

21. Pfister, G.: The Series 320 Radar, Three Dimensional Air Surveillance Radar for the 1980's, *IEEE Trans.*, vol. AES-16, pp. 626–638, September 1980.

22. Lain, C. M., and E. J. Gersten: AN/TPS-59 System, *IEEE Int. Radar Conf. Rec.*, IEEE Publ. 75 CHO 938-1 AES, pp. 527–532, Apr. 21–23, 1975.

23. AN/TPS-59 Tactical Solid State Radar, brochure, General Electric Company.

24. AN/TPS-59: First Total Solid State Radar, *ADCOM Commun. Electron. Comput. Resources Dig.* (editorial article), October–November–December 1976.

25. AN/FPS-117 Minimally Attended Solid State Radar System, brochure, General Electric Company.

26. Gostin, J. J.: The GE592 Solid State Radar, *EASCON '80 Rec.*, pp. 197–203, IEEE Publ. 80 Ch 1578-4 AES, Sept. 29, 30, Oct. 1, 1980.

27. GE-592 Solid State Radar Systems, brochure, General Electric Company.

28. Klass, P. J.: Solid State 3D Radar for NATO Tested, *Aviat. Week Space Technol.*, May 21, 1979.

29. U.S. Air Force reports on Hughes Air Defense Radar, *Flight Int.*, Dec. 4, 1982.

30. Smith, E. K., and S. Weintraub: The Constraints in the Equation for Atmospheric Refractive Index at Radio Frequencies, *Proc. IRE*, vol. 41, pp. 1035–1037, August 1953.

31. Blake, L. V.: Ray Height Computation for a Continuous Nonlinear Atmospheric Refractive-Index Profile, *Radio Sci.*, vol. 3, pp. 85–92, January 1968.

32. Millman, G. H.: Atmospheric Effects on Radio Wave Propagation, in Berkowitz, R. S. (ed.): "Modern Radar Analysis, Evaluation and System Design," John Wiley & Sons, New York, 1965, pp. 315–377.

33. Bean, B. R., and E. J. Dutton: Radio Meteorology, *Nat. Bur. Stand. Monog.* 92, pp. 59–76, March 1966.

34. Bauer, J. R., W. C. Mason, and R. A. Wilson: Radio Refraction in a Cool Exponential Atmosphere, *MIT Lincoln Laboratory, Tech. Rept.* 186, August 1958.

35. Bauer, J. R., and R. A. Wilson: Precision Tropospheric Radio Refraction Corrections for Ranges from 10–500 Nautical Miles, *MIT Lincoln Laboratory, Rept.* 33G-0015, Feb. 20, 1961.

36. Sherman, S. M.: "Monopulse Principles and Techniques," Artech House, Norwood, Mass., 1985, chap. 5, chap. 12, pp. 345–348.

37. Rhodes, D. R.: "Introduction to Monopulse," McGraw-Hill Book Company, New York, 1959; reprinted by Artech House, Norwood, Mass., 1982.

38. Kinsey, R. R.: Monopulse Difference Slope and Gain Standards, *IRE Trans.*, vol. AP-10, pp. 343–344, May 1962.

CHAPTER 21
SYNTHETIC APERTURE RADAR

L. J. Cutrona
Sarcutron, Inc.

21.1 BASIC PRINCIPLES AND EARLY HISTORY

For airborne ground-mapping radar there has been continuous pressure and desire to achieve finer resolution. Initially, this finer resolution was achieved by the application of "brute-force" techniques. Conventional radar systems of this type were designed to achieve range resolution by the radiation of a short pulse and azimuth resolution by the radiation of a narrow beam.

The range resolution problem and some of the pulse compression techniques are discussed in Chap. 10. There it is shown that techniques are available for achieving a resolution significantly finer than that corresponding to the pulse width, provided a signal of sufficient bandwidth is transmitted. Since pulse compression is adequately treated in that chapter, the present chapter will discuss pulse compression techniques only for cases in which the pulse compression technique is intimately involved with synthetic aperture techniques. This is particularly true for configurations that perform both pulse compression and azimuth compression simultaneously rather than with techniques that perform range compression and azimuth compression sequentially.

The basic technology discussed in this chapter is the exploitation of synthetic aperture techniques for improving the azimuth resolution of a mapping radar to a value significantly finer than that achievable by making use of the radiated beamwidth.

Synthetic aperture radar (SAR) is based on the generation of an effective long antenna by signal-processing means rather than by the actual use of a long physical antenna. In fact, only a single, relatively small, physical antenna is used in most cases.

In considering a synthetic aperture, one makes reference to the characteristics of a long linear array of physical antennas. In that case, a number of radiating elements are constructed and placed at appropriate points along a straight line. In the use of such a physical linear array, signals are fed simultaneously to each of the elements of the array. Similarly, when the array is used as a receiver, the elements receive signals simultaneously; in both the transmitting and the receiving modes, waveguide or other transmission-line interconnections are used, and interference phenomena are exploited to get an effective radiation pattern.

The radiation pattern of a linear array is the product of two quantities if the

radiating elements are identical. The radiation pattern of the array is the radiation pattern of a single element multiplied by an array factor. The array factor has significantly sharper lobes (narrower beamwidths) than the radiation patterns of the elements of the array. The half-power beamwidth β, in radians, of the array factor of such an antenna is given by

$$\beta = \frac{\lambda}{L} \qquad (21.1)$$

In this expression, L is the length of the physical array, and λ is the wavelength.

In the synthetic antenna* case, only a single radiating element is used in most instances. This antenna is translated to take up sequential positions along a line. At each of these positions a signal is transmitted, and the radar signals received in response to that transmission are placed in storage. It is essential that the storage be such that both amplitude and phase of received signals are preserved.

After the radiating element has traversed a distance L_{eff}, the signals in storage resemble strongly the signals that would have been received by the elements of an actual linear array. Consequently, if the signals in storage are subjected to the same operations as those used in forming a physical linear array, one can get the effect of a long antenna aperture. This idea has resulted in the use of the term *synthetic aperture* to designate this technique.

In the case of an airborne ground-mapping radar system, the antenna usually is mounted to be side-looking, and the motion of the aircraft carries the radiating element to each of the positions of the array. These array positions are the location of the physical antenna at the times of transmission and reception of the radar signals.

The designer of a synthetic aperture radar has available a number of degrees of freedom that are not available to the designer of a physical linear array. These degrees of freedom derive from the fact that the signals in storage can be selected by range and that, if desired, a different operation can be performed on the signals at different ranges. One important operation of this type is that of *focusing*.

A physical linear array can be focused to a specific range. There will then be a depth of focus surrounding this range. However, most physical linear arrays are unfocused. This is sometimes stated by saying that the antenna is "focused at infinity." In a synthetic aperture radar, however, it is possible to focus each range separately by the proper adjustment of the phases of the received signals before the summation; this results in the effective synthetic aperture. Furthermore, if desired, a different weighting can be applied to each range, although usually the same type of weighting is used at all ranges.

There is another important difference between physical linear arrays and synthetic linear arrays. This difference results in the synthetic aperture having a resolution finer by a factor of 2 than that corresponding to a real linear array of the same length. Qualitatively, the following discussion indicates the physics resulting in this factor of 2. In a more general analysis, the factor 2 arises naturally.

In a physical linear array, the transmission of the signals results in an illumination of the target area. The angle selectivity of the linear array is provided only during the reception process. During this process, the differences in phase received by each element of the linear array give the antenna pattern. In the synthetic antenna radar, on the other hand, a single element radiates and receives signals. Consequently, the round-trip phase shift is effective in forming the effective radiation pattern. This relationship is written as

---

*The terms *synthetic antenna* and *synthetic aperture* are used interchangeably in this chapter.

$$\beta_{eff} = \frac{\lambda}{2L_{eff}} \tag{21.2}$$

Here β_{eff} is the effective half-power beamwidth of the synthetic aperture, and L_{eff} is the length of the synthetic aperture.

A more detailed derivation of the resolution capability of a synthetic aperture radar will be given later in this chapter. The following derivation is that initially made by the author and his colleagues in the early days of synthetic aperture radar.

Let D represent the horizontal aperture of the physical antenna carried by an airborne ground-mapping radar. The width of the horizontal beam at range R gives the maximum value for the length of synthetic aperture that can be used at that range. Since the beamwidth of such an antenna is given by the ratio of the wavelength λ to its horizontal aperture D, the maximum length of this synthetic antenna aperture is given by

$$L_{eff} = \frac{R\lambda}{D} \tag{21.3}$$

The linear resolution in azimuth δ_α is the product of the effective beamwidth given by Eq. (21.2) and the range R:

$$\delta_\alpha = \beta_{eff}R \tag{21.4}$$

If Eqs. (21.2) and (21.3) are combined with Eq. (21.4), one obtains

$$\delta_\alpha = \frac{\lambda}{2L_{eff}}R = \frac{\lambda R}{2}\frac{D}{R\lambda} = \frac{D}{2} \tag{21.5}$$

It will be noted that Eq. (21.5) indicates an azimuth linear resolution independent of both range and wavelength. Moreover, the result indicates that finer resolution is achievable with smaller rather than larger physical apertures. This spectacular result formed much of the motivation of the research in synthetic antenna radar.

The author was first exposed to the idea of a synthetic antenna radar in 1953, during a summer study which launched a program known as Project Michigan. During that summer, the ideas relating to synthetic antennas were presented by Dr. C. W. Sherwin,[1] then of the University of Illinois, Dr. Walt Hausz of the General Electric Company, and J. Koehler, at that time with Philco Corporation. Subsequently, it came to the author's attention that Carl Wiley and the Goodyear Aircraft Company had already undertaken some work and had made substantial progress in the synthetic-antenna area.

The Pioneer Award of the IEEE Aerospace and Electronic Systems Society was given to Carl Wiley in 1985 for his work in synthetic aperture radar. His remarks from that presentation are given in Ref. 2 and relate some of the early history of SAR.

Most of the early workers considered an unfocused synthetic antenna. However, Dr. Sherwin indicated that finer resolution should be achievable by using focusing because this technique removed what would otherwise be a restriction on the maximum length of synthetic antenna that could be used. The author and

his colleagues at the University of Michigan undertook development of the focusing concept suggested by Dr. Sherwin.

21.2 FACTORS AFFECTING RESOLUTION OF A RADAR SYSTEM

In the following paragraphs a brief comparison of the conventional antenna, the unfocused synthetic antenna, and the focused antenna is given.[3,4] The language of synthetic apertures is used, and a comparison of the resolution capability for three cases is given. A more sophisticated derivation of simultaneous resolution in range and azimuth will be given later in this chapter.

Three cases are compared for their azimuth resolution capability: (1) the *conventional technique*, in which azimuth resolution depends upon the width of the radiated beam; (2) the *unfocused synthetic antenna technique*, in which the synthetic antenna length is made as long as the unfocused technique permits; and (3) the *focused synthetic antenna technique*, in which the synthetic antenna length is made equal to the linear width of the radiated beam at each range.

The linear azimuth resolution for the *conventional case* is given by

$$\text{Resolution}_{\text{conv}} = \frac{\lambda R}{D} \qquad (21.6)$$

For the *unfocused case*, the resolution is

$$\text{Resolution}_{\text{unf}} = \tfrac{1}{2}\sqrt{\lambda R} \qquad (21.7)$$

whereas for the *focused case*, the resolution is

$$\text{Resolution}_{\text{foc}} = \frac{D}{2} \qquad (21.8)$$

where λ = wavelength of radar signal transmitted
D = horizontal aperture of antenna
R = radar range

Figure 21.1 is a plot of the resolution for each of these cases as a function of radar range. This plot is for an antenna aperture of 5 ft and a wavelength of 0.1 ft.

Conventional Technique. The conventional technique for achieving azimuth resolution has been that of radiating a narrow beam. In this case the *resolution* of a target depends upon whether the target is included within the half-power points of the radiated beam, although some techniques exist for resolving targets somewhat less than a beamwidth apart.

The computation of the linear azimuth resolution for the conventional case is well known. The appropriate expression is obtained by noting that the width of the radiated beam, in radians, is given by the ratio λ/D whereas the linear width of the beam at range R is the product of this beamwidth and range. These considerations lead to the result already written as Eq. (21.6).

A consideration from antenna theory is that Eq. (21.6) applies only to the far-
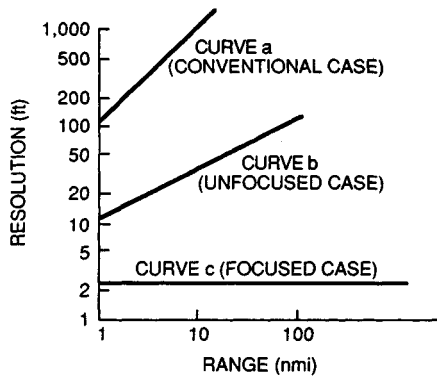
FIG. 21.1 Azimuth resolution for three cases: curve *a*, conventional; curve *b*, unfocused; curve *c*, focused.

field pattern of an antenna. The beginning of the far field occurs at a distance R_{min} for which

$$R_{min} \simeq \frac{D^2}{\lambda} \qquad (21.9)$$

It will be noted by substitution of Eq. (21.9) that the finest resolution achievable by the conventional technique is given by

$$\text{Minimum conventional resolution} = D \qquad (21.10)$$

The Unfocused Synthetic Aperture. The simpler of the synthetic antenna techniques is that which generates an unfocused synthetic aperture. In this case, the coherent signals received at the synthetic array points are integrated, with no attempt made to shift the phases of the signals before integration. This lack of phase adjustment imposes a maximum upon the synthetic antenna length that can be generated. This maximum synthetic antenna length occurs at a given range when the round-trip distance from a radar target to the center of the synthetic array differs by $\lambda/4$ from the round-trip distance between the radar target and the extremities of the synthetic aperture array.
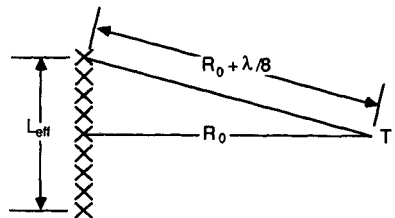
FIG. 21.2 Geometry for an unfocused synthetic antenna.

The pertinent geometry is shown in Fig. 21.2. In this figure, R_0 represents the range from a radar target to the center of the array, and L_{eff} represents the maximum synthetic antenna length such that the distance from the target to the extremities of the synthetic aperture does not exceed $R_0 + \lambda/8$.

It is evident from this geometry that

$$\left(R_0 + \frac{\lambda}{8}\right)^2 = \frac{L_{eff}^2}{4} + R_0^2 \qquad (21.11)$$

If this expression is solved for L_{eff}, subject to the assumption that $\lambda/16$ is small compared with R_0, the result is

$$L_{eff} = \sqrt{R_0\lambda} \qquad (21.12)$$

Combination of Eqs. (21.2) and (21.12) gives

$$\beta_{eff} = \frac{1}{2}\sqrt{\frac{\lambda}{R_0}} \qquad \text{rad} \qquad (21.13)$$

Multiplying this beamwidth by range results in the resolution given by Eq. (21.7).

It will be noted that for the unfocused case the transverse linear resolution is independent of the antenna aperture size, fineness of resolution is increased by the use of shorter wavelengths, resolution varies as the square root of λ, and the resolution deteriorates as the square root of range. A plot of Eq. (21.7) is given in Fig. 21.1.

The Focused Case. An expression for the resolution achievable in the focused case has been given as Eq. (21.8). It is significant that the azimuth resolution achievable for this case depends only upon the physical antenna aperture and that, in contradistinction to the conventional case, fine resolution requires the use of small rather than large antennas. Also significant is the fact that the achievable resolution for a given antenna size is independent both of the range and of the wavelength used. A graph of Eq. (21.8) is also shown in Fig. 21.1.

In order to achieve the resolution indicated by Eq. (21.8), the synthetic aperture length required is

$$L_{eff} = \frac{\lambda R}{D} \qquad (21.14)$$

The considerations used in arriving at Eq. (21.12) indicated that, unless additional processing were applied to the signals, antenna lengths such as those implied by Eq. (21.14) could not be achieved. The processing required is an adjustment of the phases of the signals received at each point of the synthetic antenna, which makes these signals of equal phase (cophase) for a given target. If this is done, the restrictions which limited the maximum antenna length to that given by Eq. (21.12) are no longer pertinent and the new limitation on the length of the synthetic antenna achievable becomes simply the linear width of the radiated beam at the range of the target.

In some cases, a resolution coarser than $D/2$ is sufficient. Then a fraction γ of the maximum focused synthetic antenna length can be used. For this case

$$L_{eff} = \frac{\gamma\lambda R}{D} \qquad (21.15)$$

and the achievable resolution is

$$\text{Resolution}_{foc} = \frac{D}{2\gamma} \qquad (21.16)$$

For situations in which the synthetic antenna length given by Eq. (21.15) is less than or equal to the synthetic antenna length for the unfocused case as given by Eq. (21.12), only a limited improvement in resolution is achievable for the focused case. However, if a resolution finer than that given by Eq. (21.7) is desired, focusing must be used. Focusing removes the restriction on synthetic aperture length that would otherwise apply.

21.3 RADAR SYSTEM PRELIMINARIES

Whether or not synthetic aperture generation is used, a number of components are required for a radar system. The use of a synthetic antenna and/or pulse compression places additional requirements on some of these components, especially with respect to coherence and stability.

It is the purpose of this section to present a block diagram of the portions of the radar system that precede the signal processor. A block diagram and several variants are described. The major variant is for the purpose of describing the transmitter-receiver portions of a radar system for the cases of synthetic antenna generation alone as compared with the case of synthetic antenna generation combined with pulse compression. The signal-processing operations will be discussed later.

The essential elements of a radar system useful in a synthetic aperture situation are shown in Fig. 21.3. The components that determine the radiated waveform are shown within the dotted lines in the upper left-hand corner of the diagram. This equipment consists of two stable oscillators. One of them is a local
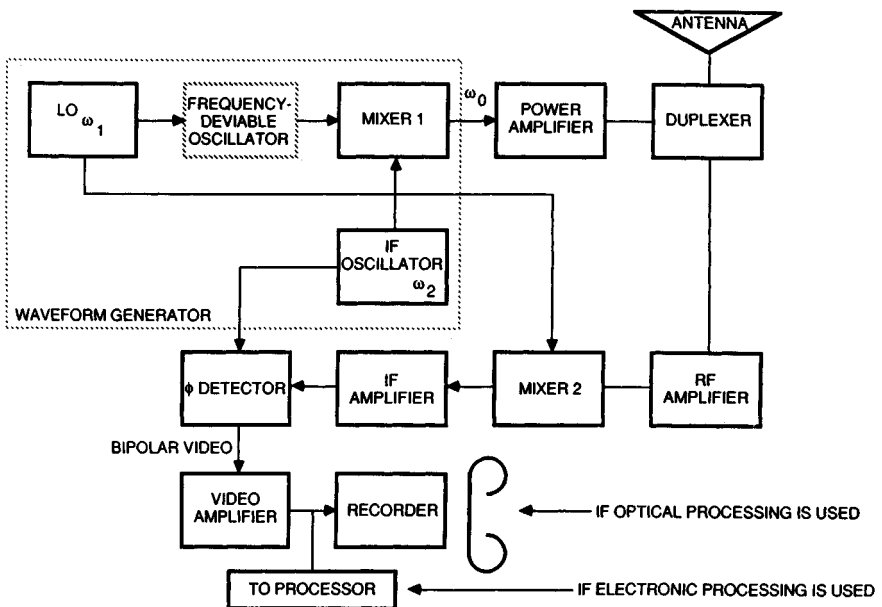
FIG. 21.3 Block diagram of a coherent radar system.

oscillator (LO) at radian frequency ω_2. The outputs of these oscillators are fed into mixer 1. In this mixer, a multiplicity of sum and difference frequencies is generated, and either the sum frequencies or the difference frequencies are selected and fed to the power amplifier.

If synthetic antenna generation without pulse compression is to be accomplished, the dotted component labeled "frequency-deviable oscillator" is not used, and the local oscillator is fed directly into mixer 1.

If pulse compression is to be combined with synthetic antenna generation, a frequency-deviable oscillator (FDO) is used for obtaining the desired waveform. In this case, the local oscillator is used to lock in the FDO. A ramp voltage is used to linearly frequency-modulate the FDO. This linearly frequency-modulated signal is then fed into mixer 1, instead of the LO signal, for the case of pulse compression. Waveforms other than linear frequency modulation may be used for pulse compression. In Fig. 21.3 the output of the video amplifier is fed to a recorder if optical processing is to be performed and/or to an electronic processor.

21.4 SIGNAL-PROCESSING THEORY

The theory of synthetic antenna generation combined with pulse compression is carried out below to show the information theoretic considerations involved and to indicate the operations necessary for achieving both azimuth synthetic antenna generation and pulse compression. A combined range-azimuth resolution function is derived. Following this treatment, an analysis of the signal-to-noise-ratio characteristics of a synthetic antenna compression radar are analyzed.

Detailed Resolution Analysis.   The analysis will be carried out in terms of an ambiguity function whose properties indicate both the azimuth and the range resolution of the system. In the analysis, some conditions are stated for which the terms affecting range resolution can be factored from the elements affecting azimuth resolution, so that the resulting ambiguity function can be written as the product of two factors, one for range and one for azimuth.

Role of the Generalized Ambiguity Function.   In this subsection, a definition of a generalized ambiguity function will be given, and its role in determining the resolution of a system will be interpreted.

To determine the generalized ambiguity function for a radar system, let a waveform $f(t)$ be radiated. We consider the operations performed upon the received signals with the objective of determining the radar reflectivity of the terrain being mapped. The function $f(t)$ may assume a variety of forms and may be a succession of short signals. If the quantity $\rho(x,y,z)$ represents the reflectivity of the terrain being mapped, the signal received by a radar system can be described by

$$s(t) = \int \int \int \rho(x,y,z) f\left[t - \frac{2R}{c}\right] dx \, dy \, dz \qquad (21.17)$$

The integration extends over the illuminated patch, and R is the range between a point (x,y,z) on the ground and the radar position $(vt,0,h)$. This equation shows that the received signal is the superposition of a large number of reflections

within the illuminated pattern of the antenna and within the range gate which arrive simultaneously at the radar antenna.

The radar design problem is one of designing an operation on $s(t)$ to recover the reflectivity function $\rho(x,y,z)$. One such operation consists of passing the signals $s(t)$ through a matched filter. The operation of subjecting $s(t)$ to this matched filter is given by

$$e_0(R,R') = \int f^*\left[t - \frac{2R'}{c}\right]s(t)\ dt \qquad (21.18)$$

In this equation, the asterisk indicates complex conjugation, and R' indicates the range from the radar antenna to the specific point (x',y',z') at which the reflectivity is to be evaluated.

Substitution of Eq. (21.17) into Eq. (21.18) gives a fourfold integral for the output, namely,

$$e_0 = \int\int\int\int \rho(x,y,z)f\left[t - \frac{2R}{c}\right]f^*\left[t - \frac{2R'}{c}\right]dt\ dx\ dy\ dz \qquad (21.19)$$

If the order of integration can be inverted so that the integration with respect to t is performed first, one can define a quantity $\chi(x,y,z;x',y',z')$. This quantity is the generalized ambiguity function, given by

$$\chi(x,y,z;x',y',z') = \int f\left[t - \frac{2R}{c}\right]f^*\left[t - \frac{2R'}{c}\right]dt \qquad (21.20)$$

In terms of the generalized ambiguity function defined by Eq. (21.17), Eq. (21.19) can be rewritten as

$$e_0 = \int\int\int \chi(x,y,z;x',y',z')\rho(x,y,z)\ dx\ dy\ dz \qquad (21.21)$$

Equation (21.21) shows that the ambiguity function can be considered as a weighting function on $\rho(x,y,z)$. The output of the radar system, therefore, is the weighted average of ρ over a domain determined by the limits of integration. If the ambiguity function is localized at some point and is essentially zero at all other points, the output will be a good representation of the radar reflectivity at that point. Otherwise, the estimate of the reflectivity at a given point will be the weighted average given by Eq. (21.21).

Although no use will be made in this section of considerations determined by the limits of integration, it should be pointed out that Eq. (21.21) states that the output estimate of reflectivity in the radar system is a weighting of the reflectivity ρ by the product of the ambiguity function and the illumination function, whereby illumination function is meant the function that determines the distribution of signal energy over the plane. Ordinarily, the antenna illumination pattern, the pulse length, and the terms appearing in the radar equation determine this illumination function. In some cases, the ambiguity function χ has peaks at more than one point. If the illumination function excludes all but one of these peaks, an unambiguous system results.

Factorization of the Ambiguity Function. Let it be assumed that $f(t)$ can be written

$$f(t) = g(t)e^{i\omega_0 t} \tag{21.22}$$

In this equation, $g(t)$ is considered a complex function having both magnitude and phase, whereas ω_0 represents a carrier frequency. If $f(t)$ with the form given by Eq. (21.22) is used in Eq. (21.20), one obtains for the ambiguity function the expression

$$\chi = \int g\left[t - \frac{2R}{c}\right] g^*\left[t - \frac{2R'}{c}\right] e^{-i\omega_0(2R/c \ - \ 2R'/c)} \, dt \tag{21.23}$$

Let $f(t)$ consist of a sequence of transmissions. It is assumed that successive transmissions may be alike or that they may be different. Thus $f(t)$ will have the characteristics of being nonzero for a sequence of time intervals and of being zero otherwise. Further, let it be assumed that the exponential term in Eq. (21.23) varies slowly during each of these transmissions. This is equivalent to a statement that the electrical path length between a target and the radar changes by a small amount during each transmission. If this assumption is valid, the exponential term in Eq. (21.23) can be considered a constant during a given transmission, although it will vary between transmissions.

The integral that is the coefficient of the exponential term in Eq. (21.23) has the form of an autocorrelation function of g with itself. This autocorrelation function for g is given by Eq. (21.24). It will be noted that the autocorrelation function of g is a function of the difference in range $R - R'$. In Eq. (21.24) the integral is carried out over the times that $g(t - 2R/c)$ overlaps $g^*(t - 2R'/c)$ for a given transmission.

$$\phi_{gg} = \int g\left[t - \frac{2R}{c}\right] g^*\left[t - \frac{2R'}{c}\right] dt = \phi_{gg}\left[\frac{2R}{c} - \frac{2R'}{c}\right] \tag{21.24}$$

If the notation given by Eq. (21.24) is used, one obtains

$$\chi = \Sigma\phi_{gg}\left[\frac{2R}{c} - \frac{2R'}{c}\right] e^{-i\omega_0(2R/c \ - \ 2R'/c)} \tag{21.25}$$

Examination of Eq. (21.25) shows that if ϕ_{gg}, the autocorrelation function for g, is the same function for each member of the sequence of transmissions, then this element can be factored out and written outside the summation term of Eq. (21.25). The expression after this common term has been factored out is

$$\chi = \phi_{gg}\left[\frac{2R}{c} - \frac{2R'}{c}\right] \Sigma e^{-i\omega_0(2R/c \ - \ 2R'/c)} \tag{21.26}$$

The summation term in Eq. (21.26) gives the azimuth resolution of the system, whereas the term ϕ_{gg} gives the range resolution. It is evident from Eq. (21.26) that the autocorrelation function of g rather than g itself determines the range resolution of the system.

A variety of waveforms have been used to achieve range resolution. Among them the two most important are those in which $g(t)$ is a short pulse and those in which $g(t)$ is a linearly frequency-modulated short pulse (chirped signal). It is, of

course, evident that any other waveform having a desirable autocorrelation function is equally possible for $g(t)$.

The form of $g(t)$ to be analyzed is that of the linearly frequency-modulated case.

Azimuth Resolution Factor of the Ambiguity Function. The azimuth resolution capability of the system can be determined from an evaluation of the sum term written as the factor in Eq. (21.26). In this sum, R represents the range from the radar to an arbitrary point on the terrain being mapped, whereas R' represents the range to a specific point for which the reflectivity is to be estimated. The geometry appropriate to the synthetic antenna case is shown in Fig. 21.4. In this diagram, it is assumed that the aircraft carries a side-looking antenna and flies at height h and with velocity v along the x axis so that the aircraft location is given by

$$x = vt$$

Consider two points being mapped having coordinates $(0,y_0,0)$ and $(x',y_0,0)$. R_0 is defined by

$$R_0 = \sqrt{y_0^2 + h^2} \tag{21.27}$$

and R and R' can be written, respectively, as

$$R = \sqrt{R_0^2 + x^2} \approx R_0 + \frac{x^2}{2R_0} \tag{21.28}$$

$$R' = \sqrt{R_0^2 + (x - x')^2} \approx R_0 + \frac{(x - x')^2}{2R_0} \tag{21.29}$$

The second expression for Eqs. (21.28) and (21.29) is an approximation that is valid whenever the inequalities $x \ll R_0$, $(x - x') \ll R_0$ are satisfied. If the approximate forms for $R - R'$ as given by the approximate expressions in Eqs.
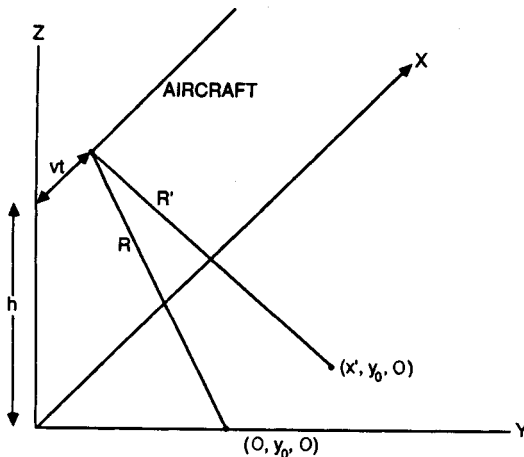
FIG. 21.4 Geometry for Eqs. (21.27) and (21.28).

(21.28) and (21.29) are used to evaluate the sum appearing in Eq. (21.26), one obtains

$$\sum e^{-i\omega_0(2R/c\,-\,2R'/c)} = \sum e^{-i(2\omega_0/c)(2xx'\,-\,x'^2)/2R_0} \tag{21.30}$$

Thus far, the summation index has not been defined. To proceed further, it is necessary to indicate the summation index and its bounds. Let it be assumed that the transmissions radiated are a sequence of pulses with time intervals between pulses that are multiples of T. Then the variable x can be given as an integral multiple of the distance vT moved between successive transmissions. This relationship is

$$x = nvT \tag{21.31}$$

If Eq. (21.31) is substituted into Eq. (21.30), one obtains

$$\sum = e^{i(2\omega_0/c)x'^2/2R_0} \sum_{-N/2}^{N/2} e^{-i4\pi(x'/\lambda R_0)nvT} \tag{21.32}$$

In writing Eq. (21.32), the summation is carried over $N + 1$ terms. The synthetic antenna length implied by these limits is given by $L = NvT$.

Inasmuch as the summation terms in Eq. (21.32) are those corresponding to a geometric progression, the sum term can be immediately evaluated. The result is

$$\sum = e^{i(2\omega_0/c)x'^2/2R_0} \frac{\sin\,[(N\,+\,1)4\pi x' vT/2\lambda R_0]}{\sin\,[4\pi x' vT/2\lambda R_0]} \tag{21.33}$$

The right-hand side of Eq. (21.33) gives the factor of the generalized ambiguity function that is responsible for the azimuth resolution. It will be noted that there are a phase term given by the exponential and a magnitude term given by the remaining terms in Eq. (21.33). The specific form of Eq. (21.33) is a consequence of the equal weighting of the signals. A weighting function can be used to shape the sidelobes in direct analogy with the use of such a technique in real antenna design.

Range Resolution Factor of the Ambiguity Function. This subsection considers the factor in the generalized ambiguity function, Eq. (21.25), that is responsible for range resolution. This factor, ϕ_{gg}, has been defined by Eq. (21.24). A specific form for $g(t)$ will be assumed, and an evaluation of ϕ_{gg} will be made for this specific waveform. The function $g(t)$ to be analyzed is that in which each radiation consists of a short, linearly frequency-modulated signal. An expression for $g(t)$ in this case is

$$g(t) = e^{i\alpha t^2} \tag{21.34}$$

The use of Eq. (21.34) in Eq. (21.24) results in

$$\phi_{gg} = e^{i\alpha[(2R/c)^2\,-\,(2R'/c)^2]} \int_{-\tau/2}^{\tau/2} e^{-i\alpha[(4R/c)t-(4R'/c)t]}\,dt \tag{21.35}$$

$$\phi_{gg} = e^{i\alpha[(2R/c)^2 - (2R'/c)^2]} \tau \frac{\sin\{\alpha\tau[(2R/c) - (2R'/c)]\}}{\alpha\tau[(2R/c - 2R'/c)]} \qquad (21.36)$$

Equation (21.36) gives the range resolution factor of the ambiguity function for a transmitted waveform of the type expressed by Eq. (21.34). It will be noted that this term consists of a phase term and an amplitude term.

Inasmuch as both the azimuth resolution factor and the range resolution factor have been evaluated, the generalized ambiguity function can be written:

$$\chi = e^{i\alpha[(2R/c)^2 - (2R'/c)^2]} \tau \frac{\sin\alpha\tau[(2R/c) - (2R'/c)]}{\alpha\tau(2R/c - 2R'/c)} e^{i(2\omega_0/c)(x'^2/2R_0)}$$

$$\times \frac{\sin[(N + 1)(2\pi x'vT)/(\lambda R_0]}{\sin[2\pi x'vT/\lambda R_0]} \qquad (21.37)$$

In interpreting Eq. (21.37), it should be noted that there are phase terms and magnitude terms. One of the magnitude terms corresponds to the range resolution capability of the system; the other, to azimuth resolution. Quantitative expressions for the resolution in each of these coordinates will be obtained below.

The resolution terms in Eqs. (21.33), (21.36), (21.37) are of the form

$$\sin Nz/\sin z \qquad (21.38)$$

with

$$z = 2\pi x'vT/\lambda R_0 \qquad (21.39)$$

for the azimuth resolution case, and

$$\sin z/z \qquad (21.40)$$

with

$$z = \alpha\tau[2R/c - 2R'/c] \qquad (21.41)$$

for the range resolution case.

If the value

$$L = (N + 1)VT \qquad (21.42)$$

is combined with Eq. (21.39) and one recognizes that

$$\alpha\tau = 2\pi B \qquad (21.43)$$

where B is the chirp signal bandwidth, one can show that the azimuth resolution δ_a and the range resolution δ_r are given by

$$\delta_a = (1.4\lambda R_0)/(\pi L) \qquad (21.44)$$

and

$$\delta_r = (1.4c)/(2\pi B) \tag{21.45}$$

Ambiguities.   It is the objective of this subsection to make some observations regarding the possibility of multiple peaks (ambiguities) in the ambiguity function as given by Eq. (21.37) and its effect on system performance as given by Eq. (21.19).

The final term on the right-hand side of Eq. (21.37) is of the form

$$\frac{\sin\left[(N+1)q\right]}{\sin q} \tag{21.46}$$

where the quantity q is defined by

$$q = \frac{2\pi x'vT}{\lambda R_0} \tag{21.47}$$

The azimuth resolution factor, therefore, has a peak whenever the quantity q takes on a value equal to an integral multiple of π rad. Thus the system is potentially ambiguous for values of x' that are solutions of

$$\frac{2\pi x'vT}{\lambda R_0} = m\pi \tag{21.48}$$

Actually, it is more meaningful to solve for the ratio x'/R_0. The angle v gives directions from the broadside at which angle ambiguities are potentially possible. This relationship is given as

$$\sin\theta = \frac{x'}{R_0} = \frac{\lambda}{D}\frac{D}{vT}\frac{m}{2} = \frac{m\beta D}{2vT} \tag{21.49}$$

In the last form of this equation, the numerator and the denominator have been multiplied by D, the horizontal aperture of the antenna, to express the result in terms of the radiated beamwidth $\beta = \lambda/D$.

Thus, the possibility of azimuth ambiguities arises as a natural consequence of the signals radiated and of the processing method. Ordinarily, these potential ambiguities in azimuth are suppressed by the illumination factor. The illumination pattern β is chosen so that the values of β corresponding to more than one value of m are not illuminated.

Possibilities also exist for ambiguities in range. The analysis carried out to the point of Eq. (21.37) was not sufficiently general to predict ambiguities in range. However, if reference is made to Eq. (21.24), it is evident that the autocorrelation function ϕ_{gg} will be periodic if $g(t)$ is periodic. Thus range ambiguities can also occur. In particular, range ambiguities will occur for ranges having a difference given by

$$\Delta R = \frac{cT}{2} \tag{21.50}$$

where T is the interpulse period.

To date, systems have been built which have avoided ambiguities by virtue of

illuminating only the part of the ambiguity diagram that excludes all but one major peak. This technique has sometimes been referred to as *ambiguity avoidance*.

For some sets of parameters, ambiguities cannot be avoided by using a radar with a single radiated beam. The use of multiple beams solves this problem. This topic is discussed in Sec. 21.5.

Signal-to-Noise-Ratio Considerations. It is the purpose of this subsection to derive expressions for signal-to-noise (S/N) ratio for radars in which pulse compression and synthetic antenna techniques are used. The signal-to-noise ratio for a radar system as a result of the reception of a single pulse is given by the well-known radar equation

$$\frac{S}{N} = \frac{P_t G_t A_r \sigma}{(4\pi)^2 R^4 k T_0 B F_n} \qquad (21.51)$$

In a pulse compression radar, signal-to-noise improvement occurs in the ratio of the uncompressed pulse length τ_i to compressed pulse length τ_0.

In a radar that achieves its azimuth resolution by the generation of a synthetic antenna, there is an additional signal-to-noise improvement factor due to the integration of a number of pulses. The number of pulses integrated is equal to the product of the pulse repetition frequency (PRF) and the time necessary to generate the synthetic antenna. In turn, this time is equal to the ratio of synthetic length L to aircraft speed v.

An expression in which the product of both factors has been written is

$$\text{Improvement factor} = \frac{\tau_i}{\tau_0} \frac{\text{PRF } L}{v} \qquad (21.52)$$

The length of synthetic antenna required to achieve azimuth resolution δ_{az} at range R and wavelength λ is given by

$$L = \frac{R\lambda}{2\delta_{az}} \qquad (21.53)$$

The substitution of Eq. (21.53) into Eq. (21.52) gives for the improvement factor

$$\text{Improvement factor} = \left[\frac{\tau_i}{\tau_0}\right] \frac{\text{PRF } R\lambda}{2v\delta_{az}} \qquad (21.54)$$

The signal-to-noise ratio including the improvement factor is obtained by multiplying together the expressions given by Eqs. (21.51) and (21.54). The result of this multiplication is

$$\frac{S}{N} = \frac{P_t G_t A_r \sigma}{(4\pi)^2 R^4 k T_0 B F_n} \frac{\tau_i}{\tau_0} \frac{\text{PRF } R\lambda}{2v\delta_{az}} \qquad (21.55)$$

Although Eq. (21.55) contains the desired information, it is useful to modify the term somewhat by expressing the antenna gain in terms of the effective area of its aperture and of the wavelength. This expression is written

$$G_t = \frac{4\pi A_r}{\lambda^2} \tag{21.56}$$

It is also desirable to collect together three terms in the numerator of Eq. (21.55), namely, P_t, the peak transmitted power; τ_i, the uncompressed pulse length; and the PRF. The product of these three factors gives the average power P_{av}. This relationship is written

$$P_{av} = P_t \tau_i \text{PRF} \tag{21.57}$$

In the design of a radar system, the bandwidth B is chosen to be the reciprocal of τ_0. Hence the product of the bandwidth and the compressed pulse width is approximately equal to unity. This relationship is written

$$B\tau_0 \approx 1 \tag{21.58}$$

Finally, it is useful to express the radar cross section σ in terms of the azimuth and range resolution, δ_{az} and δ_r, as well as in terms of the reflectivity of the terrain, ρ. The radar cross section is equal to the reflectivity of the terrain multiplied by the projected area. This projection accounts for the term $\sin \psi$. The expression for the radar cross section in terms of these parameters is given by

$$\sigma = \rho \delta_r \delta_{az} \sin \psi \tag{21.59}$$

Substitution of Eqs. (21.56) to (21.59) for the corresponding quantities in Eq. (21.55) gives

$$\frac{S}{N} = \frac{P_t \sigma_i \text{PRF} 4\pi A^2 \rho \delta_r \delta_{az} (\sin \psi) R\lambda}{(4\pi)^2 R^4 \lambda^2 k T_0 F_n (B\tau_0) 2v \delta_{az}} \tag{21.60}$$

In writing Eq. (21.60), no cancellation of terms has been made. Canceling terms that appear in both the numerator and the denominator results in

$$\frac{S}{N} = \frac{P_{av}}{8\pi} \frac{A_r^2 \rho \delta_r}{k T_0 F_n R^3 \lambda} \frac{\sin \psi}{v} \tag{21.61}$$

This is the desired result.

Equation (21.61) does not take into account factors concerned with ambiguity avoidance. The inclusion of such effects is given in Ref. 5.

Equation (21.61) shows that the signal-to-noise ratio at the output of a radar that has used pulse compression and has generated a synthetic antenna has the following properties different from conventional radar:

1. The signal-to-noise ratio is proportional to the size of the range resolution element and is independent of the size of the azimuth resolution element.

2. The signal-to-noise ratio is inversely proportional to the third power of range.

3. The signal-to-noise ratio is inversely proportional to the wavelength.

4. The signal-to-noise ratio is inversely proportional to the speed of the aircraft.

Effect of Phase Errors. In actual equipment, phase errors arise from a number of sources. Some of the instabilities arise in oscillators and other electrical components of the radar, but other sources of phase error are inhomogeneities in the atmosphere or the result of uncompensated deviation of the aircraft from linear unaccelerated motion. A number of modifications of the synthetic antenna pattern result from such uncompensated phase errors. These modifications include beam canting, beam spreading, peak gain reduction, and redistribution of the ratio of energy in the main lobe to that in the sidelobes. An analytic formulation and Monte Carlo computer simulation of the effects of phase errors for normally distributed random phase errors and for three cross-correlation functions have been given by Greene and Moller.[6]

Signal Processing. The preceding subsections have discussed a number of aspects of radar signal processing. Also discussed has been the radar system up to the point of signal processing. As part of that analysis, the waveforms of signals at a number of points in the radar system have been described. It is the purpose of this subsection to discuss a number of aspects of signal processing that are common to all mechanizations.

Many fine-resolution radar systems employ both pulse compression and synthetic antenna generation.

Theoretic Aspects of Synthetic Aperture Generation. In generating a synthetic antenna, the returns from a number of spatial positions must be combined. In doing this, one usually wishes to apply weighting to the signals for synthetic antenna pattern sidelobe-level control; in the case of focused synthetic antennas, one also wishes to adjust the phases of the signals before combination.

In the preceding discussion, the signal was represented as a function of time. For present purposes it is preferable to consider the signals as a discrete sequence numbered from 1 to N. Let S_n represent the signal received when the physical antenna is at the nth position of the antenna array. Let W_n represent weighting applied to S_n, and let ϕ_n represent the phase adjustment required for focusing.

The operation of synthetic antenna generation then consists of taking the vector sum of the signals S_n, adjusting their phases, and multiplying by weighting factors. The sum of this operation is given by

$$\Sigma S_n e^{i\phi_n} W_n = \text{focused pattern} \tag{21.62}$$

In the case of unfocused synthetic antenna generation the phase adjustments ϕ_n are not made. In this case the signal operations required have the form

$$\Sigma S_n W_n = \text{unfocused synthetic pattern} \tag{21.63}$$

There are many mechanizations possible to carry out the operations indicated by Eqs. (21.62) and (21.63). Some of them are described below. Two common techniques are digital and optical in nature.

Discussions regarding optical and digital data processing are given in Refs. 7 through 11.

Optical Techniques. The optical techniques involve the recording of the radar signals on a transparency, most frequently silver halide photographic film in any of a number of formats. Initially the successive range sweeps were placed parallel and side by side; later polar format was used. The growing understanding that we are really collecting a portion of the three-dimensional spectrum has led to use of a three-dimensional storage format.[8] This topic will be discussed further in Sec. 21.5.

The most frequently used optical processors are based on the tilted-plane optical processor described by Kozma, Leith, and Massey.[7] In this processor, both the input plane and the output plane are tilted (i.e., they are not perpendicular to the optical axis). The optical components are telescopic, and the powers of the elements in two perpendicular planes are unequal. The telescopic elements include both spherical and cylindrical elements.

The evolution of this processor is based in part upon the recognition that signal histories may be assigned focal lengths and behave to some degree as optical elements.

The azimuth along-track signals from a point target are similar to those of a zone plate. The focal length is proportional to target range. If pulse compression is used, all targets have associated with them zone plates in the range direction. These all have the same focal length. The recorded signals before processing are often referred to as *signal histories*.

Digital Processors. Digital processing has emerged as the preferred means when the amount of data to be processed is not too great. Mechanization of SAR operations is often computation intensive.

In cases not based on polar format, correlation operations have been used. These operations are usually performed using frequency-plane equivalent of correlation.

$$\int f(q)\, g(q - x)\, dx = \int F(\omega)\, G(\omega)\, \exp\,(j\omega x)\, d\omega \qquad (21.64)$$

In other cases, such as polar format, Fourier transform operations are indicated and the use of the fast Fourier transform (FFT) plays an important role.

Associated with FFT processing is the fact that algorithms exist for processing two-dimensional data with sampling points in the rectangular arrays, whereas the sample points obtained are equally spaced on radial lines. This requires formatting operations on the data points to convert them to rectangular format.[12,13]

Imagery from synthetic aperture radars is shown in Figs. 21.5 and 21.6. These images were provided by the Environmental Research Institute of Michigan (ERIM).

21.5 *ADDITIONAL SYSTEM CONSIDERATIONS*

In this section a number of considerations peculiar to synthetic aperture radar are discussed. Some are additional performance requirements on the components of the system; some are concerned with system aspects.

Antenna. The horizontal aperture of the antenna determines the finest azimuth resolution achievable in a single-beam synthetic aperture radar except for the searchlight mode. Moreover, in the signal processing it is assumed that the antenna gain is constant as a function of along-track position. Thus, it is necessary to have a degree of stabilization of antenna pointing so that the beam

FIG. 21.5 STAR-1 radar imagery, lower Lake St. Clair, upper Detroit River. Resolution, 20 ft (6 m). (*Courtesy Environmental Research Institute of Michigan.*)

rotation is some minor fraction of the beamwidth. In most cases, the antenna is side-looking, although in some cases the antenna is positioned to an angle off broadside and the system then operates in what is called the *squint* mode.

Receiver-Transmitter. The transmitter and receiver for synthetic antenna radars require maintenance of coherence of the radar signals. Consequently, there is emphasis on the stability of the oscillators and more rigorous requirements on the components. The output of coherent radar is the output of a synchronous demodulator rather than that of an envelope detector commonly used in radars. The output is bipolar video, in which a reference bias level corresponds to zero level of video output.

Storage and Recording. It is inherent that synthetic antenna radars and pulse compression radars require the storage of radar data, because the *data* for synthetic antenna generation does not occur simultaneously but is collected over some interval of time. Operations are then performed on these signals to achieve the selectivity of the radar. Moreover, each radar return participates in forming the output for a large number of points on the output map. The requirements for storage are therefore very large. Since a high volume of storage is required for a fine-resolution radar system, photographic storage is commonly used.

For digital processing, storage of the digital signals after analog-to-digital (A/ D) conversion is required. The amount of this data can be great and often limits the area over which fine resolution can be obtained. A description of what needs to be done is given in Refs. 12 and 13.

In selecting a storage medium one must consider the rate at which information must be recorded, the amount of data to be recorded, and the rate at which the
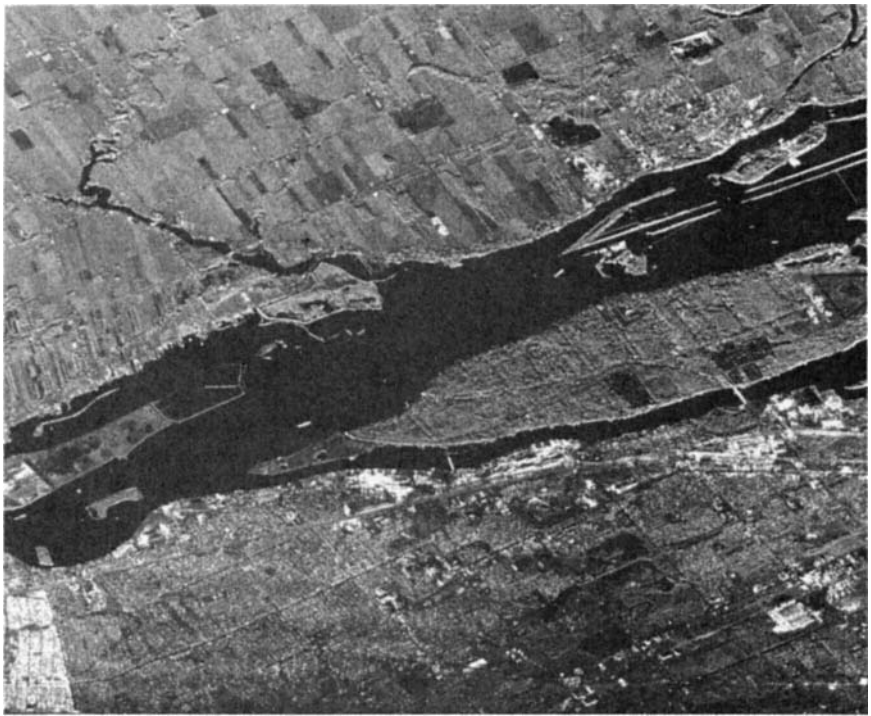
FIG. 21.6 STAR-1 radar imagery, lower Detroit River. Resolution, 20 ft (6 m). (*Courtesy of Environmental Research Institute of Michigan.*)

storage must be read out for performing the azimuth compression and the pulse compression.

Motion Compensation. In generating the synthetic antenna the signal-processing equipment assumes that the radar flies along a straight line at constant speed. In practice, the vehicle carrying the antenna is subject to deviations from unaccelerated flight. Therefore, it is necessary to have auxiliary equipment to compensate for other than straight-line motion. Motion-compensation equipment must include sensors capable of detecting the deviation of the flight path from a linear path. The output of these sensors is used in a variety of ways. For motion compensation proper, the received-signal phase must be adjusted to compensate for the displacement of the real antenna from the location of the ideal synthetic antenna being generated.

A consideration of the geometry involved shows that the phase correction that must be applied is a function of depression angle. Consequently, the correction must be made as a function of range. The rate of change correction is very rapid at steep depression angles and becomes slower at shallow depression angles.

Squint Mode. In most examples of synthetic aperture radar, the beam is directed at right angles to the ground track of the aircraft. In some cases, however, it is desirable to "squint" the antenna beam so that an area either

forward or aft of the aircraft is mapped. It is necessary to position the antenna beam so that the maximum of its radiation pattern points in the desired squint direction. Moreover, it is usually necessary to modify the signal processors to take into account the average doppler frequency shift that occurs when the antenna points in a direction other than normal to the flight path. It is, of course, also necessary to take the geometry of the squint mode into account in designing recorders and displays.

Spotlight Mode. In Sec. 21.2 Eq. (21.8) was derived for a radar in *strip-map* mode, i.e., for the case that the radar antenna is in a fixed orientation and the radar beamwidth $R\lambda/D$ is used as the length of synthetic aperture generated. One can increase the antenna length by use of *spotlight* mode. In this case the radar antenna is continuously pointed toward the region being imaged. For this case one can make a synthetic antenna length longer than $R\lambda/D$, or one can make several images and noncoherently integrate them.

Spotlight mode also makes possible the use of higher antenna gain.

Effects of Motion Errors. In generating synthetic aperture antenna images, one needs to estimate the along-track and cross-track velocities of targets in order to derive the matched filter to use in imaging. If one has an error in the radial velocity of the target, one gets a rotation in target position. If one has an error in along-track velocity, a limit is set to the achievable resolution.

Multiple-Beam Radars. The analyses leading to Eqs. (21.5) and (21.8) are correct for synthetic aperture radars which radiate only one beam. However, system parameters sometimes dictate the use of multiple beams.

The use of multiple beams is motivated by several considerations, such as ambiguity avoidance and the achievement of higher antenna gain. The achievement of a larger area coverage rate is another but less likely motivation. A great deal of flexibility is possible. The multiple beams may be arranged in the azimuth direction or the range direction, or both. With the use of multiple beams, it is possible to achieve any desired combination of unambiguous range, resolution, and area rate. The antenna area and the number of beams are determined from the performance parameters.

A more complete analysis of multibeam systems is given in Refs. 14 through 17.

ISAR. *Inverse synthetic aperture radar* (ISAR) is the term used when the motion of the object being imaged is used instead of the motion of the radar. A more general case is that in which both the radar and the object are in motion. In ISAR, the target motion is often not known to the radar. Hence, a major part of the problem is determination of the target motion to generate the matched filter needed to generate an image. A number of techniques have been studied for providing data regarding both translation and rotation of moving objects. An example of such work is that of B. Steinberg.[18]

Three-Dimensional Spectrum. The analysis starting with Eqs. (21.15) and (21.16) contains the assumption that a matched filter is applied to the radar returns from each point. This can in fact be done. It would reduce the signal-processing load if a reference function could be applied over a region. This, too, can be done, but range walk and defocusing set limits to the size of a patch which can be handled in this manner. Of the methods that have been proposed, the use of polar format and its generalization and the collection of a

portion of the three-dimensional spectrum of the scene being mapped are most significant.

If one starts with an expression such as Eq. (21.17), performs the integration with respect to τ, where τ is a dummy variable to replace t, and then makes a Fourier transform of the results, one gets

$$E_0(\omega) = \int \rho(x,y,z)|G(\omega)|^2 \exp\left[-j(2\omega/c)(2r/c - 2r'/c)\right] dx\, dy\, dz \quad (21.65)$$

In this equation $|G(\omega)|^2$ is the Fourier transform of the autocorrelation function of $g(t)$. Let the vector difference, or r, and r' be represented by q.

$$\mathbf{r} = \mathbf{r'} + \mathbf{q} \qquad (21.66)$$

so that

$$r = r' + \frac{r' \cdot q}{r'} + \frac{q^2 + (r' \cdot q/r')^2}{2r'} \qquad (21.67)$$

If one can neglect the last term, and if one writes

$$\hat{\mathbf{r}} = \mathbf{r}/r \qquad (21.68)$$

for the unit vector along r, one can write

$$r - r' = \hat{\mathbf{r}} \cdot \mathbf{q} \qquad (21.69)$$

Let a vector G be defined by

$$\mathbf{G} = (2\,\omega/c)\hat{\mathbf{r}} \qquad (21.70)$$

then

$$E_0(\omega) = \int \rho(x,y,z) G(\omega)^2 \exp\left(-j\mathbf{G} \cdot \mathbf{q}\right) dq \qquad (21.71)$$

One notes that except for the factor $|G(\omega)|^2$, Eq. (21.71) has the form of a three-dimensional spectrum of $\rho(q)$.

Equation (21.71) has been derived by Jack Walker.[8] Related developments involving use of the *projection-slice theorem* have been given by a number of other authors.[9,19]

In interpreting Eq. (21.71), it is useful to consider the point r as a general point on the object being imaged and r' as a reference point on the object. The vector q is then the vector from the reference point to all other points on the object, and the integration extends over the object.

Equation (21.71), being a three-dimensional spectrum of an object, requires that the data be taken with an origin of coordinates and a coordinate system fixed with respect to the object being imaged. Hence the effect of translation and rotation of a moving object must be compensated in order to image that object.

The three-dimensional spectrum is the proper format for storing radar data. The polar format is a special case in which the radar collection is performed in a plane. Use of the projection-slice theorem enables one to project the data along any direction. This promises to give images of *slices* of moving objects.

The projection of the radar data, followed by a two-dimensional Fourier trans-

form, can be used to form images for the most general motion of both the radar and the moving object.

REFERENCES

1. Sherwin, C. W., P. Ruina, and R. D. Rawcliffe: Some Early Developments in Synthetic Aperture Radar Systems, *IRE Trans.*, vol. MIL-6, pp. 111–115, April 1962.

2. Wiley, C.: Pioneer Award acceptance remarks, *IEEE Trans.*, vol. AES-21, pp. 433–443, May 1986.

3. Cutrona, L. J., W. E. Vivian, E. N. Leith, and G. O. Hall: A High Resolution Radar Combat-Surveillance System, *IRE Trans.*, vol. MIL-5, pp. 127–131, April 1961.

4. Cutrona, L. J., and G. O. Hall: A Comparison of Techniques for Achieving Fine Azimuth Resolution, *IRE Trans.*, vol. MIL-6, pp. 119–121, April 1962.

5. Skolnik, M. I.: "Introduction to Radar Systems," 2d ed., McGraw-Hill Book Company, 1980, p. 522, Eqs. 14 and 16.

6. Greene, C. A., and R. T. Moller: The Effect of Normally Distributed, Random Phase Errors on Synthetic Array Gain Patterns, *IRE Trans.*, vol. MIL-6, pp. 130–139, April 1962.

7. Kozma, A., E. N. Leith, and N. G. Massey: Tilted Plane Optical Processor, *Appl. Opt.*, vol. 11, pp. 1766–1777, August 1972.

8. Walker, J. L.: Range Doppler Imaging of Rotating Objects, *IEEE Trans.*, vol. AES-16, pp. 23–52, January 1980.

9. Ausherman, D. A., A. Kozma, J. L. Walker, H. M. Jones, and E. C. Poggio: Developments in Radar Imaging, *IEEE Trans.*, vol. AES-20, pp. 363–400, July 1984.

10. Cutrona, L. J., E. N. Leith, C. J. Palermo, and L. J. Porcello: Optical Data Processing and Filtering System, *IRE Trans.*, vol. IT-6, pp. 386–400, June 1960.

11. McLeod, J.: The Axicon: A New Type of Optical Element, *J. Opt. Soc. Am.*, vol. 44, pp. 592–597, August 1954.

12. Ausherman, D. A.: Digital Image Processing, *Proc. SPIE*, vol. 528, pp. 118–133, Jan. 22–23, 1985.

13. Ausherman, D. A.: Digital versus Optical Techniques in Synthetic Aperture Radar (SAR) Data Processing, *Opt. Eng.*, vol. 19, pp. 157–167, March–April 1980.

14. Cutrona, L. J.: Means to Achieve Wide Swath Widths in Synthetic Aperture Radar, *Addendum to Proc. Synth. Aperture Radar Technol. Conf.*, pp. V-9-1–V-9-21, New Mexico State University, Las Cruces, Mar. 8–10, 1978.

15. Claassen, J. P., and J. Eckerman: A System Concept for Wide Swath Constant Incident Angle Coverage, *Proc. Synth. Aperture Radar Technol. Conf.*, Pap. VI-4, pp. VI-4-1–VI-4-19, New Mexico State University, Las Cruces, Mar. 8–10, 1978.

16. Cutrona, L. J.: Comparison of Sonar System Performance Achievable Using Synthetic-Aperture Techniques with the Performance Achievable by More Conventional Means, *J. Acoust. Soc. Am.*, vol. 58, pp. 336–348, August 1975.

17. Cutrona, L. J.: Additional Characteristics of Synthetic Aperture Sonar Systems and a Further Comparison with Non-Synthetic Aperture Sonar Systems, *J. Acoust. Soc. Am.*, vol. 61, pp. 1213–1217, May 1977.

18. Steinberg, B. D.: "Microwave Imaging with Large Antenna Arrays," John Wiley & Sons, New York, 1983.

19. Munson, D. C., Jr., J. D. O'Brien, and W. K. Jenkins: A Topographic Formulation of Spotlight-Mode Synthetic Aperture Radar, *Proc. IEEE*, vol. 71, pp. 917–925, August 1983.

CHAPTER 22
SPACE-BASED RADAR SYSTEMS AND TECHNOLOGY

Leopold J. Cantafio
Space and Technology Group TRW

22.1 INTRODUCTION

Significant developments have been made in space-based radar (SBR) systems and technology since the 1970 edition of the *Radar Handbook* was published. A new rendezvous radar was developed for the space shuttle and has become operational. The unmanned orbital maneuvering vehicle (OMV) will use a new low-cost rendezvous radar that is expected to be operational during the early 1990 time period. Synthetic aperture radar (SAR) types of SBR have been used for earth and planetary exploration. Altimeters have been used on many satellites. The technology of SBR subsystems has been developed in the areas of antennas, transmitters, receivers, solid-state transmit–receive (T/R) modules, signal processors, and prime power. This chapter will review SBR systems and technology with the intent to provide a description that is not too sketchy to be substantive. Therefore, selected systems and technology will be discussed. Several SBR systems for rendezvous, earth exploration, and planetary exploration missions will be described. Systems considerations such as the space environment, orbit selection, radar tradeoffs, advantages and disadvantages, and critical issues will be discussed. Many topics, such as electronic countermeasures, will have to be omitted. This chapter should be considered a status report on the new frontier for radar systems. A more comprehensive treatment of SBR can be found in "Spacebased Radar Handbook," written and edited by the author and published by Artech House.

22.2 SBR SYSTEMS CONSIDERATIONS

Types of SBR. There are three types of radar that have been and can be based in space. SBR that is typical of Type I is the small, short-range rendezvous radar such as those used on the Shuttle, Apollo, and Gemini programs.[1-4] Type II SBR includes the earth and planetary resources radar used for mapping, scatterometers, altimeters, and subsurface probing.[5-9] Side-

looking SAR techniques are typical of mapping radars such as those used on the Seasat satellite in June 1978 and the Shuttle in November 1981 with the Shuttle Imaging Radar-A (SIR-A). Type III SBR includes the large phased array surveillance radar proposed for multimission defense, air traffic control, and disarmament functions.[10–14]

Type I SBR. Gemini and Apollo programs demonstrated the first operational experience with the rendezvous maneuver. The successful performance of the rendezvous radars in these programs effectively opened the door to many possible missions that may be performed in space. The K_u-band integrated radar and communications subsystem (IRACS), designed for the space shuttle orbiter vehicle, demonstrated the rendezvous, satellite retrieval, and station-keeping missions. The maiden voyage for this radar was aboard *Challenger* (Shuttle) STS-7 on June 22, 1983.[15] During the STS-11 flight in February 1984, the K_u-band radar assisted in the checkout of the manned maneuvering unit (MMU) operations. The radar acquired and tracked mission specialist Robert Stewart in the MMU during his 300-ft sojourn into space. The radar measured the radar cross sections (RCS) of the MMU, which varied between 2.5 and 7.5 dBsm with acquisition at a range of 100 ft and track out to the maximum range of 308 ft. Average velocity during the mission was 0.7 ft/s.

The rendezvous radar provides the tracking function for a guidance system. The rendezvous phase of the mission begins after the radar acquires the target. Thereafter, the tracking function provides data on range, range rate, and the two components of the line-of-sight inertial rate. A digital guidance computer calculates relative velocity perpendicular to the line of sight, using range and angular rate data. The closing component of velocity is obtained from the doppler or by differentiation of radar range measurements. A simplified block diagram of a typical rendezvous guidance subsystem is shown in Fig. 22.1. The radar search and acquisition mode is initiated by the guidance computer. A relatively large solid angle is searched periodically until the target is acquired in range and angle. In order to maximize the probability of detection and acquisition, the kinematics are arranged such that a long search time is available before the target escapes from
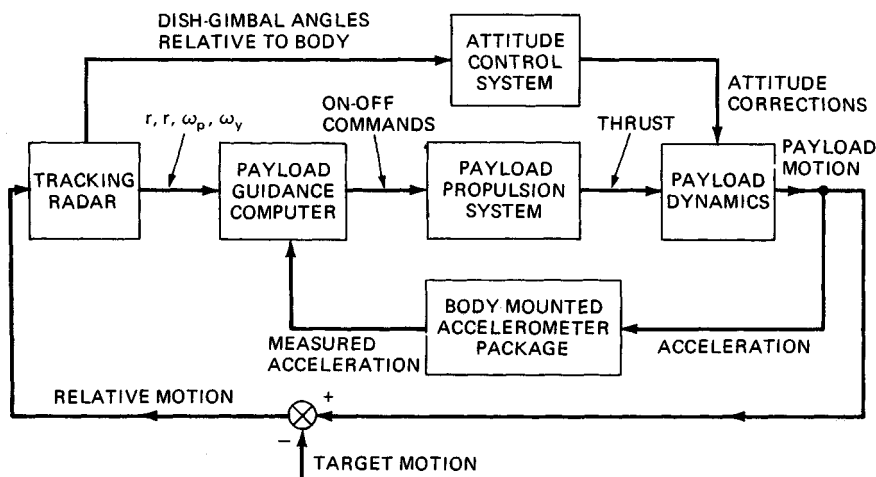
FIG. 22.1   Rendezvous guidance subsystem: simplified block diagram.[16]

TABLE 22.1   STS Rendezvous Radar Requirements*

Search	± 30° spiral scan
Acquisition	12 nmi on 0 dBsm SW-1; 300 nmi on +14 dBW transponder
Track	
Range	± 1 percent
Range rate	1 ft/s or 1 percent
Angle	8 mrad
Angle rate	0.14 mrad/s or 5 percent

*From Ref. 15.

the search sector. When detection has been accomplished, the search mode is stopped and the tracking mode is initiated by locking a tracking gate onto the target return and thereafter monopulse angle-tracking the antenna about an axis always directed toward the target. The tracking phase ends when rendezvous has been achieved within certain desired terminal accuracy on relative position and velocity. The typical requirements for the STS rendezvous radar are given in Table 22.1.[15]

At and immediately following acquisition, the relative velocity vector will generally lie in the direction of the instantaneous line of sight; however, there may be a substantial error equivalent to a relative velocity component perpendicular to the line of sight. The range at acquisition and the magnitude of the closing velocity are such that the rendezvous-phase duration can be several minutes. A reasonably long period is essential to an accurate rendezvous, since sufficient time must be allowed for smoothing the inherently noisy radar tracking data as well as for correcting measured errors. A period of as much as 10 to 20 min is still short compared with the overall mission duration. The effect of the differential earth gravity field has been shown by Hord[17] to be negligible for tracking-phase durations not exceeding 10 to 20 min. Furthermore, Wolverton[16] has shown that when the rendezvous time t_r is small compared with the product of the satellite orbital period T_0 and $(2\pi)^{-1}$, the orbital motion aspects of the rendezvous maneuver can be neglected.

Type II SBR.   Remote sensing of the earth from space began in 1960 with the launch of the first television and infrared observation satellite (Tiros) weather satellite. Remote sensing of the earth from space by radar began in 1975 with the launch of the GEOS-C by the National Aeronautics and Space Administration (NASA) and continued with the Seasat in 1978, the SIR-A on the Shuttle in 1981, and the SIR-B on the Shuttle STS-17 in 1984.

SEASAT-A SYSTEM.   The Seasat-A program was managed for the NASA Office of Applications by the California Institute of Technology Jet Propulsion Laboratory (JPL). The mission for Seasat-A was to demonstrate that measurements of ocean dynamics are feasible. The measurements included topography, surface winds, gravity waves, surface temperature, sea-ice extent and age, ocean features, and salinity. Precision of the geoid measurement was specified as ±10 cm.[18]

The Seasat-A satellite was launched at 6:12 P.M. PST on June 26, 1978. The orbital altitude was 783 km at apogee and 778 km at perigee. The retrograde polar orbit had an inclination angle of 108° and a period of 100.5 min. Three radar and two radiometer sensors were carried on the spacecraft. The coherent SAR, described in Sec. 22.3, operated at 1.275 GHz. The radar altimeter operated in the 12- to 14-GHz band and covered a 1.6-km swath directly below the spacecraft.

The wind scatterometer operated at 14.599 GHz and covered two swaths, each 400 km wide and offset on each side of the spacecraft. Four antennas were used to measure wind speed in the range from 4 to 28 m/s. The microwave radiometer had five frequency channels at 6.6 GHz, 10.6 GHz, 18 GHz, 21 GHz, and 37.6 GHz. A swath 1000 km wide, centered at the nadir, was covered. The visible and infrared (IR) radiometer covered a single swath 1800 km wide, symmetrical about the nadir.

Seasat-A collected data until Oct. 9, 1978, when a short circuit developed at the slip rings between the solar array and the power distribution bus.

The primary objectives of the SAR experiment on Seasat-A included (*a*) to obtain radar imagery of ocean wave patterns in deep oceans, (*b*) to obtain ocean wave patterns and water-land interaction data in coastal regions, and (*c*) to obtain radar imagery of sea and fresh-water ice and snow cover. The secondary objectives included (*a*) to obtain radar imagery of land surfaces; (*b*) to obtain data for mapping of the earth's surface; (*c*) to obtain data for estimates of land and sea surface roughness, ice type, differentiation of surface materials, vegetation, and landforms; (*d*) to obtain data for monitoring changes in the environment; (*e*) to obtain a demonstration of all-weather, day-night measurement capability; and (*f*) to obtain data useful for designing future high-resolution spaceborne radar systems.

GEOS-3. The Geodynamics Experimental Ocean Satellite (GEOS-3) was a remote-sensing satellite that contained five instruments in the experiment package.[19-21] These were (1) an SBR altimeter, (2) two C-band transponders, (3) an S-band transponder, (4) laser retroreflectors, and (5) a radio doppler system. The purpose of the GEOS-3 satellite was to perform experiments in support of the application of geodetic satellite techniques to geoscience investigations such as earth physics and oceanography. The SBR altimeter mission objective on the GEOS-3 satellite was to perform an in-orbit experiment that (1) determined the feasibility and utility of a space-borne radar altimeter to map the topography of the ocean surface with an absolute accuracy of ±5 m and with a relative accuracy of 1 to 2 m, (2) determined the feasibility of measuring waveheight, (3) determined the feasibility of measuring the deflection of the vertical at sea, and (4) contributed to the technology leading to a future operational altimeter satellite system with a 10-cm measurement capability.

The GEOS-C satellite (its designation was changed to GEOS-3 after successful orbit had been achieved) was launched on Apr. 9, 1975. The nominal orbit parameters were as follows: mean altitude, 843 km; inclination angle, 115°; eccentricity, 0.000; and period, 101.8 min. The GEOS-3 spacecraft was an eight-sided aluminum shell topped by a truncated pyramid. The satellite width was 132 cm (53 in), and the height was 81 cm (32 in); the weight of the GEOS-3 was 340 kg (750 lb).

Type III SBR. Before the design of a Type III SBR can begin, requirements for the surveillance radar systems must be specified. These requirements should include but not be limited to[22] (1) target radar cross section model, (2) target velocity and acceleration (maximum), (3) number of targets, (4) probability of detection, (5) probability of false alarm and false-alarm time, (6) track accuracy, (7) minimum target spacing, (8) designation error, (9) warning time, (10) length of detection fence, (11) revisit time, (12) clutter model, and (13) weather model. With these requirements as a minimum input to the design study, orbit selection can begin and parameter tradeoffs can be made. The influence of the space environment, interference, and clutter must be considered. Since the Shuttle (STS) can be a major launch vehicle for SBR, its capabilities should be examined. The

advantages and disadvantages of large surveillance radar in space should also be considered.

Target characteristics and requirements for coverage, track data rate, and revisit rate are important parameters. The radar subclutter visibility capability, antenna size, scan rate, and grazing-angle limitations also determine the orbit selected for the SBR. The space environment itself can determine the selected orbit if the natural-radiation lifetime dosage that the SBR electronics receives is too large. Finally, there is the requirement to use the least number of satellites to keep total system cost to a minimum.

Considerations

Orbit Selection.   Many factors contribute to the selection of the orbit to be used for each type of SBR and particularly for a large surveillance-type SBR. The orbit parameters of period, altitude, and velocity are the first consideration. The velocity for a satellite in a circular orbit around the earth is given by[16]

$$V_c = \sqrt{\frac{\mu}{r}} \qquad (22.1)$$

where r is the distance of the satellite from the center of the earth and μ is the product of the universal gravitational constant and the mass of the earth. The period of a satellite of the earth is given by[16]

$$T = \frac{2\pi\mu}{\sqrt{V_a^3 \, V_p^3}} \qquad (22.2)$$

where V_a is the velocity of the satellite at apogee and V_p is the velocity of the satellite at perigee. For a circular orbit, $V_a = V_p$ and the period of a circular orbiting satellite is

$$T_c = \frac{2\pi\mu}{V_c^3} \qquad (22.3)$$

Table 22.2 shows selected calculations of circular-orbit velocity and period when the radius of the earth is 20.903 $(10)^6$ ft, μ is 1.4069 $(10)^{16}$ ft^3/s^2, and 1 nmi is 6076.1 ft.

Many studies concerning the design of satellite constellations for optimal coverage have been made and reported.[23-28] Luders and Ginsberg[24] describe an analytical solution to the problem of achieving continuous coverage of latitudinally

TABLE 22.2   Selected Orbital Parameters

Altitude, nmi	Velocity, ft/s	Period, min
99	25,587	88
414	24,520	100
912	23,074	120
2,262	20,157	180
5,612	15,999	360
19,369	10,079	1,440

bounded zones of the globe. Emara and Leondes[27] solved the problem of simultaneous observations by at least four satellites by a constellation of the minimum number of satellites. Ballard[25] extended earlier work by Walker[23] and analyzed rosette constellations that provided the largest possible great-circle range between an observer anywhere on the earth's surface and the nearest subsatellite point. Single, double, triple, and quadruple visibility was provided by various constellations. Beste[26] designed satellite constellations that provided single and triple continuous coverage by the minimum number of satellites. All these studies determined coverage for satellites with sensors that observe only angles around the nadir. Electro-optical sensors and mapping radars are typical sensors that provide this coverage. However, these studies do not provide results for SBR surveillance sensors that must detect targets in clutter. These sensors typically have a *nadir hole* 20 to 30° off nadir in which the signal-to-clutter ratio (*SCR*) is too large for reliable detection. This is shown in Fig. 22.2 for a 50° maximum grazing angle and a 3° minimum grazing angle. The minimum grazing angle is a limit set by the atmospheric attenuation allocated in the SBR loss budget and the refraction angle error. To illustrate the different results that can be obtained, consider a requirement to provide continuous coverage of the earth from an orbital altitude of 10,371 km (5600 nmi). For a single sensor on each satellite with no grazing-angle limitations, a constellation of six satellites can provide the required continuous coverage from polar orbits. The satellites would be equally distributed in two orbital planes, using the study results given by Harney.[28] However, if the sensor in the SBR was limited to grazing angles between 3 and 60°, then the required coverage could be provided by a constellation of 10 satellites. This constellation consists of 1 satellite in each of 10 equally spaced orbit planes at an inclination of 49.4°, resembling the Walker 10/10/8 constellation.[25] If the grazing angles extend between 3 and 70°, then a 14-satellite constellation in a Walker 14/14/12 configuration provides a continuous global twofold coverage. The inclination angle of each orbital plane is 49.4°.

FIG. 22.2   SBR coverage and nadir hole.

Space Environment.   For a large phased array type of radar operating in space, the thermal and natural radiation environments have significant influence on the design of an SBR. Particular effects depend on the orbital altitude and the materials used in the structure.

THERMAL ENVIRONMENT EFFECTS.   In general, distortion of a phased array antenna will cause a decrease in antenna gain. Figure 22.3 shows the effect of random phase errors caused by the distortion ε when the error correlation interval is large with respect to a wavelength. It is seen from Fig. 22.3 that a 2 dB loss in gain is obtained when the distortion is about one-tenth of a wavelength. Thus for a 50-m-diameter planar corporate phased array antenna operating at a wavelength of 10 cm, the rms distortion of the plane of the array must be held to less than 1 cm if a 2 dB loss in antenna gain is to be maintained.

Thermal distortion in a 70-ft (21.34-m) diameter parabolic reflector was studied[29] at synchronous orbit. Reflector performance comparisons were made for titanium and graphite composite materials. Generally the tolerances that must be held on reflector antennas are more severe than for phased arrays for the same performance. Figure 22.4 shows the results of the analysis. Performance of the

FIG. 22.3   Antenna loss of gain due to random phase errors.

graphite composite material is superior, giving an rms distortion of about 0.076 cm. If this is one-fiftieth of the wavelength, then the antenna could operate satisfactorily at a wavelength of 3.8 cm.

Consider a 70-m-diameter-lens phased array[30,31] at an altitude of 5600 nmi as shown in Fig. 22.5. The progress of the sun angle is shown. Simulations have predicted the following maximum and minimum temperatures for selected parts of the space-fed lens antenna:

Location	Temperature, K	
	Maximum	Minimum
Ground plane	264	224
Rim	182	160
Upper stays	231	186
Lower stays	217	201
Upper dipole plane	314	201
Lower dipole plane	274	220

By choosing the proper materials, the design of this class of antenna will experience low distortions compared with those allowable. Figure 22.6[32,33] shows the loss in relative gain for a 71-m-diameter space-fed antenna as a function of the
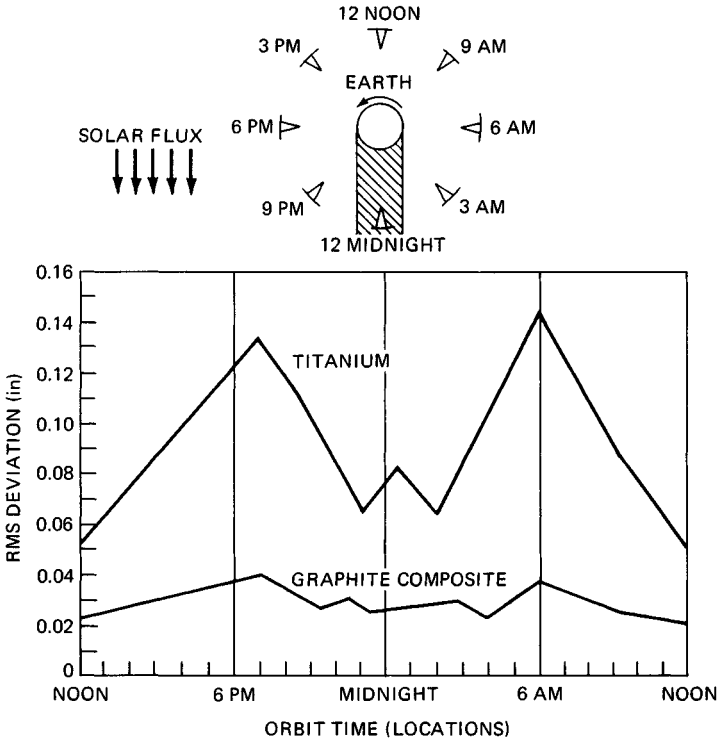
FIG. 22.4 Thermal distortion.[29]
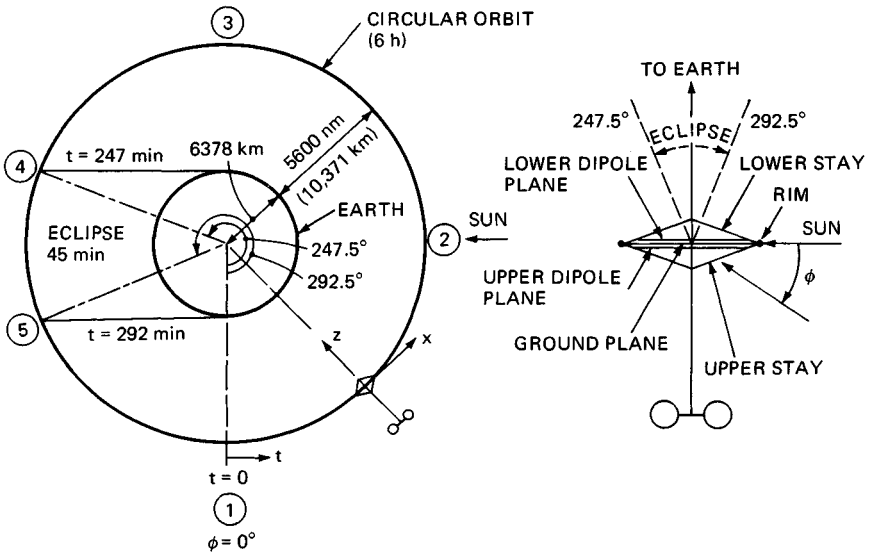
FIG. 22.5 SBR at 5600-nmi orbital altitude and the sun angle progression.[32]
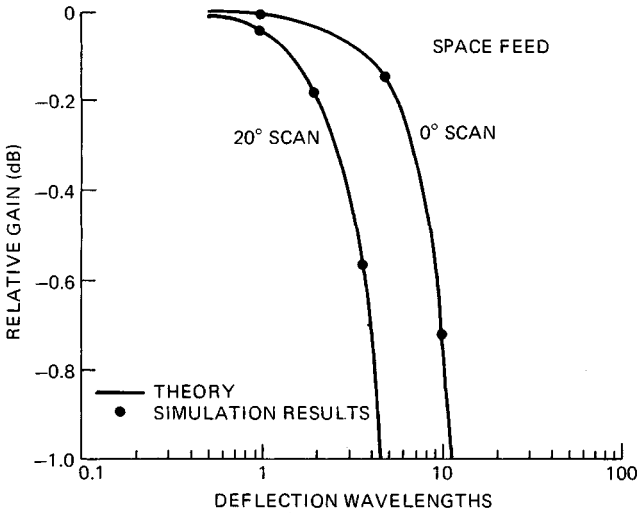
FIG. 22.6 Loss of gain due to distortion for a space-fed array.[32]

deflection or distortion in wavelengths. It is seen that the relative gain is down 1 dB when the distortion is about 5 wavelengths at a 20° scan angle.

RADIATION ENVIRONMENT EFFECTS.   SBRs can encounter in space particle radiation that may be due to both natural phenomena and nuclear detonations. The satellite must be designed to operate for a reasonable lifetime in the natural space environment. This environment is a function of orbital altitude. When the satellite is operating in midaltitude orbits, exposure to the earth's Van Allen belts will be predictable and its effect on radar electronics will be functions of the inherent hardness level of the components and the shielding used. (Reference 34 provides the trapped radiation data for proton and electron flux that has been measured as a function of altitude.) Figure 22.7[32] shows the total 5-year dose in rads (Si) that satellites in orbits between 350- and 6500-nmi altitudes will experience as a function of the aluminum shielding used. It appears that current technology in integrated-circuit hardening should produce a total dose hardness of about $5(10)^5$ rads (Si) for devices that are suitable for the SBR T/R modules. This hardness level is adequate for SBR deployment in many of the candidate orbits with a mission life in the natural environment of several years. A hardness of $5(10)^6$ rads (Si) which may be achievable is required for a 5-year mission life. Survival in a saturated nuclear environment typical of a high-altitude nuclear burst requires a hardness of 1 to $5(10)^7$ rads (Si), depending upon the specific orbit. The development and consistent fabrication of devices as hard as this are relatively uncertain.

Tradeoffs.   Obviously many tradeoffs can be made during the design of each type of SBR, depending upon the mission. In a dual-frequency surveillance and track radar performing an air traffic control (ATC) mission, as mentioned in Sec. 22.6, it is possible to trade off the length of the surveillance fence against the number of targets in track as functions of the track data rate and the radar-beam grazing angle. In a high-resolution-mapping radar mission it is possible to trade off the resolution against the orbital altitude as functions of radar wavelength and integration time. These trades are shown in detail in Refs. 12 and 13.
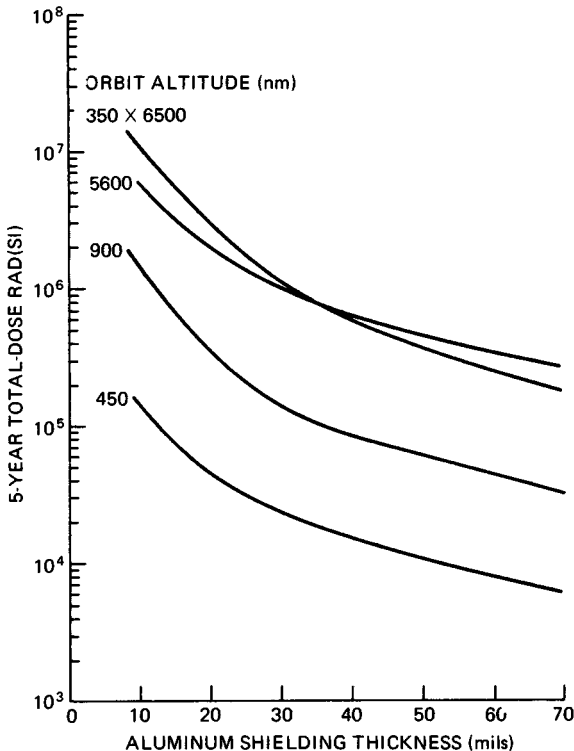
FIG. 22.7 Total dose versus shielding thickness for a 5-year mission.[32]

Clutter/Interference. SBR performance is significantly dependent upon clutter and interference, either intentional or unintentional. To illustrate the magnitude of the clutter problem, consider the ATC radar described in Sec. 22.6. When the grazing angle is 70° and the reflectivity of the ground is −15 dB, the main-beam clutter cross section is +57 dBsm. If the desired radar performance requires that a target with an RCS of +13 dBsm have an *SCR* of 25 dB, then the main-beam clutter cancellation ratio must be at least 69 dB. Therefore, SBR performance requires large clutter cancellation ratios. Reference 35 indicates that clutter cancellation ratios up to 90 dB can be obtained by using pulse doppler and displaced phase center antenna (DPCA) techniques.

Interference will enter the SBR antenna primarily through the sidelobes since the beamwidth is narrow. This interference can be either intentional noise jamming or unintentional from other radars. These effects can be reduced to acceptable levels if adaptive sidelobe cancellation techniques and sidelobe-blanking techniques are utilized.

Launcher Capabilities. The most probable launch vehicle for the SBR is the STS (shuttle). Therefore, STS capabilities to put various payloads that include one or more SBR satellites (and the propulsion systems to place them into the desired orbits) must be considered. Figure 22.8 shows the STS cargo weight as a function of orbit inclination angle for various circular orbital altitudes and orbital-
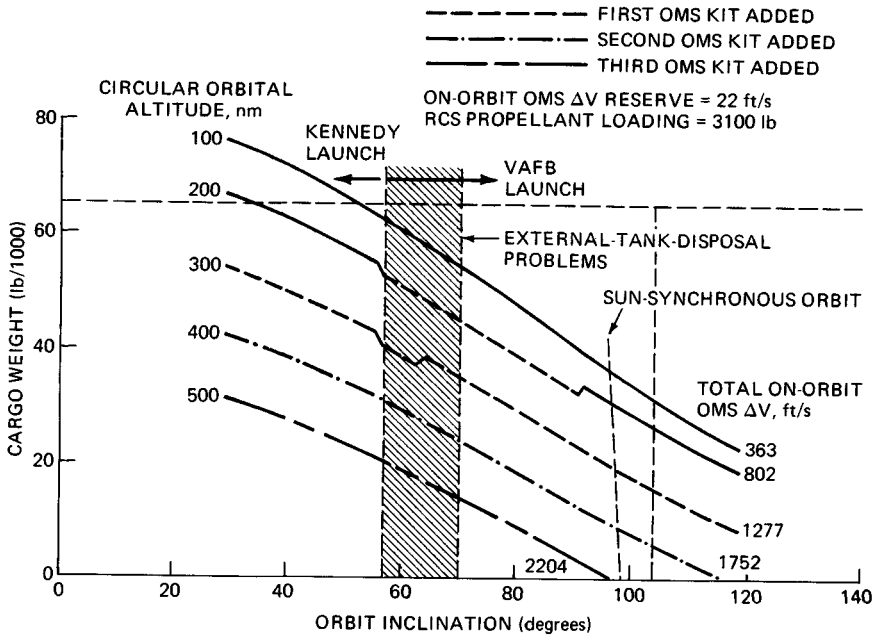
FIG. 22.8 STS (Shuttle) cargo weight versus inclination for various circular-orbit altitudes (delivery only—no rendezvous).

maneuvering-system (OMS) on-orbit velocity increments. It is seen that 64,000 lb can be delivered to a 100-nmi circular orbit inclined 50° from the Kennedy launch site in Florida. If each SBR weighs 9500 lb, then three SBR satellites can be placed into orbit along with 35,500 lb in propulsion for orbital transfer.

Advantages and Disadvantages of SBR Systems.   When sensors are required for missions involving targets in space, ocean, and air and for missile defense missions, the use of SBRs should be considered. The advantages of such radars deployed in space compared with ground-based radars are described below.

1. Coverage in both space and time is limited only by the orbit selected and the number of satellites. Large-scale continuous observation can be obtained as shown in Figs. 22.9 and 22.10.[28] In Fig. 22.9 the required number of vehicles are shown as well as the number of orbit planes in which they are distributed to provide continuous coverage of the entire earth's surface from circular polar orbits. It is seen that six vehicles in two orbit planes can be used for vehicle altitudes greater than about 6000 nmi. There is no nadir hole in the satellite coverage. Figure 22.10 illustrates the special case of equatorial orbits and the number of vehicles required for continuous coverage. This situation is limited to the use of wide swaths that extend up to the specific latitudes indicated. It is seen that four vehicles can cover a 60° swath when the vehicles are at altitudes greater than about 6000 nmi. Temporal coverage is illustrated in Fig. 22.11, which shows the maximum time for viewing ground objects from a space vehicle if the objects are tracked.[28] It can be seen that a ground object can be observed for more than 7000 s when the orbital altitude is 6000 nmi.

2. When the SBR uses an electronic scanning antenna, it is possible to per-
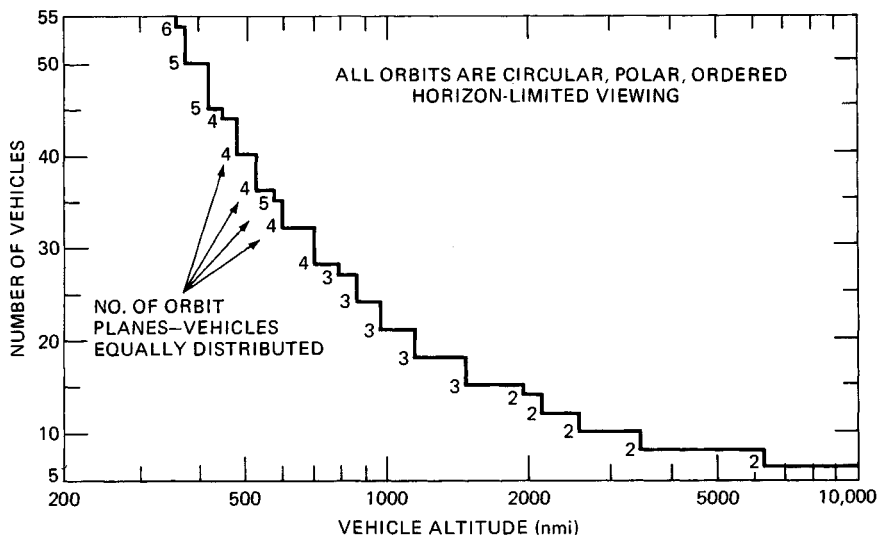
FIG. 22.9 Global coverage by polar orbits.[28]
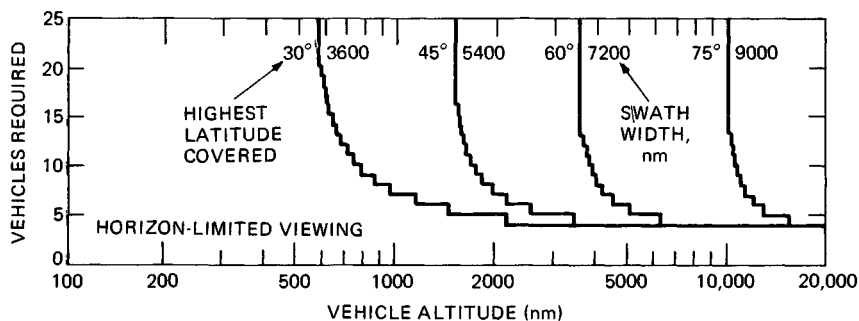
FIG. 22.10 Zonal coverage by equatorial orbits.[28]

form multiple missions. For example, a system of radar satellites can (*a*) search a fence formed completely around the continental United States (CONUS) to detect bombers at a distance from the coast, (*b*) search a fence over the poles to detect intercontinental ballistic missiles (ICBMs) before they can be detected by the Ballistic Missile Early Warning System (BMEWS), (*c*) monitor potential launch sites for space launches from any foreign country, (*d*) perform surveillance of ocean areas, (*e*) search a sea-launched ballistic missile (SLBM) detection fence, and (*f*) detect objects in space that appear to be threats to United States synchronous satellites. The number of missions is limited only by the weight and prime power available, but even these limitations can be overcome when the space shuttle is the planned launch vehicle. Therefore, the only real limitations are technology and cost.
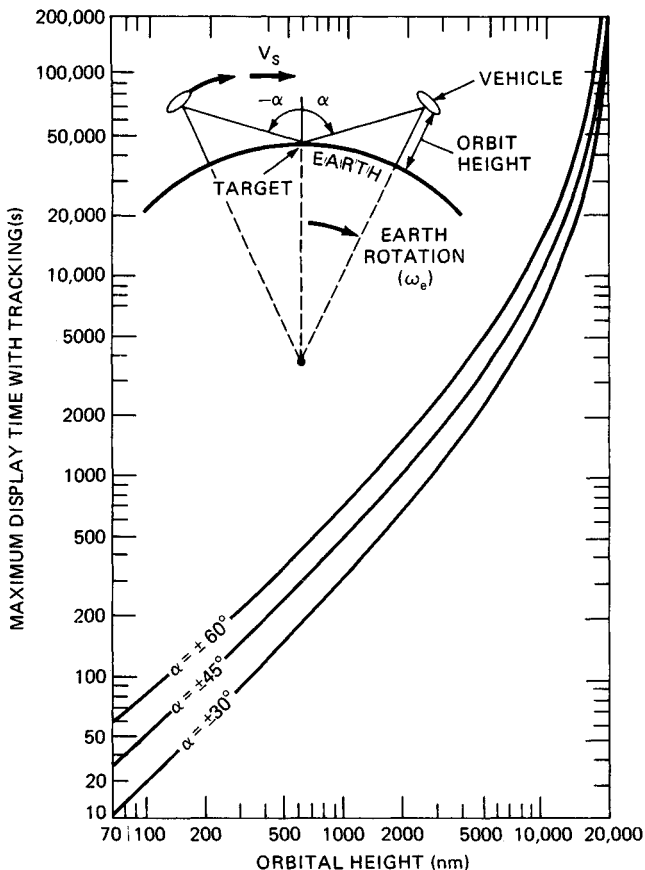
FIG. 22.11 Maximum time for viewing objects from a space vehicle if the objects are tracked.[28]

3. Atmospheric propagation problems can be minimized by proper selection of operating frequencies and favorable geometry selection.

4. No overseas stations are required if data is read out via relay satellites. Hence, the SBR system allows a country to be politically independent, and the loss of tracking stations in a foreign country has no impact on its system capabilities.

The factors that affect the pace of development of large radar systems in space are:

1. The technologies of large antenna structures in space, of large phased arrays in space, of large weights in space, and of large prime power systems in space are considered to be in their early stages.

2. The funds that can reasonably be spent on a space-based multimission ra-

dar system are to be determined. Even with the use of the Shuttle to reduce the cost per pound of payload into orbit, large investment costs are expected to be required for the SBR system.

22.3 SBR SYSTEM DESCRIPTIONS

The United States and the U.S.S.R. have deployed Type I and Type II space-based radars. This section describes some of these SBR systems.

STS Rendezvous Radar.[1,15,36] The Integrated Radar and Communications Subsystem (IRACS) was developed by Hughes Aircraft Company for use on the Space Transportation Systems (STS). The IRACS is a coherent range-gated pulsed doppler radar which searches for, acquires, and tracks other orbiting objects and provides the spatial measurement data needed to perform rapid and efficient rendezvous with those objects.

The IRACS performs both radar and communications functions for the STS. In the pulsed doppler radar mode it performs the rendezvous function just described. In the communications mode it searches for, acquires, and tracks the Tracking and Data Relay Satellite System (TDRSS) relay satellites to provide two-way communication between the space shuttle and ground tracking stations.

The IRACS hardware is subdivided into *deployed* and *inboard* assemblies. The deployed hardware is located within the Shuttle payload bay and is extended for operation through the open payload bay doors. Included in this hardware are the antenna reflector, feed, gimbals, drive motors, gyros, digital shaft encoders, rotary joints, transmitter, receiver, upconverter, first downconverter, and frequency synthesizer. The inboard hardware is located internally to the shuttle and includes the signal-processing, track-filtering, and control functions.

The K_u-band IRACS operates in the band of frequencies between 13.75 and 15.15 GHz, with radar operation between 13.75 and 14.0 GHz. There are two basic radar modes: a passive mode in which the target is noncooperative, in that no cross-section augmentation is present, and an active mode in which the target has an on-board transponder. The radar operates out to 12 nmi in the passive mode and out to 300 nmi, with a +14 dBm transponder, in the active mode. Submodes include an automatic search and angle and range track capability and external angle control operation. Under external angle control the antenna either is positioned by external slew commands or is referenced to inertial space or to the Shuttle axes. During automatic operation, angle, angle rate, range, and range rate measurements are made by the radar after track has been initiated. Under external angle control only range and range rate are measured.

The antenna is a 36-in-diameter center-fed parabola with 38.4 dB gain and 1.68° beamwidth. The five-element monopulse feed provides a sum and two orthogonal difference outputs. The difference outputs are time-multiplexed together into a *single* receiver difference channel for the angle-tracking operation. An auxiliary horn is monitored in the search mode, using the receiver difference channel, and compared with the main-antenna sum channel to prevent acquisition of large targets in the sidelobes of the main antenna. The auxiliary antenna has a peak gain which is about 20 dB less than that of the main antenna. Low-noise radio-frequency (RF) preamplifiers are used in the sum and difference channels. After amplification, at intermediate frequency (IF), the sum and difference channels are combined into a single receive channel for routing to the inboard elec-

tronics assemblies for further processing. The transmitter employs a traveling-wave tube (TWT) with 44 dB gain to amplify the coherent synthesizer output to 50 W of peak power. For short-range operation (down to 100 ft) the TWT is bypassed to reduce the power on the target. Five RF frequencies are used in the radar mode to decorrelate Swerling 1 (slowly fluctuating) target returns and improve detection. A 16-point digital Fourier transform (DFT) processor is employed to coherently integrate multiple pulse returns and to provide fine-resolution measurements of target relative velocity. The deployed assemblies weighed 135 lb, and the prime power was 460 W.

Seasat-A Synthetic Aperture Radar.[18,37]   The Seasat-A was a focused SAR consisting of five subsystems: (1) spacecraft radar antenna, (2) spacecraft radar sensor, (3) spacecraft-to-ground data link, (4) ground data recorder and formatter, and (5) ground data processor. The antenna was a microstrip array of eight panels that were fed by a corporate-feed network and operated at 1275 MHz. Details of the Seasat-A antenna are discussed in Sec. 22.4. The solid-state radar transmitter generated a nominal peak power of 800 W with a linear frequency modulation (LFM) derived from a stable local oscillator (stalo). The antenna illuminated a 100-km-swath width at the surface of the earth with an antenna elevation beamwidth of 6° that was oriented at an angle of 20° with respect to the nadir. Upon reception of the reflected signal by the receiver in the radar sensor, the return signal was amplified by a sensitivity-time-controlled RF amplifier. This signal and a fraction of the radar stalo were then combined and transmitted to a ground station by an analog data link. At the ground station, the data line demodulator recovered the radar sensor stalo and the radar return signal. The recovered synchronously demodulated video radar signal was then converted into digital form by the radar data recorder and formatter subsystem. Upon conversion, the signal was buffered and recorded by a high-density magnetic tape recorder. Subsequently, the radar data processor converted the digital recorded data into a two-dimensional map of the radar cross section of the area observed by the antenna. The SAR system generated a 25-m-resolution radar map in elevation (across track) by time-gated compressed radar return signals and in azimuth (along track) by focusing the coherent radar returns during the data-processing interval in the earth-based signal processor. Total SAR on-orbit weight was 223 kg; required radar prime power was 624 W. Table 22.3 gives the characteristics of the Seasat SAR.

Shuttle Imaging Radar.[36]   The technology developed for the Seasat-A SAR formed the basis for the shuttle imaging radar (SIR) series, SIR-A and SIR-B. Minor differences in the antenna will be discussed in Sec. 22.4. The L-band radar transmitter was utilized with slight bandwidth changes so that resolution was 40 m on SIR-A and 20 m on SIR-B. Swath width was 50 km for both radars. Orbital altitudes were 240 km and 220 km, respectively, so that radar range and incidence angles were different.

GEOS-C SBR System Characteristics.[8,19–21]   The GEOS-C radar altimeter was a precision K_u-band (13.9-GHz) SBR altimeter developed primarily to measure ocean surface topography and sea state. It was a complex multimode radar system with two distinct radar gathering modes (global and intensive modes) and two corresponding self-test–calibration modes for use in on-orbit functional test and instrument calibration. The key performance features were its capability to (1) provide precise satellite-to-ocean surface-height measurements [precision

TABLE 22.3   Synthetic Aperture Radar

Antenna	
Type	Planar phased array (10.74 m × 2.16 m)
Beamwidth	1.1° azimuth, 6° elevation (1 dB points)
Look angle	20° depression, 90° with respect to the velocity vector
Gain	34.7 dB
Polarization	Horizontal
Weight	113 kg
Transmitter	
Type	Solid-state transistor
Efficiency	38 percent
RF carrier	1275 MHz
Peak power	800 W (nominal), 1125 W (maximum)
Pulse length	33.8 μs
PRF	1463, 1540, 1645 pps
Duty cycle	0.05 (maximum)
Average power	44.5 W (nominal), 62.6 W (maximum)
Waveform	Pulse, LFM, 19-MHz Bandwidth
Receiver	
Noise temperature	550 K
Bandwidth	22 MHz
System input noise	−127.42 dBW
AGC time constant	5 s
STC gain variation	9 dB
Stalo stability	3×10^{-10} in 5 ms
Recorder	25 kb/s digital
System weight	110 kg (excluding antenna)
Total prime power	624 W (maximum)
Resolution	25 m
Swath width	100 km
Swath length	2000 km per pass
Swath orientation	Right side of orbit path
Signal-to-noise ratio	9 dB (nominal)

of 50 cm in the global mode (GM) and 20 cm in the intensive mode (IM) at an output rate of one per second] for use in mapping the shape of the ocean surface and (2) provide data which can be processed to estimate peak-to-trough ocean waveheight (waveheights in the range of 2 to 10 m can be estimated to an accuracy of 25 percent). Several key areas of technology included in the design are (1) high-frequency logic circuitry with a 160-MHz clock and four-phase division for 1.56-ns resolution, (2) a wideband (100-MHz) linear FM pulse compression system with a compression ratio of 100:1 and a compressed pulse width of 12.5 ns, (3) high-speed sample-and-hold circuitry for accurate sampling of wideband (50-MHz) noisy video return signals, and (4) design and packaging of high-voltage (12-kV) power supplies for space application.

The instrument weighs 68 kg (150 lb) and occupies a volume of 0.119 m³ (4.2 ft³) including the antenna, which is a 0.6-m (24-in) diameter parabolic dish with

a 2.6° beamwidth and a 36 dB gain. The instrument is packaged in two basic sections: an RF section and an attached electronics section, which are both mounted to a center-cylindrical disk baseplate with a diameter of 0.65 m (26 in). The major subsystems contained in the RF section are (1) the IM transmitter (chirp generator, upconverter, 1-W driver TWT and high-voltage power supply, 2-kW output TWT and high-voltage power supply), (2) the GM transmitter (a 2-kW peak-power magnetron and high-voltage power supply), (3) the RF switch assembly (RF switches, waveguide runs, calibrate attenuation path, and TR switch), and (4) the receiver front end (downconverter-preamplifier). The major subsystems contained in the attached electronics section are (1) the IF receiver (IF amplifiers, filters, pulse compressor, detectors), (2) the signal processor (AGC, acquisition, and tracking functions implemented with analog and digital circuitry on multilayer board assemblies), (3) the frequency synthesizer, (4) the mode control circuitry, (5) the calibrate-test circuitry, and (6) the low-voltage power supply. The nominal power required for operation was 71 W for the global mode and 126 W for the intensive mode (16 waveform samplers).

U.S.S.R. Cosmos 1500 Side-Looking Radar.[38,39] The U.S.S.R. launched the Cosmos 1500 oceanographic satellite on Sept. 28, 1983, into a nominal 650-km polar orbit. The satellite was the first of a series intended to provide continuous world ocean observations for civil and military missions. The sensors provide side-looking radar (SLR), radiometric, and visual coverage of oceans and ice zones for land- and sea-based users through an operational distribution network.[38] Table 22.4 summarizes the parameters and performance of the real-beam SLR. The radar operates at a frequency of 9500 MHz with a magnetron transmitter that has a peak power output of 100 kW. The antenna is a slotted waveguide that is 11 m long and 4 cm high. Cosmos 1500 has demonstrated many significant capabilities, including (1) routine automatic picture transmission of SLR images of earth; (2) mapping of inhomogeneities of Antarctic and Greenland ice cover that were previously not detected; (3) radar images of polar regions of multiyear and first-year ice zones; (4) mapping of elongated zones of ice-cover continuity disturbances; (5) tracking of sea-ice drift by using a series of radar images of the same water area; (6) detection of oil slicks, wind fields, and currents; and (7) guidance of ships trapped in arctic ice during October–November 1983.

The orbit of Cosmos 1500 allowed complete earth coverage each 1.41 days for the optical sensors and each 5.9 days for the radar sensor. Subsequent launches of the Cosmos 1500 type of satellite have occurred.

22.4 TECHNOLOGY

The desire to develop large radars in space has stimulated progress in several new technologies such as (1) large deployable parabolic and phased array antennas, (2) lightweight, low-cost monolithic microwave integrated circuit (MMIC) transmit/receive modules, (3) high-level prime power systems, (4) efficient on-board signal processors, (5) large lightweight space structures, (6) lightweight, low-cost phase shifters, (7) radiation-hardened electronic devices, (8) materials with a low thermal coefficient of expansion, and (9) advanced calibration and self-test techniques. Some of these technologies are briefly reviewed here.

TABLE 22.4 Cosmos 1500 SBR Parameters and Performance

Type	Real-beam side-looking radar (460-km swath)
Frequency/wavelength	9500 MHz/3.15 cm
Antenna	
Type	Slotted waveguide
Size	11.085 m × 40 mm
No. of slots	480
Illumination	Cosine on a pedestal
Beamwidth	0.20° × 42°
Gain	35 dB
Sidelobes	−22 dB to −25 dB
Waveguide	Copper, 23 × 10-mm cross section
Polarization	Vertical
Swing angle	35° from nadir
Noise temperature	300 K
Transmitter	
Type	Magnetron
Power	100 kW peak, 30 W average
Pulse width	3 μs
PRF	100 pps
Loss	1.7 dB
Receiver	
Type	Superheterodyne
Noise power	−140 dBW
Loss	1.7 dB
Pulses integrated	8 noncoherent
LNA noise temperature	150 to 200 K
LNA gain	15 dB
Dynamic range	30 dB
IF	30 MHz ± 0.1 MHz
Input power	400 W
Range	700 km (minimum), 986 km (maximum)
SNR	0 dB on $\sigma^0 = -20$ dB

Antennas. The development of SBR is strongly dependent upon the technology of large space-deployable antennas. Large antennas must be used since the radar ranges are significantly greater than usual and the prime power in the radar is limited. The vacuum of space and the zero-g environment permit the deployment of antennas with low mass per unit antenna area. Antennas with large diameters, up to 1 km, have been discussed by United States developers.[40–48] In the U.S.S.R., antennas with diameters in the 1- to 10-km range have been discussed.[49] In addition to being large and deployable, the SBR antenna must maintain its desired shape whether it be parabolic or planar. As shown earlier (Fig. 22.3), small deviations can cause a significant loss in antenna gain. Stable configurations are obtained by using low-coefficient-of-thermal-expansion (CTE) materials. Characteristics of selected materials for stable RF systems are shown in Table 22.5. Data includes CTE, density, modulus, conductivity, and attenuation of WR 75 waveguide fabricated

TABLE 22.5 Potential Material Selection for Thermally Stable RF System

Material	Expansion coefficient, in/(in/°F) $\times 10^{-6}$	Density, lb/in^3	Young's modulus, $\times 10^6$ lb/in^2	Thermal conductivity, $\dfrac{Btu \cdot in}{h \cdot ft^2 \cdot °F}$	WR75 attenuation, dB/ft at 11.95 GHz
Aluminum	13.1	0.10	10	1513	0.049
Titanium	5.1	0.16	16	444	0.274
Invar	1.1	0.29	20	93	0.370
Beryllium	6.8	0.07	40–44	1138	0.082
Graphite/ epoxy	0.03	0.06	17–25	75 (axial), 7.3 (transverse)	1.560 (bare), 0.040 (coated)
Gold	6.8	0.70	. . .	2064	0.048
Copper	7.8	0.32	. . .	2944	0.040
Silver	11.0	0.38	. . .	3101	0.039
Rhodium	4.7	0.45	. . .	611	0.087
Kevlar®* 49	− 1.1 longitudinal, + 33 radial	0.052	19	0.334 (axial), 0.285 (transverse)	

*®Du Pont trademark.

out of each material. In the following discussion selected antenna designs are described to illustrate the state of the art in large antennas in space.

United States Space-Deployable Antennas. A large space-deployable antenna that the United States deployed in space was the Lockheed-NASA ATS-6 parabolic reflector, launched in 1974. It was 9.1 m in diameter with a tolerance of 1.52 mm rms and a specific weight of 1.4 kg/m^2.[41,50] The ATS-6 antenna embodies the flex-rib technique. During the years subsequent to that launch, Lockheed has evolved flex-rib deployment technology to additional reflector designs, the polyconic and the maypole designs.[41]

Harris developed the radial-rib double-mesh design and in 1970 built a 12.5-ft-diameter antenna.[40] This was followed by the TDRSS 4.88-m-diameter antenna and three generic antenna designs including the radial-rib, TRAC, and hoop-column concepts. The weight-versus-diameter capabilities of these three designs are shown in Fig. 22.12.[47] As part of the NASA deployable antenna flight experiment (DAFE) design study, Harris estimated that a 50-m-diameter reflector assembly would have an overall weight of 819 kg (1805 lb).

The specific mass of this design is 0.417 kg/m^2, and the estimated surface error was 4 mm rms. In a parallel DAFE competition, the Grumman Aerospace Corporation designed a 50-m-diameter phased array lens antenna that would have a specific mass of 0.522 kg/m^2.

The DAFE studies were conducted for NASA Marshall Space Flight Center (MSFC) by Harris and Grumman in a competition during the period from August 1980 to September 1981. The primary objectives of the study were (1) to demonstrate, by a flight experiment, the capability to launch, deploy, retract, and return to earth a large (50-m-diameter) space frame; and (2) to verify, by flight experiment, the capability of the space frame to attain and maintain the dimensional
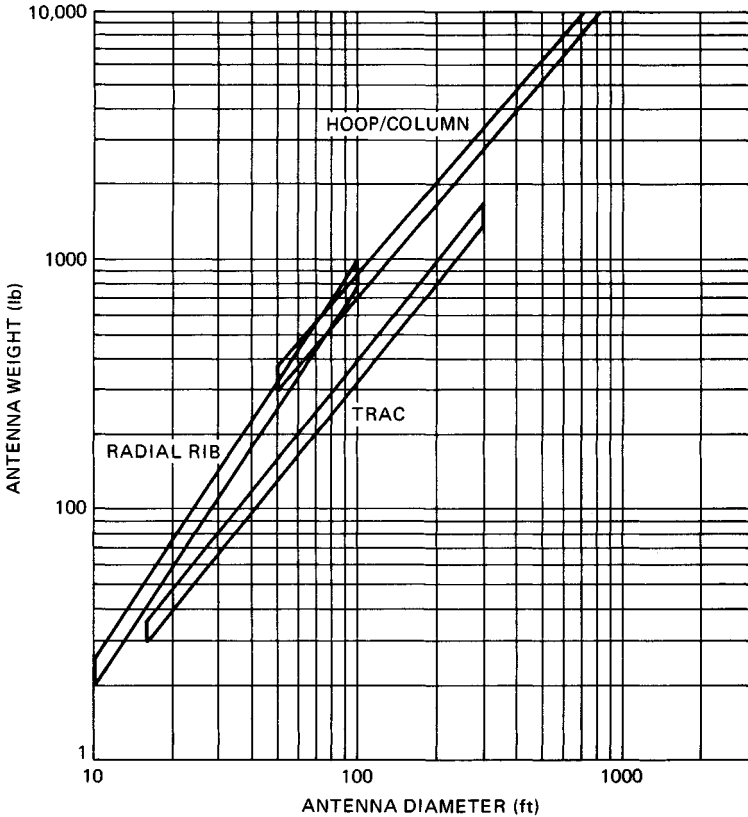
FIG. 22.12 Weight versus diameter for three generic antenna designs.[47]

precision required to operate as a spaceborne antenna. Both contractors devised orbiter-attached experiments that would maximize program outputs while minimizing orbiter and experiment risks. Although many flight configurations were designed, overall results were similar for both phased array and parabolic antennas. Both contractors also devised measurement techniques that would provide a 50-mil rms accuracy required for the measurement of antenna deformation.

General Dynamics has designed space-erectable antennas and parabolic graphite-epoxy reflectors for space applications.[43,46] A 2.44-m-diameter reflector was built and tested. It has a surface tolerance of 0.0635 mm rms and a specific mass of 4.4 kg/m². The space-erectable designs had a specific mass of 0.49 kg/m²; however, the tolerance was on the order of 10 mm rms. Therefore, the space-erectable antenna designs were configured primarily for relatively low-frequency operations.

TRW has developed an advanced antenna concept under work sponsored by JPL as part of the NASA large space systems technology (LSST) program.[51] The feasibility of stowing large, solid antenna reflectors in the Shuttle was examined. The antennas would be designed to operate in the 10- to 100-GHz range and main-
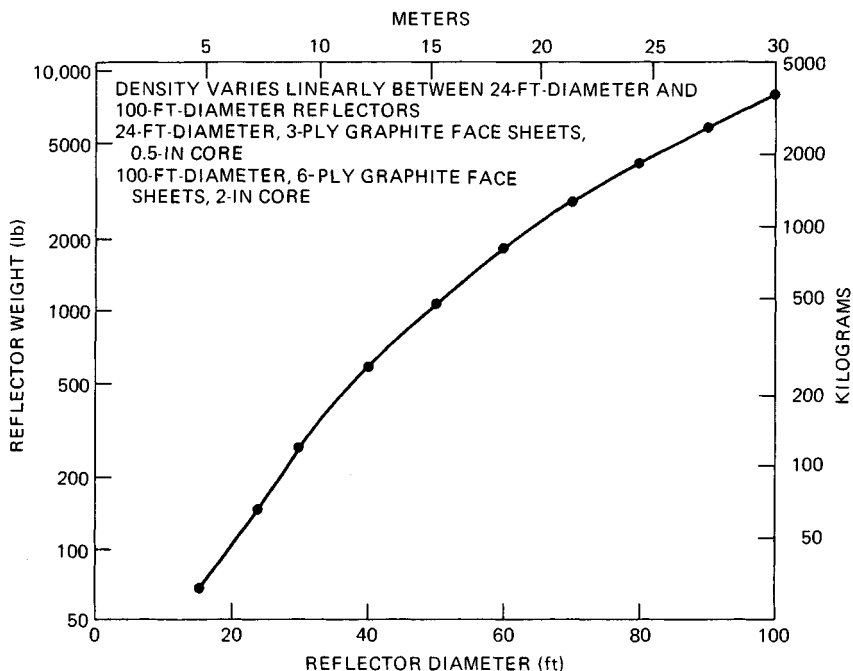
FIG. 22.13   TRW antenna reflector weight estimate.[51]

tain rms deviation on the order of 10^{-5}-diameter fabrication error. Thermal deviation for a 100-ft-diameter antenna was estimated to be 0.0034 in rms. The weight of antenna reflectors was estimated for diameters of 16 to 100 ft. Figure 22.13 shows the plot of reflector weight excluding the weight of feeds and subreflectors. The basic construction assumed a graphite-epoxy-aluminum honeycomb-sandwich configuration.

Antenna systems have been studied and fabricated under the LSST program. Lockheed Missiles and Space Company (LMSC) has demonstrated, in a simulated zero-gravity environment, the technology for a large space-deployable antenna.[52] LMSC fabricated a 22.5° sector of a 55-m-diameter *wrap-rib* parabolic antenna and deployed it in a ground-based zero-*g* facility. The surface of the antenna is a knit mesh of 1.2-mil gold-plated molybdenum wire that is contoured by graphite-epoxy ribs. Each rib weighs 9.1 kg, is 27.5 m long, and is lenticular in shape. This shape allows the ribs to collapse as they are wrapped around a central hub for stowage prior to deployment. The ribs resume their required structural shape as they unwind (under constraint), thereby stretching the mesh into the proper parabolic shape. The Lockheed development program was initiated to demonstrate the readiness of large-diameter offset reflector technology through development of ground-testable, flight-representative full-size hardware.

The Seasat-A antenna (designed by Ball) is a 10.74- by 2.16-m microstrip array that is deployed after orbit insertion. The operating wavelength is 23.5 cm. This antenna is very similar to the SIR-A. Both are significant developments in large deployable antennas.[6,37] The SIR-B antenna is similar except that it was mechan-

TABLE 22.6   Characteristics of Seasat, SIR-A, SIR-B, and SIR-C Antennas

	Seasat	SIR-A	SIR-B	SIR-C
Frequency	1275 MHz	1278 MHz	1282 MHz	1275 and 5300 MHz
Bandwidth (1.5:1 VSWR)	22 MHz	8 MHz	16 MHz	> 20 MHz
Gain	34.9 dB	33.6 dB	33.0 dB	37.0 dB (L band); 43.0 dB (C band)
Polarization	Horizontal linear	Horizontal linear	Horizontal linear	Horizontal linear and vertical linear
Beamwidths				
H plane	6.2°	6.2°	6.2°	Adjustable through amplitude and phase, 0.99° L band; 0.24° C band
E plane	1.1°	1.4°	1.1°	
Beam-pointing angle	20.5°	47°	15 to 60° (mechanical steering)	Tilted to 35°, then ±25° electronically steered
Size (deployed)	10.74 × 2.16 m	9.4 × 2.16 m	10.74 × 2.16 m	12.06 × 4.2 m
Size (folded)	1.34 × 2.16 m		4.1 × 2.16 m	4.1 × 4.2 m
Weight	103 kg	181 kg	306 kg	900 kg
Support structure	Graphite-epoxy 3D truss	Rigid aluminum 3D truss	Rigid aluminum 2D and 3D truss	Graphite-epoxy 2D truss
Fold mechanisms	Multifold (spring-loaded)	Fixed	Two folds (motor-driven)	Two folds (motor-driven)
Number of radiating elements	1024	896	1024	864 (L band); 5184 (C band)
Number of panels	8	7	8	9
Feed system	Microstrip, coaxial and suspended substrate	Microstrip, coaxial	Microstrip, coaxial	Microstrip, coaxial, waveguide
W/A, kg/m^2	4.44	8.9145	13.1906	17.7683

ically steerable. The SIR-C antenna is electronically steerable and dual-frequency. Table 22.6 summarizes the RF and mechanical characteristics of the Ball Seasat-class antennas.

Ball Aerospace Systems Division designed an antenna for the low-altitude space-based radar (LASBR) mission.[53] The 13.8- by 63.6-m array is a direct extension of space-proven Seasat and SIR-A technology with stringent constraints of array two-way sidelobe and beam skirt performance. The design features a single-axis deployable truss fabricated from graphite-epoxy microstrip honeycomb panels and passive 3-bit hybrid phase shifters at each of the 49,152 elements. The loss and weight penalty ($W/A = 4.02$ kg/m^2) of a corporate-feed network is compensated for by using transmit and receive gain at each of 384 subpanels.

U.S.S.R. Cosmos 1500 Antenna.   On Sept. 28, 1983, the U.S.S.R. launched the Cosmos 1500 satellite with an SLR for the all-weather probing of the surface and ice cover of the earth's seas and oceans.[38] (The SLR was mentioned in Sec. 22.3.) The antenna is a slotted-waveguide array of 480 slots with a length of 11.085 m and a height of 4 cm. The operational wavelength is 3.15 cm. The beamwidth of the antenna is 0.2 by 42°, providing a gain of 35 dB. The antenna is constructed out of a copper waveguide that measures 23 by 10 mm in cross section. The slots are in the wide wall, with variable spacing to provide a cosine on a pedestal amplitude distribution. Figure 22.14 shows the antenna during deployment. The five sections of the antenna are mated and held in position by spring-loaded locks on the ends that are operated by release mechanisms at the end of the deployment cycle. Helical springs are provided on the flange faces along the wide wall for electrically tight joints. Relative leakage power between sections is down by 50 dB. After deployment, the antenna can be rotated through 35° from the nadir.

Transmit/Receive Modules.   During the advanced development stage of large phased array space-based radars, the use of small, low-cost, lightweight low-power T/R modules was proposed in active array configurations.[54] Goals for these T/R modules[31] included costs of less than $100 each in large mass production and a size of 1 in^2, using 0.5 to 1 W of power. Each module contains a phase shifter, drivers, logic switches, power amplifiers, low-noise receiver, and other components. They can also include a means for sensing and compensation of element displacement error. Further information on solid-state transmitters and transceiver module characteristics is found in Chap. 5.
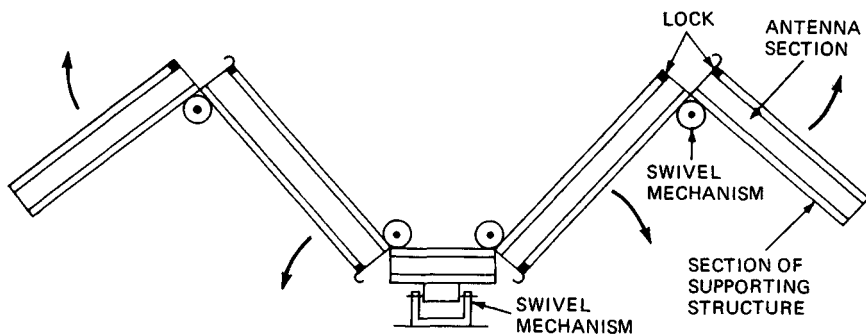
FIG. 22.14   Schematic of the antenna module for Cosmos 1500.

On-Board Processors. The bandwidth requirements for the satellite communications data link can be reduced when an on-board processor is utilized. The major functions of an on-board processor in a large SBR can include pulse compression, doppler filtering, adaptive beamforming, calibration, range walk correction, video integration, constant false-alarm rate (CFAR), monopulse error signal, burst waveform weighting, sidelobe blanking, and editing out interference.

In addition to providing these functions, the processor must be low-powered, have low mass, operate for many years without manual repair, and have radiation-hardened memories. Technology using 16K random-access-memory (RAM) chips and very-large-scale-integration (VLSI) computers can reduce "typical" system power and mass from 3 kW and 2000 lb, respectively, to 400 W and 400 lb.[55] The Defense Advanced Research Projects Agency (DARPA) and others have been working on the development of an advanced on-board signal processor (AOSP) and have made considerable progress.[10] The concern here is to develop a very reliable and survivable on-board computer using gallium arsenide circuitry that can resist the electromagnetic pulse and other radiation effects produced by nuclear detonations. AOSP program technical goals include (1) prime power of 100 W, (2) weight of 100 lb, (3) volume of 2 ft³, (4) 5-year life with a probability of survival of 95 percent, and (5) an input rate of 50 million words per second.[32,56]

Prime Power. The performance of any SBR will ultimately be limited by the prime power system. The most frequently utilized source of prime power for satellites is the solar-battery configuration. High-efficiency GaAs solar cells have demonstrated efficiencies of 18 percent.[57] With the addition of other subsystems, including panels, rotary joints, slip rings, battery, power control, and distribution equipment, the specific power density of the prime power system is on the order of 13 to 24 W/kg. Solar-battery systems are limited and have several disadvantages that will be discussed later.

Space nuclear prime power systems offer certain advantages to SBR, and they have been launched into space by the United States since 1961, beginning with the SNAP-3A. Of the nuclear power systems that were placed into orbit between 1961 and 1977, only one was a nuclear reactor, SNAP-10A. Since then the technology has advanced[58–61] to the extent that it is estimated that an SP-100 type of nuclear reactor would have a mass of 2770 kg and a power output of 100 kW, thereby providing a specific power density of 36 W/kg.

Two baseline deployment configurations of solar-battery and nuclear prime power systems were designed with two power levels, 25 and 100 kW, for the same deployment altitude and are shown in Fig. 22.15.[61] It is seen that the solar systems are larger than the nuclear. As the power level increases, the increased size of the solar system becomes more pronounced. In comparing overall lengths, the 100-kW solar system is 2.4 times the length of the nuclear system. The weight of the solar system depends upon the orbital altitude and the operational requirement during eclipse. For a continuous-operation solar array at geosynchronous altitude, the 100-kW solar system weight is estimated to be 3970 kg.

In comparing advantages and disadvantages, the solar-battery system is based on known technology, and extrapolation to a larger power output is considered to be an engineering design task. The nuclear reactor design re-
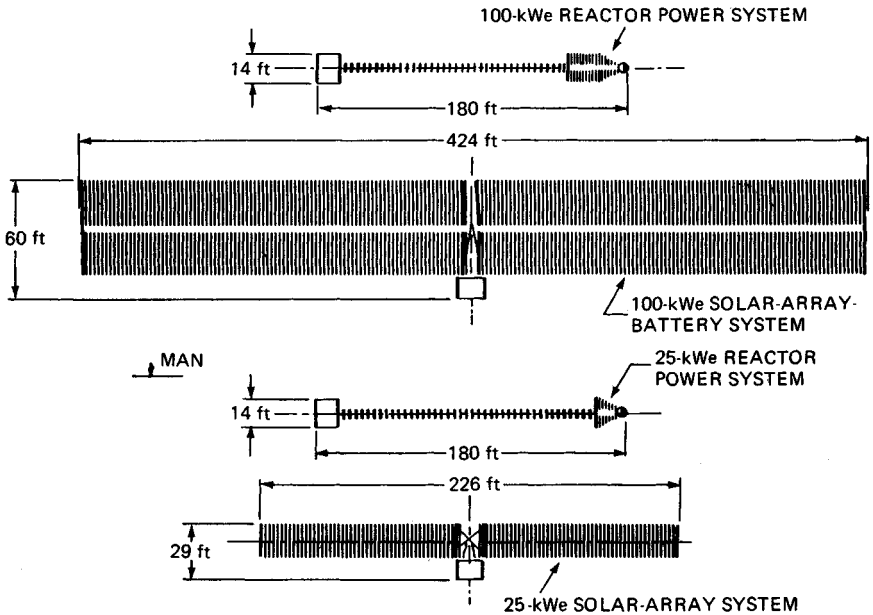
FIG. 22.15 Baseline prime power systems: deployed configurations.[61]

quires engineering development. The advantages of nuclear prime power sys-
tems include (1) reduced mass and size at the higher power levels; (2) no per-
turbation by natural background in low earth orbit (LEO) and geosynchronous
orbit (GEO); (3) no need for alignment, gimballing, slip rings, and long-life
batteries, suggesting that the nuclear system will have significantly enhanced
reliability; (4) reduced effect on SBR antenna, i.e., multipath and sidelobes;
(5) nuclear-hard compared with solar systems; (6) reduced optical and radar
signature; (7) reduced cost by a factor of 3; (8) continuous availability of
power; (9) no orientation requirements; (10) no maneuver limitations; (11) no
power degradation, i.e., beginning-of-life–end-of-life (BOL-EOL) power level;
and (12) no large, flexible structure.

The issue of safety was addressed in 1980 by the United Nations Working
Group report,[59] which studied the safety of nuclear power sources (NPS) in
space. That group reaffirmed the conclusion that NPS can be used safely in
space. It placed responsibility on the launching nation to (1) conduct safety tests
and evaluations consistent with international standards; (2) provide the United
Nations with detailed design and test data of the NPS at launch time; and (3)
when reentry of the NPS becomes reasonably certain, provide the United Na-
tions with details of orbiting parameters, probable impact regions, power history,
inventory of nuclear fuel, and radiation dosage at 1 m for survival sections. The
working group noted that U-235-fueled reactors required 400 years' decay time to
reduce fission product activity by a factor of 1000. It implied that a minimum or-
bit altitude of 300 nmi should be used.

It is obvious from a technical point of view that nuclear prime power systems
should be utilized for large SBR systems whenever high power is required.

22.5 CRITICAL ISSUES

A succinct treatment of selected critical issues is given here. The critical issues in the development of SBR include (1) system cost, (2) system survivability or vulnerability, (3) system calibration, (4) antenna deployment and distortion, (5) onboard processing, and (6) nuclear prime power.

SBR System Costs.   The author uses a cost-estimating ratio for SBR satellites of $64,000 per kilogram in 1988 dollars that is based upon informal study of many satellites that have been placed into orbit. Launch costs are not included; they depend upon the launch vehicle. Informal study of many satellite launches has resulted in the data shown in Fig. 22.16, which gives a launch cost for several types of vehicles when launched from two United States launch sites, the Eastern Test Range (ETR) and the Western Test Range (WTR). It can be seen that polar orbit costs are greater than launches from the ETR due east and that it is more economical (on a dollars-per-pound basis) to launch large payloads on STS and Titan class vehicles.

Survivability and Vulnerability.   SBR system survivability and vulnerability must be demonstrated and tested. The natural space radiation environment will cause a significant total dose on a T/R module depending upon its shielding. Table 22.7 is a summary of the total dose for a 5-year period for circular orbits at altitudes of 450, 900, and 5600 nmi.[32] The T/R module in the analysis has an area of 1 in^2, and it has been assumed that the total dose values will be double those expected in order to account for the particle radiation that penetrates both sides of the module package. Some shielding may be provided by the chip substrate; however, this has been ignored.

22.6 SBR FUTURE POSSIBILITIES

Rendezvous Radar Missions.   All satellite rendezvous missions have been performed by manned vehicles. In the foreseeable future, the majority of rendezvous missions might be conducted by unmanned vehicles such as the OMV. The planned list of missions for the OMV includes (1) large-observatory servicing at the shuttle, (2) payload placement, (3) payload retrieval, (4) payload reboost, (5) payload deboost to reentry, (6) payload viewing, (7) subsatellite mission, (8) multiple-payload mission, (9) in situ servicing mission, (10) STS transfer to space station, and (11) base support. Details of these missions may be found in the NASA OMV request for proposal.[62] The initial design of the OMV is modular so as to permit upgrading its capability to operate from the space station and to accommodate the following growth missions by the addition of appropriate kits or elements to the system: (1) logistic support, (2) debris collection mission, (3) extended on-orbit operation, (4) satellite buildup, (5) satellite refueling, (6) servicing mission, and (7) space station reboost.

A rendezvous radar that is low in cost and light in weight will be used to perform these future OMV missions. Such an OMV radar might have the major performance characteristics shown in Table 22.8. The rendezvous radar set (RRS) will be an X-band coherent, range-gated, pulse doppler radar with redundant
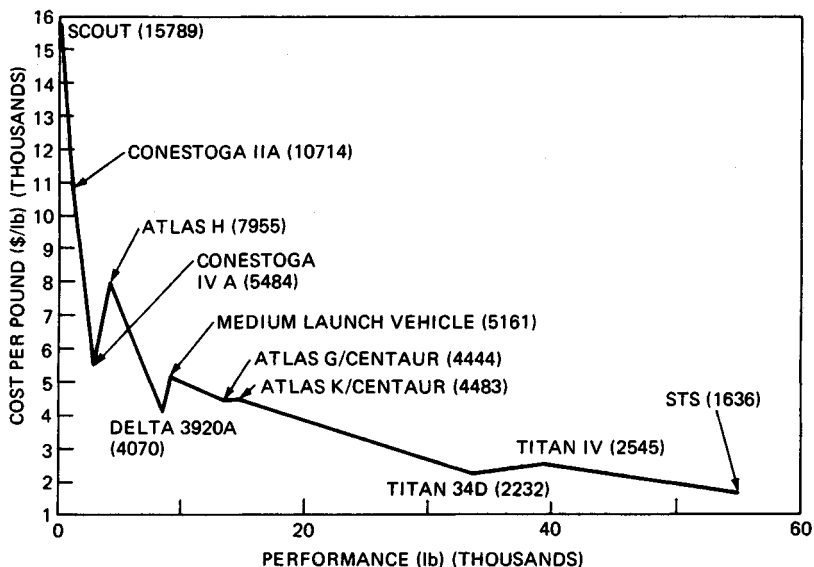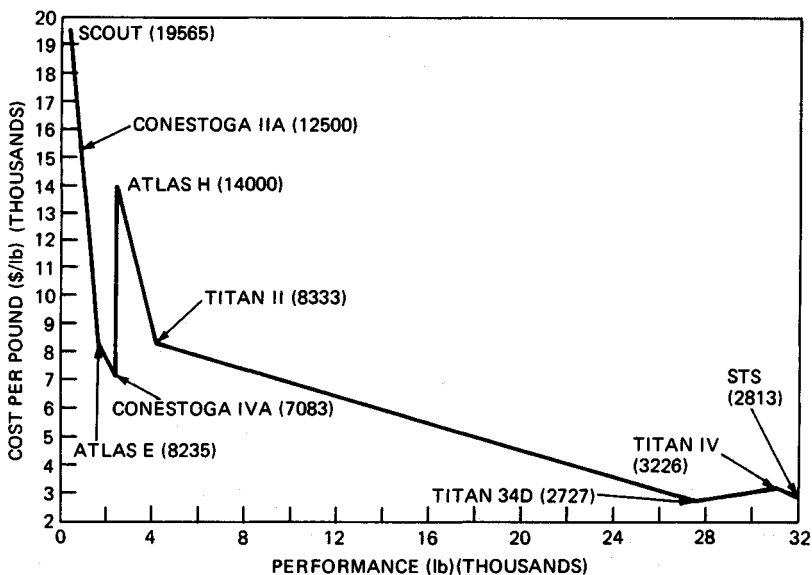
FIG. 22.16 Cost per pound to low earth orbit. (*a*) Eastern Test Range launch due east. (*b*) Western Test Range launch into polar orbit.

TABLE 22.7 Space Radiation Environment Summary*

SBR orbit, nmi	5-year total dose, rads (Si); aluminum shielding thickness		
	15 mils	25 mils	50 mils
450	$2(10)^5$	$6(10)^4$	$2(10)^4$
900	$2(10)^6$	$4(10)^5$	$2(10)^5$
5600	$6(10)^6$	$4(10)^6$	$(10)^6$

*From Ref. 32.

TABLE 22.8 OMV Radar Characteristics

Frequency	9.5 to 9.8 GHz
PRF	6.67 kHz
Pulse width	0.05, 0.2, 1.5, and 15 µs
Transmitter peak power	2 W (GaAs FET with ≈ 30-dB gain)
Receiver noise figure	< 4 dB, GaAs FET LNA
Antenna	Planar slotted array, linear polarization
Antenna size	14 by 15 by 1 in
Antenna gain and beamwidth	30.5 dB (at 9.65 GHz) and 5.0°
Search scan	± 20° cone, with 5-min scan time
Angle accuracy (3σ)	20 mrad
Range accuracy (3σ)	Greater of 20 ft or 2 percent of range
Range rate accuracy (3σ)	Greater of 0.1 ft/s or 2 percent of range rate
Deployed assembly weight	26 lb
Inboard assembly weight	50 lb (redundant total)
Electronics volume	≈ 2 ft^3 (redundant total)
Prime power	< 60 W

electronics and redundant gimbal motor windings. The OMV system computer initiates the acquisition-search function to permit detection of a 1-m^2 Swerling 1 target at a 4.5-nmi range (with 99 percent probability of detection and a false-alarm rate of one alarm per hour.) Monopulse tracking is performed to within a minimum range of 35 ft. Peak power is programmed over a 50 dB range during the rendezvous maneuver to minimize the RF radiation intensity on sensitive targets. Pulse frequency agility is utilized; up to 30 carrier frequency changes in 10-MHz steps over the 300-MHz operating band are used to decorrelate Swerling 1 target fluctuations. At each dwell, 128 pulses are coherently integrated in the fast Fourier transform (FFT) processor prior to noncoherent integration of the FFT outputs. Up to 30 FFT outputs can be integrated.

Initial configuration of the Space Station will have limited tracking-system requirements that include tracking of cooperative vehicles within a 37-km control zone.[63] This is based upon the assumption that all vehicles will provide accurate position and velocity data to the space station tracking system by way of the space-to-space link. Automatic tracking of an extravehicular (EV) astronaut is not required. For growth configurations of the Space Station, the tracking system

must expand to meet additional requirements. More co-orbiting vehicles, nonco-operative or disabled vehicles, automatic tracking of EV astronauts, and sensors for berthing and docking operations will require additional tracking capabilities within the system. Some type of short-range radar may be required to track vehicles which do not have global positioning system (GPS) capability or vehicles which have been disabled. Results of preliminary tradeoffs on multiple-target tracking radars indicate[36] that either a K_a-band or an X-band phased array radar would be the preferred approach.

Remote-Sensing Missions.[64]   SBR will participate in many remote-sensing missions for observation of the earth and the planets. The SIR series is expected to have the capability to image the earth's surface by using all polarization states (*HH, VV,* and *HV*) and with multiple frequency bands.

A number of SBR SAR missions have been considered by various countries for various purposes. An example is the Canadian Radarsat, which employs a C-band SBR SAR primarily for monitoring polar ice dynamics for use in ship routing; the SAR will have a 200-km swath.

In planetary exploration areas, SBR imaging systems are key elements for the exploration of two bodies that are continuously cloud-covered, Venus and Titan. In the exploration of Venus during the late 1970s, a radar sensor on the Pioneer Venus Orbiter provided low-resolution (40 to 100 km) images of the planet. A U.S.S.R. Venera satellite[65] produced radar images of part of the northern hemisphere of Venus with a resolution of 1300 m. The United States Venus radar mission has the objective of providing global coverage with a resolution of 150 m. In the exploration of Titan, a satellite of Saturn, the larger distance to the earth will put a very tight limit on data rate transmission, which directly impacts mapping coverage and resolution. A radar can be placed into orbit around Saturn, and on selected orbits the spacecraft will fly by Titan. These flybys will be targeted so that during each flyby a different region of Titan will be mapped with an SBR SAR. The Titan radar mapper will have a very wide swath (600 to 800 km) to obtain a global map during the small number of flybys. Real aperture imaging will provide a resolution of 6 to 40 km. A synthetic aperture mode can be used to observe limited regions with a resolution of about 200 m.

Other missions using radar are planned for ocean scatterometry and altimeters. Scatterometers are used to obtain accurate measurement of global surface winds for oceanography and meteorology. Wind speed with errors of about 2 m/s and a wind direction error of less than 16° are sought. SBR altimeters expect to measure altitude with an error of 5 cm from a 1300-km polar orbit inclined 65° over the ocean. Over the solid surface of Mars, a 37-GHz altimeter on the Mars orbiter mission expects to gather global high-resolution topographic mapping data with a height resolution of 15 m.

The overall goal of the Earth Observing System (EOS) is to advance the scientific understanding of the entire earth system on the global scale through developing a deeper understanding of the components of that system, interactions among them, and how the system is changing.[66] International space station elements include the following satellites in polar and equatorial orbits (1) a NASA EOS platform at 824 km, sun-synchronous, 1:30 P.M. equator-crossing time, ascending-node orbit; (2) a European Space Agency (ESA) platform at 824 km, sun-synchronous, 10:00 A.M. equator-crossing time, descending-node orbit; and (3) the manned space station in a 335- to 460-km 28.5° inclined orbit. Instruments planned for the satellites include radar, radiometers, IR, optics, and ultraviolet (UV). These sensors will measure parameters such as winds, clouds, rain, liquid-moisture content, geologic parameters,

ocean currents, etc. Radars will be used to make atmospheric and geological observations. Two of the radars proposed are the tropical-rain mapping radar (TRAMAR) and the land, ocean, and rain radar altimeter (LORRA).[67]

Global Air Traffic Surveillance.[12] ATC is increasingly a matter of global concern, and the explosive growth of aircraft density in, around, and between major metropolitan areas in Europe and North America is common knowledge. If a United Nations organization were responsible for ATC for 120 to 130 nations in the world, it is conceivable that as many as 84,000 commercial aircraft could require ATC in the twenty-first century. A rosette constellation of SBR satellites at an orbital altitude of 5600 nmi (10,371 km) in a 14/14/12 Walker orbit[23] inclined at 49.4° provides continuous worldwide visibility by at least two satellites simultaneously. (One satellite is deployed in each of 14 equally spaced orbit planes.) Each satellite provides radar coverage between grazing angles from 3 to 70°. The major subsystems in the satellite include (1) radar, (2) communications, (3) guidance and control, and (4) electrical power subsystems. Details of these subsystems are found in Ref. 12. The radar parameters are shown in Table 22.9. Each T/R module would have a peak power of 0.155 W and an average power of 15 mW and would weigh 5 g.

TABLE 22.9 Radar Parameters for Global Air Traffic Surveillance*

Antenna	
Type	Corporate-fed active phased array
Diameter	100 m
Frequency	2 GHz
Wavelength	0.15 m
Polarization	Circular
Number of elements	576,078
Number of modules	144,020
Element spacing	0.7244 wavelength
Beamwidth	1.83 mrad
Directive gain	66.42 dB
Maximum scan angle	22.4°
Receiver	
Type	Distributed solid-state monolithic T/R module
Bandwidth	500 kHz
System noise temperature	490 K
Compressed pulse width	2 μs
Transmitter	
Type	Distributed solid-state monolithic T/R module
Peak power	22.33 kW
Pulse width	2000 μs
Maximum duty	0.20
Frequency	2 GHz
Signal Processor	
Type	Digital
Input speed	50 million words per second

*From Ref. 12.

Military SBR Systems. Brookner and Mahoney[11] derived a satellite radar architecture for performing the basic surveillance missions for the fleet defense and air defense of the CONUS. The system was an L-band, corporate-fed phased array radar in orbit constellations of 3 to 12 satellites at altitudes from 600 to 2000 nmi. At the highest orbital altitude, a 10- by 30-m phased array that contained 15,000 radiating elements or modules was designed. The modules delivered an average power of 6 kW, and the radar required a prime power of 30 kW.

REFERENCES

1. Hughes Aircraft Company: "K$_u$-Band Integrated Radar and Communications Equipment for the Space Shuttle Orbiter Vehicle," preliminary design review, vol. 1, Mar. 14–24, 1978.

2. RCA Government and Commercial Systems, Aerospace Systems Division, Burlington, Mass.: "The Apollo LM Rendezvous Radar and Transponder," *Rept. LTM* 3300-14D, February 1971.

3. Quigley, W. W.: Gemini Rendezvous Radar, *Microwave J.*, pp. 39–45, June 1965.

4. Fenner, R. G., and R. F. Broderick: Spaceborne-Radar Applications, chap. 34 in Skolnik, M. I. (ed.): "Radar Handbook," McGraw-Hill Book Company, New York, 1970.

5. Elachi, C., et al.: Spaceborne Synthetic Aperture Imaging Radars: Applications, Techniques and Technology, *Proc. IEEE*, vol. 70, pp. 1174–1209, October 1982.

6. Elachi, C., and J. Granger: Spaceborne Imaging Radars Probe "in Depth," *IEEE Spectrum*, vol. 19, pp. 24–29, November 1982.

7. Williams, F. C., et al.: The Pioneer Venus Orbiter Radar, *1976 WESCON Sess. 4*, Los Angeles, Sept. 14–17, 1976.

8. Hofmeister, E. L., et al.: GOES-C Radar Altimeter, vol. 1, "Data Users Handbook," General Electric Company, Utica, N.Y., May 1976.

9. Soviet Radar Records Venus Surface Imager, *Aviat. Week Space Technol.*, vol. 119, p. 18, Oct. 24, 1983.

10. Ulsamer, E.: In Focus—Approach Set on Space Radars, *Air Force Mag.*, vol. 67, pp. 17–18, February 1984.

11. Brookner, E., and T. F. Mahoney: Derivation of a Satellite Radar Architecture for Air Surveillance," *IEEE EASCON '83 Conf. Rec.*, Washington, Sept. 19–21, 1983.

12. Cantafio, L. J., and J. S. Avrin: Satellite-Borne Radar for Global Air Traffic Surveillance, *IEEE ELECTRO '82 Prof. Sess. Rec.*, Boston, May 25–27, 1982.

13. Cantafio, L. J.: Space Based Radar Concept for the Proposed United Nations International Satellite Monitoring Agency, *Mil. Microwaves Conf.*, London, Oct. 24–26, 1984.

14. "The Implication of Establishing an International Satellite Monitoring Agency," UN Publ. Sales No. E.83.IX.3.

15. Griffin, J. W., et al.: K$_u$-Band—The First Year of Operation, *IEEE Int. Radar Conf. Rec.*, pp. 330–339, Arlington, Va., May 6–9, 1985. (IEEE Cat. No. 85CH2076-8.)

16. Wolverton, R. W., et al.: "Flight Performance Handbook for Orbital Operations," John Wiley & Sons, New York, 1961.

17. Hord, R. A.: Relative Motion in the Terminal Phase of Interception of a Satellite or Ballistic Missile, *NACA TN* 8399, September 1958.

18. Functional Requirements for the Seasat-A Synthetic Aperture Radar System, *Jet Propulsion Lab., FR No. FM*511774, rev. dated Aug. 2, 1976, Pasadena, Calif.

19. "Geodynamics Experimental Ocean Satellite Project of the Earth and Ocean Physics

Applications Program," NASA brochure, NASA Wallops Flight Center, Wallops Island, Va., 1975.

20. New Satellite to Measure Ocean Surface Topography and Sea State, *NASA News Release* 75-88, Washington, Mar. 31, 1975.

21. GEOS-C Mission Plan, NASA TK-6340-001, rev. 3, NASA Wallops Flight Center, Wallops Island, Va., Dec. 18, 1974.

22. Cantafio, L. J.: Satellite-Borne Radar, *Lecture IX, Adv. Radar Technol. Short Course*, scheduled by Technology Service Corporation, San Diego, Apr. 22, 1983.

23. Walker, J. G.: Continuous Whole Earth Coverage by Circular Orbit Satellite Patterns, *R. Aircr. Estab. Tech. Rept.* 77044, Mar. 24, 1977.

24. Luders, R. D., and L. J. Ginsberg: Continuous Zonal Coverage—A Generalized Analysis, *AIAA Pap.* 74-842, *AIAA Mech. Control of Flight Conf.*, Anaheim, Calif., Aug. 5–9, 1974.

25. Ballard, A. H.: Rosette Constellation of Earth Satellites, *IEEE Trans.*, vol. AES-16, pp. 656–673, September 1980.

26. Beste, D. C.: Design of Satellite Constellations for Optimal Continuous Coverage, *IEEE Trans.*, vol. AES-14, pp. 466–473, May 1978.

27. Emara, E. T., and C. T. Leondes: Minimum Number of Satellites for Three-Dimensional Continuous Worldwide Coverage, *IEEE Trans.*, vol. AES-13, pp. 108–111, March 1977.

28. Harney, E. D.: "Space Planners Guide," U.S. Air Force Systems Command, Publ. 0-774-405, 1965.

29. Fager, J. A.: Application of Graphite Composites to Future Spacecraft Antennas, *AIAA Pap.* 76-328, *Sixth Commun. Satellite Syst. Conf.*, Apr. 6–8, 1976.

30. Schultz, J. L., and P. Nosal: Space-Based Radar, *Horizons*, vol. 15, no. 1, p. 10, Grumman Aerospace Corporation, 1979.

31. Fawcette, J.: Large Radar Satellite Proposed, *Microwave Syst. News*, vol. 8, pp. 17–20, September 1978.

32. Mrstik, A. V., et al.: RF Systems in Space—Space-Based Radar Analysis, *General Research Corporation, RADC TR-83-91, Final Tech. Rept.*, vol. II, April 1983.

33. Ludwig, A. C., et al.: RF Systems in Space—Space Antennas Frequency (SARF) Simulation, *General Research Corporation, RADC TR-38-91, Final Tech. Rept.*, vol. I, April 1983.

34. Kendrick, J. B. (ed.): "TRW Space Data," 3d ed., 1967.

35. Barton, D. K.: A Half Century of Radar, *IEEE Trans.*, vol. MTT-32, pp. 1161–1169, September 1984.

36. Tu, K., et al.: Space Shuttle Communications and Tracking System, *Proc. IEEE.*, vol. 75, pp. 356–370, March 1987.

37. Brejcha, A. G., L. H. Keeler, and G. G. Sanford: The Seasat-A Synthetic Aperture Radar Antenna, *Synth. Aperture Radar Technol. Conf.*, Las Cruces, N. Mex., Mar. 8–10, 1978.

38. Kalmykov, A. I., et al.: Side-Looking Radar of Kosmos-1500 Satellite, *Issled. Zemli Kosmosa*, no. 3, May–June 1985.

39. Soviets Plan to Launch New Spacecraft, *Aviat. Week Space Technol.*, vol. 127, p. 27, Oct. 19, 1987.

40. Bearse, S. V.: Knitted Antenna Solving Knotty Problems, *Microwaves*, p. 14, March 1974.

41. Large Furlable Antenna Study, *Lockheed Missiles and Space Company, Rept. LMSC-D384797*, Jan. 20, 1975.

42. Cummings, Freeman, and Benz: Deployable Parabolic Antenna, U.S. Patent 3,789,375, Dec. 18, 1973, assigned to Rockwell Inc.

43. Fager, J. A., and R. Garriott: Large Aperture Expandable Truss Microwave Antenna, *IEEE Trans.*, vol. AP-17, pp. 452–458, July 1969.

44. Das, A., and J. A. Delaney: Spacecraft Phased Array Configurations, *IEEE Trans.*, vol. AP-17, pp. 522–524, July 1969.

45. Final Report—Spaceborne Radar Study, *Grumman Aerospace Corporation, AFSC-ESD contract F19628-74-R-0140, Rept. 74-21AF-I*, June 28, 1974.

46. Hagler, T.: Building Large Structures in Space, *Astronaut. Aeronaut.*, vol. 14, pp. 56–61, May 1976.

47. Deployable Antenna Flight Experiment—Preliminary Definition Study, *Harris Corporation, Third Q. Rev.*, June 24, 1981.

48. Deployable Antenna Flight Experiment Definition Study: Mid-Term Review, *Grumman Aerospace Corporation, NAS*-8-33932, Mar. 20, '981.

49. Bujakes, V. I., et al.: Infinitely Built-Up Radio Telescope, *Pap. IAF-77-67, IAF XXVIII Cong.*, Prague, Sept. 25–Oct. 1, 1977.

50. Ulsamer, E.: ATS-6, NASA's Huge Transmitter in the Sky, *Air Force Mag.*, vol. 57, August 1974.

51. Archer, J. S.: Advanced Sunflower Antenna Concept Development, *LSST First Ann. Tech. Rev.*, NASA LRC, Nov. 7–8, 1979.

52. Lockheed Tests Large Space Antenna, *Aviat. Week Space Technol.*, vol. 120, p. 70, Apr. 30, 1984.

53. Larson, T. R., A Microstrip Honeycomb Array for the Low Altitude Space Based Radar Mission, *Ball Aerospace Systems Division, Rept. F*81-06, August 1981.

54. Final Report—Spaceborne Radar Study, *Grumman Aerospace Corporation, Rept. 74-21-AF-1, prepared for ADSC-ESD (XRS) contract F*19623-74-R-0140, June 28, 1974.

55. Thimlar, M. E., et al.: Future Space-Based Computer Processors, *Aerosp. Am.*, vol. 22, pp. 78–82, March 1984.

56. Works, G. A.: Advanced Onboard Signal Processor, *IEEE EASCON '80 Rec.*, p. 233, 1980.

57. GaAs Solar Cells, *Hughes Aircraft Company, presented at Space Power Conf.*, Los Angeles, Jan. 13–14, 1981.

58. Emigh, C. R.: Reactor Technology, January–March 1980, *Los Alamos Scientific Laboratory, Prog. Rept. LA*-8403-*PR-UC*-80, June 1980.

59. Buden, D., et al.: Space Nuclear Reactor Power Plants, *LASL Informal Rept. LA*-8223-*MS-UC*-33, January 1980.

60. Buden, D., et al.: Selection of Power Plant Elements for Future Reactor Space Electric Power Systems, *LASL Rept. LA*-7858, September 1979.

61. Kelley, J. H.: Minutes of Space Nuclear Power Service Working Group Meeting, Sept. 24, 1982, issued Oct. 8. 1982.

62. Orbital Maneuvering Vehicle, *NASA George C. Marshall Space Flight Center, Request for Proposal* 1-6-*pp*-01438, November 1985.

63. Dietz, R. H.: Space Station Communications and Tracking Systems, *Proc. IEEE*, vol. 75, pp. 371–382, March 1987.

64. Carver, K. R., C. Elachi, and F. T. Ulaby: Microwave Remote Sensing from Space, *Proc. IEEE*, vol. 73, pp. 970–996, June 1985.

65. Bogomolov et al.: Venera 15 and 16 Synthesized Aperture Radar in Orbit around Venus, *Izv. Vyssh. Uchebn. Zaved., Radiofiz.*, vol. 28, pp. 259–274, March 1985.

66. NASA Announcement of Opportunity: The Earth Observing System (EOS), *A.O.* No. *OSSA*-1-88, Jan. 19, 1988.

67. Lorra/Tramar Design—Feasibility Study, *Malibu Res. and TRW*, MRA p. 214-3, May 11, 1988.

CHAPTER 23
METEOROLOGICAL RADAR

Robert J. Serafin
National Center for Atmospheric Research*

23.1 INTRODUCTION

As this handbook is being written, dramatic changes are taking place in the field of radar meteorology. While the majority of radar engineers are familiar with current operational meteorological radars, few are aware of the advances that have been made in the past two decades. For example, doppler radar meteorology, using modern digital signal-processing techniques and display technology, has moved ahead so rapidly that the United States is now planning to replace its existing operational weather radar network with a next-generation doppler system (NEXRAD). This system will provide quantitative and automated real-time information on storms, precipitation, hurricanes, tornadoes, and a host of other important weather phenomena, with higher spatial and temporal resolution than ever before.[1] A second network of doppler radars, in airport terminal areas, will provide quantitative measurements of gust fronts, wind shear, microbursts, and other weather hazards for improving the safety of operations at major airports in the United States.[1,2] Next-generation doppler radars that use flat-plate antennas, color displays, and solid-state transmitters are now available for commercial aircraft. And many of these new technologies are being deployed in countries throughout the world.

In the research arenas, multiple-doppler radars are used for deriving three-dimensional wind fields.[3] Airborne doppler radar[4,5] has been used to duplicate these capabilities, thus providing for great mobility. Polarization diversity techniques[6] are used for discriminating ice particles from water, for improved quantitative precipitation measurement, and for detecting hail. And there is a new family of radars, ultrahigh-frequency (UHF) and very-high-frequency (VHF) fixed-beam systems that are being used to obtain continuous profiles of horizontal winds.[7] These examples are illustrative of the vitality of the field.

This chapter is intended to introduce the reader to meteorological radar and particularly those system characteristics that are unique to meteorological applications. In this regard, it should be noted that most meteorological radars appear

23.1

similar to radars used for other purposes. Pulsed and pulsed doppler systems are common. Parabolic dish antennas, focal-point feeds, and low-noise solid-state receivers are used. Magnetrons, phase-locked magnetrons, klystrons, traveling-wave tubes, and other forms of transmitters are used.

The major distinction between meteorological radar and other kinds of radars lies in the nature of the targets. Meteorological targets are distributed in space and occupy a large fraction of the spatial resolution cells observed by the radar. Moreover, it is necessary to make quantitative measurements of the received signal's characteristics in order to estimate such parameters as precipitation rate, precipitation type, air motion, turbulence, and wind shear. In addition, because so many radar resolution cells contain useful information, meteorological radars require high-data-rate recording systems and effective means for real-time display.[8,9] Thus, while many radar applications call for discrimination of a relatively few targets from a clutter background, meteorological radars focus on making accurate estimates of the nature of the *weather clutter* itself. This poses some challenging problems for the radar system designer to address.

The discussion here will refer to a number of useful texts and references for the reader to use. However, Battan's text,[10] revised in 1973, deserves special mention for its clarity and completeness and remains a standard for courses in radar meteorology that are taught in universities around the world. Doviak and Zrnić[11] place special emphasis on doppler meteorological radar. Chapter 24 in the first "Radar Handbook," by Bean et al.,[12] addresses the problem of weather effects on radar. Finally, perhaps the broadest and most complete set of references on progress in the field can be found in the *Proceedings* and *Preprints* of the series of radar meteorology conferences sponsored by the American Meteorological Society (AMS). These documents can be found in most technical libraries and also can be obtained through the offices of the AMS in Boston.

23.2 THE RADAR RANGE EQUATION FOR METEOROLOGICAL TARGETS

The received power from distributed targets can be derived from any of a variety of expressions that are applicable to radar in general. A simple form, with which to begin, is given below:

$$P_r = \frac{\beta \sigma}{r^4} \tag{23.1}$$

where β is a constant dependent upon radar system parameters, r is the range, and σ is the radar cross section.

It is in the calculation of σ for meteorological targets that the radar range equation differs from that for point targets. σ may be written

$$\sigma = \eta V \tag{23.2}$$

where η is the radar reflectivity in units of cross-sectional area per unit volume and V is the volume sampled by the radar. η can be written as

$$\eta = \sum_{i=1}^{N} \sigma_i \tag{23.3}$$

where N is the number of scatterers per unit volume and σ_i is the backscattering cross section of the ith scatterer. In general, the meteorological scatterers can take on a variety of forms, which include water droplets, ice crystals, hail, snow, and mixtures of the above.

Mie[13] developed a general theory for the energy backscattered by a plane wave impinging on spherical drops. This backscattered energy is a function of the wavelength, the complex index of refraction of the particle, and the ratio $2\pi\alpha/\lambda$, where α is the radius of the spherical particle and λ is the wavelength.

When the ratio $2\pi\alpha/\lambda - 1$, the Rayleigh approximation[10] may be applied, and σ_i becomes

$$\sigma_i = \frac{\pi^5}{\lambda^4}|K|^2 D_i^6 \qquad (23.4)$$

where D_i is the diameter of the ith drop and

$$|K|^2 = \left|\frac{m^2 - 1}{m^2 + 2}\right|^2 \qquad (23.5)$$

where m is the complex index of refraction. At temperatures between 0 and 20°C, for the water phase, and at centimeter wavelengths

$$|K|^2 \approx 0.93 \qquad (23.6a)$$

and for the ice phase

$$|K|^2 \approx 0.20 \qquad (23.6b)$$

Equation (23.3) can now be written as

$$\eta = \frac{\pi^5}{\lambda^4}|K|^2 \sum_{i=1}^{N} D_i^6 \qquad (23.7)$$

and the radar reflectivity factor Z defined as

$$Z \quad \sum_{i=1}^{N} D_i^6 \qquad (23.8)$$

In radar meteorology, it is common to use the dimensions of millimeters for drop diameters D_i and to consider the summation to take place over a unit volume of size 1 m^3. Therefore, the conventional unit of Z is in mm^6/m^3. For ice particles, D_i is given by the diameter of the water droplet that would result if the ice particle were to melt completely.

It is often convenient to treat the drop or particle size distribution as a continuous function with a number density $N(D)$, where $N(D)$ is the number of drops per unit volume, with diameters between D and $D + dD$. In this case, Z is given by the sixth moment of the particle size distribution,

$$Z = \int_0^\infty N(D)D^6 dD \qquad (23.9)$$

If the radar beam is filled with scatterers, the sample volume V is given[10] approximately by

$$V \approx \frac{\pi\theta\phi r^2 c\tau}{8} \tag{23.10}$$

where θ and ϕ are the azimuth and elevation beamwidths, c is the velocity of light, and τ is the radar pulsewidth.

Substituting Eqs. (23.10), (23.2), and (23.4) into Eq. (23.1) gives

$$\begin{aligned}
P_r &= \frac{\beta\pi}{r^4} \frac{\theta\phi r^2 c\tau}{8} \frac{\pi^5}{\lambda^4} |K|^2 \sum_{i=1}^{N} D_i^6 \\
&= \frac{\beta\pi^6\theta\phi\, c\tau\, |K|^2}{8\lambda^4 r^2} Z \\
&= \frac{\beta' Z}{r^2}
\end{aligned} \tag{23.11}$$

This simple expression illustrates that the received power is a function only of β' (a constant dependent upon radar system parameters), is proportional to the radar reflectivity factor Z, and is inversely proportional to r^2.

In actual fact, the antenna gain is not uniform over the beamwidth, and the assumption of a uniform gain can lead to errors in the calculation of Z. Probert-Jones[14] took this into account, assumed a gaussian shape for the antenna beam, and derived the following equation for the received power:

$$P_r = \frac{P_t G^2 \lambda^2 \theta\phi\, c\tau}{512(2\ln 2)\pi^2 r^2} \sum_{i=1}^{N} \sigma_i \tag{23.12}$$

where $2\ln 2$ is the correction due to the gaussian-shaped beam.

By using the relationships in Eqs. (23.7) and (23.8), Eq. (23.12) can be written in terms of the reflectivity factor Z as

$$P_r = \frac{P_t G^2 \theta\phi c\tau\pi^3 |K|^2 Z}{512(2\ln 2) r^2 \lambda^2} \tag{23.13}$$

One must be careful to use consistent units in Eq. (23.13). If meter-kilogram-seconds (mks) units are used, the calculation of Z from Eq. (23.13) will have dimensions of m^6/m^3. Conversion to the more commonly used units of mm^6/m^3 requires that the result be multiplied by the factor 10^{18}. Because Z values of interest can range over several orders of magnitude, a logarithmic scale is often used, where

$$dBZ = 10 \log Z \tag{23.14}$$

Equation (23.13) can be used to measure the reflectivity factor Z when the antenna beam is filled, when the Rayleigh approximation is valid, and when the scatterers are in either the ice or the water phase. Because all these conditions

are not always satisfied, it is common to use the term Z_e, the effective reflectivity factor, in place of Z. When Z_e is used, it is generally understood that the above conditions are assumed. Practitioners in the field of radar meteorology often use Z_e and Z interchangeably, albeit incorrectly.

Finally, it is important to note the range of Z values that are of meteorological significance. In nonprecipitating clouds, Z values as small as -40 dBZ are of interest. In the optically clear boundary layer, Z values of the order -20 dBZ to 10 dBZ are of interest. In rain, Z may range from about 20 dBZ to as much as 60 dBZ, with a 55 to 60 dBZ rain being of the type that can cause severe flooding. Severe hailstorms may produce Z values higher than 70 dBZ. Operational radars are generally designed to detect Z values ranging from 10 to 60 dBZ, while research applications usually aim for the maximum dynamic range possible. In light of the above, operational radars often employ sensitivity time control (STC) to compensate for inverse r^2 dependence, but research radars usually do not use STC owing to the attendant loss of sensitivity at short ranges.

23.3 DESIGN CONSIDERATIONS

Three of the more significant factors that affect the design of meteorological radars are attenuation, range-velocity ambiguities, and ground clutter. The combination of these three, along with the need to obtain adequate spatial resolution, leads to a wavelength selection in the range of 3 to 10 cm for most meteorological applications.

Attenuation Effects. Attenuation has at least two negative effects on meteorological radar signals. First, because of attenuation it becomes difficult, if not impossible, to make quantitative measurements of the backscattered energy from precipitation which is at greater range (and at the same azimuth and elevation angles) than precipitation closer to the radar. This inability to precisely measure the backscattering cross section makes quantitative measurements of precipitation rates more difficult.

Second, if the attenuation due to precipitation or the intervening medium is sufficiently great, the signal from a precipitation cell behind a region of strong absorption may be totally obliterated, leading to potentially disastrous effects. One example of the potentially serious consequences of very strong absorption is the impact it might have on airborne storm avoidance radars, most of which are in the 3-cm band, although some use a 5-cm wavelength. Metcalf[15] has examined ground-based radar data from the storm that was responsible for the 1977 crash of Southern Airways Flight 242 in northwest Georgia. The crew had relied on its on-board radar for penetration of a severe storm. Metcalf shows strong evidence that the region penetrated by the aircraft, while appearing to be free of echo, had actually been obliterated because of severe attenuation. Severe storms can also produce very strong absorption at 5-cm wavelength, as noted by Allen et al.[16]

In some meteorological radar applications, it is desirable to attempt to measure attenuation along selected propagation paths. This is done because absorption is related to liquid-water content and can provide useful information for the detection of such phenomena as hail, in accordance with the dual-wavelength technique described by Eccles and Atlas.[17]

In the following subsections, quantitative expressions relating attenuation to precipitation are given. Much of this is taken from Bean, Dutton, and Warner.[12]

Battan's textbook[10] is also an excellent source for additional information on the absorbing properties of precipitation.

Attenuation in Clouds. Cloud droplets are regarded here as those water or ice particles having radii smaller than 100 μm, or 0.01 cm. For wavelengths of incident radiation well in excess of 0.5 cm, the attenuation becomes independent of the drop-size distribution. The generally accepted equations for attenuation by clouds usually show the moisture component of the equations in the form of the liquid-water content (grams per cubic meter). Observations indicate that the liquid-water concentration in clouds generally ranges from[18] 1 to 2.5 g/m³, although Weickmann and aufm Kampe[19] have reported isolated instances of cumulus congestus clouds with water contents of 4.0 g/m³ in the upper levels. In ice clouds, it rarely exceeds 0.5 and is often less than 0.1 g/m³. The attenuation due to cloud drops may be written[12]

$$K = K_1 M \tag{23.15}$$

where K = attenuation, dB/km
K_1 = attenuation coefficient, dB/(km · g · m³)
M = liquid-water content, g/m³

$$M = \frac{4\pi\rho}{3}\sum_{i=1}^{N} a_i^3 \tag{23.16}$$

$$K_1 = 0.4343\frac{6\pi}{\lambda}\text{Im}\left(-\frac{m^2 - 1}{m^2 + 2}\right) \tag{23.17}$$

where the a_i are droplet radii, ρ is the density of water, and Im is the imaginary part. Values of K_1 for ice and water clouds are given for various wavelengths and temperatures by Gunn and East in Table 23.1.

Several important facts are demonstrated by Table 23.1. The decrease in attenuation with increasing wavelength is clearly shown. The values change by about an order of magnitude, for a change of λ from 1 to 3 cm. The data presented here also shows that attenuation in water clouds increases with decreasing temperature. Ice clouds give attenuations about two orders of magnitude smaller

TABLE 23.1 One-Way Attenuation Coefficient K_1 in Clouds in dB/(km · g · m³)*

		Wavelength, cm			
Temperature, °C		0.9	1.24	1.8	3.2
Water cloud	20	0.647	0.311	0.128	0.0483
	10	0.681	0.406	0.179	0.0630
	0	0.99	0.532	0.267	0.0858
	−8	1.25	0.684	0.34 (extrapolated)	0.112 (extrapolated)
Ice cloud	0	8.74×10^{-3}	6.35×10^{-3}	4.36×10^{-3}	2.46×10^{-3}
	−10	2.93×10^{-3}	2.11×10^{-3}	1.46×10^{-3}	8.19×10^{-4}
	−20	2.0×10^{-3}	1.45×10^{-3}	1.0×10^{-3}	5.63×10^{-4}

*After Gunn and East.[20]

than water clouds of the same water content. The attenuation of microwaves by ice clouds can be neglected for all practical purposes.[10]

Attenuation by Rain. Ryde and Ryde[21] calculated the effects of rain on microwave propagation and showed that absorption and scattering effects of raindrops become more pronounced at the higher microwave frequencies, where the wavelength and the raindrop diameters are more nearly comparable. In the 10-cm band and at shorter wavelengths the effects are appreciable, but at wavelengths in excess of 10 cm the effects are greatly decreased. It is also known that suspended water droplets and rain have an absorption rate in excess of that of the combined oxygen and water-vapor absorption.[22]

In practice, it has been convenient to express rain attenuation as a function of the precipitation rate R, which depends on both the liquid-water content and the fall velocity of the drops, the latter in turn depending on the size of the drops.

Ryde[23] studied the attenuation of microwaves by rain and deduced, by using Laws and Parsons[24] distributions, that this attenuation in decibels per kilometer can be approximated by

$$K_R = \int_0^{r_0} [R(r)]^\alpha dr \qquad (23.18)$$

where K_R = total attenuation, dB
K = function of frequency[25]
$R(r)$ = rainfall rate along path r
r_0 = length of propagation path, km
α = function of frequency[10]

Medhurst[26] shows that $\alpha = 1$ is a good assumption in many cases. The path loss per mile, according to Ryde, for the three carrier frequency bands of 4, 6, and 11 GHz, is shown in Fig. 23.1.

The greatest uncertainty in predictions of attenuation caused by rainfall, when theoretical formulas are used as a basis for calculation, is the extremely limited knowledge of drop-size distribution in rains of varying rates of fall under differing climatic and weather conditions. There is little evidence that a rain with a known rate of fall has a unique drop-size distribution, although studies on this problem seem to indicate that a certain most probable drop-size distribution can be attached to a rain of a given rate of fall.[27] Results of this study are shown in Table 23.2, which gives the percentage of total volume of rainfall occupied by raindrops of different diameters (centimeters) and varying rainfall rates (millimeters per hour). On the basis of these results, the absorption cross section of raindrops of different sizes is shown in Table 23.3. This table gives the decibel attenuation per kilometer in rains of different rates of fall for radio wavelengths between 0.3 and 10 cm.

Since the total-attenuation cross section[28] depends on the temperature (because of its effects on the dielectric properties of water), it is important to evaluate the attenuation of rains whose drops are at different temperatures from those in the preceding tables. Table 23.4 contains the necessary data relative to the change of attenuation with temperature and is to be used with Table 23.3.

To determine total attenuation caused by rainfall through a particular storm, something must be known about the nature of the storm itself and, consequently, about how its rainfall rates and drop sizes are distributed in three dimensions.

A systematic vertical variation of R, decaying with height above a measured surface value, seems to be appropriate in rainfall of a widespread (continuous) nature.[29] Widespread rainfall is usually triggered by a relatively large-scale mechanism, such as a frontal or monsoon situation. A vertical variation of R of the form
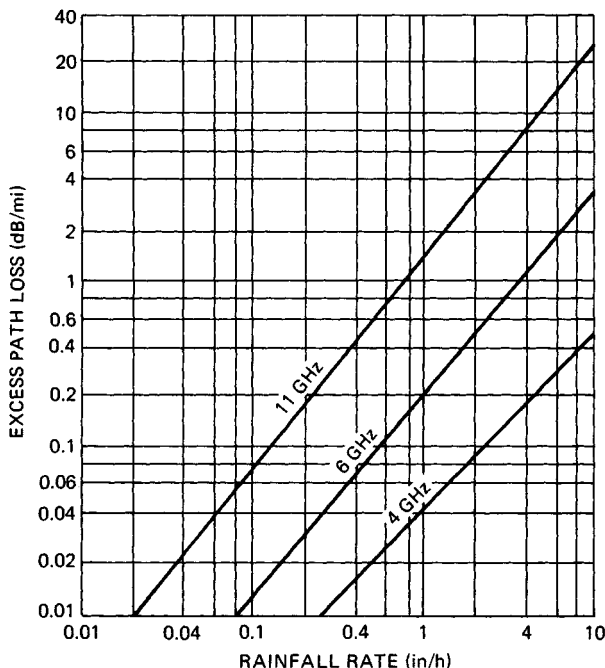
FIG. 23.1 Theoretical rain attenuation versus rainfall rate.

$$R = R_0 e^{-dh^2} \qquad (23.19)$$

can be assumed to be appropriate under continuous-rainfall conditions.[29] In Eq. (23.19), R_0 is the surface rainfall rate, h is the height above the earth's surface, and d is a constant, equal to about 0.2.

Convective-type precipitation, however, shows a quite different nature. The presence of the *virga* (precipitation aloft but evaporating before reaching the surface) associated with so many shower-type clouds indicates that Eq. (23.19) is not especially representative of shower rainfall. Dennis[30] has done considerable work in examining rainfall determinations in shower-type activity. His observations show that the reflectivity factor $Z(mm^6/m^3)$ of an element of a vertical slice taken through a spherical shower cell is well represented by a regression line of the form

$$Z = c_1(r_0 - r)^{c_2} \qquad (23.20)$$

In Eq. (23.20), r is the distance from the center of the cell of radius r_0, and c_1 and c_2 are positive constants.

Attenuation by Hail. Ryde[23] concluded that the attenuation caused by hail is one-hundredth of that caused by rain, that ice-crystal clouds cause no sensible attenuation, and that snow produces very small attenuation even at the excessive rate of fall of 5 in/h. However, the scattering by spheres surrounded by a concentric film of different dielectric constant does not give the same effect that

TABLE 23.2 Drop-Size Distribution*

Drop diameter D, cm	Precipitation rate p, mm/h							
	0.25	1.25	2.5	12.5	25	50	100	150
	Percentage of a given volume containing drops of diameter D							
0.05	28.0	10.9	7.3	2.6	1.7	1.2	1.0	1.0
0.10	50.1	37.1	27.8	11.5	7.6	5.4	4.6	4.1
0.15	18.2	31.3	32.8	24.5	18.4	12.5	8.8	7.6
0.20	3.0	13.5	19.0	25.4	23.9	19.9	13.9	11.7
0.25	0.7	4.9	7.9	17.3	19.9	20.9	17.1	13.9
0.30	.....	1.5	3.3	10.1	12.8	15.6	18.4	17.7
0.35	.....	0.6	1.1	4.3	8.2	10.9	15.0	16.1
0.40	.....	0.2	0.6	2.3	3.5	6.7	9.0	11.9
0.45	.....	.....	0.2	1.2	2.1	3.3	5.8	7.7
0.50	.....	.....	.....	0.6	1.1	1.8	3.0	3.6
0.55	.....	.....	.....	0.2	0.5	1.1	1.7	2.2
0.60	.....	.....	.....	.....	0.2	0.5	1.0	1.2
0.65	.....	.....	.....	.....	.....	0.2	0.7	1.0
0.70	.....	.....	.....	.....	.....	.....	.....	0.3

*From Burrows and Attwood.[27]

TABLE 23.3 Attenuation in Decibels per Kilometer for Different Rates of Rain Precipitation at Temperature 18°C*

Precipitation rate p, mm/h	Wavelength λ, cm								
	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 1.0$	$\lambda = 1.25$	$\lambda = 3.0$	$\lambda = 3.2$	$\lambda = 10$
0.25	0.305	0.230	0.160	0.106	0.037	0.0215	0.00224	0.0019	0.0000997
1.25	1.15	0.929	0.720	0.549	0.228	0.136	0.0161	0.0117	0.000416
2.5	1.98	1.66	1.34	1.08	0.492	0.298	0.0388	0.0317	0.000785
12.5	6.72	6.04	5.36	4.72	2.73	1.77	0.285	0.238	0.00364
25.0	11.3	10.4	9.49	8.59	5.47	3.72	0.656	0.555	0.00728
50	19.2	17.9	16.6	15.3	10.7	7.67	1.46	1.26	0.0149
100	33.3	31.1	29.0	27.0	20.0	15.3	3.24	2.80	0.0311
150	46.0	43.7	40.5	37.9	28.8	22.8	4.97	4.39	0.0481

*From Burrows and Attwood.[27]

Ryde's results for dry particles would indicate.[23,31] For example, when one-tenth of the radius of an ice sphere of radius 0.2 cm melts, the scattering of 10-cm radiation is approximately 90 percent of the value that would be scattered by an all-water drop.

At wavelengths of 1 and 3 cm with $2a = 0.126$ (a = radius of drop), Kerker, Langleben, and Gunn[31] found that particles attained total-attenuation cross sections corresponding to all-melted particles when less than 10 percent of the ice particles was melted. When the melted mass reached about 10 to 20 percent, the attenuation was about twice that of a completely melted particle. These calculations show that the attenuation in the melting of ice immediately under the 0°C isotherm can be substantially larger than in the snow region just above and, under

TABLE 23.4   Correction Factor (Multiplicative) for Rainfall Attenuation*

Precipitation rate p, mm/h	λ, cm	0°C	10°C	18°C	30°C	40°C
0.25	0.5	0.85	0.95	1.0	1.02	0.99
	1.25	0.95	1.00	1.0	0.90	0.81
	3.2	1.21	1.10	1.0	0.79	0.55
	10.0	2.01	1.40	1.0	0.70	0.59
2.5	0.5	0.87	0.95	1.0	1.03	1.01
	1.25	0.85	0.99	1.0	0.92	0.80
	3.2	0.82	1.01	1.0	0.82	0.64
	10.0	2.02	1.40	1.0	0.70	0.59
12.5	0.5	0.90	0.96	1.0	1.02	1.00
	1.25	0.83	0.96	1.0	0.93	0.81
	3.2	0.64	0.88	1.0	0.90	0.70
	10.0	2.03	1.40	1.0	0.70	0.59
50.0	0.5	0.94	0.98	1.0	1.01	1.00
	1.25	0.84	0.95	1.0	0.95	0.83
	3.2	0.62	0.87	1.0	0.99	0.81
	10.0	2.01	1.40	1.0	0.70	0.58
150	0.5	0.96	0.98	1.0	1.01	1.00
	1.25	0.86	0.96	1.0	0.97	0.87
	3.2	0.66	0.88	1.0	1.03	0.89
	10.0	2.00	1.40	1.0	0.70	0.58

*From Burrows and Attwood.[27]

some circumstances, greater than in the rain below the melting level. Further melting cannot lead to much further enhancement, apparently, and may lead to a lessening of the reflectivity of the particle by bringing it to sphericity or by breaking up the particle. Melting of ice particles produces enhanced backscatter, and this effect gives rise to the radar-observed *bright band* near the 0°C isotherm.

Attenuation by Fog.   The characteristic feature of a fog is the reduction in visibility. Visibility is defined as the greatest distance in a given direction at which it is just possible to see and identify with the unaided eye (1) in the daytime a prominent dark object against the sky at the horizon and (2) at night a known, preferably unfocused, moderately intense light source.[32]

Although the visibility depends upon both drop size and number of drops and not entirely upon the liquid-water content, in practice the visibility is an approximation of the liquid-water content and therefore may be used to estimate radiowave attenuation.[33] On the basis of Ryde's work, Saxton and Hopkins[34] give the figures in Table 23.5 for the attenuation in a fog or clouds at 0°C temperature. The attenuation varies with the temperature because the dielectric constant of water varies with temperature; therefore, at 15 and 25°C the figures in Table 23.5 should be multiplied by 0.6 and 0.4, respectively. It is immediately noted that cloud or fog attenuation is an order of magnitude greater at 3.2 cm than at 10 cm. Nearly another order-of-magnitude increase occurs between 3.2 and 1.25 cm.

Range and Velocity Ambiguities.   The unambiguous doppler frequency or Nyquist frequency for a fixed pulse-repetition-frequency (PRF) radar is given by

$$\Delta f = \pm \text{ PRF}/2 \qquad (23.21)$$

TABLE 23.5   Attenuation Caused by Clouds or Fog Temperature = 0°C*

Visibility, m	Attenuation, dB/km		
	$\lambda = 1.25$ cm	$\lambda = 3.2$ cm	$\lambda = 10$ cm
30	1.25	0.20	0.02
90	0.25	0.04	0.004
300	0.045	0.007	0.001

*From Saxton and Hopkins.[34]

where PRF is the pulse repetition frequency. The unambiguous range interval is given by

$$\Delta r = \frac{c}{2\text{PRF}} \tag{23.22}$$

and the product $\Delta f \Delta r$ is simply

$$\Delta f \Delta r = \frac{c}{2} \tag{23.23}$$

Since the doppler shift f and the target radial velocity v are linearly related by the expression

$$v = \frac{\lambda}{2} f \tag{23.24}$$

it follows that the product of unambiguous velocity and unambiguous range is

$$\Delta v \Delta r = \frac{\lambda c}{4} \tag{23.25}$$

and is maximized by maximizing λ, the transmitted wavelength.

Ground Clutter Effects.   Many meteorological radar applications call for the detection of precipitation echoes in the presence of ground clutter. Airborne weather radars during takeoff or landing are particularly susceptible. Another application, in which ground clutter is serious, relates to the detection of low-level wind shear.

While ground clutter cannot be eliminated, its effects can be mitigated through careful design. The most straightforward approach is to use antennas with low sidelobes, particularly in elevation. A second approach is through the use of shorter wavelengths. Shorter wavelengths result in improved signal-to-clutter ratios owing to the fact that the backscattered weather signal power is inversely proportional to λ^4 while the ground clutter return is only weakly dependent on wavelength. If one assumes that the clutter signal is wavelength-independent and the antenna beamwidth is fixed, Eq. (23.13) may be used to show that the weather-signal-power to clutter-power ratio is inversely proportional to λ^2.

Typical Weather Radar Designs. There is no universal weather radar system design that can serve all purposes. Airborne weather radars are constrained by size and weight limitations. Ground-based radars may be constrained by cost considerations. Severe storm warning radars require long range and high unambiguous velocity, and they must penetrate very heavy rain, thus dictating long wavelengths. Radars designed for studies of nonprecipitating clouds may use short wavelengths[35,36] (8 mm or even 3 mm) in order to achieve sufficient sensitivity to detect small cloud particles of the order of 100 μm and smaller. And FM-CW radars[37] have been used to obtain very-high-range resolution for detection of very thin layers in the clear air.

However, most meteorological radars are conventional pulsed or pulsed doppler systems. Ground-based radars used for severe storm research or warning will normally use S-band ($\approx$3 GHz) or C-band ($\approx$5.5 GHz) transmitters. Airborne storm avoidance radars will use either C-band or X-band ($\approx$10 GHz) transmitters.

A 1° beamwidth is commonly used for longer-range radars. Admittedly, this is somewhat arbitrary, but the choice of 1° is based upon several decades of experience. A 1° beam will provide resolution of 2 km at a range of 120 km. Because thunderstorms contain important spatial features, such as heavy precipitation shafts and updraft cores, with horizontal dimensions of the order 1 to 5 km, a 1° beam is reasonably well matched to the phenomena being observed. Shorter-range and airborne weather radars often employ beamwidths of between 2 and 3°.

Operational weather radars normally are capable of short- and long-pulse operation in the range of 0.5 μs to about 6 μs. Through pulse-width diversity, high resolution is obtained, usually at short range, while for long-range detection longer pulses provide increased sensitivity and tend to equalize the along-beam and cross-beam resolutions.

Equation (23.13) shows that the received power is directly proportional to the pulse width τ. The noise power N is conventionally given by

$$N = \kappa TB \qquad (23.26)$$

where κ = Boltzmann's constant, 1.38×10^{-23} W/(H$_z$ · K)
$\quad T$ = receiver noise temperature, K
$\quad B$ = receiver noise bandwidth

For a matched receiver

$$B \approx \frac{1}{\tau} \qquad (23.27)$$

The signal-to-noise ratio is therefore given by the proportionality

$$\frac{P_r}{N} \propto \frac{\tau}{\kappa TB} \approx \frac{\tau^2}{\kappa T} \qquad (23.28)$$

Thus, for distributed targets and with the pulse volume filled with scatterers, the signal-to-noise ratio for a single pulse is proportional to the pulse width squared. This assumes that the peak power is unchanged and that the average power increases linearly with τ. If the transmitter's average power is fixed, the signal-to-noise ratio will be proportional to τ.

PRFs for meteorological radars range from as low as several hundred s^{-1} for

long-range detection to several thousand s^{-1} for shorter-wavelength systems attempting to achieve high unambiguous velocities. Generally speaking, most meteorological doppler radars are operated in a single mode, compromising the radar's ability to unambiguously resolve either range or velocity. More recent designs, however, may use a dual pulse repetition period[38] (PRT) to resolve both range and velocity. Another approach[39] is to employ a transmitted-pulse sequence with random phases from pulse to pulse. Range ambiguities cannot be totally eliminated, but their effects can be significantly mitigated through these approaches.

To discuss design details of all types of meteorological radars is beyond the scope of this chapter. However, it will be useful to include some of the important characteristics of the NEXRAD radar, which illustrate the performance of a modern operational weather radar ca. 1989. Table 23.6 contains some of the more relevant NEXRAD design features.

TABLE 23.6   Some Relevant NEXRAD System Characteristics

Transmitted power (klystron)	700,000 W
Pulse width	1.6, 4.8 μs
Range (doppler mode)	230 km
Unambiguous velocity (doppler mode)	±50 m/s
Range (nondoppler mode)	460 km
Clutter rejection	50 dB
Beamwidth	1°
System sensitivity	(−8 dBZ at 50 km)

23.4 SIGNAL PROCESSING

It can be shown[8,11] that the received signal from meteorological targets is well represented by a narrowband gaussian process. This is a direct consequence of the fact that (1) the number of scatterers in the pulse volume is large ($>10^6$); (2) the pulse volume is large compared with the transmitted wavelength; (3) the pulse volume is filled with scatterers, causing all phases on the range from 0 to 2π to be returned; and (4) the particles are in motion with respect to one another due to turbulence, wind shear, and their varying fall speeds.

The superposition of the scattered electric fields from such a large number of particles (each with random phase) gives rise, through the central limit theorem, to a signal with gaussian statistics. Because the particles are in motion with respect to one another, there is also a doppler spread, often referred to as the variance of the doppler spectrum. Finally, since all the particles within the sample volume are moving with some mean or average radial velocity, there is a mean frequency of the doppler spectrum which is shifted from the transmitted frequency.

The power spectral density of a meteorological signal is depicted schematically in Fig. 23.2 and can be interpreted as follows. The received power is simply the integral under the curve and is given by

$$P_r = \int S(f)df = \int S(v)dv \tag{23.29}$$
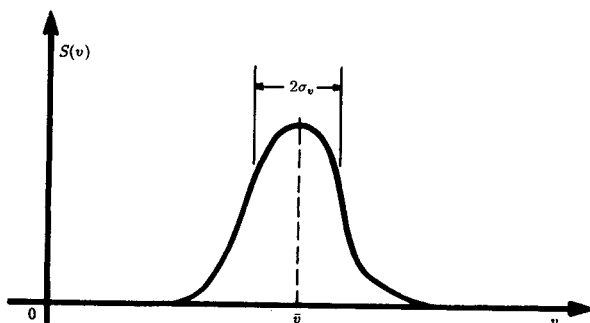
FIG. 23.2 The doppler spectrum. Received power, radial velocity, and spectrum width can be calculated and are directly related to meteorological variables.

where f and v are related by $f = (2/\lambda)v$.

The mean velocity is given by the first moment of the spectrum

$$\bar{v} \quad \frac{\displaystyle\int vS(v)dv}{\displaystyle\int S(v)dv} \tag{23.30}$$

The second central moment σ_v^2 is given by

$$\sigma_v^2 \quad \frac{\displaystyle\int (v - \bar{v})^2 S(v)dv}{\displaystyle\int S(v)dv} \tag{23.31}$$

where σ_v is the velocity width. Radar meteorologists refer to σ_v^2 as the spectrum variance because of its computational equivalence to the variance of a continuously distributed random variable. In short, $S(v)$ is analogous to a probability density function for v. The term *spectrum variance* will be used to refer to σ_v^2, and the term *spectrum width* to refer to σ_v. It is clear, therefore, that the doppler spectrum contains the information necessary to measure important signal parameters.

In the most general case, quadrature phase detection is used to obtain the real and imaginary parts of the complex signal envelope.[8] These are usually digitized in a large number of range gates (≈ 1000) at the radar's pulse repetition frequency. The resultant complex time series in each gate can then be processed by using a fast Fourier transform (FFT) to obtain an estimate of the doppler spectrum from which the mean velocity and spectrum variance can be obtained.

A more efficient estimation technique is described by Rummler.[40] This estimator makes use of the fact that the complex autocorrelation function of the signal has the general form

$$R(x) = P_r \rho(x) e^{j\frac{4\pi\bar{v}}{\lambda}x} \tag{23.32}$$

where $\rho(x)$ is the correlation coefficient and x is a dummy variable.

It follows that $\bar{v}$, the mean velocity, is given by

$$\bar{v} = \frac{\lambda}{4\pi x}\text{arg } [R(x)] \qquad (23.33)$$

It can also be shown that

$$\sigma_v^2 \approx \frac{\lambda^2}{8\pi^2 x^2}\left[1 - \frac{R(x)}{R(o) - N}\right] \qquad (23.34)$$

where N is the noise power.

This estimator is widely used for mean-frequency estimation with doppler meteorological radars. The estimates are unbiased in the presence of noise when the doppler spectrum is symmetrical. Its greatest appeal, however, is due to its computational simplicity. For a pulsed radar, with a pulse repetition period (PRT) T, $R(T)$ is obtained from the simple expression[8]

$$R(T) = \frac{1}{N}\sum_{(k=0)}^{(N-1)} s_{k+1}s_k^* \qquad (23.35)$$

where the s_k are the complex signal samples (sampled at the radar PRT) in a given range gate and s_k^* is the complex conjugate. It is clear that this algorithm requires only N complex multiplications for a time series of N samples while the FFT requires $N \log_2 N$. This *pulse-pair algorithm*, as it is often called, therefore not only is an excellent estimation technique but is less complex and costly than comparable FFT processors. In addition, FFT estimates of mean velocity and spectrum width are biased by receiver noise. If the FFT approach is used, the bias due to noise can be removed by estimating the noise threshold in the spectral domain and truncating the derived spectrum.[41]

For most applications, the pulse-pair processor has become the technique of choice. However, in some research applications it remains advantageous to have access to the full doppler spectrum. Very fast and programmable digital signal-processing chips make it possible for radar meteorologists to have their cake and eat it too. Flexibility due to programmability permits tailoring of the processor's characteristics to the application from day to day or even beam to beam and range gate to range gate. Until recently, most pulse-pair or FFT processors for meteorological radars have been hard-wired and therefore inflexible.

Measurement Accuracy. Because the received signals are sample functions from gaussian random processes, the doppler spectrum and its moments cannot be measured exactly in any finite period of time. Consequently, all measurements will be somewhat in error, with the error being a function of the properties of the atmosphere, the radar wavelength, and the time allocated to the measurement.

The theoretical development of signal estimator statistics is found in Denenberg, Serafin, and Peach[42] for the FFT technique. Doviak and Zrnić[11] cover the subject quite completely. Following are some useful expressions for the mean square error of mean power and mean velocity estimates.

Power Estimation. It is well known that for a gaussian process,[43] using square-law signal detection, samples of the mean power P_r of the process are exponentially distributed with variance P_r^2. Given a time T_0 allocated to the measurement and a signal bandwidth σ_f(Hz), there will be approximately $\sigma_f T_0$ independent samples of the square of the signal envelope. It follows, therefore, that an estimate $\hat{P}_r$ of the mean power for this process will have a variance or mean square error given by

$$\text{var}(\hat{P}_r) \approx \frac{P_r^2}{\sigma_f T_0} \tag{23.36}$$

Substituting for σ_f from the expression $\sigma_f = 2\sigma_v/\lambda$, where σ_v is the width of the doppler spectrum, Eq. (23.36) becomes

$$\text{var}(\hat{P}_r) \approx \frac{\lambda P_r^2}{2\sigma_v T_0} \tag{23.37}$$

This expression is valid for high signal-to-noise cases.

Velocity Estimation. Denenberg, Serafin, and Peach[42] give the following expression for the variance of mean-frequency estimates of the doppler spectrum

$$\text{var}(\hat{f}) = \frac{1}{P_r^2 T_0}\int f^2 S^2(f + \overline{f})df \tag{23.38}$$

This is an interesting result, showing that the variance of the estimate $\hat{f}$ is a function only of the shape of the doppler spectrum and the integration time T_0. If the spectrum has a gaussian shape, with variance σ_f^2, Eq. (23.38) becomes

$$\text{var}(\hat{f}) = \frac{\sigma_f}{4\sqrt{\pi}T_0} \tag{23.39}$$

Noting that $\text{var}(\hat{v}) = (\lambda/2)^2 \text{ var}(\hat{f})$, we can write

$$\text{var}(\hat{v}) = \frac{\lambda\sigma_v}{8\sqrt{\pi}T_0} \tag{23.40}$$

If we multiply numerator and denominator by σ_v, Eq. (23.40) becomes

$$\text{var}(\hat{v}) = \frac{\lambda\sigma_v^2}{8\sqrt{\pi}\sigma_v T_0} = \frac{\sigma_v^2}{4\sqrt{\pi}\sigma_f T_0} \tag{23.41}$$

Thus, it is seen that the variance of the mean velocity estimate $\hat{v}$ is directly proportional to the variance of the doppler spectrum and inversely proportional to the number of independent samples. Note also that var ($\hat{v}$) is proportional to λ, indicating that, for the same processing time T_0 and for the same σ_v, the variance of the estimate can be reduced by reducing the wavelength, which increases the number of independent samples.

Equations (23.38), (23.39), (23.40), and (23.41) are applicable in high signal-

to-noise-ratio cases. Zrnić[44] gives the following expression for the variance of the mean-frequency estimate $\hat{f}$ for the pulse-pair estimation technique and a gaussian-shaped spectrum

$$\text{var}(\hat{f}) = \frac{1}{8\pi^2 T_0 \rho^2(T)T}\left\{2\pi^{3/2}\sigma_f T + \frac{N^2}{S^2} + 2\frac{N}{S}[1 - \rho(2T)]\right\} \qquad (23.42)$$

where ρ is the correlation coefficient and N/S is the noise-to-signal ratio. Equation (23.42) applies to a single PRF with interpulse period T and assumes that all pulses in the interval T_0 are used in the estimation algorithm. It reduces exactly to Eq. (23.39) for large S/N and for narrow spectra, i.e., $\rho(T) \approx 1$. The reader is referred to Zrnić[44] for further details regarding the estimation of other moments of the doppler spectrum.

Processor Implementations. In nondoppler radars it is common to use log-video receivers along with sensitivity time control (STC) for inverse r^2 correction in order to achieve the widest dynamic range possible. For signal power estimation, the log-video signal is digitized and averaged or, in the most rudimentary of systems, may be used to modulate an analog PPI or other type of radarscope directly. Most modern meteorological radars, however, use digital averaging along with digital color displays for added quantitative precision. Note that when the logarithm is averaged, the estimate will be biased downward by as much as 2.5 dB.[45] This bias must be removed in order to accurately estimate the received signal power.

For doppler radars it has been common to use both linear and logarithmic receivers, with the log channel used for reflectivity estimation and the linear channel for doppler parameter estimation. This approach, however, often results in saturation of the linear channel and therefore some distortion of the doppler spectrum.[46]

Most modern designs now attempt to maintain linearity in the receiver through the use of a dynamic automatic gain control (AGC), whereby the receiver gain is adjusted from range gate to range gate through the use of rapidly switched attenuators. The estimate needed to select the proper attenuator may come from an independent log channel or may be based upon a short segment of the signal. Another approach[47] is to delay the signal for a period of the order of a microsecond while an estimate of signal strength can be made and the proper attenuator setting can be established. Clearly, such rapid switching in the receiver requires careful design in order to avoid the effects of switching transients. An approach that avoids transient effects is to use parallel IF strips, each with moderate dynamic range and fixed gains, and to sample the signal in the channel that is best matched to the signal strength.

In all these approaches, it is possible to achieve wide linear dynamic range of the order of 80 dB or greater and to use floating-point digital arithmetic. The reflectivity, mean doppler velocity, and spectrum width can all be estimated digitally from the floating-point linear channel samples.

23.5 OPERATIONAL APPLICATIONS

As has been demonstrated, meteorological radars measure backscattered power and radial velocity parameters. The challenge to the radar meteorolo-

gist is to translate these measurements, their spatial distributions, and their temporal evolution into quantitative assessments of the weather. The level of sophistication used in interpretation varies broadly, ranging from human interpretation of rudimentary gray-scale displays to computer-based algorithms and modern color-enhanced displays to assist human interpreters. Expert system approaches[48] that attempt to reproduce human interpretive logical processes can be employed effectively. Baynton et al.,[49] Wilson and Roesli,[50] and Serafin[1] all show how modern meteorological radars are used for forecasting the weather. The degree to which automation can be applied is evident in the NEXRAD radar system design, where the meteorological products shown in Table 23.7 will be automated.[51]

TABLE 23.7 NEXRAD Automated Products

Doppler radar data archive of storm phenomena
Precipitation analysis
Wind analysis
Tornado analysis
Fine-line analysis
Tropical cyclone analysis
Mesocyclone analysis
Thunderstorm analysis
Turbulence analysis
Icing analysis
Hail analysis
Freezing-melting analysis
Interpretive techniques
Multiple-radar mosaics

Precipitation Measurement. Among the more important parameters to be measured is rainfall, having significance to a number of water resource management problems related to agriculture, fresh-water supplies, storm drainage, and warnings of potential flooding.

The rainfall rate can be empirically related to the reflectivity factor[12] by an expression of the form

$$Z = aR^b \qquad (23.43)$$

where a and b are constants and R is the rainfall rate, usually in millimeters per hour. Battan[10] devotes three full pages of his book to the listing of dozens of Z-R relationships derived by investigators at various locations throughout the world, for various weather conditions and in all seasons of the year. The fact that no universal expression can be applied to all weather situations is not surprising when one notes that rainfall drop-size distributions are highly variable. For many conditions,[10] the drop-size distribution can be represented by an exponential function

$$N(D) = N_0 e^{-\Lambda D} \qquad (23.44)$$

where N_0 and Λ are constants. If $N(D)$ is known, the reflectivity factor can be calculated from Eq. (23.9). By using the terminal-fall speed data of Gunn and Kinzer,[52] the rainfall rate can also be obtained and Z directly related to R.

Clearly, a single-wavelength, single-polarization radar can measure only a single parameter Z and must assume Rayleigh scattering. Since the rainfall rate depends upon two parameters, N_0 and Λ, it is not surprising that Eq. (23.43) is nonuniversal. Despite this fact, Battan[10] lists four expressions as being "fairly typical" for the following four types of rain:

$$\text{Stratiform rain}^{53} \qquad Z = 200 \, R^{1.6} \qquad (23.45)$$

$$\text{Orographic rain}^{54} \qquad Z = 31 \, R^{1.71} \qquad (23.46)$$

$$\text{Thunderstorm rain}^{55} \qquad Z = 486 \, R^{1.37} \qquad (23.47)$$

$$\text{Snow}^{56} \qquad Z = 2000 \, R^2 \qquad (23.48)$$

Stratiform refers to widespread, relatively uniform rain. *Orographic* rain is precipitation that is induced or influenced by hills or mountains. In each of the above expressions, Z is in mm^6/m^3 and R is in mm/h. In Eq. (23.48), R is the precipitation rate of the melted snow.

For a more complete treatment of this topic, the reader is referred to Battan.[10] Wilson and Brandes[57] give a comprehensive treatment of how radar and rain-gauge data can be used to complement one another in measurements of precipitation over large areas. Bridges and Feldman[58] discuss how two independent measurements (reflectivity factor and attenuation) can be used to obtain both parameters of the drop-size distribution and therefore precisely determine the rainfall rate. Seliga and Bringi[59] show how the measurement of Z at horizontal and vertical polarization also can produce two independent measurements and therefore provide more accurate rainfall rate measurements. Zawadzki[60] argues, however, that other factors contribute far more to the variability of precipitation rate than does the drop-size distribution. He states, therefore, that dual-parameter estimation techniques are not likely to be successful in many cases. Wilson and Brandes[57] state that cumulative precipitation measurements with radar, in storm situations, can be expected to be accurate to a factor of 2 for 75 percent of the time. Accuracies over large areas can be improved to about 30 percent with the addition of a surface rain-gauge network. It is this author's opinion that no single topic in radar meteorology has received more attention than rainfall rate measurement. Although useful empirical expressions have evolved, a completely satisfactory approach remains to be discovered.

Severe Storm Warning. One of the primary purposes of weather radars is to provide timely warnings of severe weather phenomena such as tornadoes, damaging winds, and flash floods. Long-term forecasting of the precise location and level of severity of these phenomena, through numerical weather prediction techniques, is beyond the state of the art. Operational radars, however, can detect these phenomena and provide warnings (of up to 30 min) of approaching severe events; they can also detect the rotating mesocyclones in severe storms that are precursors to the development of tornadoes at the earth's surface.[61]

Tornado Detection. A single doppler radar can only measure the radial component of the vector wind field. Hence, exact measurements of vector winds at a point are generally not possible. However, rotating winds or vortices can be detected and their intensities measured by simply measuring the change in radial
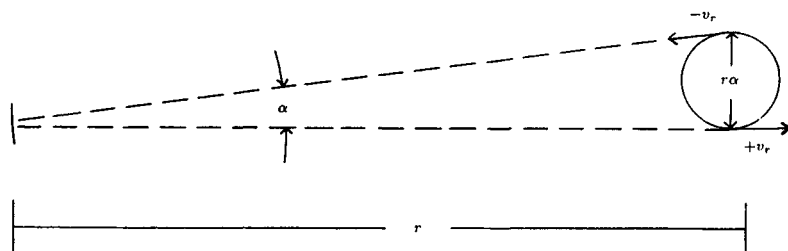
FIG. 23.3 Measurement of rotation or azimuthal shear in a mesocyclone. The azimuthal shear is given by $\Delta v/\Delta x = 2v_r/r\alpha$.

velocity with azimuth angle as shown in Fig. 23.3. The radar scans in azimuth and detects a couplet in radial velocity at constant range. The azimuthal shear is given simply by the expression

$$\frac{dv_r}{dx} \approx \frac{2v_r}{r\alpha} \tag{23.49}$$

where x is in the direction orthogonal to the radius r and d is the angle subtended by the circulation at range r.

Because mesocyclones, which spawn tornadoes, can be many kilometers in diameter, radars with 1° beams have the spatial resolution to detect mesocyclones at ranges in excess of 60 km. It should be clear that any mean translational motion would change the absolute values of the measured radial velocities but would not affect the shear measurement. Armstrong and Donaldson[62] were the first to use shear for severe storm detection. Azimuthal shear values of the order of 10^{-2} s^{-1} or greater and with vertical extent greater than the diameter of the mesocyclone are deemed necessary for a tornado to occur.[63]

Detection of the tornado vortex itself is not generally possible, since its horizontal extent may be only a few hundred meters. Detection of the radial shear, therefore, is not possible unless the tornado is close enough to the radar to be resolved by the beamwidth. In cases where the tornado falls entirely within the beam, the doppler spectral width[64] may be used to estimate tornadic intensity. In some cases, both a mesocyclone and its incipient tornado can be detected. Wilson and Roesli[50] show an excellent example of a tornado vortex signature (TVS) embedded within a larger mesocyclone.

Microbursts. Fujita and Caracena[65] first identified the microburst phenomenon as the cause of an airliner crash that took place in 1975. The microburst and its effects on an aircraft during takeoff or landing are depicted in Fig. 23.4. The microburst is simply a small-scale, short-duration downdraft emanating from a convective storm. This "burst" of air spreads out radially as it strikes the ground, forming a ring of diverging air about 0.3 to 1 km deep and of the order of 2 to 5 km in diameter. Aircraft, penetrating a microburst, experience first an increase in head wind and then a continuous, performance-robbing decrease in head wind, which can cause the plane to crash if encountered shortly before touchdown or just as the aircraft is taking off. More complete descriptions of microbursts and their effects on aviation safety are given by Fujita[66,67] and McCarthy and Serafin.[68]

Strong Downdraft

Increasing Headwind   Increasing Tailwind

Outflow   Outflow

(1–2 mi)
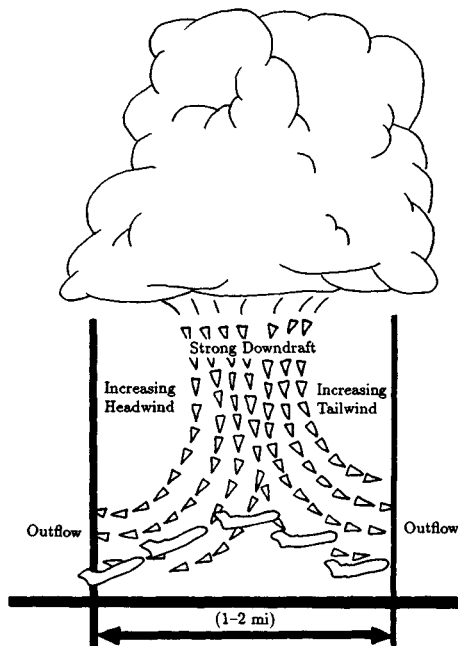
FIG. 23.4 Artist's depiction of a microburst and its effect on an aircraft during takeoff. The loss of airspeed near to the ground can be extremely hazardous.

Microburst detection, like tornado detection, is accomplished by estimating shear. However, in the case of the microburst, it is the radial shear of the radial velocity that is typically measured. Human interpretation of microburst signatures in color-enhanced radial velocity displays is easily accomplished with trained observers.[50] Radial velocity differences of 10 to 50 m/s are observed in microbursts. A radial velocity difference of 25 m/s over the length of a jet runway ($\approx$3 km) is of serious concern.

One principal problem concerning microbursts is their short lifetimes, which are of order 15 min. The duration of peak intensity is only 1 or 2 min. The Classify, Locate, and Avoid Wind Shear (CLAWS) project[69] in 1984 clearly demonstrated that a 2-min advance warning using doppler radar and human interpreters can be achieved. The use of doppler radars operationally, however, will require completely automated detection algorithms. A second major problem is ground clutter. Since the phenomenon occurs near the ground and oftentimes in very light or no precipitation, ground clutter mitigation is necessary.

C band seems to be the preferred operational frequency for several reasons. First, a C-band antenna will be physically smaller than an S-band antenna for the same beamwidth, an important consideration for use near airports. Second, since long-range detection is not of importance, attenuation effects are not of primary concern. Third, C band offers improved signal-to-clutter performance. X band is not the frequency of choice owing to more serious range-velocity ambiguities and the more severe attenuation that can occur in very heavy rain. It is expected that deployment of a national network of doppler radars near airports will begin in the early 1990s.

Hail.   The NEXRAD radar will make use of a hail-detection algorithm similar to that discussed by Witt and Nelson.[70] This algorithm combines high reflectivity factor with echo height and upper-level radial velocity divergence to detect the occurrence of hail. Eventually, polarization diversity techniques may improve quantitative hail detection. Aydin, Seliga, and Balaji[71] propose a hail-detection technique using reflectivity measurements at orthogonal polarizations. This technique depends upon the fact that the ratio of horizontal to vertical reflectivity is unity (≈ 0 dB) when hail is present. This differs sharply from heavy rain, where this ratio can be as large as 6 dB. The combination of absolute reflectivity factor at horizontal polarization and ratio of reflectivities at horizontal and vertical polarizations (differential reflectivity) gives unique signatures for hail and heavy rain, each of which is characterized by high reflectivity factor. The difference in the differential reflectivity signatures is easily explained. Large raindrops assume pancakelike shapes as they fall and thus scatter back horizontally polarized electric fields more strongly than vertically polarized electric fields. Hailstones, while irregular in shape, appear to tumble while they fall and therefore exhibit no preferred orientation on average.

Wind Measurement.   Lhermitte and Atlas[72] were the first to show how a single doppler radar can be used to measure vertical profiles of horizontal wind. This technique can be used if the precipitation and the wind are uniform in the region scanned by the radar. The method depends upon an analysis of the radial velocity measured during a complete scan in azimuth with elevation angle fixed. At any slant range r, the diameter of the region scanned is $r \cos \alpha$, and the height of the measurement is $r \sin \alpha$, where α is the elevation angle (see Fig. 23.5). If β is the azimuth angle, V_h is the horizontal wind speed, and V_f is the fall speed of the particles, the radial velocity at range r is given by

$$V_r(\beta) = V_h \cos \beta \cos \alpha + V_f \sin \alpha \qquad (23.50)$$
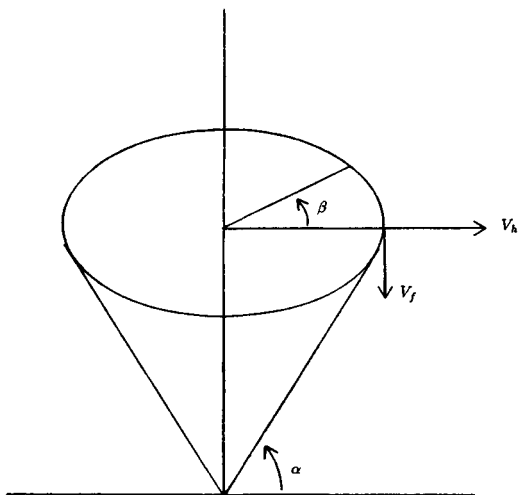
FIG. 23.5   Velocity-azimuth-display geometry for measuring horizontal wind with a single doppler radar. Measurement of the radial velocity for a complete azimuthal scan (β) permits measurement of horizontal winds.

A harmonic analysis can be used to obtain V_h, the horizontal wind speed, the wind direction, and V_f, the particle fall speed. The technique is referred to as the velocity-azimuth-display (VAD) technique. Browning and Wexler[73] later showed how the technique could be extended to measure other parameters of the wind field including divergence and deformation. Baynton et al.[49] show how the VAD can be applied in real time by using a color-enhanced radial velocity display.

   Thunderstorm Prediction.   Wilson and Schreiber[74] illustrate how modern meteorological doppler radar can be used to detect locations where new thunderstorm development is likely to occur. Modern radars have sufficient sensitivity to detect clear-air discontinuities in the lower 2 to 4 km of the atmosphere. Principally, this detection occurs in the summer months. The backscattering mechanism may be due to index-of-refraction inhomogeneities caused by turbulence in the lower layers and/or by insects. Wilson and Schreiber have found that about 90 percent of the thunderstorms that occur in the Front Range of the Rockies in the summertime develop over such boundaries. Since these boundaries can be detected before any clouds are present and because it is possible to infer the air mass convergence that is taking place along these boundaries through doppler measurements, more precise prediction of thunderstorm occurrence appears to be possible. From the radar designer's standpoint, such applications dictate the use of antennas with very low sidelobes and signal processors with significant ground clutter rejection capability. The NEXRAD radar system, with 50 or more dB of clutter rejection, is well suited to this eventual operational task.

23.6 RESEARCH APPLICATIONS

Operational meteorological radars are designed for reliability and simplicity of operation while providing the performance needed for operational applications. Research radars are considerably more complex, since cutting-edge research requires more detailed and more sensitive measurements of a multiplicity of variables simultaneously. In the research community, multiple-parameter radar studies, multiple-doppler radar network studies, and plans for airborne and spaceborne radars are all receiving considerable attention.

   Multiple-Parameter Radar.   It has been noted earlier that doppler radar provides a significant increase in the useful information that can be obtained from meteorological targets. The detection of hail, through the use of polarization diversity, adds additional information, and multiple wavelength provides yet another input related to the eventual interpretation of the size, water-phase state, and types of hydrometeors in all classes of clouds and precipitation. Very-short-wavelength radars are useful for probing newly developing clouds, while longer-wavelength radars are necessary for the study of severe storms. Researchers often need a wide range of these capabilities simultaneously. The capabilities desired of multiple-parameter meteorological radars are presented in the collection of papers edited by Hall.[6]

   From the radar engineering standpoint, the challenge is considerable, requiring radar designers to develop fully coherent, polarization-diverse, and wavelength-diverse radars. Figure 23.6 is a photograph of the S (10-cm)- and X (3-cm)-band polarization-diverse doppler radar operated by the National Center for Atmospheric Research (NCAR). The system permits simultaneous measurements of the reflectivity factor on 2 wavelengths—the doppler parame-

FIG. 23.6 The CP-2, multiple-parameter radar at the National Center for Atmospheric Research, Boulder, Colorado. (*Courtesy of the National Center for Atmospheric Research.*)

ters on a single wavelength, S band, and polarization-diverse measurements at both wavelengths. The antenna beams are matched with approximately 1° beamwidths. The peak transmitted power at S band is 1 MW and 50 kW at X band. The pulse widths are approximately 1 μs, and the PRF is typically 1000 s^{-1}. The system is characteristic of the technologies currently in place in the research community in this field.

Multiple Radars. A single doppler radar measures only a single radial component of velocity. Lhermitte[3] was among the first to describe how two or more doppler radars could be used, scanning together, to obtain the full three-dimensional air motion fields in precipitation. This pioneering work led the way toward the use of networks of doppler radars for studies of individual clouds and larger-scale cloud systems. For the first time, it became possible to examine the three-dimensional structure of vector air motion in precipitation. Figure 23.7 illustrates an air motion field obtained by multiple-doppler radar observations in an individual convective storm cell. Shown are the horizontal
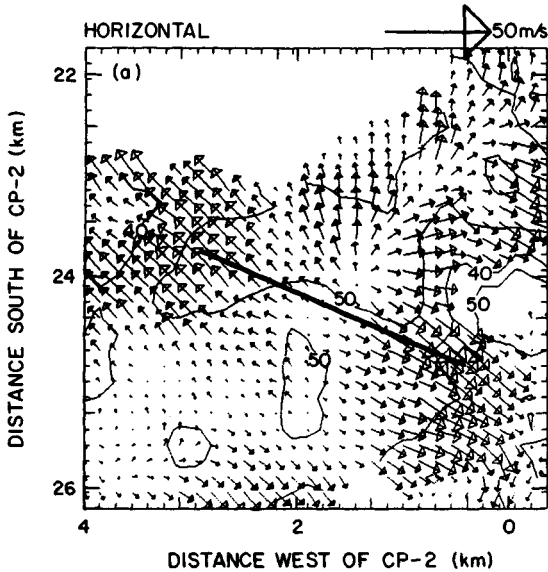
FIG. 23.7 Vector wind fields in a horizontal plane derived from dual-doppler radar observations of a summertime convective storm near Denver, Colorado. The divergent outflow is from a microburst. The dark, solid line is shown to indicate the length of a typical jet aircraft runway. (*Courtesy of the National Center for Atmospheric Research.*)

vector fields in a plane approximately 100 m above the surface. The phenomenon being measured is a low-level divergent outflow (or microburst). Figure 23.8 shows another example of air motion fields in a vertical plane orthogonal to an intense squall line in California.[75]

Rapid Scanning. The use of multiple-doppler radars has provided dramatic new information on the internal winds in large precipitating systems—information that can be obtained in no other way. Despite the power of this technique, the spatial resolution in the derived three-dimensional motion fields is generally not better than of the order of 2 km. The reasons for this are several. The finite beamwidth limits the resolution available at longer ranges. At shorter ranges, the large solid angle that must be scanned in order to cover all regions of a storm requires total scanning times of the order of 3 to 5 min even for ideally situated storms. This is a consequence of the on-target time necessary for accurate radial velocity measurements. Finally, the storm itself is evolving and moving during this measurement time.

Some research applications require faster scanning. These applications include the study of finer-scale storm features, interactions between the kinematics and hydrometeor growth processes in the storms, and studies of electric-charge separation in clouds. Brook and Krehbiel[76] were the first to discuss a very-rapid-scanning radar (although nondoppler) for effectively obtaining snapshots of convective storms. Keeler and Frush[77] discuss design considerations for a rapid-scanning doppler radar. Any rapid-scanning approach generally must encompass
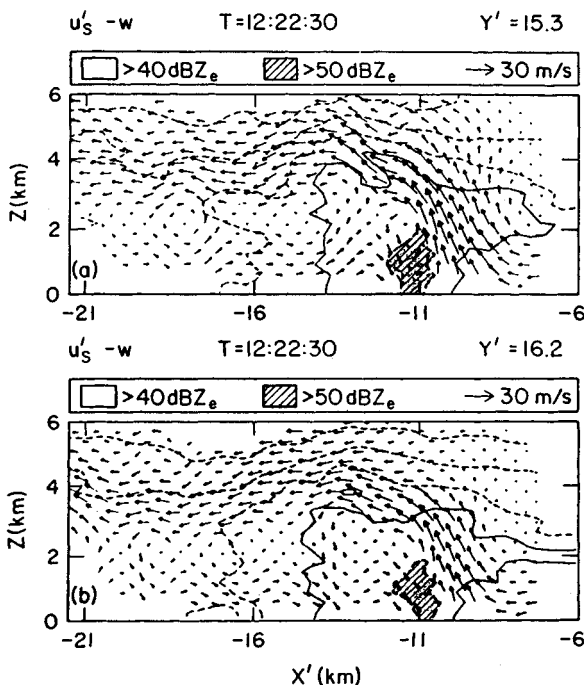
FIG. 23.8 Vertical cross sections of vector air motion in planes orthogonal to a California wintertime squall line.[75]

two features: (1) there must be relatively broadband transmissions to increase the independent samples available within the spatial resolution cell, thus reducing the dwell time; and (2) the antenna must be scanned—either very rapidly mechanically or electronically. An alternative approach might use several simultaneous beams and receivers.

Airborne and Space-Borne Radars. Hildebrand and Mueller[4] and Mueller and Hildebrand[5] have quantitatively demonstrated that it is possible for an airborne meteorological doppler radar to measure internal kinematic fields that are comparable to those obtained from ground-based systems. This powerful technique permits the use of a mobile platform, which therefore allows measurements over regions not accessible by ground-based systems. Moreover, the mobility of the aircraft permits longer-term observations of rapidly moving but long-lived storms and cloud systems. Figure 23.9 shows a photograph of the antenna of the airborne doppler radar mounted on the tail of the P-3 aircraft operated by the National Oceanic and Atmospheric Administration (NOAA). The antenna, covered by a cylindrical radome in flight, scans in range-height-indicator (RHI) mode, in vertical planes orthogonal to the aircraft fuselage. The aircraft is flown on orthogonal tracks in order to synthesize dual-doppler observations and therefore to obtain vector winds.

A point should be made here regarding the use of two doppler radars for measurements of three-dimensional winds. Since in principle two independent looks
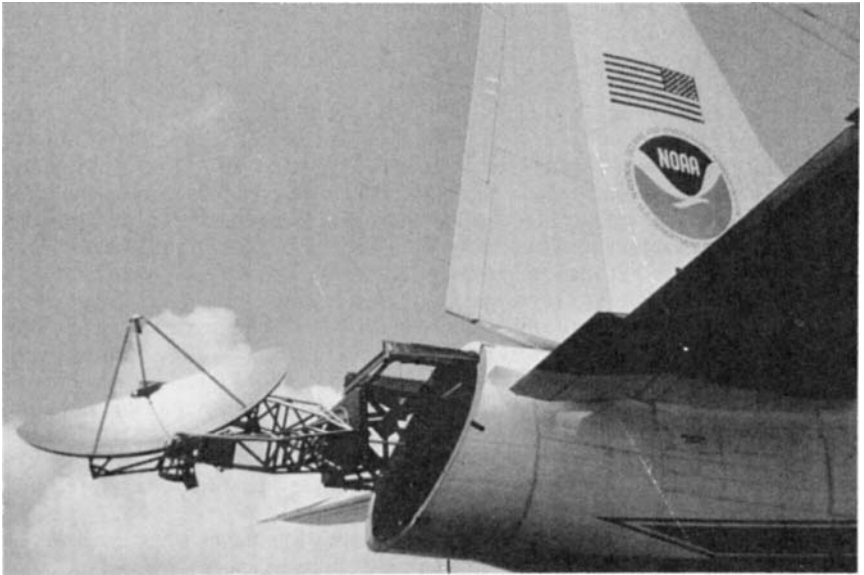
FIG. 23.9 Tail-mounted doppler radar antenna on the P-3 research and reconnaissance aircraft operated by the National Oceanic and Atmospheric Administration (NOAA).

can measure only two components of vector air motion, the assumption of mass continuity is invoked. The equation of continuity ($\nabla \cdot \overline{V} = 0$) is used to obtain the third-dimensional component, where $\overline{V}$ is the vector air motion. The vertical air motion is calculated from vertical integration of the continuity equation.

Walther, Frush, and Hildebrand[78] describe a next-generation airborne doppler radar system that consists of two antennas mounted in the tail—one pointed forward from the orthogonal plane by an angle of the order of 30° and one pointed rearward, also by about 30°. With such a system, each antenna scans in a conical surface—one cone pointing forward, one rearward—thus permitting synthesis of a dual-doppler radar system along the aircraft track. Because the aircraft need not fly orthogonal tracks, the time required for measurements of cloud systems is dramatically reduced. Moreover, severe storms (which could otherwise not be penetrated along an orthogonal track) can be observed fully by an aircraft outside the regions of severe weather.

Among the more significant challenges facing researchers today is the need to make global measurements of precipitation. Understanding of the global climate requires that quantitative measurements of precipitation be made throughout the world, particularly in the tropics and over the oceans. Satellite observations appear to offer the only practical mechanism for obtaining these measurements. Meneghini and Atlas[79] describe a concept for a dual-wavelength radar for precipitation measurements from space.

Clear-Air Radars.   Another form of doppler radar that has become popular in the research community is the so-called wind profiler. Wind profilers usually

take the form of VHF and UHF fixed-beam systems, pointing vertically and at angles approximately 15° from the zenith. Such radars[7] can make doppler measurements throughout the range of altitudes from a few hundred meters to 15 km above the surface, depending upon the wavelength selected and the power-aperture product available. Very powerful radars of this type are referred to as Mesosphere, Stratosphere, Troposphere (MST) radars because of their ability to make measurements throughout most of these atmospheric regions. Powerful MST radars are operated at many laboratories around the world. Major facilities are located at Kiruna, Sweden; the Massachusetts Institute of Technology in Cambridge, Massachusetts; Arecibo, Puerto Rico; Jicamarca, Peru; and at the University of Kyoto in Japan.

These clear-air radars receive energy backscattered from index-of-refraction inhomogeneities due to atmospheric turbulence. The antenna systems usually take the form of phased arrays. Transmitters are generally in the form of high-powered, fully coherent transmitting tubes. One exception is at the University of Kyoto, where the antenna-transmitter system consists of more than 400 radiating elements, each with its own solid-state transmitter. This approach allows for full electronic scanning of the beam. A network of 400-MHz wind profilers in the central United States is also expected to use solid-state transmitters, but electronic scanning will not be possible.

The meteorological community is excited about these devices because of their ability to measure winds continuously. This capability permits the observation of smaller-scale temporal and spatial wind-field features than can be obtained from the global 12-hourly rawindsonde (balloon) networks. These smaller-scale measurements are important for understanding local and regional weather and for effective forecasting on these scales.

It is important to recognize that two-beam systems can measure horizontal winds if the wind field is uniform and if vertical velocities are negligible. A three-beam system can measure all three velocity components if the wind is uniform. Four- and five-beam systems allow one to determine the quality of the measurements by detecting the presence of nonuniformity. Carbone, Strauch, and Heymsfield[80] and Strauch et al.[81] address the issue of wind measurement error in detail.

The reader is referred to the review paper by Röttger and Larsen[82] for a thorough treatment of wind-profiler technology.

Synthetic Aperture Radar and Pulse Compression. Metcalf and Holm[83] and Atlas and Moore[84] have considered the use of synthetic aperture radar (SAR) in order to obtain high-resolution measurements from mobile airborne or space-borne platforms. In general, both papers conclude that the cross-beam resolution possible is inherently limited by the decorrelation time of the targets due to their turbulent motion. Consequently, SAR offers little advantage over real aperture systems for meteorological applications from aircraft. However, space-borne systems can effectively use SAR because of the high speed of the orbiting spacecraft.

Pulse compression is not generally used for meteorological applications because peak power is not usually a limitation on system performance. Keeler and Frush,[77] however, point out that pulse compression can be of benefit in some rapid-scanning applications. In situations where signals are very weak (such as for MST applications), pulse compression is used to increase system sensitivity by increasing the average power of the system.

A note of caution is in order when considering pulse compression for meteorological radars. This relates to the matter of range sidelobes. Careful design is necessary to minimize these sidelobes, just as antenna sidelobes should be minimized, in order to mitigate the effects of interpretive errors caused by wide-dynamic-range distributed weather targets.

REFERENCES

1. Serafin, R. J.: New Nowcasting Opportunities Using Modern Meteorological Radar, *Proc. Mesoscale Analysis Forecast. Symp.*, pp. 35–41, European Space Agency, Paris, 1987.

2. McCarthy, J., J. Wilson, and T. T. Fujita: The Joint Airport Weather Studies (JAWS) Project, *Bull. Am. Meteorol. Soc.*, vol. 63, pp. 15–22, 1982.

3. Lhermitte, R. M.: Dual-Doppler Radar Observations of Convective Storm Circulations, *Preprints, 14th Conf. Radar Meteorol.*, pp. 139–144, American Meteorological Society, Boston, 1970.

4. Hildebrand, P., and C. Mueller: Evaluation of Meteorological Airborne Doppler Radar, Part I: Dual-Doppler Analyses of Air Motions, *J. Atmos. Ocean. Technol.*, vol. 2, pp. 362–380, 1985.

5. Mueller, C., and P. Hildebrand: Evaluation of Meteorological Airborne Doppler Radar, Part II: Triple-Doppler Analysis of Air Motions, *J. Atmos. Ocean. Technol.*, vol. 2, pp. 381–392, 1985.

6. Hall, M. (ed.): Special papers: Multiple Parameter Radar Measurements of Precipitation, *Radio Sci.*, vol. 19, 1984.

7. Strauch, R. G., D. A. Merritt, K. P. Moran, K. B. Earnshaw, and D. Van De Kamp: The Colorado Wind Profiling Network, *J. Atmos. Ocean. Technol.*, vol. 1, pp. 37–49, 1984.

8. Serafin, R. J., and R. Strauch: Meteorological Radar Signal Processing, in "Air Quality Meteorology and Atmospheric Ozone," American Society for Testing and Materials, Philadelphia, 1977, pp. 159–182.

9. Gray, G. R., R. J. Serafin, D. Atlas, R. E. Rinehart, and J. J. Boyajian: Real-Time Color Doppler Radar Display, *Bull. Am. Meteorol. Soc.*, vol. 56, pp. 580–588, 1975.

10. Battan, L. J.: "Radar Observation of the Atmosphere," University of Chicago Press, 1973.

11. Doviak, R. J., and D. S. Zrnić: "Doppler Radar and Weather Observations," Academic Press, Orlando, Fla., 1984.

12. Bean, B. R., E. J. Dutton, and B. D. Warner: Weather Effects on Radar, in Skolnik, M. (ed.): "Radar Handbook," McGraw-Hill Book Company, New York, 1970, pp. 24-1–24-40.

13. Mie, G.: Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen [Contribution to the optics of suspended media, specifically colloidal metal suspensions], *Ann. Phys.*, vol. 25, pp. 377–445, 1908.

14. Probert-Jones, J. R.: The Radar Equation in Meteorology, *Q. J. R. Meteorol. Soc.*, vol. 88, pp. 485–495, 1962.

15. Metcalf, J. I.: Airborne Weather Radar and Severe Weather Penetration, *Preprints, 19th Conf. Radar Meteorol.*, pp. 125–129, American Meteorological Society, Boston, 1980.

16. Allen, R. H., D. W. Burgess, and R. J. Donaldson, Jr.: Severe 5-cm Radar Attenuation of the Wichita Falls Storm by Intervening Precipitation, *Preprints, 19th Conf. Radar Meteorol.*, pp. 87–89, American Meteorological Society, Boston, 1980.

17. Eccles, P. J., and D. Atlas: A Dual-Wavelength Radar Hail Detector, *J. Appl. Meteorol.*, vol. 12, pp. 847–854, 1973.

18. Donaldson, R. J., Jr.: The Measurement of Cloud Liquid-Water Content by Radar, *J. Meteorol.*, vol. 12, pp. 238–244, 1955.

19. Weickmann, H. K., and H. J. aufm Kampe: Physical Properties of Cumulus Clouds, *J. Meteorol.*, vol. 10, pp. 204–221, 1953.

20. Gunn, K. L. S., and T. W. R. East: The Microwave Properties of Precipitation Particles, *Q. J. R. Meteorol. Soc.*, vol. 80, pp. 522–545, 1954.

21. Ryde, J. W., and D. Ryde: "Attenuation of Centimeter Waves by Rain, Hail, Fog, and Clouds," General Electric Company, Wembley, England, 1945.

22. Bean, B. R., and R. Abbott: Oxygen and Water Vapor Absorption of Radio Waves in the Atmosphere, *Geofis. Pura Appl.*, vol. 37, pp. 127–144, 1957.

23. Ryde, J. W.: The Attenuation and Radar Echoes Produced at Centimetre Wavelengths by Various Meteorological Phenomena, in "Meteorological Factors in Radio Wave Propagation," Physical Society, London, 1946, pp. 169–188.

24. Laws, J. O., and D. A. Parsons: The Relationship of Raindrop Size to Intensity, *Trans. Am. Geophys. Union*, 24th Annual Meeting, pp. 452–460, 1943.

25. Schelleng, J. C., C. R. Burrows, and E. B. Ferrell: Ultra-Short-Wave Propagation, *Proc. IRE*, vol. 21, pp. 427–463, 1933.

26. Medhurst, R. G.: Rainfall Attenuation of Centimeter Waves: Comparison of Theory and Measurement, *IEEE Trans.*, vol. AP-13, pp. 550–564, 1965.

27. Burrows, C. R., and S. S. Attwood: "Radio Wave Propagation, Consolidated Summary Technical Report of the Committee on Propagation, NDRC," Academic Press, New York, 1949, p. 219.

28. Humphreys, W. J.: "Physics of the Air," McGraw-Hill Book Company, New York, 1940, p. 82.

29. Atlas, D., and E. Kessler III: A Model Atmosphere for Widespread Precipitation, *Aeronaut. Eng. Rev.*, vol. 16, pp. 69–75, 1957.

30. Dennis, A. S.: Rainfall Determinations by Meteorological Satellite Radar, *Stanford Research Institute, SRI Rept.* 4080, 1963.

31. Kerker, M., M. P. Langleben, and K. L. S. Gunn: Scattering of Microwaves by a Melting Spherical Ice Particle, *J. Meteorol.*, vol. 8, p. 424, 1951.

32. "Glossary of Meteorology," vol. 3, American Meteorological Society, Boston, 1959, p. 613.

33. Best, A. C.: "Physics in Meteorology," Sir Isaac Pitman & Sons, Ltd., London, 1957.

34. Saxton, J. A., and H. G. Hopkins: Some Adverse Influences of Meteorological Factors on Marine Navigational Radar, *Proc. IEE (London)*, vol. 98, pt. III, p. 26, 1951.

35. Pasqualucci, F., B. W. Bartram, R. A. Kropfli, and W. R. Moninger: A Millimeter-Wavelength Dual-Polarization Doppler Radar for Cloud and Precipitation Studies, *J. Clim. Appl. Meteorol.*, vol. 22, pp. 758–765, 1983.

36. Lhermitte, R.: A 94-GHz Doppler Radar for Cloud Observations, *J. Atmos. Ocean. Technol.*, vol. 4, pp. 36–48, 1987.

37. Richter, J. H.: High-Resolution Tropospheric Radar Sounding, *Proc. Colloq. Spectra Meteorol. Variables, Radio Sci.*, vol. 4, pp. 1261–1268, 1969.

38. Tang Dazhang, S. G. Geotis, R. E. Passarelli, Jr., A. L. Hansen, and C. L. Frush: Evaluation of an Alternating PRF Method for Extending the Range of Unambiguous Doppler Velocity, *Preprints, 22d Conf. Radar Meteorol.*, pp. 523–527, American Meteorological Society, Boston, 1984.

39. Laird, B. G.: On Ambiguity Resolution by Random Phase Processing, *Preprints, 20th Conf. Radar Meteorol.*, p. 327, American Meteorological Society, Boston, 1981.

40. Rummler, W. D.: Introduction of a New Estimator for Velocity Spectral Parameters, *Tech. Memo. MM-68-4121-5*, Bell Telephone Laboratories, Whippany, N.J., 1968.

41. Hildebrand, P. H., and R. H. Sekhon: Objective Determination of the Noise Level in Doppler Spectra. *J. Appl. Meteorol.*, vol. 13, pp. 808–811, 1974.

42. Denenberg, J. N., R. J. Serafin, and L. C. Peach: Uncertainties in Coherent Measurement of the Mean Frequency and Variance of the Doppler Spectrum from Meteorological Echoes, *Preprints, 15th Conf. Radar Meteorol.*, pp. 216–221, American Meteorological Society, Boston, 1972.

43. Davenport, W. B., Jr., and W. L. Root: "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Publishing Company, New York, 1958.

44. Zrnić, D. S.: Estimation of Spectral Moments for Weather Echoes, *IEEE Trans.*, vol. GE-17, pp. 113–128, 1979.

45. Marshall, J. S., and W. Hitschfeld: The Interpretation of the Fluctuating Echo for Randomly Distributed Scatterers, pt. I, *Can. J. Phys.*, vol. 31, pp. 962–994, 1953.

46. Frush, C.: Doppler Signal Processing Using IF Limiting, *Preprints, 20th Conf. Radar Meteorol.*, pp. 332–337, American Meteorological Society, Boston, 1981.

47. Mueller, E. A., and E. J. Silha: Unique Features of the CHILL Radar System, *Preprints, 18th Conf. Radar Meteorol.*, pp. 381–386, American Meteorological Society, Boston, 1978.

48. Campbell, S. D., and S. H. Olson: Recognizing Low-Altitude Wind Shear Hazards from Doppler Weather Radar: An Artificial Intelligence Approach, *J. Atmos. Ocean. Technol.*, vol. 4, p. 518, 1987.

49. Baynton, H. W., R. J. Serafin, C. L. Frush, G. R. Gray, P. V. Hobbs, R. A. Houze, Jr., and J. D. Locatelli: Real-Time Wind Measurement in Extratropical Cyclones by Means of Doppler Radar, *J. Appl. Meteorol.*, vol. 16, pp. 1022–1028, 1977.

50. Wilson, J., and H. P. Roesli: Use of Doppler Radar and Radar Networks in Mesoscale Analysis and Forecasting, *ESA J.*, vol. 9, pp. 125–146, 1985.

51. Bonewitz, J. D.: The NEXRAD Program—An Overview, *Preprints, 20th Conf. Radar Meteorol.*, pp. 757–761, American Meteorological Society, Boston, 1981.

52. Gunn, R., and Kinzer, G. D.: The Terminal Velocity of Fall for Water Droplets in Stagnant Air, *J. Meteorol.*, vol. 6, pp. 243–248, 1949.

53. Marshall, J. S., and W. M. K. Palmer: The Distribution of Raindrops with Size, *J. Meteorol.*, vol. 4, pp. 186–192, 1948.

54. Blanchard, D. C.: Raindrop Size Distribution in Hawaiian Rains, *J. Meteorol.*, vol. 10, pp. 457–473, 1953.

55. Jones, D. M. A.: 3 cm and 10 cm Wavelength Radiation Backscatter from Rain, *Proc. Fifth Weather Radar Conf.*, pp. 281–285, American Meteorological Society, Boston, 1955.

56. Gunn, K. L. S., and J. S. Marshall: The Distribution with Size of Aggregate Snowflakes, *J. Meteorol.*, vol. 15, pp. 452–466, 1958.

57. Wilson, J. W., and E. A. Brandes: Radar Measurement of Rainfall—A Summary, *Bull. Am. Meteorol. Soc.*, vol. 60, pp. 1048–1058, 1979.

58. Bridges, J., and J. Feldman: An Attenuation Reflectivity Technique to Determine the Drop Size Distribution of Water Clouds and Rain, *J. Appl. Meteorol.*, vol. 5, pp. 349–357, 1966.

59. Seliga, T. A., and V. N. Bringi: Potential Use of Radar Differential Reflectivity Measurements at Orthogonal Polarizations for Measuring Precipitation, *J. Appl. Meteorol.*, vol. 15, pp. 69–76, 1976.

60. Zawadzki, I.: Factors Affecting the Precision of Radar Measurements of Rain, *Preprints, 22d Conf. Radar Meteorol.*, pp. 251–256, American Meteorological Society, Boston, 1984.

61. Burgess, D., et al.: Final Report on the Joint Doppler Operational Project (JDOP), 1976–1978, *NOAA Tech. Memo. ERL NSSL-86*, 1979.

62. Armstrong, G. M., and R. J. Donaldson, Jr.: Plan Shear Indicator for Real-Time Doppler Identification of Hazardous Storm Winds, *J. Appl. Meteorol.*, vol. 8, pp. 376–383, 1969.

63. Donaldson, R. J., Jr.: Vortex Signature Recognition by a Doppler Radar, *J. Appl. Meteorol.*, vol. 9, pp. 661–670, 1970.

64. Zrnić, D. S., and R. J. Doviak: Velocity Spectra of Vortices Scanned with a Pulse Doppler Radar, *J. Appl. Meteorol.*, vol. 14, pp. 1531–1539, 1975.

65. Fujita, T., and F. Caracena: An Analysis of Three Weather-Related Aircraft Accidents, *Bull. Am. Meteorol. Soc.*, vol. 58, pp. 1164–1181, 1977.

66. Fujita, T.: "The Downburst," Satellite and Mesometeorology Research Project, Department of the Geophysical Sciences, University of Chicago, 1985.

67. Fujita, T.: "The DFW Microburst," Satellite and Meteorology Research Project, Department of the Geophysical Sciences, University of Chicago, 1986.

68. McCarthy, J., and R. Serafin: The Microburst: Hazard to Aviation, *Weatherwise*, vol. 37, no. 3, pp. 120–127, 1984.

69. McCarthy, J., J. Wilson, and M. Hjelmfelt: Operational Wind Shear Detection and Warning: The CLAWS Experience at Denver and Future Objectives, *Preprints, 23d Conf. Radar Meteorol.*, pp. 22–26, American Meteorological Society, Boston, 1986.

70. Witt, A., and S. P. Nelson: The Relationship between Upper-Level Divergent Outflow Magnitude as Measured by Doppler Radar and Hailstorm Intensity, *Preprints, 22d Conf. Radar Meteorol.*, pp. 108–111, American Meteorological Society, Boston, 1984.

71. Aydin, K., T. A. Seliga, and V. Balaji: Remote Sensing of Hail with a Dual Linear Polarization Radar, *J. Clim. Appl. Meteorol.*, vol. 25, pp. 1475–1484, 1986.

72. Lhermitte, R. M., and D. Atlas: Precipitation Motion by Pulse Doppler Radar, *Proc. Ninth Weather Radar Conf.*, pp. 218–223, American Meteorological Society, Boston, 1961.

73. Browning, K. A., and R. Wexler: A Determination of Kinematic Properties of a Wind Field Using Doppler Radar, *J. Appl. Meteorol.*, vol. 7, pp. 105–113, 1968.

74. Wilson, J. W., and W. E. Schreiber: Initiation of Convective Storms at Radar-Observed Boundary Layer Convergence Lines, *Mon. Weather Rev.*, vol. 114, pp. 2516–2536, 1986.

75. Carbone, R. E.: A Severe Frontal Rainband, Part I: Storm-Wide Hydrodynamic Structure, *J. Atmos. Sci.*, vol. 39, pp. 258–279, 1982.

76. Brook, M., and P. Krehbiel: A Fast-Scanning Meteorological Radar, *Preprints, 16th Conf. Radar Meteorol.*, pp. 26–31, American Meteorological Society, Boston, 1975.

77. Keeler, R. J., and C. L. Frush: Rapid-Scan Doppler Radar Development Considerations, Part II: Technology Assessment, *Preprints, 21st Conf. Radar Meteorol.*, pp. 284–290, American Meteorological Society, Boston, 1983.

78. Walther, C., C. Frush, and P. Hildebrand: The NCAR Airborne Doppler Radar, Part III: Overview of Radar Design Details, *Preprints, 23d Conf. Radar Meteorol.*, vol. I, pp. 155–158, American Meteorological Society, Boston, 1986.

79. Meneghini, R., and D. Atlas: Simultaneous Ocean Cross-Section and Rainfall Measurements from Space with a Nadir-Looking Radar, *J. Atmos. Ocean. Technol.*, vol. 3, pp. 400–413, 1986.

80. Carbone, R. E., R. Strauch, and G. M. Heymsfield: Simulation of Wind Profilers in Disturbed Conditions, *Preprints, 23d Conf. Radar Meteorol.*, vol. I, pp. 44–47, American Meteorological Society, Boston, 1986.

81. Strauch, R. G., B. L. Weber, A. S. Frisch, C. G. Little, D. A. Merritt, K. P. Moran, and D. C. Welsh: The Precision and Relative Accuracy of Profiler Wind Measurements, *J. Atmos. Ocean. Technol.*, vol. 4, pp. 563–571, 1987.

82. Röttger, J., and M. F. Larsen: Clear Air Radar Techniques, invited paper, *40th Conf.*

Radar Meteorol., American Meteorological Society, Boston, 1987. Published in Atlas, D. (ed.): "Radar Meteorology," American Meteorological Society, Boston, 1989.

83. Metcalf, J. I., and W. A. Holm: Meteorological Applications of Synthetic Aperture Radar, final report, Project A-2101, Engineering Experiment Station, Georgia Institute of Technology, 1979.

84. Atlas, D., and R. K. Moore: The Measurement of Precipitation with Synthetic Aperture Radar, *J. Atmos. Ocean. Technol.*, vol. 4, pp. 368–376, 1987.

CHAPTER 24
HF OVER-THE-HORIZON RADAR*

J. M. Headrick
Naval Research Laboratory

24.1 INTRODUCTION

Beyond-the-horizon ranges up to thousands of nautical miles can be achieved by radar operation in the high-frequency (HF) band (3 to 30 MHz). The longer-range performance is achieved by using sky-wave propagation. Ground-wave propagation over the sea is useful for short but still over-the-horizon distances. HF radar development over the past several decades has led to this capability,[1,2] and several operational systems are deployed.[3,4,5] Targets of interest are the same as for microwave radar and include aircraft, missiles, and ships. The wavelengths used are of the same order as ocean gravity waves, and this correspondence makes HF radar able to provide information on the waveheight directional spectrum and, by inference, surface winds and ocean currents.[6] In addition, this sensor is useful for observing various forms of high-altitude atmospheric ionization such as that due to aurora, meteors, and missile launches.[7,8] The wavelengths used and the nature of the transmission path make the spatial resolution coarse when compared with much-higher-frequency radars; however, the doppler resolution can be fine.

For effective radar operation, environmental parameters need to be determined in real time; transmission-path information is generally derived from adjunct vertical and oblique sounders as well as by using the radar itself as a sounder. An ionospheric electron-density model complex enough to enable adequate sounding interpretation is required. Ionospheric or transmission-path statistical forecasts are necessary for radar design and for development of the model, not for real-time operation. In addition, other users in the spectrum must be observed continuously and operating frequencies selected to avoid interference. In the context of transmission-path assessments for frequency selection, real time is measured as that interval in which there are no important changes in the ionosphere.

The magnitude and doppler distribution of the earth-surface backscatter is a major factor in setting system dynamic-range and signal-processing characteristics. Backscatter from the sea can be employed as a reference level and is a generally used diagnostic tool.

Figure 24.1 provides examples of radar coverage and spatial resolution appro-

priate for aircraft and ship targets. In the upper left the ray-path sketch shows that at higher-elevation radiation angles the rays escape, causing a skip zone of range coverage, that energy is returned to the earth until the reflection height horizon is reached, and that useful range coverage will lie between these limits. Although only one hop is shown, a multiplicity will exist, and energy may at times circle the earth. In the upper-right sketch, the ray paths show that different range extents are illuminated by using different operating frequencies, the longer ranges requiring the higher frequencies. In the example, 500 nmi is shown as the range coverage extent illuminated by one operating frequency. The trailing edge of an extent may vary as a function of radar parameters and target size, but the start is set by frequency selection and immediately follows the skip zone. Directly below is a plan view of the coverage area. It shows nine different areas that might be illuminated by a separate transmit beam of 8° width. It is likely that the ionosphere will vary across the 64° azimuth scan, so that if a single operating frequency is used, the range to the transmitter footprint will change with azimuth; however, in general a different operating frequency could be selected for each 8° to obtain the desired illumination. Each transmitter footprint is filled by 16 contiguous receive beams, each ½° wide. At the lower left, one illumination sector is shown divided into receive resolution cells with each cell being approximately 10 nmi on the side.
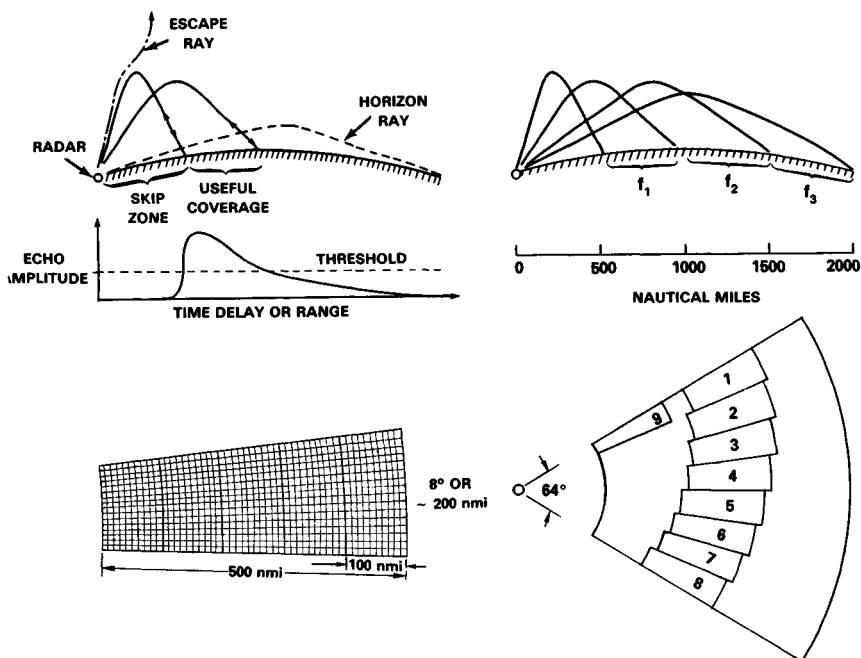
FIG. 24.1 At the upper left an example of single-hop sky-wave propagation with the useful coverage in range is sketched. On the upper right extended range coverage is obtained by using three frequencies. An azimuthal scan is shown in plan at the lower right, and at the lower left are given the receiver resolution cells in one transmitter footprint.

24.2   RADAR EQUATION

A form of the radar equation, Eq. (24.1), can be used to point to aspects of HF radar that are significantly different from radars that use higher frequencies. These differences include adaptation to environment, frequency and waveform selection, radar cross section, path losses, multipath effects, noise, interference, antenna gain, spatial resolution, and sky clutter.

$$\frac{S}{N} = \frac{P_{\text{av}}\, G_t G_r T \lambda^2 \sigma F_p}{N_o\, L(4\pi)^3 R^4}$$

(24.1)

where S/N = output signal-to-noise ratio
$\quad P_{\text{av}}$ = average transmitted power, W
$\quad G_t$ = transmitter antenna gain
$\quad G_r$ = receiver antenna gain
$\quad T$ = effective processing time, s
$\quad \lambda$ = wavelength
$\quad \sigma$ = target radar cross section
$\quad F_p$ = propagation-path factor
$\quad N_o$ = noise power per hertz
$\quad L$ = transmission-path and system losses
$\quad R$ = distance between radar and target

These parameters are explained as follows:

1. *Antennas, G_t and G_r:* A common convention for HF-band radars is to include earth effects in the antenna performance characterization, and that convention will be used here. For example, a half-wave dipole in free space has a maximum gain over an isotrope of 2.15 dB. If that antenna is oriented vertically, just above but not touching a perfectly conducting earth, its maximum gain will be increased by a factor of 4, or 6 dB, to 8.15 dB at 0° elevation angle. Since the earth is never perfect, its conductivity and dielectric constant are factors in determining antenna performance. The electrical properties of the earth are a much stronger factor for vertical polarization than for horizontal; however, terrain features and surface roughness are important for both polarizations.

2. *Coherent processing time, T:* HF radar is generally a look-down type that has earth backscatter at the same ranges as targets. Doppler processing is used to separate targets from earth backscatter.

3. *Wavelength (λ):* The wavelength or operating frequency must be selected so that energy is refracted by the ionosphere to illuminate the desired area of the earth. The spectrum of the emissions must be constrained not to interfere with other users. Since both the ionosphere and the HF-band occupancy distributions are time-varying parameters, adaptive radar management is required.

4. *Radar cross section (RCS), σ:* The radar cross section of conventional targets will generally be a function of frequency, polarization, and aspect angle. Often some sort of average value is used in analysis. Clutter levels will be large rel-

ative to most targets and therefore are important in radar design. For the RCS of earth clutter, the surface scattering coefficient σ^0 is multiplied by the resolution cell size A and the cosine of the grazing angle. The important resolution cell size factors, receiver antenna beamwidth and spectral bandwidth, are not explicitly contained in Eq. (24.1)

5. *Propagation factors (F_p):* Several propagation phenomena, including Faraday polarization rotation, multipath, and ionospheric focusing, may need inclusion in the equation. With linearly polarized radiation, Faraday rotation will result in a polarization mismatch varying with time and distance. That is, the target-incident-energy polarization will rotate as distance changes. Since many targets have RCSs that vary with polarization, an important result is that the most favorable polarization will illuminate the target recurrently.

6. *Noise (N_0):* For radars operating in the HF band it is possible to design antennas and receivers with low enough noise figures that environmental noise is dominant.

7. *Losses (L):* The loss term contains the two-way losses along the path traversed including ionospheric absorption and ground-reflection losses as well as any radar system losses. Ionospheric losses, while predicted on a statistical basis, constitute a major unknown in real-time radar operation.

8. *Range (R):* The range in the equation is the distance along the virtual path between target and radar. The ionospheric reflection height needs to be used to convert this range to great-circle ground distance. The apparent range to a particular target may take on more than one value since multiple paths may exist.

24.3 TRANSMITTERS

Most of the radar designs and missions require transmitter average power levels between 10 kW and 1 MW. Antennas are generally arrays of elements, and the common trend is to drive each element with a separate amplifier. This approach permits beam steering at a low level in the amplifier chain. Power control and amplitude-shaping requirements indicate a linear amplifier design. Since the radar uses doppler filter signal processing to separate the targets from the clutter, the clutter returned on the phase and amplitude noise sidebands radiated by the transmitter should be below that of desired targets. This can put a stringent condition on the emitted signal-to-noise ratio of the transmitter. The signal-to-noise ratio of the initial signal synthesizer must meet the requirements. The lower-level signal amplification can generally be designed to add no noise. Mechanical vibration in the high-power amplifiers can add noise, and care must be exercised in the air or liquid coolant flow system design.

The active element in each final transmitter stage can be either a traditional vacuum tube[9] or a solid-state device.[10] If the radar is to perform wide-area surveillance, frequent frequency changes are required in order to cover the various range extents. In addition, relative phase or time-delay changes are required in each amplifier chain to accomplish azimuthal steering. Reference 11 contains a discussion on modern high-efficiency amplifier designs for broadcasting; while some of the techniques are useful in radar transmitters, the final amplifiers tend to be narrowband. A broad-bandwidth performance and a tolerance to a variable standing-wave ratio load are desired features in a wide-area surveillance radar. Since the antenna elements will be wideband,

harmonic filters may be required. As an example, one transmitter and harmonic filter combination might have a 5- to 9-MHz passband and a stopband for 10 Mhz and higher frequencies; a second combination might pass up to 17 MHz and reject 18 Mhz and higher, and the design would continue in this manner to the highest frequency of operation.

24.4 ANTENNAS

A single antenna can be duplexed and used for both transmit and receive. The Naval Research Laboratory (NRL) magnetic-drum recording equipment (MADRE) antenna is an example.[1] This 100-m-wide by 40-m-high aperture provided sufficient gain and angular resolution for aircraft tracking in the upper part of the HF band: the frequencies used in daytime. A horizontal aperture of twice the width would provide similar resolution at night, where frequencies would be in the lower part of the HF band. However, it is desirable to have better azimuthal resolution for location accuracy and to reduce the clutter amplitude, and horizontal apertures of 3 km and even wider can be useful.[12,13] It is common practice to use separate transmit and receive antennas with the transmitter floodlighting many simultaneously formed receive beams. In elevation, desirable radiation angles run between 0 and 40°, with the specific angles depending on range and reflection height. Any sensitivity gained by directivity in elevation directly improves radar performance. This is in contrast with azimuth directivity for transmitting, where an increase in gain is accompanied by a decrease in area coverage. The AN/FPS 118 over-the-horizon (OTH) radar[3] does not employ steerable directivity in elevation but covers all necessary radiation angles with one broad elevation beam. This choice permits the antenna to have a relatively small vertical dimension. The AN/FPS 118 radar module uses a nominal 7.5° transmitter footprint filled with five receive beams; this combination is stepped as a surveillance barrier within a potential coverage area 500 to 1800 nmi in distance by 60° in width.

The antennas and power amplifiers used in HF broadcast station service have much in common with an HF radar transmitter-antenna combination; that is, the broadcast-service aim is to obtain a specified level of illumination over some selected area. The multiple-band and steerable broadcast antennas reviewed in the *IEEE Transactions on Broadcasting* Special Issue on Short-Wave Broadcasting,[11] those described in Johnson and Jasik,[14] pages 26–29 through 26–35, and those given in "Shortwave Antennas,"[15] Chaps. 13 and 14, are examples. It will be noticed that many of the broadcast-service antennas employ large vertical apertures. However, antennas used for radar have an added severe demand to minimize mechanical motion that would cause signal modulation, and this requirement is easier to meet with low-antenna-height designs.

Among the many factors that influence antenna design are:

1. The transmitter-receiver antenna gain product must be large enough to make its required contribution to sensitivity.

2. The receive beamwidth must be narrow enough to provide the required location accuracy.

3. The receive beamwidth must restrict clutter levels to values permitted by system dynamic range and slow-target detection requirements.

24.5 CLUTTER: THE ECHO FROM THE EARTH

Early in HF radar experiments it was noted that the clutter received by sky-wave paths represented a large signal and provided an indication of the earth-surface area illumination. Extensive observations made at the Naval Research Laboratory viewing alternately Atlantic Ocean areas and central United States areas indicated that, averaged over a wide area, sea clutter power levels were about an order of magnitude higher than those from an area of similar size in the central United States. Later observations indicated even less backscatter from ice-covered Greenland areas. The backscattered energy from land is topography-dependent; as an example, a city in the central plains provides a larger echo than its surroundings. In contrast with such abrupt changes, the seascattering coefficient σ^0 for HF radio waves is fairly uniform; that is, it changes gradually with range and azimuth. The sea echo power is proportional to the resolution cell area, to a good approximation. The nature of the sea echo needs some description. For the seascattering coefficient to be approximately constant, long time averaging (minutes) is required. Sky-wave transmission characteristics are notably variable; however, the sea echo can be used as an amplitude reference when care is exercised. An elementary description of sea echo behavior and some applications follows.

Ocean wave generation and propagation are a complex and not completely understood subject; a similar statement can be made about electromagnetic scattering from the waves. However, for the scale size of HF-band wavelengths the sea is a surface that is only slightly rough, and the method of Rice[16] can be used to explain reflection. Backscatter from the sea can be considered a resonant interaction. The disorganized-looking ocean waves are thought of as being the sum of an infinite number of Fourier surfaces, each being a sinusoidally corrugated sheet with a different wavenumber and direction.[17] The principal backscatter for a grazing-incidence electromagnetic wave comes from that component sheet that has a wavelength equal to one-half of the radar wavelength and that is either directly approaching the radar or receding from it.[18,19] The surface scattering coefficient, or RCS per unit surface area (σ^0), is much larger for vertical than for horizontal polarization. The echo from the sea with horizontal polarization can be neglected at the smaller grazing angles. The water wavelengths between 10 and 100 m are gravity waves that in deep water follow the dispersive relation

$$v = \left[\frac{gL}{2\pi}\right]^{1/2} \tag{24.2}$$

where v = water-wave phase velocity
      g = acceleration of gravity
      L = water wavelength

Ocean waves are excited by the surface winds. If a wind blows at a constant velocity long enough and over sufficient fetch, a steady-state condition will be achieved where the wind provides just enough energy to the water to supply that lost in breaking and other dissipation. It will completely arouse the water waves with a phase velocity nearly equal to the wind speed. It will also completely arouse all the ocean waves of lower velocities. Figure 24.2a gives an example of the spectrum as derived from a measurement of waveheight versus time at a particular point. Pierson and Moskowitz have derived the following relation for a fully developed spectrum based upon empirical data.[20]
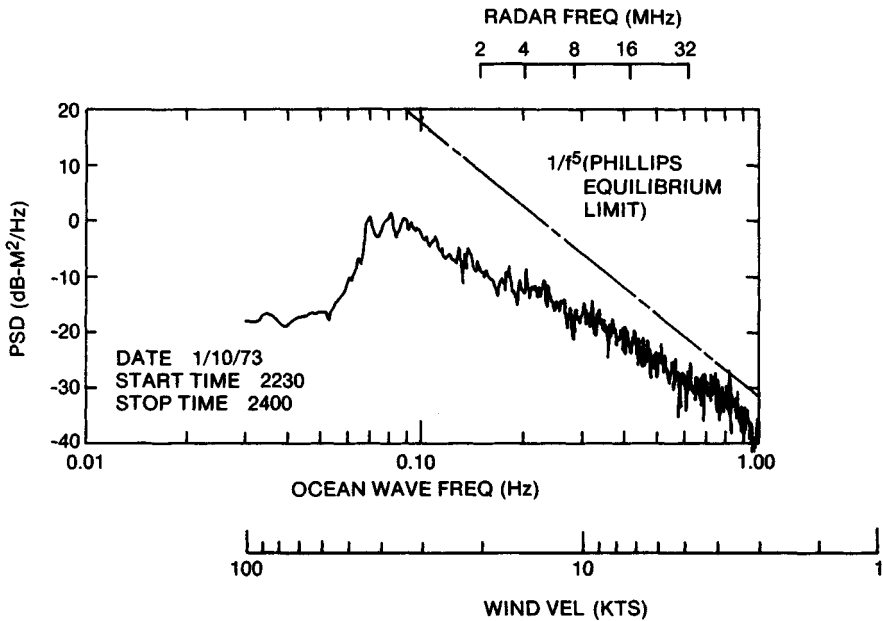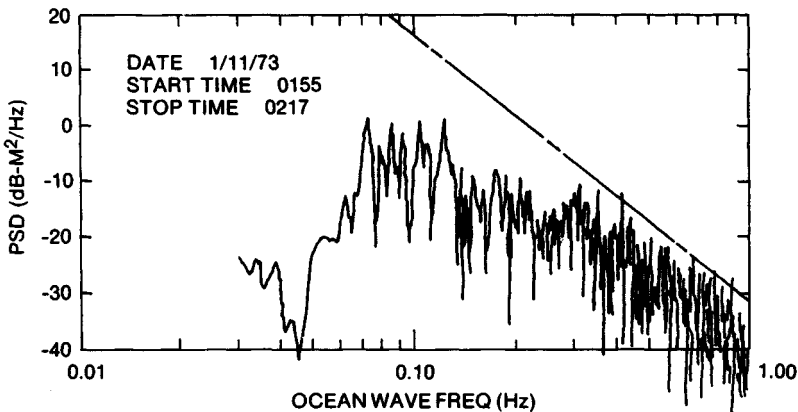
FIG. 24.2 (*a*) Ocean-waveheight power spectral density (PSD) is given as a function of water-wave frequency. The straight line is a saturation asymptote, and for this example the limit is approached at the higher frequencies. The scale marked "wind velocity" can be used to deduce that winds up to 40 kn have excited waves with frequencies as low as 0.08 Hz but that either the length of time or the fetch, or both, have not been sufficient for full development. In this example the individual spectral analyses, requiring about 100 s of data, have been averaged for 1½ h. (*b*) The spectral density is given as in *a* but is averaged over a shorter time, 23 min. Both higher peaks and deeper voids are seen. The scale across the top of *a* gives the radar operating frequency that experiences resonant backscatter, and it can be seen that at many frequencies this level is far below the long-term average.

$$S(\omega) = C_1 g^2 \omega^{-5} \exp\left[-C_2\left(\frac{g}{\omega v}\right)^4\right]$$ (24.3)

where S = waveheight squared per hertz
ω = water-wave angular frequency
v = wind speed
C_1, C_2 = constants

The exponential term approximates the decay in the spectrum above the wind maximum velocity. The ω^{-5} asymptotic term is a feature of major interest and use. This saturation effect suggests that the sea echo can provide an amplitude reference, and an important consequence is that with such a reference path losses may be estimated. The inference is that for vertically polarized waves from a radar, looking along the wind direction, σ^0 is constant for a specific radar operating frequency if the sea is fully developed. Further, following the analysis of Barrick,[19] σ^0 will have the same value for all water-wave frequencies along the wind direction where the sea is fully developed. In general, the scattering coefficient will be proportional to the resonant waveheight squared. The term *resonant* refers to the Fourier component of the surface where the water wavelength multiplied by the cosine of the grazing angle equals one-half of the radar wavelength. This type of scatter is frequently refered to as *Bragg scatter*. These models are helpful in developing an understanding of the sea surface and the radiowave interaction; however, a very long time average is implied when the water-wave spectrum is described as the product of two terms, one of which is frequency to the minus fifth power and the other an exponential with frequency in the argument. A single 100-s look at the waveheight-squared spectrum with a frequency resolution of $\frac{1}{100}$ Hz would yield a very jagged spectrum; Fig. 24.2*b* provides a short look to contrast with the long time average of Fig. 24.2*a*.

The variability and value of σ_{vv}^0 was examined with the San Clemente Island ground-wave radar.[21] The *vv* subscripts indicate that both transmission and reception were vertically polarized. This radar had several valuable and unique features: a transmission path out over the open sea, multiple-frequency operation in a repetition period, calibrated antennas, known transmitter power, and ground truth in the form of ocean-wavelength recordings. When looking into an approximately 20-kn wind, values of σ_{vv}^0 were found to be constant within a few decibels for operating frequencies where the water-wave spectrum was approximately fully developed; these observations provided a confirmation of Barrick's first-order theory.[19] It is emphasized that a constant level in scattering coefficient implies long averaging times. By using the antenna gain conventions stated earlier and a semi-isotropic sea directional spectrum, the value of σ_{vv}^0 was calculated as -29 dB, and the measured values were grouped between -7 and $+3$ dB of this value over a 5- to 20-MHz-frequency span. This experiment provided a direct measure of the sea-surface scattering coefficient and has exposed characteristics that should be considered when using σ^0 as a reference.

Some other features need consideration. Water-wave directions neither will be confined to the wind direction nor will they be semi-isotropic, but they will spread throughout 360°, with the spreading function depending upon frequency and other variables. Oceanographers generally treat the directional wave spectrum in the wind direction half plane only.[22] But HF radar has sufficient sensitivity to expose the waves running against the wind that have an RCS more than two orders of magnitude below those running with the wind. If the wave directional-spectrum distribution of Long and Trizna is used, the maximum value

of σ^0 for a saturated sea is -27 dB in the upwind or downwind direction (longitudinal sea) and is -39 dB in the crosswind direction (transverse sea).[23] Figure 24.3 gives the nominal shape of the scattering coefficient for these two conditions. The statements so far have all related to vertical polarization, which will have the largest scattering coefficient. Sky-wave radar resolution cell sizes are generally large enough that Faraday rotation can be expected to cause illumination of a resolution cell to be over all polarization angles and consequently to reduce the average value of σ^0 by 3 or 4 dB, depending on the elevation radiation angle. If -30 dB is used for the effective scattering coefficient and a resolution cell is considered to be a square 10 nmi on a side, the surface area A is 85 dBsm and the RCS is $85 - 30 = 55$ dBsm. A 12 dB path enhancement will be effective owing to constructive multipath addition, which will increase the resolution cell effective RCS to 67 dBsm. The receiver and processor must be able to handle both the high-level signal due to this large RCS and those much smaller signals due to targets. An HF radar must be designed to accommodate such clutter levels even though they will not exist all the time or at any one time over all areas, especially at the lower operating frequencies.
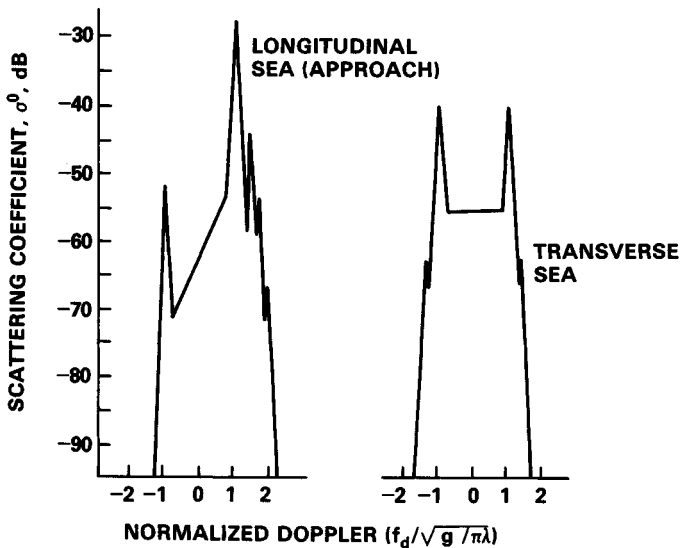
FIG. 24.3 The nominal maximum values of the seascattering coefficient are given as a function of doppler frequency for a coherent processing time of 51 s. A longitudinal, directly approaching or receding sea gives the largest value. A transverse, cross to the radar look direction, gives two peaks equal in height.

By scanning over a radar's total potential coverage area and using the ratio of the two resonant doppler frequency sea responses, a map of sea principal direction can be constructed.[23,24] By inference, surface winds can be mapped. Surface-wind maps are a regular contribution by the Jindalee radar.[25] Actual waveheight estimates can be made on the basis of the multiple scatter features of the sea echo spectra.[21,24,26–28] A major difficulty with these methods lies in eliminating corruption of the echo doppler spectrum by ionospheric path propagation.

Trizna has reported a method for estimating σ^0 from three simple and direct measurements made on the radar echo spectrum, and this method is relatively insensitive to propagation-path variation.[29,30] All the methods for estimating sea state or scattering coefficient require long coherent processing times plus incoherent averaging of a number of coherent times in order to achieve a distinct and repeatable amplitude-doppler signature. This type of radar operation will frequently be incompatible with other radar missions. However, the sea echo is a large signal, and it can be obtained with an adjunct oblique sounder operating in an appropriate radar mode.

In summary, the sea echo power in a resolution cell (1) is generally the largest in-band echo signal; (2) generally exists in the open ocean even in relative calm; (3) varies as the square of resonant waveheight, which is frequently limiting at the higher frequencies; and (4) varies with direction, being greatest for seas running toward or away from the radar.

24.6 RADAR CROSS SECTION

The RCS of the sea has been treated in Sec. 24.5. A number of natural scatterers have been described at the 1981 Symposium on the Effect of the Ionosphere on Radiowave Systems.[31] Findings on the nature of auroral echoes are given by Greenwald,[8] and an auroral echo-scattering model has been developed by Elkins.[32] Chapter 4 of Ref. 7 treats the RCS associated with rockets and their exhausts. Here attention will be confined to echoes from airborne and surface targets.

Aircraft and ships have dimensions that put them in the resonant scattering region. The smallest aircraft and cruise missiles will be in the Rayleigh scattering region for the lower half of the HF band. The RCS has aspect sensitivity but strongly depends upon the target's gross dimensions. For an aircraft the span of the wings, the fuselage length, the tail and elevator span, the vertical stabilizer and rudder height, and their relative locations are the features that determine the RCS. Target shaping of a scale size much less than a wavelength will have little effect. For bodies with high-conductivity surfaces the scattering cross section can be calculated by using numerical methods.[33] Facilities exist where good scale-model measurements can be made, and the Ohio State University compact range is an example.[34]

Rough but useful RCS estimates can be made by using the behavior of a few "canonical" shapes. Figure 24.4 is a family of plots giving RCS versus radar frequency for an oblong-shaped conducting body. The straight line marked 90° $\lambda/2$ dipole gives the RCS of a resonant, conducting half-wavelength rod, where the rod is parallel with the electric field. This geometry gives the maximum RCS for the rod. The upper scale of the abscissa gives the one-half-wavelength dimension of the frequency given on the lower scale. The curve marked 90° is the RCS of the oblong-shaped conducting body of 11-m length and 1-m thickness; again the target long dimension is aligned with the electric field. The maximum RCS coincides with the nominal half-wavelength dimension or with the first resonance. The curves marked 45, 15, and 0° give the RCS as the target is rotated to these angles in the plane that contains the electric vector. The little sketches give, at the left, the body shape, then the RCS patterns at nominal ½ wavelength, 1 wavelength, and ¾ wavelength in order to help visualize how the RCS will change as the as-
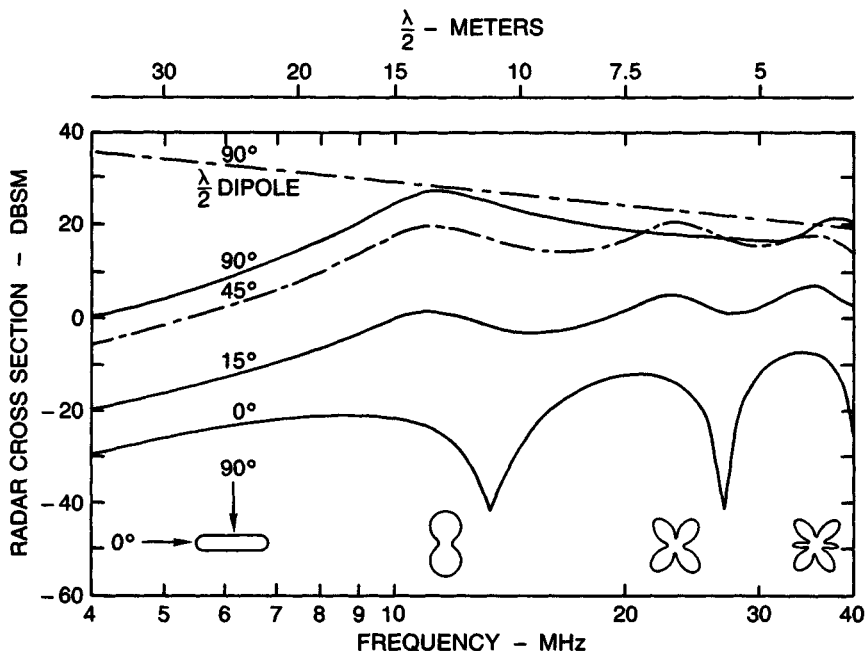
FIG. 24.4 The RCS is given versus frequency for an 11-m-long and 1-m-thick oblong conducting target. The E vector and the 11-m dimension are in the same plane; 0° (nose-on), 15°, 45°, and 90° (broadside) curves are given. The top dashed curve is for a resonant dipole at 90°. The little sketches at the first, second, and third resonances of the 11- by 1-m target show how the RCS behaves with an illumination aspect.
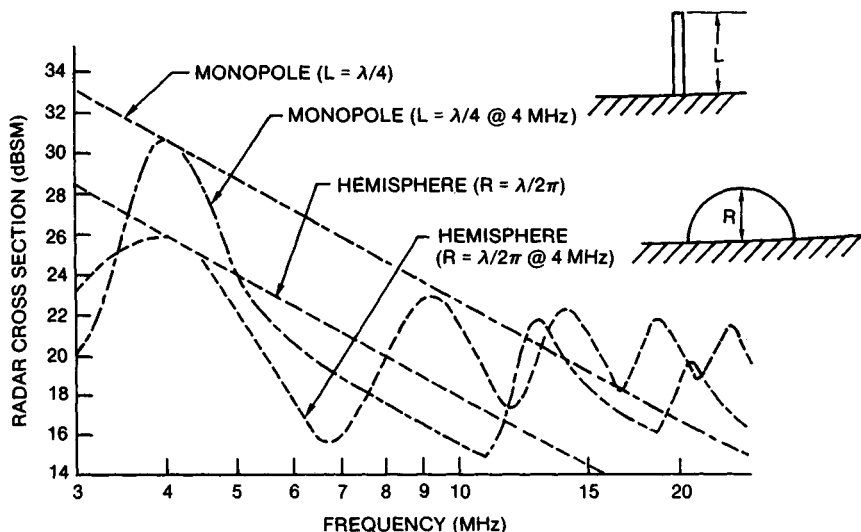
FIG. 24.5 The vertical-polarization RCS of a hemisphere and a monopole on a perfect conductor are plotted against the radar operating frequency.

pect angle is varied. For targets of other lengths with approximately the same shape factor, the response can be determined by sliding the curve along the $\lambda/2$ line and making the first resonance coincide with the line at the $\frac{1}{2}$-wavelength point. Aircraft with their wing-fuselage cruciform shape will have an RCS that varies with the aspect angle but not greatly. As has been mentioned, Faraday rotation will ensure periodic illumination with the most favorable polarization.

Figure 24.5 gives the vertical-polarization RCS of a rod and a hemisphere sticking out of a perfectly conducting surface. With these canonical shapes an estimate of RCS can be made for surface craft. For small vessels the mast height will be of most importance.[35] For surface targets where the maximum RCS is with vertical polarization, the 12 dB sky-wave RCS enhancement mentioned for the sea echo will occur.

24.7 NOISE AND INTERFERENCE

In the HF band, receiving systems can be designed such that external noise is dominant. The major source of noise at the lower frequencies is lightning discharges ionospherically propagated from all over the world (*sferics*). At the high end of the band, extraterrestrial or galactic noise may be larger than that due to sferics. Receive sites in an area of extensive electrical equipment use can find human-made noise dominant. The HF band is well occupied by other users, and channel selection can be such that other transmitter emissions constitute the background "noise" level.

The widely used source on noise is International Radio Consultative Committee (CCIR) Report 322.[36] This report is based upon measurements made at 16 locations throughout the world. The measurement and data analysis was performed to exclude individual collection site local thunderstorm contributions. Spaulding and Washburn[37] have added data from the U.S.S.R., and a revised CCIR report is available. The noise-level medians as a function of frequency are given in the form of worldwide maps by season and 4-h time block. Lucas and Harper[38] have provided a numerical representation of CCIR Report 322 useful for computer computations, and this has been revised by adding the work of Spaulding and Washburn. The numerical maps of median values are accompanied with decile values to indicate distributions over days of the season. These noise maps provide the level that an omnidirectional antenna would receive. The common method of use is to treat the noise as isotropic even though it must be azimuth- and elevation-angle-dependent. Examination of maps indicates that tropical rain forests and other regions of concentrated thunderstorm activity are major sources of noise. Also, examination suggests that a denser distribution of data collection sites would improve the data. It would be very useful to have noise source maps so that antenna patterns and explicit propagation effects could be taken into account. Some work toward source maps has been done by Ortenburger and Kramer,[39] but there is no known generally available database in this form.

Even though available noise description has limitations, it does provide a reference level for radar design. An HF radar is generally designed to take advantage of what the environment permits; that is, the receiver noise figure should be good enough to make environmental noise the limitation. An example of the CCIR Report 322 data will be discussed. Figure 24.6 was drawn from Lucas and Harper.[38] Noise power in a 1-Hz band relative to 1 W (dBW) is given as a func-
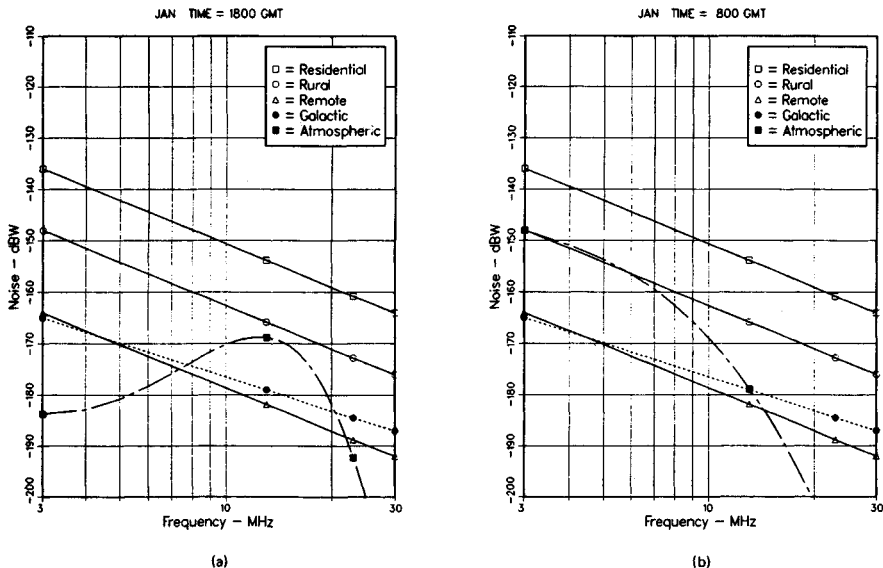
FIG. 24.6   Noise power per hertz is given for 38.65° north latitude and 76.53° west longitude and winter. (*a*) 1800 UTC is given as a daytime example. (*b*) 0800 UTC is given as a nighttime example. Universal time coordinated (UTC) and Greenwich mean time (GMT) are interchangeable in this treatment.

tion of frequency for three different sources of noise. The practice in use is to select the largest. This is a winter daytime example at a United States east coast location. The three straight lines are estimates of anthropogenic (formerly man-made) noise for three different types of sites. The shape of the anthropogenic curves is described by the equations

$$N_0 = -136 - 12.6 \ln (f/3) \qquad \text{residential}$$
$$N_0 = -148 - 12.6 \ln (f/3) \qquad \text{rural}$$
$$N_0 = -164 - 12.6 \ln (f/3) \qquad \text{remote}$$

where the frequency f is in megahertz and ln indicates the natural logarithm. The above trends with frequency approximate many measurements of human-made noise; however, ideally the curve would be based on measurements at the particular radar site. The galactic-noise curve should be selected when it is the largest and when there is a path through the ionosphere; the path will not exist for the lower operating frequencies in the daytime. The atmospheric noise rises from low frequencies to about 12 MHz and then rapidly falls. Figure 24.6*b* is for nighttime. All the curves are the same as in Fig. 24.6*a* except for atmospheric noise. At 10 MHz the night and day levels are the same; below 10 MHz the noise decreases with decreasing frequency in daytime and increases at night. Above 10 MHz daytime levels are greater than those at night. These effects can be partially explained by the very lossy long-range paths in day that attenuate the long-range noise at the lower frequencies and by there being few or no sky-wave paths to noise terrestrial sources at the higher frequencies at night. Later it will be seen that in general nighttime noise will be greater than daytime noise for sky-wave

illumination of a selected range. The general trends of atmospheric noise in other seasons are similar to those of winter. However, there can be large differences in level at other locations on the earth.

Other effects that can control radar performance are sometimes mistaken for the *passive* noise discussed above. One of these is the spread-in-doppler clutter from localized high-density irregularities in sky ionization; this is sometimes refered to as *active* or multiplicative noise. The occurrence of this type of clutter is greater at night and is much more prevalent in the auroral zones and around the magnetic equator. As stated earlier, Elkins[32] has developed a model for HF auroral clutter that can be used to predict target obscuration when the transmission path is through the auroral region. Lucas has provided spread-F maps for inclusion in ionospheric models so that spread doppler clutter can be predicted.[40] Ionospheric irregularities that scatter back to the radar receiver occur much more often at night than by day at any latitude. Their effects can be reduced with spatial resolution.

Anyone engaged in extensive HF radar performance analyses should have the numerical description of noise maps in their computer data files. When only a few performance predictions are needed, CCIR Report 322 can be used manually.

24.8 SPECTRUM USE

The waveforms that can be effective for HF radar are in general similar to those used at the higher frequencies and are selected for similar reasons. However, the transmission path is dispersive, and waves experience polarization rotation with frequency; because of these effects, bandwidths are limited to the order of 100 kHz without correction. The more restrictive constraint on emissions is that of noninterference to and by other services.

Frequency spans in the HF spectrum are allocated for various types of service such as broadcasting, point-to-point communications, maritime mobile, aeronautical mobile, standard frequency and time, and amateur. The variability of the sky-wave transmission medium requires different operating frequencies at different times. A single point-to-point circuit can require as many as five different frequencies spread over a wide range if the circuit is to be reliable over all hours of the day and seasons of the year and through the solar activity cycle. If a radar is to perform surveillance over large areas by ionospheric refraction at all times of day, seasons, and degrees of solar activity, frequency channels distributed over a large part of the HF band are required, although only a single channel may be used at any one time. When the HF band is scanned with a spectrum analyzer at a particular hour, it can be seen that the gross features of occupancy are remarkably stationary over the days of a season. This is due to broadcast stations, fixed-service point-to-point transmitters, and many other spectrum users having regular schedules. Figure 24.7 examines one particular segment and time. These observations were made with 5-kHz-bandwidth filters. When narrower-bandwidth filters with steep skirts are used, several channels 5 to 10 kHz wide with no detectable users are generally found within any 1-MHz span. The maximum frequency that will still reflect energy back to the earth during the day may be twice that at night; therefore, the occupancy tends to be denser at night than during the day.
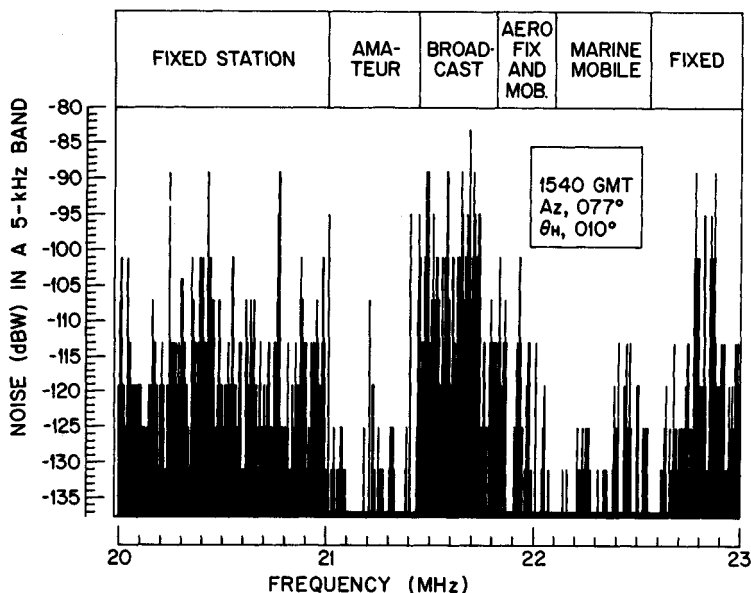
FIG. 24.7   An example of spectrum occupancy measured at the MADRE radar.

The part of the spectrum that is useful for sky-wave propagation is densely populated. Even out-of-band signal levels are a consideration in receiver front-end design, where it is convenient to have bandwidths much wider than that of the radar signal. There are a large number of broadcast stations that have 500-kW transmitters and antennas with more than 20-dB gain. Measurements made on the middle Atlantic coast of the United States show HF broadcast-band signals with strengths of 5 to 10 mV/m. These ambient levels must be accommodated in receiver design since a wideband front end is desirable for rapid and frequent frequency changes.

The practice in allocations for HF radar operation is to permit use of broad bands of the spectrum with a requirement to cause no interference to an existing service and to provide a lockout feature for channels that need protection. An integral part of an HF radar is a channel occupancy analyzer that provides a real-time description of spectrum availability.

24.9 *SKY-WAVE TRANSMISSION MEDIUM*

Solar radiation and particle emission bombardment are the cause of ionization in the earth's upper atmosphere. Even though there is no incident radiation at night, the ionization never completely decays; that is, there is always an ionosphere. The electron-density distribution is the major control over the propagation of HF radio waves. Ground illumination over the horizon is enabled by refraction in the ionosphere. When an oblique-incidence radio wave traverses a path where electron density is increasing with altitude, the ray is bent away from the vertical; if

the gradient in density is sufficient, the wave will "reflect" back back to the earth, providing long-distance illumination. The lower the radio-wave frequency, the smaller the required gradient. Since some ionization in the upper atmosphere always exists, it is always possible to illuminate the earth over the horizon if there is freedom in frequency selection. Ionospheric outages do not exist in the sense that long-distance illumination is impossible. Path outages are due to deficiencies in frequency channel allocations and insufficient radiated power. Additional factors that can affect radar performance are ionization irregularities that degrade path quality and backscatter from spread-in-doppler ionization gradients that can obscure targets.

The solar activity that drives the ionization of the earth's atmosphere is variable on a diurnal, seasonal, and long-term basis with a superimposed random component. Current prediction and analysis methods depend upon a statistical description of the ionosphere. A large amount of vertical-incidence reflection-height versus frequency-sounding (ionosonde) data has been collected over several decades, and from this data most descriptions of the ionosphere are derived. Davies[41] and Chap. 10 of the "Handbook of Geophysics and the Space Environment"[42] can be read for information on ionospheric radio-wave transmission. The radar designer needs a statistical description that will permit matching the design to the required frequency span, power levels, and vertical radiation-angle gain. The radar operator needs a model with enough sophistication to permit full interpretation of the real-time soundings for both operating parameter selection and data analysis.

The regions of the ionosphere that are considered necessary to model for an understanding of transmission paths are as follows:

1. *D region:* This region occupies the lowest altitudes considered. It ranges from 50 to 90 km, where electron density rapidly increases with altitude in the daytime. The maximum ionization in the D region occurs near the subsolar point and will be greatest during periods of highest solar activity (sunspot maximum). The D region may not be explicit in some ionospheric models where its effects are accounted for with an empirically derived path-loss calculation. Most models have this nondeviative absorption as a median value plus a distribution.

2. *E region:* This ionization region extends between about 90 and 130 km in altitude with a maximum near 110 km when sunlit. In addition, there may be anomalous ionization referred to as sporadic E. This latter ionization layer is thin in altitude, may be either smooth or patchy, is seasonally and diurnally variable but not well correlated with solar activity, and has marked variation with latitude.

3. *F Region:* This is the highest-altitude region of interest for sky-wave propagation, and it is also the region of greatest electron density. In the daylight hours there may be two components that should be recognized, especially in summer. The F1 region lies between 130 and 200 km and, like the E region, is directly dependent upon solar radiation; it reaches maximum intensity about 1 h after local noon. The F2 region is variable in both time and geographical location. The altitudes of the F2 region peaks are considered to lie between 250 and 350 km in the middle latitudes. The F2-region ionization shows marked day-to-day variations and in general is not the regular sun follower that the E and F1 regions are. Most models have a statistical description of F2 maximum electron density (or critical frequency) in the form of a median and upper and lower decile values.

Goodman and Reilly discuss shortwave prediction methodologies on pages 230 to 237 of Ref. 11. The ionospheric models that have been extensively used in

HF radar performance analysis are in programs called ITSA-1, ITS-78, RADAR C, IONCAP, and AMBCOM.[43-47] Lucas[48] provides some detail on these models and their origins. In summary, they all draw on the large database of recorded ionospheric soundings made during the International Geophysical Year of 1957–1958 and the International Year of the Quiet Sun of 1964–1965. All the models rely heavily on data taken from the maximum and minimum solar activity years of one solar cycle, with supplement from an adjacent cycle of lower activity. Linear interpolation between the two extremes is used for conditions of intermediate solar activity. This limited database might be considered a serious deficiency since there can be considerable difference in the measure of solar activity from one cycle to another. However, while the degree of atmospheric ionization appears to be a strong function of position in the solar cycle, it has only a weak dependence on solar activity measure. The commonly used indices or measures of solar activity are the sunspot number (SSN) and the radiated-microwave flux density. On a yearly average basis these indices relate well to the median ionospheric description; however, short-term predictions are more difficult to relate to solar activity. In general, the models listed are part of a prediction method. Some of the prediction methods have not been well documented although widely distributed; also, users frequently "improve" upon a model and prediction method to suit their specific needs. As an example, the model RADAR C of Ref. 45 is the basic building block of Thomason, Skaggs, and Lloyd in NRL Report 8321;[49] however, they have added a D region, a collision-frequency distribution, an earth's magnetic field, a topside electron distribution, an auroral electron-density modification,[50] and other features that make the model more generally useful. The ionospheric model as described in NRL Report 8321 will be used for examples here.

A form for display of the data is that of vertical virtual height versus frequency sounding and a true-height plot. *Virtual height* is defined as the speed of light times the time delay for the ionospheric echo; *true height* is the actual distance to reflection height. The critical frequency is the highest frequency that is reflected. Reference 46 has a figure that illustrates how the stored median ionosphere compares with a set of actual soundings; it is shown here as Fig. 24.8. Each of these soundings was made at the same hour but on a different day of the month. In this figure each trace shows the virtual height of a vertical sounding versus probing frequency for the ordinary ray. When vertical soundings are made with an ionosonde that uses a linearly polarized antenna, the ionosphere will be birefractive and provide two traces that are called *ordinary* and *extraordinary*. With right-hand and left-hand circularly polarized antennas the two responses can be separated. Figure 24.9a and b gives an example in the mid-Atlantic off the east coast from the data file. Such plots can be obtained at a selected earth location for a specified level of solar activity (SSN), month, and time of day. Similarly to the illustration given in Fig. 24.8, if a collection of experimental soundings is taken at the geographical location, SSN, and times of day of Fig. 24.9 over a month, the medians will approximate the curves of Fig. 24.9 closely. In addition, the upper and lower decile values of the measured monthly families of F2 critical frequencies will deviate from the median by about ±25 percent. For radar performance calculations median values will be used. However, when designing a radar, distribution should be considered for lowest operating-frequency selection. Distributions are important in communications when a limited number of channels are assigned, but since the radar operating practice is to select a near-optimum frequency, distribution is not important. The model uses three parabolas to approximate the electron distribution with altitude. Electron density and
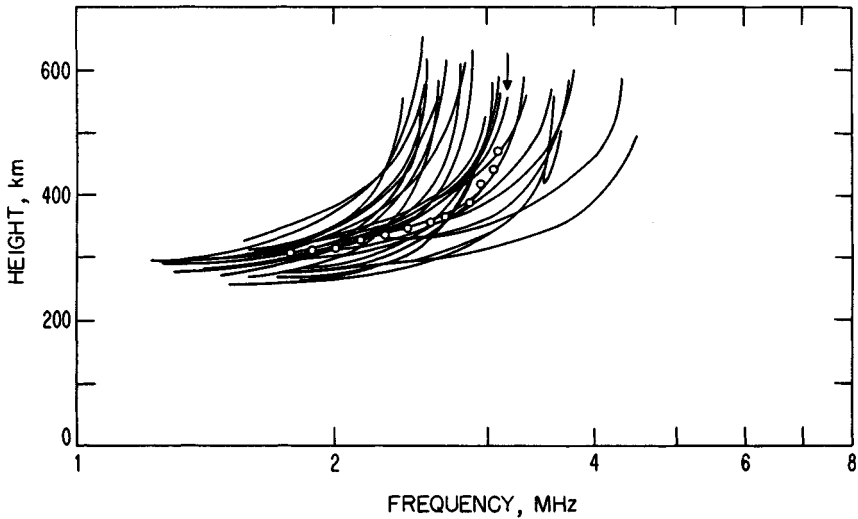
FIG. 24.8 A predicted vertical-incidence ionogram is compared with observed ionograms. The arrow points to the predicted 3.2-MHz critical frequency, and the little circles give points on the predicted median sounding. The measured median of this ensemble is about 3 MHz.

critical or plasma frequency are sometimes used almost interchangeably; when this occurs, the relation implied is $N_e = f^2/81$, where N_e is free-electron density, in numbers per cubic meter, and f is the frequency, in hertz. For the little tables shown in Figs. 24.9 and 24.10, FC is the critical frequency, in megahertz; HC is the height of maximum ionization or the nose of the parabola, in kilometers; and YM is the semithickness of the parabola, in kilometers. ES gives the sporadic-E distribution as M (median), L (lower), and U (upper) decile critical frequencies, in megahertz. This table shows the form of the stored data in the ionospheric model, and with it the various profiles can be generated. All these constants can be adjusted to fit diagnostic observations.

In Fig. 24.9 it is seen that the F2 critical is 6.5 MHz in the daytime and 4 MHz at night for summer. Figure 24.10 gives similar data but for winter, where the day and night comparison is from 8.4 MHz to 3.5 MHz. Figures 24.11 and 24.12 give summer and winter plasma-frequency contours for the same location versus time of day. These plots are provided to show how abrupt the night-to-day and day-to-night transitions are; during winter dawns the critical frequency changes from 2 to 5 MHz in 1 h. The radar-frequency management task is most difficult in these time periods; during most of the day and most of the night changes are relatively slow. This data indicates the median diurnal frequency variations required for a particular path: more than 2:1 in winter and somewhat less in summer. Day-to-day variability will impose greater extremes.

In this section the predictable and random variability of the transmission path has been indicated. When extensive and detailed radar performance calculations are to be made, a computer-stored base is required.
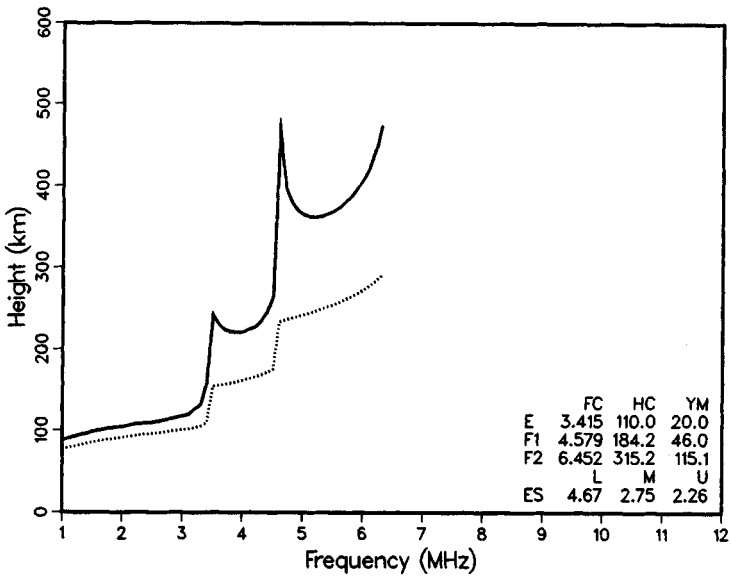
(a)
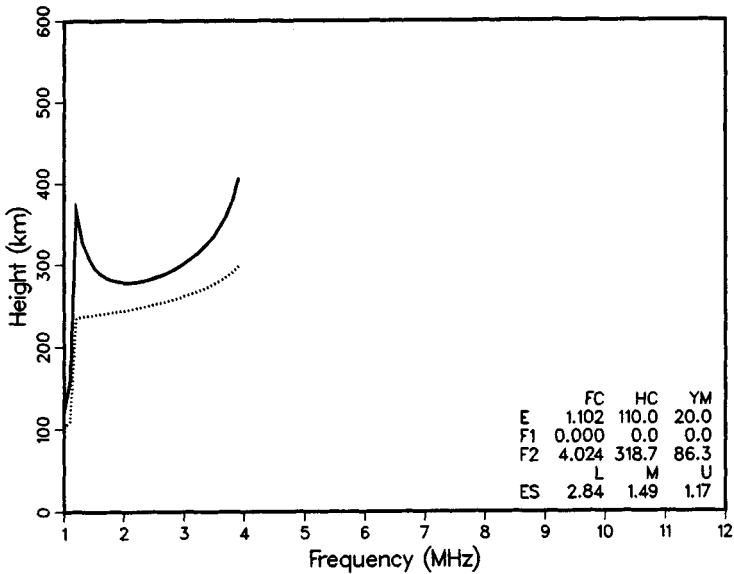
(b)

FIG. 24.9 The virtual (solid line) and true (dotted line) reflection heights are given for July, SSN = 50, and a mid-Atlantic-coast radar refraction area. (*a*) 1800 UTC is a daytime example, first hop. (*b*) 0800 UTC is a nighttime example.
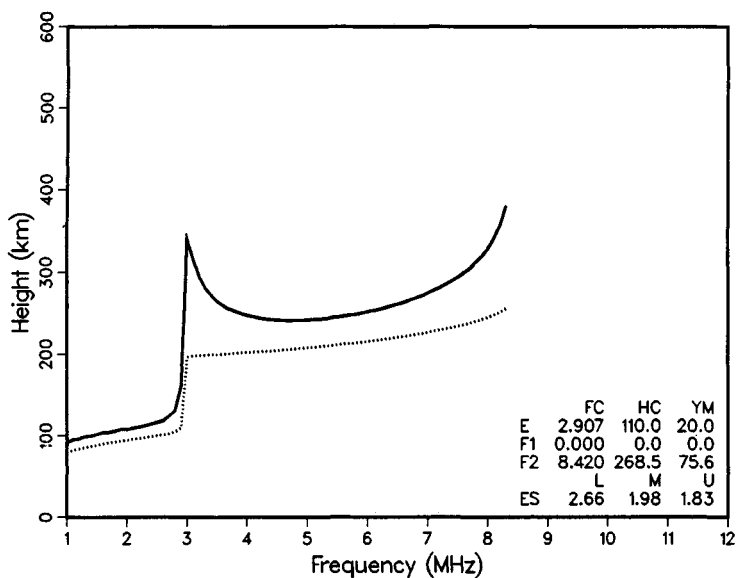
(a)
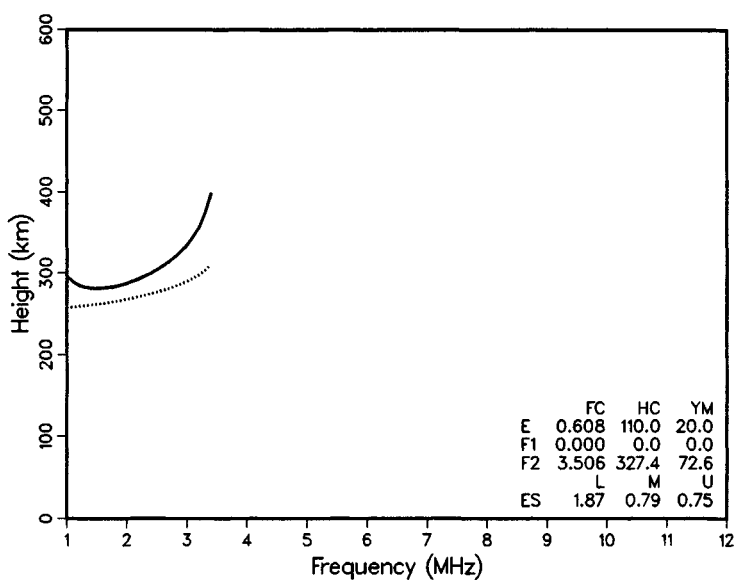
(b)

FIG. 24.10 Predicted ionograms as in Fig. 24.9, but in January. (*a*) 1800 UTC for day. (*b*) 0800 UTC for night.
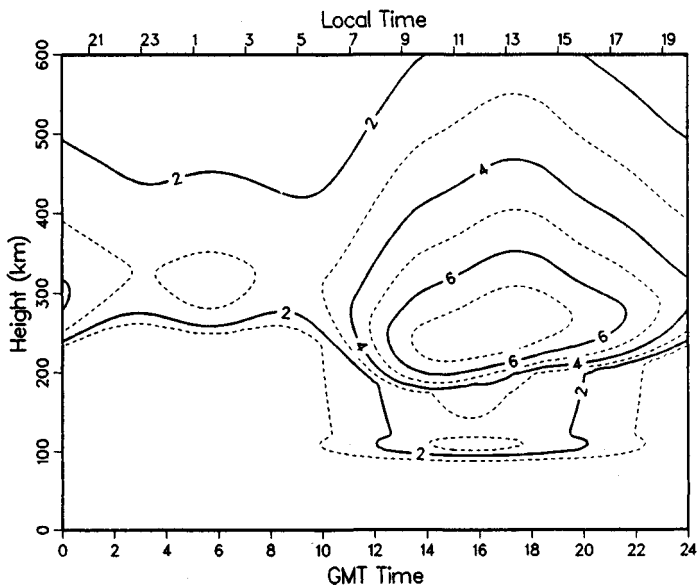
FIG. 24.11 Plasma- (critical-) frequency contours are given as a function of time of day for July; SSN = 50, latitude = 37.55°N, and longitude = 60.56°W.
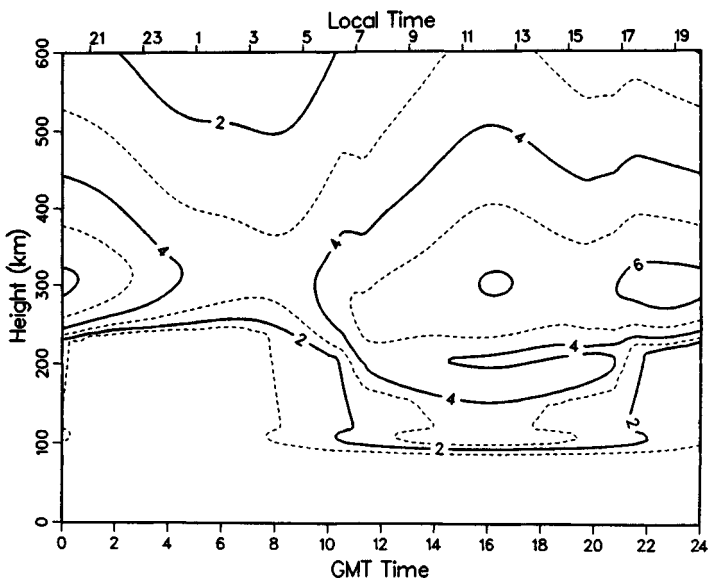
FIG. 24.12 Plasma-frequency contours are given as a function of time of day for January. Other controls are as in Fig. 24.11.

24.10 SKY-WAVE RADAR PERFORMANCE

The performance index used in this section will be the signal-to-noise ratio (*SNR*) indicated when using the transmission-medium model treated in Sec. 24.9 and the CCIR Report 322 noise discussed in Sec. 24.7. An additional need is a method to determine path effects. The detail with which path tracing is treated can vary greatly. A geometrical optics code that envolves integration along the ray as described by Jones and Stephenson[51] can provide paths in three dimensions, including delays and losses for both ordinary and extraordinary rays. When the details of electron distributions are uncertain, such comprehensive calculations are excessive. A number of other methods of path determination are available: for example, a path approximation technique goes with each of the ionospheric models mentioned in Sec. 24.9. For the performance exhibits treated here, the NRL-ITS Radar Performance Model will be used. This has been called RADAR C not quite correctly, but it is a lineal descendant of RADAR C.

NRL Memorandum Report 2500[52] describes the basic technique that will be used for path determinations. A simple closed-form virtual path trace, Snell's law for a spherically symmetric medium, is sequenced through elevation radiation angles in 1° increments. This process is incremented in 1-MHz steps over the radar's operating band. A vertical sounding of the ionosphere 700 km downrange has been used as the electron distribution for all one-hop paths, and a sounding 1400 km downrange is used for two-hop paths. Figures 24.9 and 24.10 gave a night and day example of the ionosphere 700 km downrange from a radar located at 38.65°N and 76.50°W looking east. Figure 24.13 gives constant plasma-frequency contours versus range from the radar for 0800 UTC, SSN 50, January
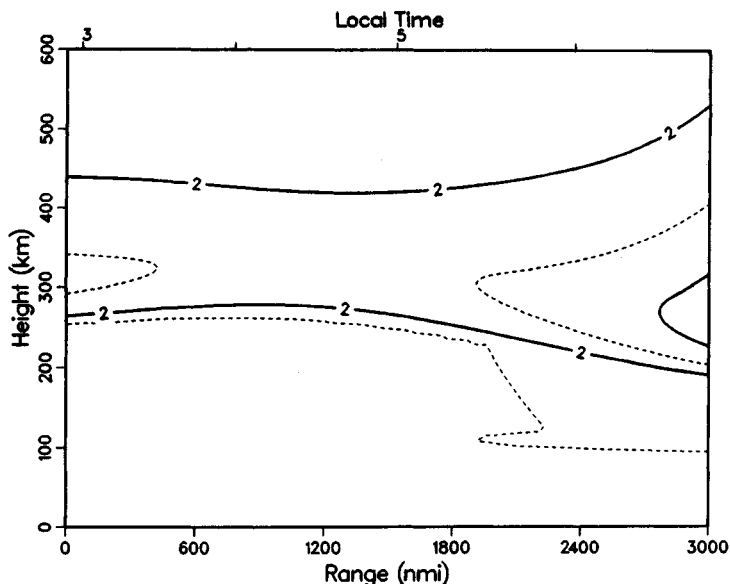
FIG. 24.13 Plasma-frequency contours are given, extending from the radar to a point east 3000 nmi downrange; January night example.

(night); and Fig. 24.14 for 1800 UTC (day). For the night case, the concentric spherical assumption from the 700-km downrange position will give paths that are slightly long for one-hop ranges. In the two-hop ranges the no-gradient assumption causes more distortion. In general, errors of this nature have little impact on performance prediction. However, near-real-time analysis for virtual range and azimuth correction to great-circle distance and bearing (grid registration) requires that tilt or gradient effects be taken into account. The daytime example has little horizontal gradient, and the simplifying assumption makes little difference. When better accuracy is desired, the correct vertical profile can be used for each radiation angle; also, gradients can be simulated by making the ionosphere nonconcentric with the earth. Both of these measures or something more complete should be used in radar performance assessment and management.
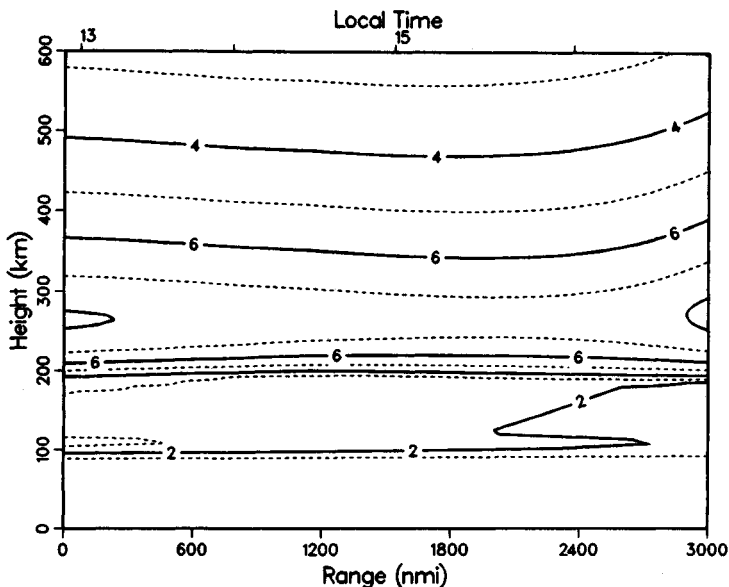
FIG. 24.14 Plasma-frequency contours are given, extending from the radar to a point east 3000 nmi downrange; January day example.

Figure 24.15 shows a performance prediction in the form of an oblique sounding. A typical sky-wave radar will be equipped with a vertical sounder and an oblique backscatter sounder for transmission-path analysis and to aid in radar-frequency management. Of course, the radar itself is an oblique sounder, but its sounding data is restricted to the frequency, waveform, and scan program of its primaiy surveillance task. An adjunct oblique sounder can present information in the form of Fig. 24.15 on earth backscatter echoes. In this prediction, SNR in decibels is plotted as a function of operating frequency and great-circle time delay or ground range. The numbers just above the abscissa (at 1-ms delay) are the noise powers in decibels below 1 W/Hz. For this plot the UTC time is 1800, SSN = 50, P_t = 200 kW, G_tG_r = 50 dB, T = 1 s, and σ = 20 dBsm. Figure 24.16 gives the corresponding night plot. The shape of these displays is quite similar to what would be seen with a diagnostic oblique sounding; the levels would gener-
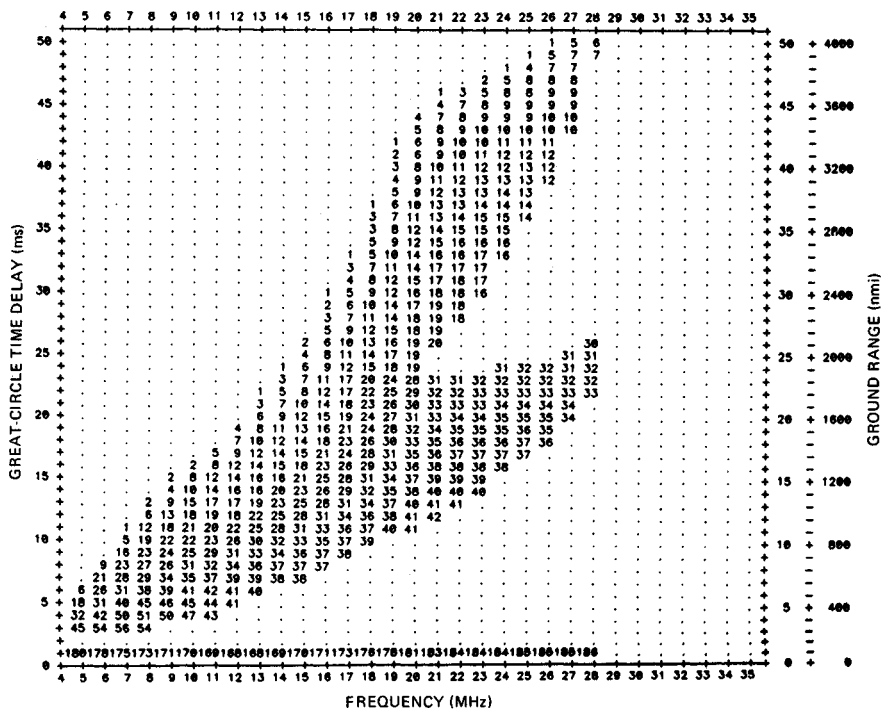
FIG. 24.15 *SNR* is given as a function of frequency and range in the form of a typical oblique backscatter sounding; January, 1800 GMT (day), SSN 50, location 38.65°N and 76.53°W, bearing 90°.

ally be greater since the resolution cell area times the surface scattering coefficient is generally much larger than 20 dBsm. Some of the night-day contrasts, such as available frequencies and difference in noise level for the same range, are evident. Also note that at night the 5-MHz lower frequency limit does not provide coverage closer than about 500 nmi. It should be remembered that this is a median SSN 50 calculation, and if consistent performance for ranges as close as 500 nmi is required during nights, a lower frequency limit should be selected to deal with periods of lower solar activity and the critical frequency distribution. The plots show that operation on a single frequency provides less than ±3 dB variation over a 500-nmi range interval. Also, if frequency selection had been made with a 2-MHz granularity instead of the 1 MHz used, the *SNR* would be reduced by only a decibel or so.

The performance-estimating aids that follow come from analyses as described above. After calculations as for oblique sounding, a range-ordered table of parameters is made. Parameter selections are made on the basis of the best *SNR* in each nominal 50-nmi interval, but the selection is adjusted to come from the adjacent lower frequency to avoid an optimistic bias. Then parameter plots are made as a function of range. The variables shown are losses, frequency, noise, and elevation radiation angle. The choice of range as the independent variable may seem artificial, but it is a useful approach for performance examination. With these curves the impact on radar *SNR* performance can be estimated for

GREAT-CIRCLE TIME DELAY (ms) *(vertical axis, values 0 to 50)*

GROUND RANGE (nmi) *(right vertical axis, 0 to 4000)*

FREQUENCY (MHz) *(horizontal axis, 4 to 35)*
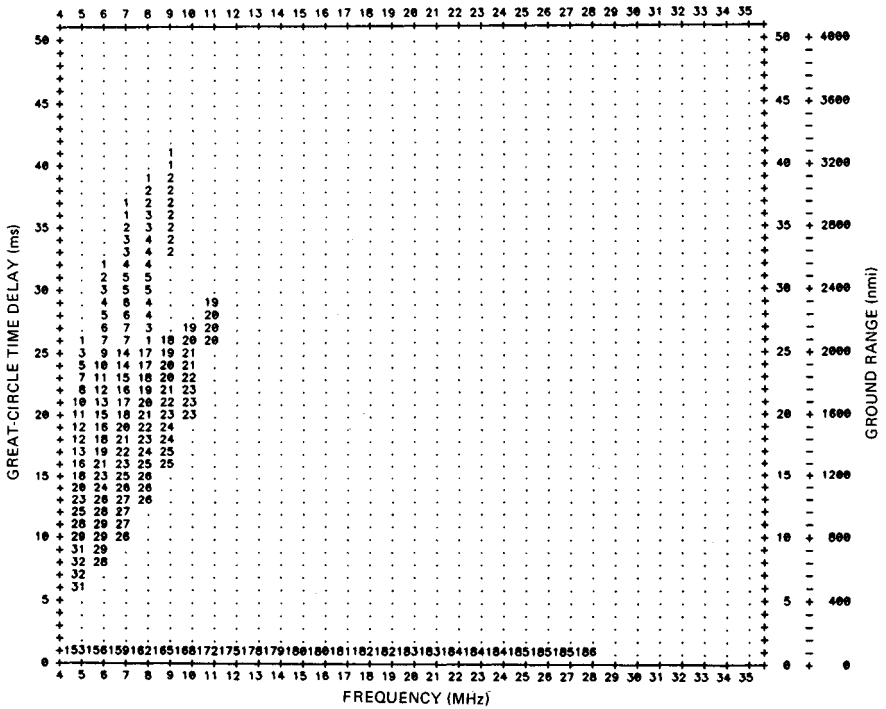
FIG. 24.16 *SNR* is given as a function of frequency and range in the form of an oblique sounding, as in Fig. 24.15, but for 0800 GMT (night).

selected antenna gain patterns, transmitted powers, target RCS, and coherent integration times (CIT). Figure 24.17 is an example for January with low solar activity in daytime. The R^4 loss is the fourth power of range to the target in meters given in decibels. $R^4 + L$ adds nondeviative absorption, deviative absorption, sporadic-E obscuration, and ground-reflection losses if there is more than one hop. The sharp increase in loss just before 2000 nmi is caused by transition from one to two hops; for two hops the lossy D region is transited twice as many times, ground-reflection loss is added, and required operation at a lower frequency increases loss. The jagged curve in the transition region is due to the parameter selection process; in radar operation the frequency would be selected to minimize transition effects. The frequency, radiation angle, and noise power per hertz that go with this site and look direction are also plotted. An example will be treated. It will be convenient to write the radar equation (24.1) in decibels:

$$SNR = P_{av} + G_t + G_r + \lambda^2 + T + \text{RCS} + F_p - (4\pi)^3 - (R^4 + L) - N$$

Select 1000 nmi as the range. Then the frequency is 17.5 MHz (wavelength = 17.1 m and $\lambda^2 = 25$ dB), noise power = -175 dB, and $R^4 + L = 261$ dB. Choose 53 dBW for P_{av}, 20 dB for G_t, 30 dB for G_r, 0 dBs for T, 20 dBsm for RCS, and 6 dB for F_p.

$$SNR = 53 + 20 + 30 + 25 + 0 + 20 + 6 - 33 - 261 - (-175) = 35 \text{ dB}.$$
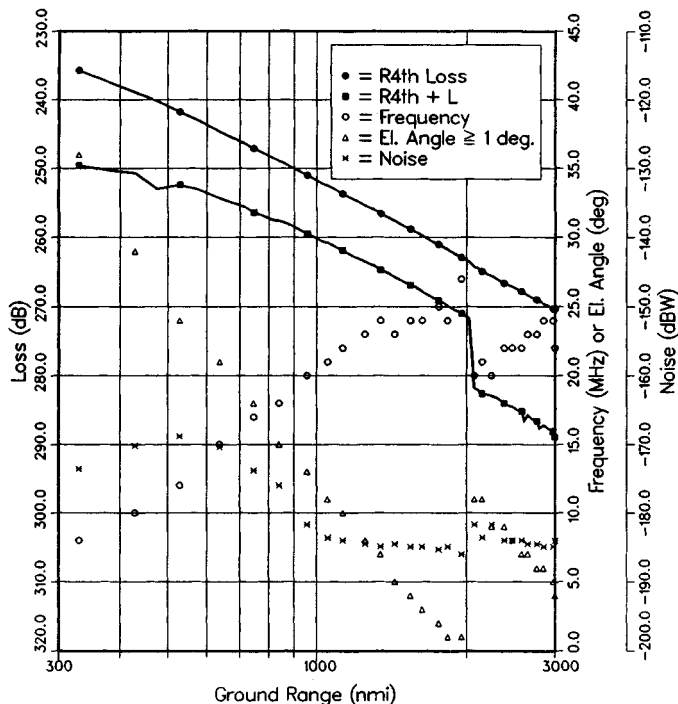
FIG. 24.17 Radar performance-controlling variables are given as a function of range; January, 1800 UTC, SSN 50.

Figure 24.18 shows the performance indicated with these assumptions for all ranges. A path factor enhancement of 6 dB has been chosen as an estimate of constructive multipath interference for an aircraft target that is effective in target detection. The beamwidth has been taken to be 5.7° and the surface scattering coefficient to be −35 dB, and with a 12 dB path enhancement the clutter level has been plotted. The clutter-to-noise ratio (*CNR*) at 1000 nmi is about 82 dB. For the constant beamwidth assumed, the clutter-to-signal ratio increases with range and is 47 dB at 1000 nmi. Large clutter-to-signal ratios are typical of HF radar; some form of doppler filtering is used to separate targets from clutter.

In Figs. 24.19 through 24.26 the performance-estimating curves are given for four seasons, night and day, and high and low solar activity. The permissible frequency selection is set between 5 and 28 MHz, and antenna radiation is not considered below an elevation angle of 1°. The analyses were made for a radar off the mid-Atlantic coast of the United States and should be a good approximation for any location where transmission paths are through the middle magnetic latitudes. This OTH performance presentation can be used to decide on the antenna patterns and powers required for specific targets and missions, or it can be used to exhibit periods of enhanced or degraded performance for an existing design. In looking at the performance index curves with radar range as the independent variable,
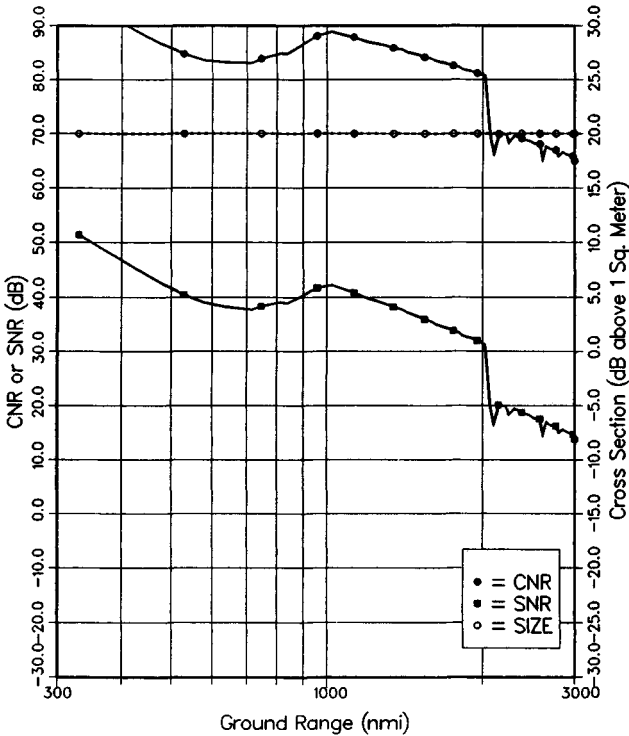
FIG. 24.18 A specific example of *SNR* and clutter-to-noise ratio (*CNR*) is determined by using Fig. 24.17. The target RCS is labeled "size" and is considered constant.

1. Summer shows much greater losses than other seasons.
2. Except for summer, night losses are only slightly less than day losses.
3. Night noise is much greater than day noise.
4. For a specific range, optimum frequencies vary by 3:1.

Several qualifiers should be kept in mind. At other geographic locations, the appropriate CCIR noise should be selected or, better yet, measured noise used. For radars that use auroral zone paths, specific analyses are required and target obscuration by spread-in-doppler clutter must be considered. The performance estimates from the figures assume that the radar design and waveforms are such that external noise is the control. The use of a single description for night and day gives a fair representation, but the transition from night to day is very abrupt and requires careful frequency management in radar operation. The ionospheric description that has been used is for what has been termed the *quiet ionosphere*; there will be a few hours per year when performance is very inferior to that predicted.

(a)

(b)
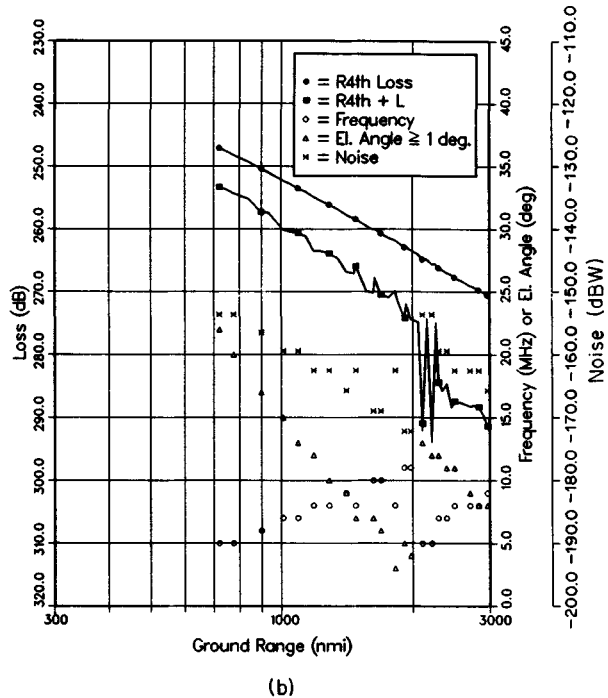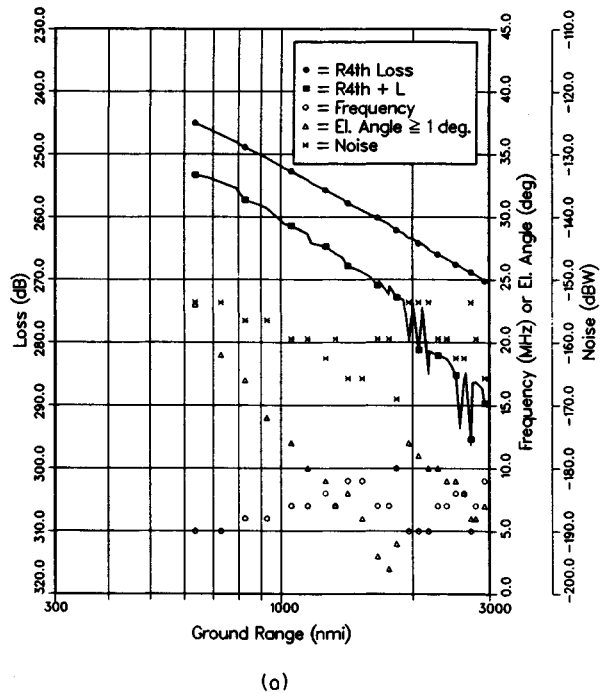
FIG. 24.19 Radar performance estimate; January, 0800 UTC. (*a*) SSN = 10. (*b*) SSN = 100.
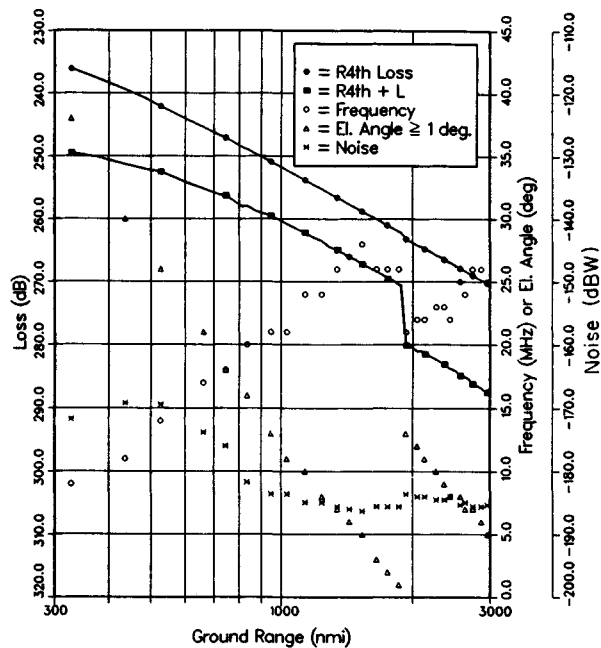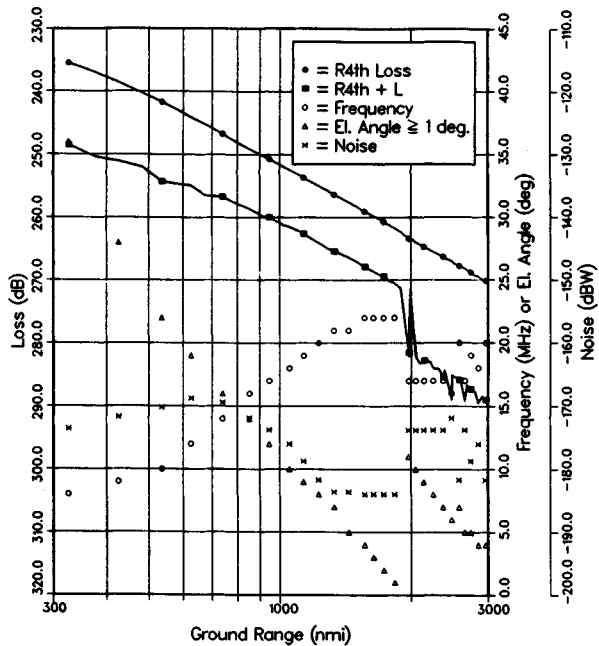
FIG. 24.20 Radar performance estimate; January, 1800 UTC. (*a*) SSN = 10. (*b*) SSN = 100.
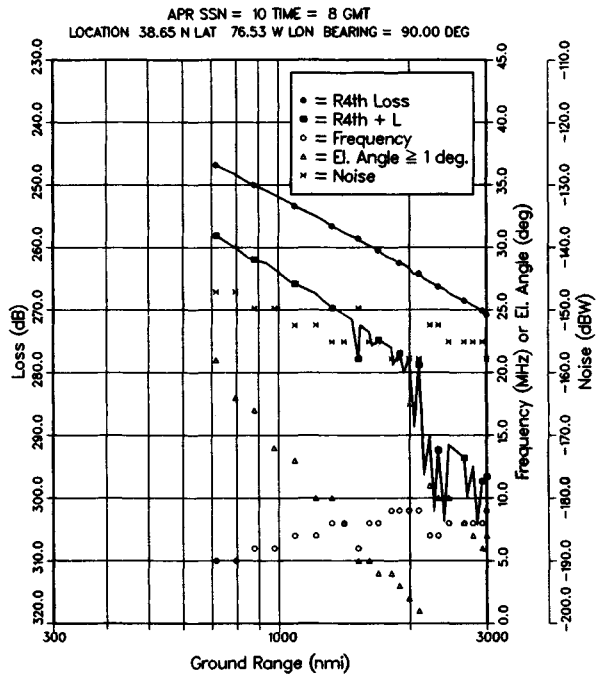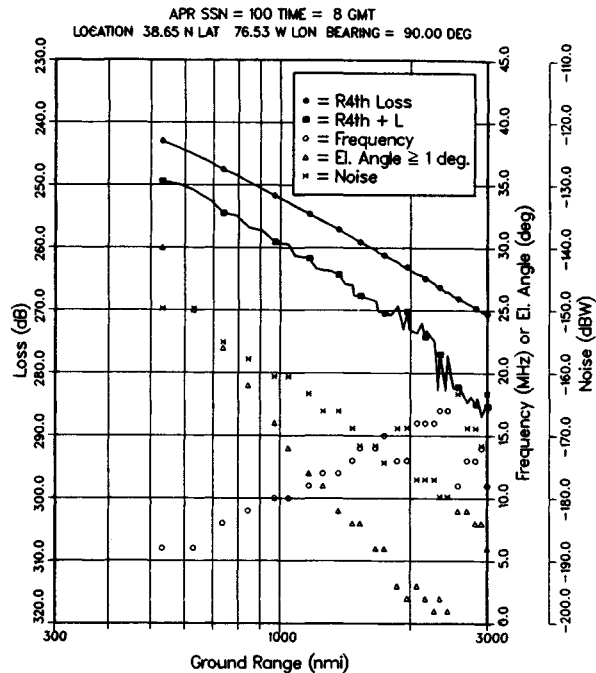
FIG. 24.21 Radar performance estimate; April, 0800 UTC. (*a*) SSN = 10. (*b*) SSN = 100.
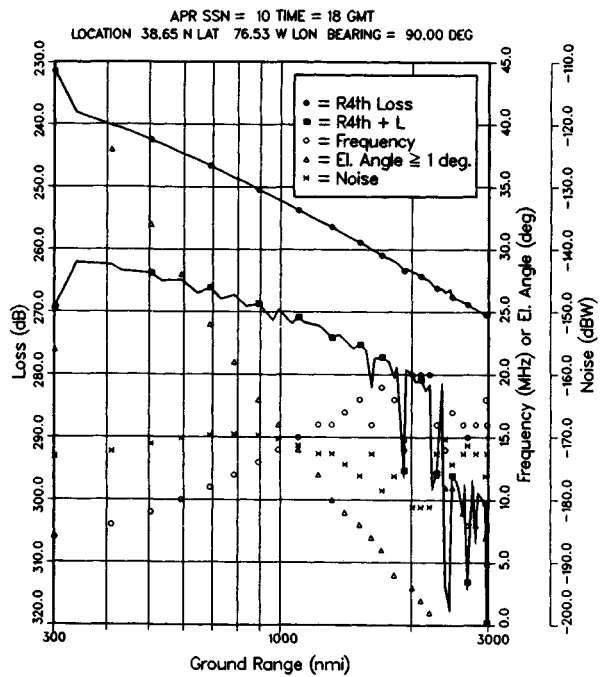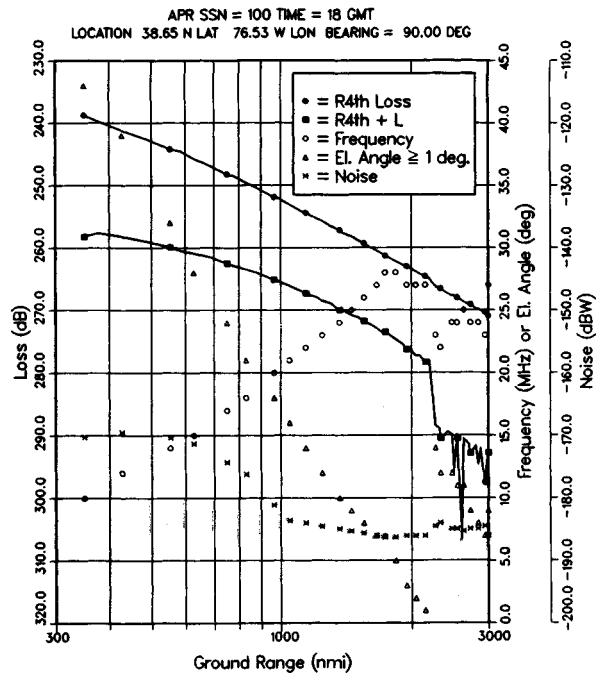
FIG. 24.22   Radar performance estimate; April, 1800 UTC. (a) SSN = 10. (b) SSN = 100.

(a)

(b)
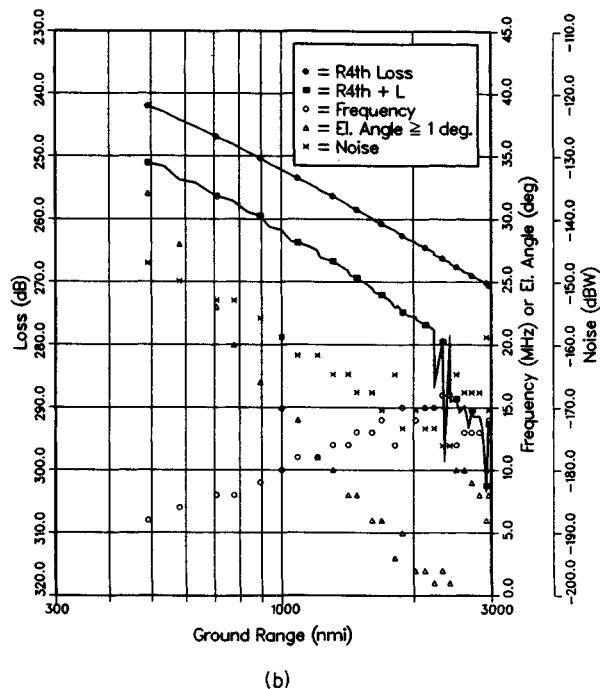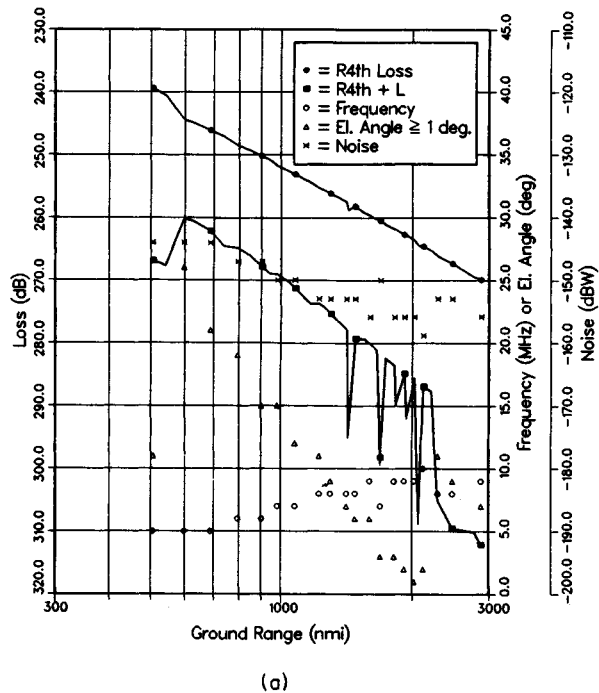
FIG. 24.23 Radar performance estimate; July, 0800 UTC. (*a*) SSN = 10. (*b*) SSN = 100.
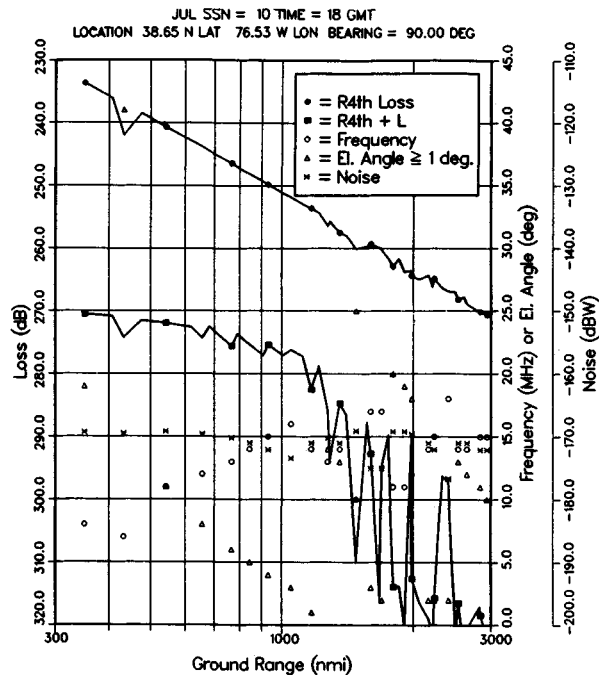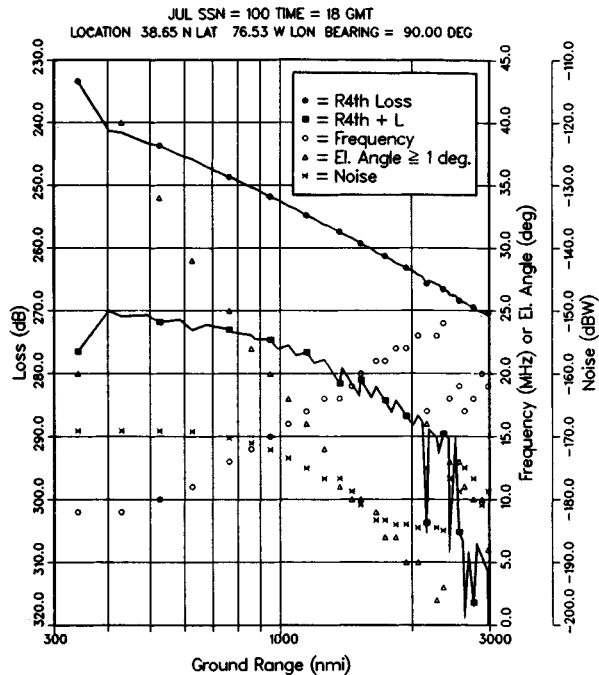
FIG. 24.24   Radar performance estimate; July, 1800 UTC. (a) SSN = 10. (b) SSN = 100.
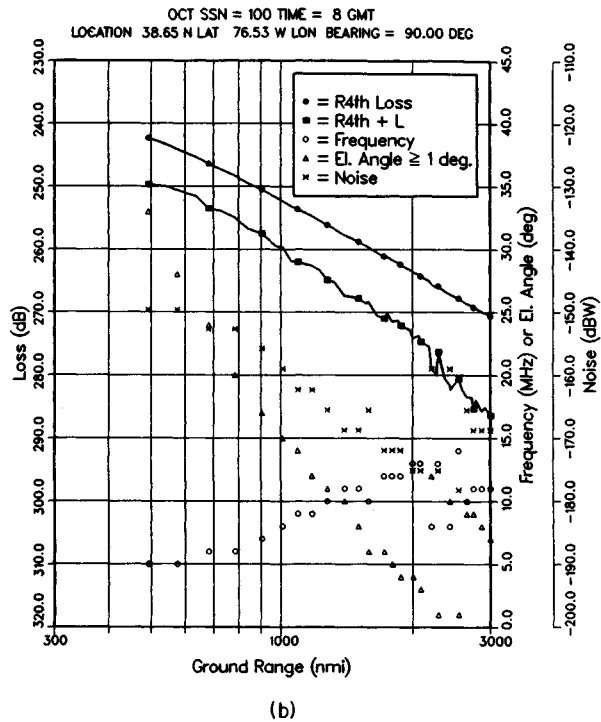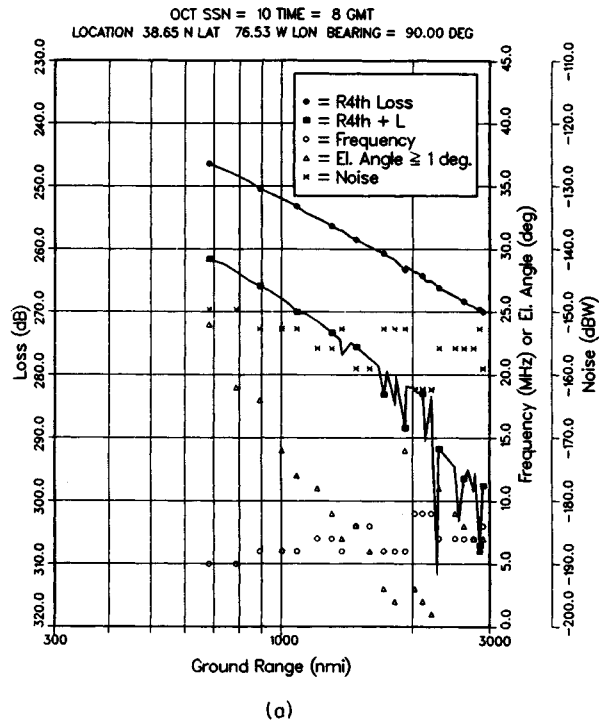
FIG. 24.25 Radar performance estimate; October, 0800 UTC. (*a*) SSN = 10. (*b*) SSN = 100.
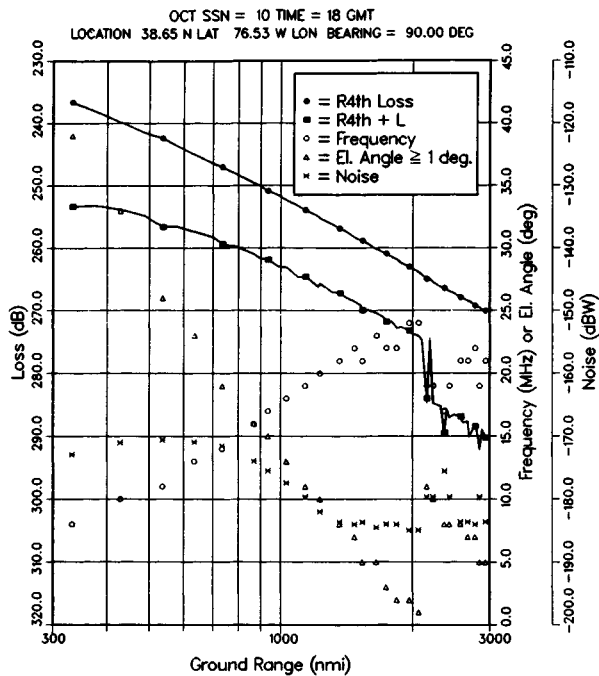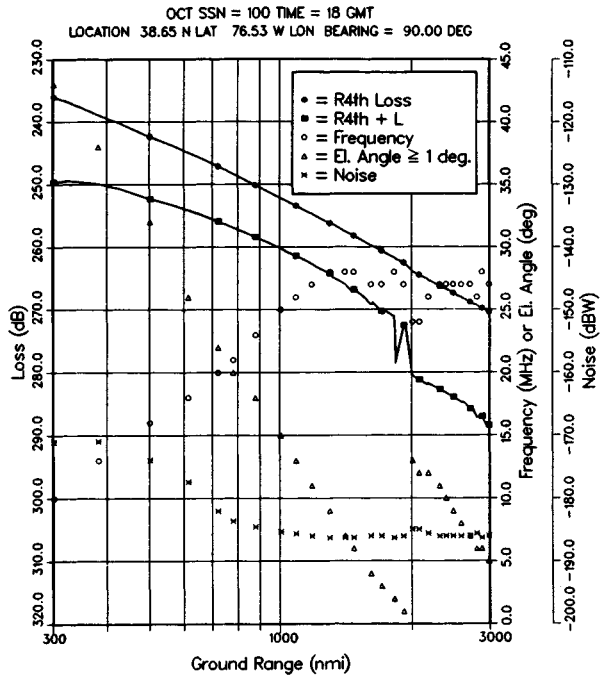
FIG. 24.26 Radar performance estimate; October, 1800 UTC. (*a*) SSN = 10. (*b*) SSN = 100.

24.11 RECEIVER-PROCESSOR

The prediction method indicated above is based upon long-term medians of vertical soundings, measured path losses, and observed noise. The statistical distributions are for a particular hour over days of the month or season. These kinds of statistics are insufficient to define the requirements for detection and tracking. An example will be given on the basis of a long dwell on a target using a constant frequency. The data will be used to indicate the required dynamic range and processor size and to show the input of the detection and tracking process. The amplitude levels in Fig. 24.27 are given in decibels relative to an arbitrary reference. Figure 24.27a gives a short time history of received power amplitude versus doppler frequency in one range gate. The waveform repetition frequency (wrf) was 20 Hz. Noise (N) samples were taken at wrf/2, target samples (T) on a target peak, and approach (A) and recede (R) on the resonant ocean wave peaks; N, T, A, and R are plotted in Fig. 24.27b. For this processing with a CIT of 12.8 s, the doppler filter bandwidth is a nominal 0.08 Hz, and at least 256 doppler filters should be used. The distance between the minimum noise points and the maximum clutter points is of the order of 100 dB, which indicates the dynamic range requirement if small targets are to be seen. For digital processing, an analog-to-digital (A/D) converter of at least 16-bit accuracy is in order. If data is processed

(a)
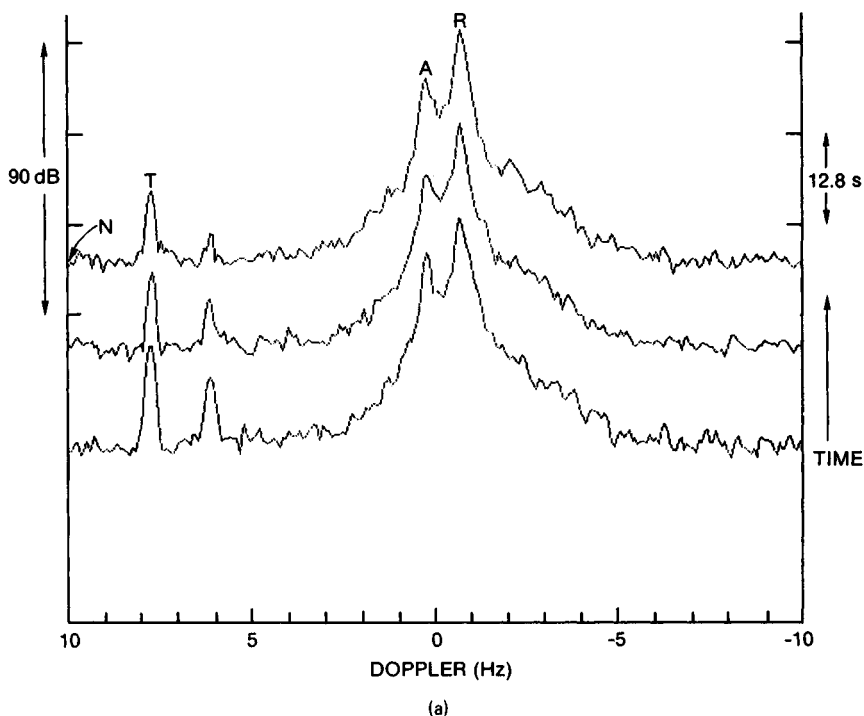
FIG. 24.27 (a) Amplitude versus doppler is plotted for a sequence of three 12.8-s coherent dwells. Indicated are a target marked T, the approach and recede resonant sea clutter peaks marked A and R, and the position for taking the noise sample marked N.
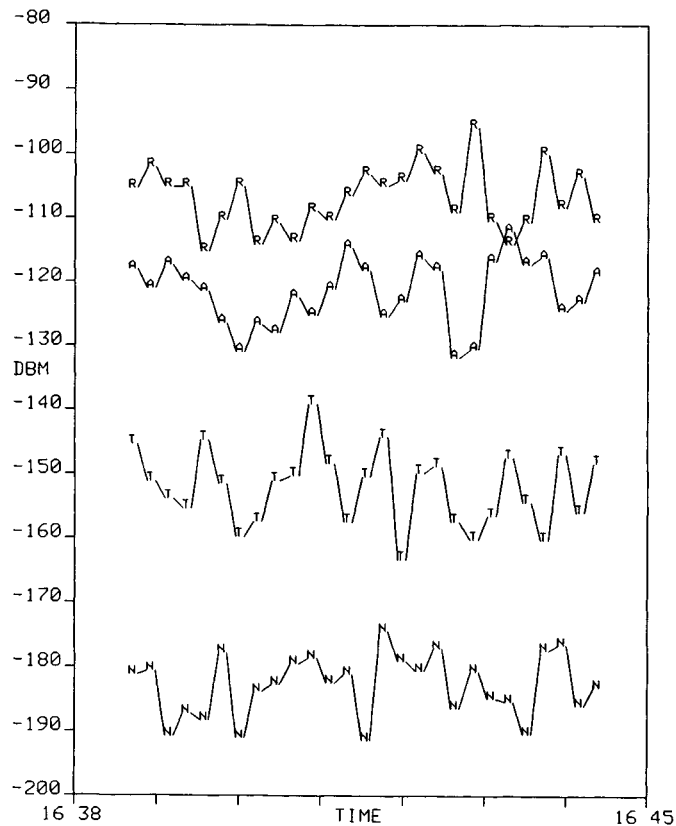
FIG. 24.27 (*Continued*) (*b*) Target, clutter peaks, and noise are plotted versus time for a longer sequence of data as in *a*.
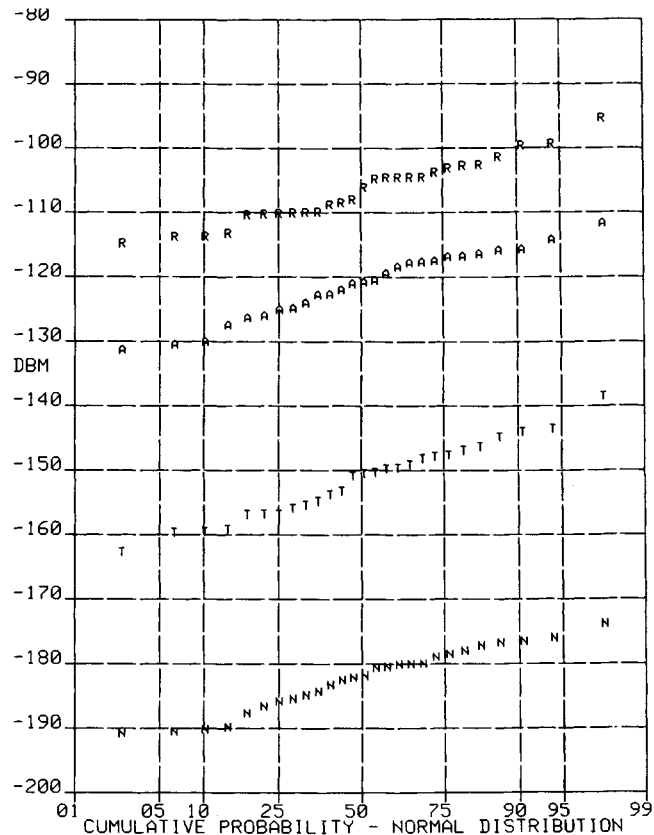
FIG. 24.27 (*Continued*) (*c*) Amplitude distributions of target, clutter peaks, and noise are plotted.

for a radar operating as shown in Fig. 24.1, there will be a 500-nmi range extent at a 10-kHz resolution bandwidth requiring about 60 range gates; I and Q processing to expose unambiguously target dopplers between $\pm$wrf/2 requires a sample rate of 40 kHz. For the Fig. 24.1 example, 16 simultaneous receiver-processor channels are required for the multiple receive beams. Figure 24.27c gives the corresponding power-level distributions. These approximately log-normal distributions are typical. The wide-area surveillance application makes automatic detection and tracking very desirable; the single transmitter footprint shown in Fig.24.1 has 800 receive range-azimuth cells. Tracker requirements differ from those of other sensors in that it is generally necessary to have thresholds that permit many natural responses. It is common practice to defer target declaration until a track is recognized and thereby to reduce the false-alarm rate.

24.12 GROUND-WAVE RADAR PERFORMANCE

Ground-wave propagation as defined here will include all but sky-wave paths. That is, the paths or illumination considered are direct line-of-sight and by sea-
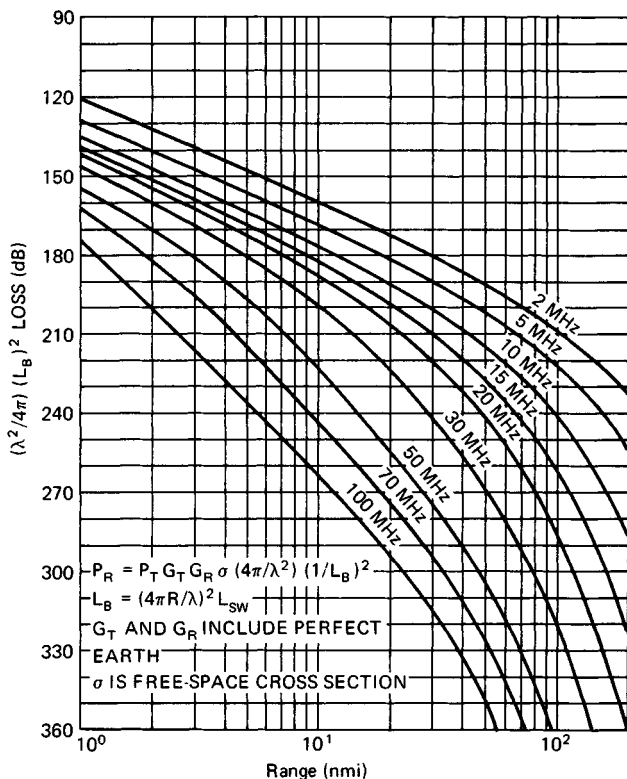
FIG. 24.28 Curves for estimating ground-wave radar performance are given as a function of range and are parametric in frequency. The surface is considered smooth, target and antenna heights are 2 m, conductivity is assumed to be 5 S/m, and the dielectric constant is 80.

surface reflection when radar and target are above the horizon and illumination in the penumbra and shadow region by a surface-attached wave. The feature that causes this propagation to be considered is that vertical polarization provides useful illumination down to the sea surface beyond the optical horizon. Figure 24.28 gives an example of ground-wave radar performance, parametric in frequency for the case in which both the radar antenna and the target are near the sea surface. These curves are for a smooth surface and use a 4/3 earth radius to approximate atmospheric refraction effects. The propagation code is due to Berry

$$\frac{F_D}{\sqrt{\dfrac{G}{\pi\lambda}}}$$
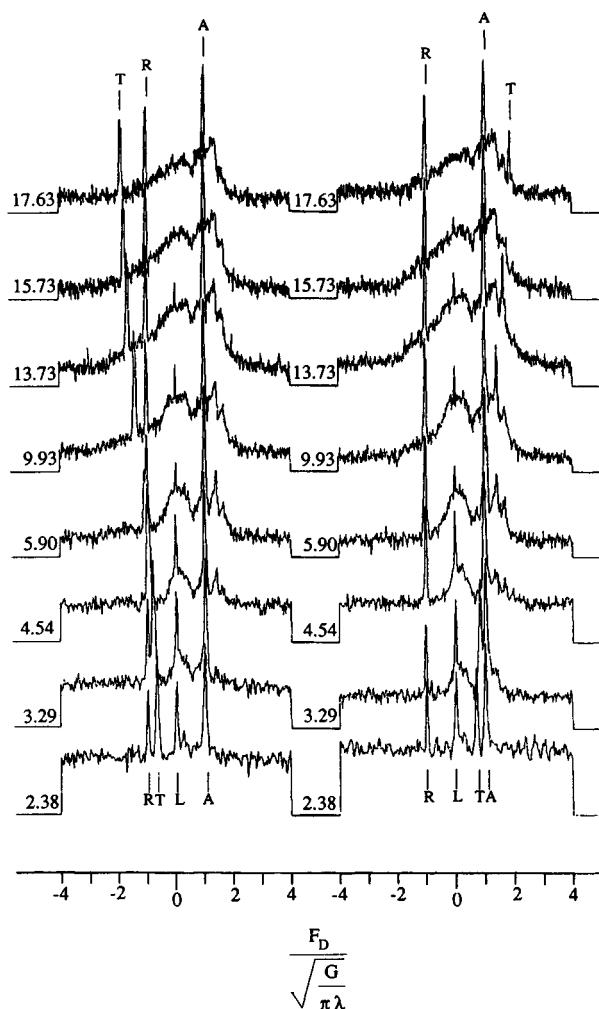
FIG. 24.29 Examples of a target T are shown when approaching (right) and receding (left) in the presence of the sea echo. The format is of received power versus normalized doppler for seven radar operating frequencies. The resonant wave or Bragg peaks are marked A for approaching and R for receding. The peak at zero frequency is due to a stationary target in an antenna sidelobe.

and Chrisman,[53] and it is quite flexible, permitting antenna and target altitudes, surface conductivity and permittivity, polarization, and frequency to be specified. Sea roughness can be taken into account by using the work of Barrick.[54] Path-loss descriptors such as shown in Fig. 24.28 can be used to estimate performance. For example, consider a radar at 5 MHz with an average power of 10 kW (40 dBW), a transmit-receive antenna gain product of 15 dB, and a target at 100 nmi with an RCS of 20 dBsm; then the received power

$$P_r = 40 + 15 + 20 - 222 = -147 \text{ dBW}$$

By using the January nighttime noise as given in Fig. 24.6b,

$$SNR = P_r - N = -147 + 153 = 6 \text{ dB}$$

And if 10-s coherent processing time is used,

$$SNR = 16 \text{ dB}$$

Figure 24.29 provides a display of a 13-kn target and the sea echo as seen by a ground-wave radar. This is a family of received power versus doppler frequency plots over seven operating frequencies and for a target both approaching and receding. The abscissa units are in doppler normalized to the resonant wave or Bragg frequency; therefore, the resonant wave responses peak at ±1. The amplitude range for each plot is 60 dB. The peaks at zero-doppler frequency are due to land in an antenna sidelobe. The target doppler coincides with a resonant line at 4.93 Mhz; it is between the lines below that frequency and outside them for frequencies above. The approach resonant wave peak is about 20 dB larger than the recede peak, indicating an almost direct sea into the radar. The processing used in developing these displays was 200-s CIT and 30-min averaging. This order of time for processing is appropriate for surface-target speeds and ground-wave radar; it also shows in doppler detail the form of the sea echo.

REFERENCES

1. Headrick, J. M., and M. I. Skolnik: Over-the-Horizon Radar in the HF Band, *Proc. IEEE*, vol. 62, pp. 664–673, June 1974.

2. Barnum, J. R.: Ship Detection with High Resolution HF Skywave Radar, *IEEE J. Ocean. Eng.*, vol. OE-11, pp. 196–210, April 1986.

3. Boutacoff, D. A.: Backscatter Radar Extends Early Warning Times, *Def. Electron.*, vol. 17, pp. 71–83, May 1985.

4. Sinnott, D. H.: The Jindalee Over-the-Horizon Radar System, *Conf. Air Power in the Defence of Australia*, Australian National University, Research School of Pacific Studies, Strategic and Defence Studies Centre, Canberra, July 14–18, 1986.

5. Wylder, J.: The Frontier for Sensor Technology, *Signal*, vol. 41, pp. 73–76, March 1987.

6. Guest editorial and invited papers in special issue on high-frequency radar for ocean and ice mapping and ship location: *IEEE J. Ocean. Eng.*, vol. OE-11, April 1986.

7. Kolosov, A. A. (ed.): "Fundamentals of Over-the-Horizon Radar," in Russian, *Radio i svyaz*, 1984. Also a translation by W. F. Barton, Artech House, Norwood, Mass., 1987.

8. Greenwald, R. A., K. B. Baker, R. A. Hutchins, and C. Hanuise: An HF Phased Array Radar for Studying Small-Scale Structure in the High Latitude Ionosphere, *Radio Sci.*, vol. 20, pp. 63–79, January–February 1985.

9. ITT Avionics Division, Electro-Physics Laboratories: EPL Model ATL-75 Transmitter for Radar and Communication, *IR&D Project Rept.* 274, *Results of Performance Measurements*, January 1975.

10. Hoft, D. J., and Fuat Agi: Solid State Transmitters for Modern Radar Applications, *CIE Int. Radar Conf. Rec.*, pp. 775–781, Nov. 4–7, 1986.

11. Guest editorial and invited papers in special issue on shortwave broadcasting, *IEEE Trans. Broadcast.*, vol. 34, June 1988.

12. Sweeney, L. E.: Spatial Properties of Ionospheric Radio Propagation as Determined with Half-Degree Azimuthal Resolution, *Stanford Electron. Lab. Tech. Rept.* 155 *SU-SEL*-70-034, Stanford University, June 1970.

13. Lynch, J. T.: Aperture Synthesis for HF Radio Signals Propagated via the F-Layer of the Ionosphere, *Stanford Electron. Lab. Tech. Rept.* 161 *SU-SEL*-70-066, Stanford University, September 1970.

14. Johnson, R. C., and H. Jasik (eds.): "Antenna Engineering Handbook," 2d ed., McGraw-Hill Book Company, New York, 1984.

15. Kurashov, A. G. (ed.): "Shortwave Antennas," 2d ed., in Russian, *Radio i svyaz*, January 1985.

16. Rice, S. O.: Reflection of Electromagnetic Waves from Slightly Rough Surfaces, in Kline, M. (ed.): "Theory of Electromagnetic Waves," Interscience Publishers, New York, 1951, pp. 351–378.

17. Pierson, W. J., G. Neumann, and R. W. James: Practical Methods for Observing and Forecasting Ocean Waves by Means of Wave Spectra and Statistics, *H.O. Pub.* 603, chap. 1, Hydrographic Office, U.S. Navy, 1960.

18. Crombie, D. D.: Doppler Spectrum of the Sea Echo at 13.56 Mcs, *Nature*, vol. 175, pp. 681–682, 1955.

19. Barrick, D. E.: First Order Theory and Analysis of MF/HF/VHF Scatter from the Sea, *IEEE Trans.*, vol. AP-20, pp. 2–10, January 1972.

20. Pierson, W. J., and L. Moskowitz: A Proposed Spectral Form for Fully Developed Wind Seas Based on the Similarity Theory of S. A. Kitaigordskii, *J. Geophys. Res.*, vol. 69, no. 24, pp. 5181–5190, 1964.

21. Barrick, D. E., J. M. Headrick, R. W. Bogle, and D. D. Crombie: Sea Backscatter at HF: Interpretation and Utilization of the Echo, *Proc. IEEE*, vol. 62, pp. 673–680, June 1974.

22. Clancy, R. M., J. E. Kaitala, and L. F. Zambresky: The Fleet Numerical Oceanography Center Global Spectral Ocean Wave Model, *Bull. Am. Meteorol. Soc.*, vol. 67, no. 5, May 1986.

23. Long, A. E., and D. B. Trizna: Mapping of North Atlantic Winds by HF Radar Sea Backscatter Interpretation, *IEEE Trans.*, vol. AP-21, pp. 680–685, September 1973.

24. Ahearn, J. L., S. R. Curley, J. M. Headrick, and D. B. Trizna: Tests of Remote Skywave Measurment of Ocean Surface Conditions, *Proc. IEEE*, vol. 62, pp. 681–686, June 1974.

25. Anderson, S. J.: Remote Sensing with the Jindalee Skywave Radar, *IEEE J. Ocean. Eng.*, vol. OE-11, pp. 158–163, April 1986.

26. Trizna, D. B., J. C. Moore, J. M. Headrick, and R. W. Bogle: Directional Sea Spectrum Determination Using HF Doppler Radar Techniques, *IEEE Trans.*, vol. AP-25, pp. 4–11, January 1977.

27. Barrick, D. E.: Extraction of Wave Parameters from Measured HF Radar Sea-Echo Spectra, *Radio Sci.*, vol. 12, no. 3, p. 415, 1977.

28. Lipa, B.: Derivation of Directional Ocean-Wave Spectra by Integral Inversion of Second-Order Radar Echoes, *Radio Sci.*, vol. 12, no. 3, p. 425, 1977.

29. Trizna, D. B.: Estimation of the Sea Surface Radar Cross Section at HF from Second-Order Doppler Spectrum Characteristics, *Naval Res. Lab. Rept.* 8579, May 1982.

30. Pilon, R. O., and J. M. Headrick: Estimating the Scattering Coefficient of the Ocean Surface for High-Frequency Over-the-Horizon Radar, *Naval Res. Lab. Memo. Rept.* 5741, May 1986.

31. Trizna, D. B., and J. M. Headrick: Ionospheric Effects on HF Over-The-Horizon Radar, in Goodman, J. M. (ed.): *Proc. Effect Ionosphere on Radiowave Syst.*, ONR/AFGL-sponsored, pp. 262–272, Apr. 14–16, 1961.

32. Elkins, T. J.: A Model for High Frequency Radar Auroral Clutter, *RADC Rept. TR-80-122*, March 1980.

33. Burke, G. J., and A. J. Poggio: Numerical Electromagnetic Code (NEC)—Method of Moments, *NOSC Tech. Doc.* 116, 1981.

34. Walton, E. K., and J. D. Young: The Ohio State University Compact Radar Cross Section Measurement Range, *IEEE Trans.*, vol. AP-32, pp. 1218–1223, November 1984.

35. Bogle, R. W., and D. B. Trizna: Small Boat Radar Cross Sections, *Naval Res. Lab. Memo. Rept.* 3322, July 1976.

36. CCIR (International Radio Consultative Committee): World Distribution and Characteristics of Atmospheric Radio Noise, *CCIR Rept.* 322, International Telecommunications Union, 1964.

37. Spaulding, A. D., and J. S. Washburn: Atmospheric Radio Noise: Worldwide Levels and Other Characteristics, *NTIA Rept.* 85-173, National Telecommunications and Information Administration, April 1985.

38. Lucas, D. L., and J. D. Harper: A Numerical Representation of CCIR Report 322 High Frequency (3–30 Mcs) Atmospheric Radio Noise Data, *Nat. Bur. Stand. Note* 318, Aug. 5, 1965.

39. Ortenburger, L. N., D. A. Schaefer, F. W. Smith, and A. J. Kramer: Prediction of HF Noise Directivity from Thunderstorm Probabilities, *GTE Sylvania Rept. EDL-M1379*, 1971.

40. Lucas, D. L.: Predictions of Backscatter Clutter Power in the Radar C Computer Program, final report on NRL Contract N0014-84-C-2451; CU5-36903, University of Colorado, Boulder, June 9, 1986.

41. Davies, K.: "Ionospheric Radio Propagation," *Nat. Bur. Stand. Monog.* 80, Apr. 1, 1965.

42. Jursa, A. S. (ed.): "Handbook of Geophysics and the Space Environment," Air Force Geophysics Laboratory, AFSC, U.S. Air Force, 1985.

43. Lucas, D. L., and G. W. Haydon: Predicting Statistical Performance Indexes for High Frequency Telecommunications Systems, *ESSA Tech. Rept. IER* 1 *ITSA* 1, U.S. Department of Commerce, 1966.

44. Barghausen, A. L., J. W. Finney, L. L. Proctor, and L. D. Schultz: Predicting Long-Term Operational Parameters of High-Frequency Sky-Wave Communications Systems, *ESSA Tech. Rept. ERL* 110-*ITS* 78, U.S. Department of Commerce, 1969.

45. Headrick, J. M., J. F. Thomason, D. L. Lucas, S. McCammon, R. Hanson, and J. Lloyd: Virtual Path Tracing for HF Radar Including an Ionospheric Model, *Naval Res. Lab. Memo. Rept.* 2226, March 1971.

46. Teters, L. R., J. L. Lloyd, G. W. Haydon, and D. L. Lucas: Estimating the Performance of Telecommunication Systems Using the Ionospheric Transmission Channel—

Ionospheric Communications Analysis and Prediction Program Users Manual, *Nat. Telecom. Inf. Adm. NTIA Rept.* 83-127, July 1983.

47. Hatfield, V. E.: HF Communications Predictions 1978 (An Economical Up-to-Date Computer Code, AMBCOM), *Solar Terrestrial Preduction Proc.*, vol. 4, in Donnelley, R. F. (ed.): "Prediction of Terrestrial Effects of Solar Activity," National Oceanic and Atmospheric Administration, 1980.

48. Lucas, D. L.: Ionospheric Parameters Used in Predicting the Performance of High Frequency Skywave Circuits, Interim Report on NRL Contract N00014-87-K-20009, Account 153-6943, University of Colorado, Boulder, Apr. 15, 1987.

49. Thomason, J., G. Skaggs, and J. Lloyd: A Global Ionospheric Model, *Naval Res. Lab. Rept.*, 8321, Aug. 20, 1979.

50. Miller, D. C., and J. Gibbs: Ionospheric Analysis and Ionospheric Modeling, *AFCRL Tech. Rept.* 75-549, July 1975.

51. Jones, R. M., and J. J. Stephenson: A Versatile Three-Dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere, *Office Telecom. Rept.* 75-76, October 1975.

52. Lucas, D. L., J. L. Lloyd, J. M. Headrick, and J. F. Thomason: Computer Techniques for Planning and Management of OTH Radars, *Naval Res. Lab. Memo. Rept.* 2500, September 1972.

53. Berry, L. A., and M. E. Chrisman: A FORTRAN Program for Calculation of Ground Wave Propagation Over Homogeneous Spherical Earth for Dipole Antennas, *Nat. Bur. Stand. Rept.* 9178, 1966.

54. Barrick, D. E.: Theory of HF and VHF Propagation across the Rough Sea, pts. 1 and 2, *Radio Sci.*, vol. 6, pp. 517–533, May 1971.

CHAPTER 25
BISTATIC RADAR

Nicholas J. Willis
Technology Service Corporation

25.1 CONCEPT AND DEFINITIONS

Bistatic radar employs two sites that are separated by a considerable distance. A transmitter is placed at one site, and the associated receiver is placed at the second site. Target detection is similar to that of monostatic radar: target illuminated by the transmitter and target echoes detected and processed by the receiver. Target location is similar to but more complicated than that of a monostatic radar: total signal propagation time, orthogonal angle measurements by the receiver, and some estimate of the transmitter location are required to solve the transmitter-target-receiver triangle, called the *bistatic triangle*. Continuous-wave (CW) waveforms can often be used by a bistatic radar because site separation, possibly augmented by sidelobe cancellation, provides sufficient spatial isolation of the *direct-path* transmit signal.

When separate transmit and receive antennas are at a single site, as is common in CW radars, the term *bistatic* is not used to describe such a system since the radar has characteristics of a monostatic radar. In special cases, the antennas can be at separate sites and the radar is still considered to operate monostatically. For example, an over-the-horizon (OTH) radar can have site separation of 100 km or more. But that separation is small compared with the target location of thousands of kilometers,[1,2] and the radar operates with monostatic characteristics.

When two or more receive sites with common spatial coverage are employed and target data from each site is combined at a central location, the system is called a *multistatic radar*. Thinned, random, distorted, and distributed arrays,[3-6] interferometric radars,[7-10] the radio camera,[11,12] and the multistatic measurement system[13,14] are sometimes considered a subset of multistatic radars. They usually combine data coherently from each receiver site to form a large receive aperture. Multiple transmitters can be used with any of these configurations. They can be located at separate sites or colocated with the receive sites. Three range-only monostatic radars combined in a radar net are sometimes called a *trilateration radar*. The trilateration concept applies to multistatic radars that measure target location by time-difference-of-arrival (TDOA) or differential doppler techniques.

The foregoing definitions are broad and traditional[1,15,16] but are by no means uniformly established in the literature. Terms such as *quasi-bistatic, quasi-*

monostatic, pseudo-monostatic, tristatic, polystatic, real multistatic, multi-bistatic, and *netted bistatic* have also been used.[17-20] They are usually special cases of the broad definitions given above.

Passive receiving systems, or electronic support measure (ESM) systems, often use two or more receiving sites. Their purpose is typically to detect, identify, and locate transmitters such as monostatic radars. They are also called *emitter locators*. Target location is by means of combined angle measurements from each site (e.g., triangulation), TDOA, and/or differential doppler measurements between sites. These systems usually are not designed to detect and process the echoes from targets illuminated by the transmitter. They can, however, be used with a bistatic or multistatic radar to identify and locate a suitable transmitter to initialize radar operations. Thus, while they have many requirements and characteristics common to multistatic radars, they are not radars and will not be considered here.

25.2 HISTORY

Early experimental radars in the United States, the United Kingdom, France, the Soviet Union, Germany, and Japan were of the bistatic type, where the transmitter and receiver were separated by a distance comparable to the target distance.[21-26] These bistatic radars used CW transmitters and detected a beat frequency produced between the direct-path signal from the transmitter and the doppler-frequency-shifted signal scattered by a moving target. This effect was called CW wave interference.[1] The geometry was similar to that of the forward-scatter (or near-forward-scatter) configuration, where the target position is near the baseline joining transmitter and receiver. Much of the early bistatic radar technology was derived from existing communications technology—separated sites, CW transmissions, and frequencies ranging from 25 to 80 MHz.[27] These early bistatic radars were typically configured as fixed, ground-based fences to detect the *presence* of aircraft: a major, emerging threat in the 1930s. The problem of extracting target *position* information from such radars could not readily be solved with techniques available at the time.[1]

Many of the early United States bistatic radar experiments were conducted by the Naval Research Laboratory (NRL).[1] In 1922 NRL researchers detected a wooden ship using a CW wave interference radar operating at 60 MHz. An NRL proposal for further work was rejected. In 1930 an aircraft was accidentally detected when it passed through a 33-MHz direction-finding beam received by an aircraft on the ground. Interest was revived, and in 1932 CW wave interference equipment detected an aircraft up to 80 km from the transmitter. In 1934 this work was disclosed in a patent, granted to Taylor, Young, and Hyland.[21]

In the Soviet Union an operational system, the RUS-1, evolved from an experimental bistatic CW radar.[24] By the time of the German invasion in 1941, 45 systems had been built and deployed to the Far East and the Caucasus. They were subsequently replaced by the RUS-2 and RUS-2C, both pulsed radars. The RUS-2 used two trucks, one for the transmitter and one for the receiver, separated by about 300 m to provide receiver isolation. Although the RUS-2 used two sites, separation was not sufficient to define the configuration as bistatic. The French also deployed a bistatic CW radar in a two-fence con-

figuration prior to World War II, thus providing a coarse estimate of target course and speed.[24]

The Japanese deployed about 100 bistatic CW radar fences, called Type A, starting in 1941.[26] These remained in use until the end of World War II. Type A operated between 40 and 80 MHz with 3 to 400 W of transmitter power. Maximum detection ranges of up to 800 km on aircraft were achieved, with one system operating between Formosa (Taiwan) and Shanghai. Target location along the forward-scatter baseline was never achieved with this system.

A variation of these fence configurations was developed by the Germans during World War II.[17] They built a bistatic receiver, known as the Klein Heidelberg, that used a British Chain Home radar as the transmitter. The receiver gave warning of the onset of Allied bombing raids when the planes were over the English Channel, without endangering the German ground sites. This bistatic radar appears to be the first operational configuration to use a noncooperative transmitter.

The Chain Home radars themselves operated with separate transmitter and receiver sites, but again with separation small compared with target distance. However, they had a standard, reversionary mode in which, in the presence of electronic countermeasures (ECM) or a transmitter failure, a receiver site could operate with a transmitter at an adjacent site, hence becoming bistatic.[28]

The invention of the duplexer at NRL in 1936 provided a means of using pulsed waveforms with a common transmit and receive antenna. This single-site configuration is the familiar monostatic radar, and it greatly expanded the utility of radar, particularly for use by aircraft, ships, and mobile ground units. As a consequence bistatic radars became dormant.

It was not until the early 1950s that interest in bistatic radars was revived for aircraft detection.[1,29–31] The United States AN/FPS-23 was designed as a gap-filler fence for the Distant Early Warning (DEW) line in the arctic. It was installed in the mid-1950s but was later removed.[24] The Canadians also developed a bistatic radar for their McGill fence.[29,32] The United States Plato and Ordir ballistic missile detection systems were designed as the first multistatic radars; they combined range sum and doppler information from each receiver site to estimate target position. They were not deployed.[24,32]

The Azuza, Udop, and Mistram interferometeric radars, a variant of multistatic radars, were installed at the United States Eastern Test Range for precision measurement of target trajectories. They used a single CW transmitter, multiple receivers at separate, precisely located sites, and cooperative beacon transponders on the target.[9,10] The SPASUR, a satellite fence interferometric radar, was also implemented with a single CW transmitter and multiple receivers but with enough performance to detect satellite-skin echoes.[7,8]

A major development at this time was the semiactive homing missile seeker, in which the large, heavy, and costly transmitter could be off-loaded from the small, expendable missile onto the launch platform (Chap. 19). While these seekers are clearly a bistatic radar configuration, missile engineers have developed a different lexicon to describe their technology and operation, e.g., semiactive versus bistatic, illuminator versus transmitter, rear reference signal versus direct-path signal, etc. The missile and radar communities continue to go their separate ways.

In the 1950s and early 1960s bistatic radar system theory was codified.[15] Bistatic radar cross-section theory was developed, and measurements were taken.[33–41] Bistatic clutter measurements were also taken.[42,43] The name *bistatic radar* was coined by K. M. Siegel and R. E. Machol in 1952.[34]

Bistatic radars received renewed interest in the 1970s and 1980s as counters to retrodirective jammers and attacks by antiradiation missiles (ARMs). Retrodirective jamming levels can be reduced by selecting a geometry such that the receive site lies outside the jammer's main beam, which is directed at the transmit site. The effectiveness of an ARM attack can be reduced by removing the transmitter from the battle area into a "sanctuary," which is less vulnerable to attack. Several air defense field test programs explored these capabilities and the problems inherent in bistatic operation, such as time synchronization, coverage, and clutter suppression.[18,44-50]

Other bistatic radar concepts were identified and tested at this time, such as *clutter tuning* from an airborne transmitter and receiver.[51-53] One potential implementation of this concept allows the receiver to generate a synthetic aperture radar (SAR) map of modest resolution directly on its velocity vector—an impossible task for the monostatic SAR. Clutter tuning combined with the sanctuary concept protects the transmitter while allowing the receiver platform to fly toward the target with no radar emissions.

The concept of using a small bistatic receiver that "hitchhikes" off airborne radars was also developed and successfully tested.[54] It alerts and cues autonomous short-range air defense and ground surveillance systems to improve survivability and acquisition performance. This hitchhiking concept was extended to other transmitters of opportunity, including a commercial television station that served as a bistatic transmitter. Initial attempts to detect aircraft were only marginally successful.[55]

Bistatic radars using space-based transmitters and receivers that are either space-based, airborne, or ground-based have been studied.[3,56-59] Limited field tests were conducted by using a communication satellite as the transmitter and a ground-based receiver to detect aircraft.[58] Since the effective radiated power of the satellite was modest and the transmitter-to-target ranges were large, detection ranges were small, <4 km, unless a very large receive aperture was used.

A pulse doppler bistatic radar was developed and tested to protect military aircraft on the ground from intruders.[60] It was configured for near-forward-scatter operation. Five small portable transmitter-receiver units, typically separated by 65 m, were located around the aircraft, with one transmitter servicing an adjacent receiver. In field tests the radar detected moving targets, including high-speed vehicles and intruders creeping at 2 cm/s.

The Multistatic Measurement System (MMS) was installed at the United States Kwajalein Missile Range in 1980 to track ballistic missile skin echoes.[14] The TRADEX L-band and ALTAIR ultrahigh-frequency (UHF) monostatic radars are used to illuminate the targets, and the bistatic echoes, collected at two unmanned stations located about 40 km from the radars, are combined coherently at a central site. The system is projected to measure three-dimensional position and velocity with accuracies better than 4 m and 0.1 m/s, respectively, throughout reentry.[13]

Other multistatic radar concepts have been studied. They include the Doppler Acquisition System (DAS), which used multiple transmitters and receivers,[61] and Distributed Array Radar (DAR) concepts, with large[3] and small[5] spatial separation between receive sites. The DAS combines data from each site noncoherently; the DAR, coherently.

Bistatic radars have been analyzed, proposed, and in some cases developed for other than military applications. Such applications include high-resolution imaging at short ranges (in the near field of the antennas) for use by robotics in an industrial environment;[62] airport ground vehicle and aircraft collision warning

and avoidance using a baseband bistatic radar;[63] planetary surface and environment measurements using a satellite-based transmitter and an earth-based receiver[64–67] or a planet-based transmitter and a satellite-based receiver;[68] geological probing of horizontally stratified, underground layers from a transmitter and receiver on the surface, usually operating at frequencies from 100 to 1000 MHz;[69] ocean wave spectral measurements (wavelength, frequency, and direction of travel) using a Loran-A system;[70] and detection and soundings of tropospheric layers, ionospheric layers, and high-altitude, clear-air atmospheric targets using ground-based sites.[16,71,72]

25.3 COORDINATE SYSTEM

A two-dimensional north-referenced coordinate system[73] is used throughout this chapter. Figure 25.1 shows the coordinate system and parameters defining bistatic radar operation in the x,y plane. This is sometimes called the *bistatic plane*.[74] The bistatic triangle lies in the bistatic plane. The distance L between the transmitter and the receiver is called the *baseline range* or simply the *baseline*. The angles θ_T and θ_R are, respectively, the transmitter and receiver look angles. They are also called angles of arrival (AOA) or lines of sight (LOS). Note that the bistatic angle $\beta = \theta_T - \theta_R$. It is also called the cut angle or the scattering angle. It is convenient to use β in calculations of target-related parameters and θ_T or θ_R in calculations of transmitter- or receiver-related parameters. Development of three-dimensional bistatic coordinate systems for some applications is available elsewhere.[16,46,48,75,76]

A useful relationship is that the bisector of the bistatic angle is orthogonal to
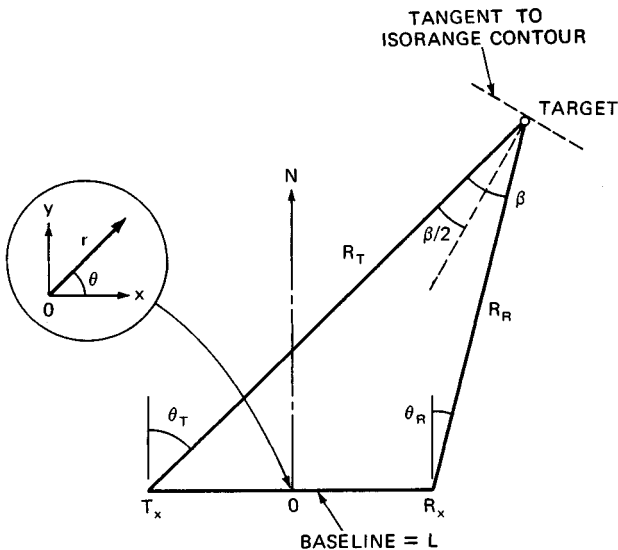
FIG. 25.1 Bistatic radar north coordinate system for two dimensions.[73]

the tangent of an ellipse with foci at the transmitter and receiver sites and passing through the target position. Such an ellipse is called an *isorange contour*. The tangent is often a good approximation to an isorange contour within the bistatic footprint, the area common to the transmit and receive beams.

Geometry often distinguishes bistatic from monostatic radar operation. In these distinguishing cases, equivalent monostatic operation is obtained by setting $L = O$, or $R_T = R_R$ and $\beta = 0$.

25.4 RANGE RELATIONSHIPS

Range Equation.[1,3,5,15,16,77,78] The range equation for a bistatic radar is derived in a manner completely analogous to that for a monostatic radar. With this analog, the bistatic radar maximum-range equation can be written as

$$(R_T R_R)_{max} = \left(\frac{P_T G_T G_R \lambda^2 \sigma_B F_T^2 F_R^2}{(4\pi)^3 K T_s B_n (S/N)_{min} L_T L_R} \right)^{1/2} \tag{25.1}$$

where R_T = transmitter-to-target range
  R_R = receiver-to-target range
  P_T = transmitter power
  G_T = transmit antenna power gain
  G_R = receive antenna power gain
  λ = wavelength
  σ_B = bistatic radar target cross section
  F_T = pattern propagation factor for transmitter-to-target path
  F_R = pattern propagation factor for target-to-receiver path
  K = Boltzmann's constant
  T_s = receive system noise temperature
  B_n = noise bandwidth of receiver's predetection filter
  $(S/N)_{min}$ = signal-to-noise power ratio required for detection
  L_T = transmit system losses (> 1) not included in other parameters
  L_R = receive system loss (> 1) not included in other parameters

Equation (25.1) is related to the corresponding monostatic radar range equation by the following: $\sigma_M = \sigma_B$, $L_T L_R = L_M$, and $R_T^2 R_R^2 = R_M^4$. More specific formulations of the maximum-range equation, as given in Chap. 2, also apply to the bistatic radar case. Equation (25.1) is used in this chapter because it more clearly illustrates the utility of constant S/N contours (ovals of Cassini) and other geometric relationships. The right side of Eq. (25.1) is called the bistatic maximum-range product κ.

Ovals of Cassini. Equation (25.1), with $(R_T R_R)_{max} = \kappa$, is the maximum-range oval of Cassini. It can be used to estimate the signal-to-noise S/N power ratio at any R_T and R_R simply by dropping the "max" and "min" designation for $(R_T R_R)$ and S/N respectively. Then when Eq. (25.1) is solved for S/N,

$$S/N = \frac{k}{R_T^2 R_R^2} \tag{25.2}$$

where S/N = signal-to-noise power ratio at ranges R_T, R_R, and

$$k = \frac{P_T G_T G_R \lambda^2 \sigma_B F_T^2 F_R^2}{(4\pi)^3 K T_S B_n L_T L_R} \qquad (25.3)$$

The term k is the bistatic radar constant. The constants k and κ are related as

$$k = \kappa^2 (S/N)_{min} \qquad (25.4)$$

Equation (25.2) represents one form of the ovals of Cassini. They can be plotted on the bistatic plane when R_T and R_R are converted to polar coordinates (r, θ), as shown on Fig. 25.1:

$$R_T^2 R_R^2 = (r^2 + L^2/4)^2 - r^2 L^2 \cos^2 \theta \qquad (25.5)$$

where L is the baseline range. Figure 25.2 is such a plot for k arbitrarily set to $30L^4$.
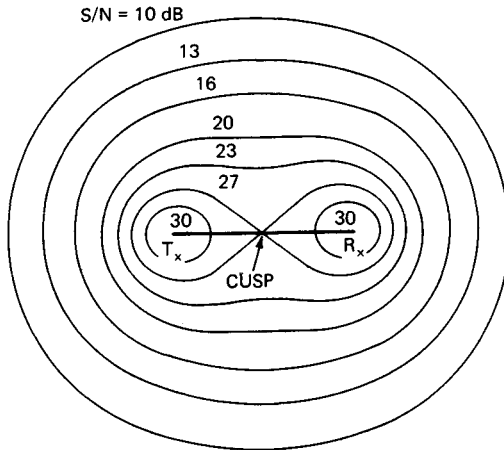
FIG. 25.2 Contours of constant signal-to-noise ratio, or ovals of Cassini, where the baseline $= L$ and $k = 30L^4$.

The ovals of Fig. 25.2 are contours of constant signal-to-noise ratio on any bistatic plane. They assume that an adequate line of sight (LOS) exists on the transmitter-to-target path and the receiver-to-target path and that σ_B, F_T, and F_R are invariant with r and θ, which is usually not the case. But this simplifying assumption is useful in understanding basic relationships and constraints. As S/N or L increases, the ovals shrink, finally collapsing around the transmitter and receiver sites. The point on the baseline where the oval breaks into two parts is called the cusp. The oval is called a lemniscate (of two parts) at this S/N. When $L = 0$, $R_T R_R = r^2$, which is the monostatic case where the ovals become circles.

Operating Regions. Ovals of Cassini define three distinct operating regions for a bistatic radar: receiver-centered region, transmitter-centered region, and receiver-transmitter-centered region, or simply the cosite region. Critical to the selection of these operating regions is the value of the bistatic radar constant k that is available. Many of the terms in Eq. (25.3) are transmitter-controlled. It is convenient to define three transmitter configurations that control k: dedicated, cooperative, and noncooperative. The *dedicated* transmitter is

defined as being under both design and operational control of the bistatic radar system; the *cooperative* transmitter is designed for other functions but found suitable to support bistatic operations and can be controlled to do so; and the *noncooperative* transmitter, while suitable for bistatic operations, cannot be controlled. The bistatic receiver is sometimes said to hitchhike off a cooperative or noncooperative transmitter, usually a monostatic radar.

Table 25.1 summarizes useful bistatic radar applications permitted by operating regions and transmitter configurations. The two omitted entries on the "Transmitter-centered" row are operational constraints: a dedicated or cooperative transmitter can usually gather nearby data in a monostatic radar mode more easily than can a remote, bistatic receiver. The two omitted entries on the "Cosite" row are technical constraints: to generate a sufficiently large bistatic radar constant for cosite operation the transmitter design and operation must be optimized for bistatic radar use; hence the dedicated transmitter is often the only viable cosite configuration. Exceptions to this rule include exploiting HF groundwave propagation and occasional atmospheric ducting.

Isorange Contours. The transmitter-to-target-to-receiver range measured by a bistatic radar is the sum $(R_T + R_R)$. This sum locates the target somewhere on the surface of an ellipsoid whose two foci are the transmitter and receiver

TABLE 25.1 Bistatic Radar Applications

Bistatic radar operating regions	Range relationships	Transmitter configuration		
		Dedicated	Cooperative	Noncooperative
Receiver-centered	$R_T \gg R_R$ k small	• Air-to-ground attack (silent penetration) • Semiactive homing missile (lock on after launch)	• Short-range air defense • Ground surveillance • Passive situation awareness	• Passive situation awareness
Transmitter-centered	$R_R \gg R_T$ k small	. . . .	. . . .	• Intelligence data gathering • Missile launch alert
Cosite	$R_T \sim R_R$ k larger	• Medium-range air defense • Satellite tracking • Range instrumentation • Semiactive homing missile (lock on before launch) • Intrusion detection	. . . .	. . . .

sites. The intersection of the bistatic plane and this ellipsoid produces the familiar ellipses of constant range sum, or *isorange contours*.

Since the (constant range sum) isorange contours and the (constant S/N) ovals of Cassini are not colinear, the target's S/N will vary for each target position on the isorange contour. This variation can be important when target returns are processed over a bistatic range cell, defined by two concentric isorange contours with separation $\Delta R_B \approx c\tau/2 \cos{(\beta/2)}$, where τ = compressed pulse width. The S/N over an isorange contour, $(S/N)_i$ is

$$(S/N)_i = \frac{4k(1 + \cos\beta)^2}{[(R_T + R_R)^2 - L^2]^2} \qquad (25.6)$$

where the denominator defines the isorange countour and the bistatic angle β defines the target's position on the isorange contour.

The maximum bistatic angle, β_{max}, on an isorange contour is $2\sin^{-1}[L/(R_T + R_R)]$, where $L/(R_T + R_R)$ is the eccentricity of the isorange contour. The minimum bistatic angle, β_{min}, is zero for all isorange contours. For example, when $L/(R_T + R_R)] = 0.95$, $\beta_{max} = 143.6°$ and $(S/N)_i$ at β_{max} is 20 dB less than at β_{min}.

25.5 AREA RELATIONSHIPS

Location.[1,15,16,18,46–48,73,79–83] Target position relative to the receive site (θ_R, R_R) is usually required in a bistatic radar. The receiver look angle θ_R is measured directly, or target azimuth and elevation measurements are converted directly to θ_R. Beam-splitting techniques can be used to increase the measurement accuracy.

The receiver-to-target range R_R cannot be measured directly, but it can be calculated by solving the bistatic triangle (Fig. 25.1). A typical solution in elliptical coordinates is[1]

$$R_R = \frac{(R_T + R_R)^2 - L^2}{2(R_T + R_R + L \sin\theta_R)} \qquad (25.7)$$

The baseline L can be calculated from coordinates provided by a dedicated transmitter or measured by an emitter location system. The range sum $(R_T + R_R)$ can be estimated by two methods. In the direct method the receiver measures the time interval ΔT_{rt} between reception of the transmitted pulse and reception of the target echo. It then calculates the range sum as $(R_T + R_R) = c\Delta T_{rt} + L$. This method can be used with any transmitter configuration, given an adequate LOS between transmitter and receiver.

In the indirect method synchronized stable clocks are used by the receiver and (dedicated) transmitter. The receiver measures the time interval ΔT_{tt} between transmission of the pulse and reception of the target echo. It then calculates the range sum as $(R_T + R_R) = c\Delta T_{tt}$. A transmitter-to-receiver LOS is not required unless periodic clock synchronization is implemented over the direct path.

For the special case of a bistatic radar using the direct range sum estimation method, where $L \gg c\Delta T_{rt}$, Eq. (25.7) can be approximated as

$$R_R = \frac{c\Delta T_{rt}}{1 + \sin\theta_R} \qquad (25.8)$$

This approximation does not require an estimate of L. The error in Eq. (25.8) is less than 10 percent for $0° < \theta_R < 180°$ and $L > 4.6\, c\Delta T_{rt}$.

Other target location techniques are possible.[16,18] The transmitter beam-pointing angle θ_T can be used in place of θ_R. Unless the transmitter is also a monostatic radar tracking the target, target location accuracy is degraded, since beam splitting is sacrificed. A hyperbolic measurement system can be used, in which a receiver measures the difference in propagation times from two separate transmitters. The locus of target position now lies on a hyperbola, and the inter-section of the receiver's AOA (angle of arrival) estimate with the hyperbola es-tablishes the target position. Use of a third transmitter provides a full hyperbolic fix on the target. A theta-theta location technique uses the angles θ_T and θ_R and an estimate of L, where θ_T is typically provided by a monostatic radar, which acts as a cooperative bistatic transmitter.

For an elliptic location system, target location errors typically *increase* as the target approaches the baseline, ignoring S/N changes. The principal source of er-rors is the geometry inherent in Eq. (25.7). Additional errors occur when the di-rect range sum estimation method is used. They include interference from the direct-path signal (analogous to eclipsing), pulse instability, and multipath ef-fects. Compounding the eclipsing problem is interference from range sidelobes when pulse compression is used by the transmitter. If linear FM pulse compres-sion is used, Hamming or cosine-squared time-domain weighting by the receiver improves near-in sidelobe suppression by about 5 dB, when compared with the same type of frequency-domain weighting.[79]

For a hyperbolic location system, target location errors *decrease* as the target approaches the line joining the two transmitters. For a theta-theta location sys-tem, the error is a minimum when the target lies on the perpendicular bisector of the baseline with $\beta = 45°$ and increases elsewhere.[18] When successive data mea-surements (or redundant data) are available to a bistatic or multistatic radar, tar-get state estimates can be made with Kalman or other types of filters.[80,81]

Coverage. Bistatic radar coverage, like monostatic radar coverage, is determined by both sensitivity and propagation. Bistatic radar sensitivity is set by the contour of constant $(S/N)_{min}$ and the oval of Cassini. Bistatic radar propagation requires a suitable path between the target and both sites and must include the effects of multipath, diffraction, refraction, shadowing, absorption, and geometry. The first five effects are usually included in the pattern propagation and loss factor terms of Eq. (25.1). The geometry effect is treated separately.

For given target, transmitter, and receiver altitudes the target must simulta-neously be within LOS to both the transmitter and the receiver sites. For a smooth earth these LOS requirements are established by coverage circles cen-tered at each site. Targets in the area common to both circles have an LOS to both sites as shown in Fig. 25.3. For a 4/3 earth model, the radius of these cov-erage circles, in kilometers, is approximated by[16]

$$r_R = 130(\sqrt{h_t} + \sqrt{h_R}) \qquad (25.9)$$
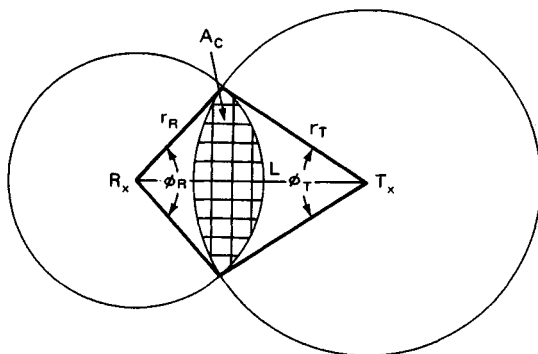
FIG. 25.3 Geometry for common-coverage area A_C.

$$\text{and } r_T = 130(\sqrt{h_t} + \sqrt{h_T}) \tag{25.10}$$

where h_t = target altitude, km
$\quad h_R$ = receive antenna altitude, km
$\quad h_T$ = transmit antenna altitude, km

If the receiver establishes synchronization via the direct-path link, then an adequate LOS is also required between transmitter and receiver. In this case $h_t = 0$ and $r_R + r_T \geq L$, where L is the baseline range. Thus,

$$L \leq 130(\sqrt{h_R} + \sqrt{h_T}) \tag{25.11}$$

If synchronization is accomplished by stable clocks, this LOS is not required and the system must satisfy only the requirements of Eqs. (25.9) and (25.10).

The common-coverage area A_C is shown in Fig. 25.3 as the intersection of the two coverage circles and is

$$A_C = \tfrac{1}{2}[r_R{}^2(\phi_R - \sin \phi_R) + r_T{}^2(\phi_T - \sin \phi_T)] \tag{25.12}$$

where ϕ_R and ϕ_T are shown on Fig. 25.3 and are

$$\phi_R = 2 \cos^{-1}\left(\frac{r_R{}^2 - r_T{}^2 + L^2}{2r_R L}\right) \tag{25.13}$$

$$\phi_T = 2 \cos^{-1}\left(\frac{r_T{}^2 - r_R{}^2 + L^2}{2r_T L}\right) \tag{25.14}$$

Terrain and other types of masking or shadowing degrade both monostatic and bistatic coverage. For ground-based bistatic transmitters and receivers the degradation can be severe.[84] For this reason some air defense bistatic radar concepts use an elevated or airborne transmitter.[44,45,48,54] As a general rule bistatic coverage is less than monostatic coverage in both single and netted configurations.

Clutter Cell Area.[42,51,59,73,85-89] The main-lobe bistatic clutter cell area A_c is defined, in the broadest sense, as the intersection of the range resolution cell, the doppler resolution cell, and the bistatic main-beam footprint. The range and doppler resolution cells are defined by isorange and isodoppler contours, respectively. The bistatic footprint is the area on the ground, or clutter surface, common to the one-way transmit and receive beams, where the beamwidths are conventionally taken at the 3-dB points. Three clutter cell cases are usually of interest: beamwidth-limited, range-limited, and doppler-limited.

Beamwidth-Limited Clutter Cell Area. The beamwidth-limited clutter cell area $(A_c)_b$ is the bistatic footprint. It has been evaluated for specific antenna pattern functions and specific geometries by numerical integration techniques.[42,85,86] At small grazing angles a two-dimensional approximation to $(A_c)_b$ is a parallelogram shown as the single-hatched area in Fig. 25.4 with area

$$(A_c)_b = \frac{R_R \Delta \theta_R R_T \Delta \theta_T}{\sin \beta} \tag{25.15}$$

where $R_R \Delta \theta_R$ is the cross-range dimension of the receive beam at the clutter cell, $R_T \Delta \theta_T$ is the corresponding dimension for the transmit beam, and $\Delta \theta_R$ and $\Delta \theta_T$ are, respectively, the 3 dB beamwidth of the receive and transmit beams. Respective transmit and receive beam rays are assumed to be parallel, which is a reasonable approximation when the range sum is much greater than the baseline range, $R_T + R_R \gg L$. The cell area is a minimum at $\beta = 90°$.
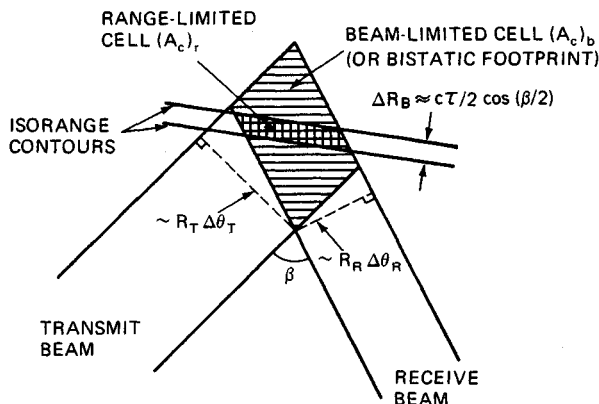
FIG. 25.4 Geometry for clutter cell areas.

Range-Limited Clutter Cell Area. The range-limited clutter cell area $(A_c)_r$ has been evaluated at small grazing angles for all geometries of interest.[87] At small grazing angles and at large range sums ($R_T + R_R \gg L$), a two-dimensional approximation to $(A_c)_r$ is a parallelogram shown as the double-hatched area in Fig. 25.4 with area

$$(A_c)_r = \frac{c\tau R_R \Delta \theta_R}{2 \cos^2 (\beta/2)} \tag{25.16}$$

where τ is the radar's compressed pulse width. The isorange contours are assumed to be straight lines within the bistatic footprint. For this example the

cross-range dimension of the transmit beam $R_T\Delta\theta_T$ is greater than that of the receive beam $R_R\Delta\theta_R$, so that the clutter cell is determined by the intersection of the receive beam and the range cell. For a given geometry one or the other beam will usually determine the clutter cell area. In either case the cell area increases as β increases. For small range sums, the cell shape is trapezoidal or triangular at small β and is rhomboidal or hexagonal at large β.[87]

An exact expression for $(A_c)_r$ has been developed,[88] again for two dimensions, with one beam and the range cell determining the clutter cell area. Equation (25.16) gives results that are within a few percent of the exact results for $\beta < 90°$. The error increases significantly for $\beta \gg 90°$ and $\theta_R < -80°$.

Doppler-Limited Clutter Cell Area. The doppler-limited clutter cell area $(A_c)_d$ has been determined by numerical integration techniques when it is bounded by a range resolution cell.[51,89] No convenient algebraic expression has been developed for the cell area since the doppler cell size and orientation with respect to the baseline change as the transmitter and receiver velocity vectors and look angles change. In the special case where the transmitter and receiver velocity vectors are equal and the bistatic angle is large, the isorange and isodoppler contours are essentially parallel, creating very large clutter cell areas.[59]

25.6 DOPPLER RELATIONSHIPS

Figure 25.5 defines the geometry and kinematics for bistatic doppler when the target, transmitter, and receiver are moving. The target has a velocity vector of magnitude V and aspect angle δ referenced to the bistatic bisector. The transmitter and receiver have velocity vectors of magnitude V_T and V_R and aspect angles δ_T and δ_R referenced to the north coordinate system of Fig. 25.1, respectively.
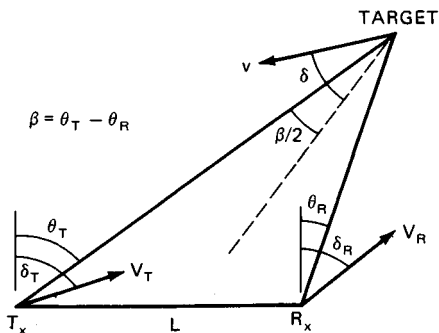
FIG. 25.5 Geometry for bistatic doppler.

Target Doppler. When the transmitter and receiver are stationary $(V_T = V_R = 0)$, the target's bistatic doppler at the receive site f_B is

$$f_B = (2V/\lambda) \cos \delta \cos (\beta/2) \qquad (25.17)$$

When $\beta = 0°$, Eq. (25.17) reduces to the monostatic case, where δ is now the angle between the velocity vector and the radar-to-target LOS, which is colinear

with the bistatic bisector. When $\beta = 180°$, the forward-scatter case, $f_B = 0$ for any δ. Equation (25.17) shows that:

- For a given δ, the magnitude of the bistatic target doppler is never greater than that of the monostatic target doppler when the monostatic radar is located on the bistatic bisector.
- For all β, when $-90° < \delta < +90°$, the bistatic doppler is positive; under this definition a closing target referenced to the bistatic bisector generates a positive, or *up*, doppler.
- For all β, when the target's velocity vector is normal to the bistatic bisector ($\delta = \pm 90°$) the bistatic doppler is zero; the vector is tangent to a range-sum ellipse drawn through the target position (a contour of zero target doppler).
- For all $\beta < 180°$, when the target's velocity vector is colinear with the bistatic bisector, the magnitude of the bistatic doppler is maximum; the vector is also tangent to an orthogonal hyperbola drawn through the target position, which is a contour of maximum target doppler.

Isodoppler Contours.   When the target is stationary and the transmitter and receiver are moving (e.g., airborne), the bistatic doppler shift at the receiver site f_{TR} is

$$f_{TR} = (V_T/\lambda) \cos (\delta_T - \theta_T) + (V_R/\lambda) \cos (\delta_R - \theta_R) \qquad (25.18)$$

where the terms are defined on Fig. 25.5.

The locus of points for constant doppler shift on the earth's surface is called an *isodoppler contour*, or *isodop*. In the monostatic case and a flat earth, these isodops are conic sections in three dimensions and radial lines emanating from the radar in two dimensions. In the bistatic case the isodops are skewed, depending upon the geometry and kinematics. They are developed analytically for two dimensions and a flat earth by setting f_{TR} = constant in Eq. (25.18) and solving for θ_R (or θ_T if appropriate).

Figure 25.6 is a plot of bistatic isodops in a two-dimensional bistatic plane, i.e., where the transmitter and receiver are at zero or near-zero altitude, for the following conditions:[90]

$$V_T = V_R = 250 \text{ m/s}$$
$$\delta_T = 0°$$
$$\delta_R = 45°$$
$$\lambda = 0.03 \text{ m}$$

The dimension of the grid on the bistatic plane is arbitrary; that is, the isodops are invariant with scale. On the left and right sides of Fig. 25.6 the isodops approximate radial lines, which are pseudo-monostatic operating points.

25.7 TARGET CROSS SECTION[1,16,33–41,91–102,104,105]

The bistatic radar cross section (RCS) of a target σ_B is a measure, as is the monostatic radar cross section σ_M, of the energy scattered from the target in the
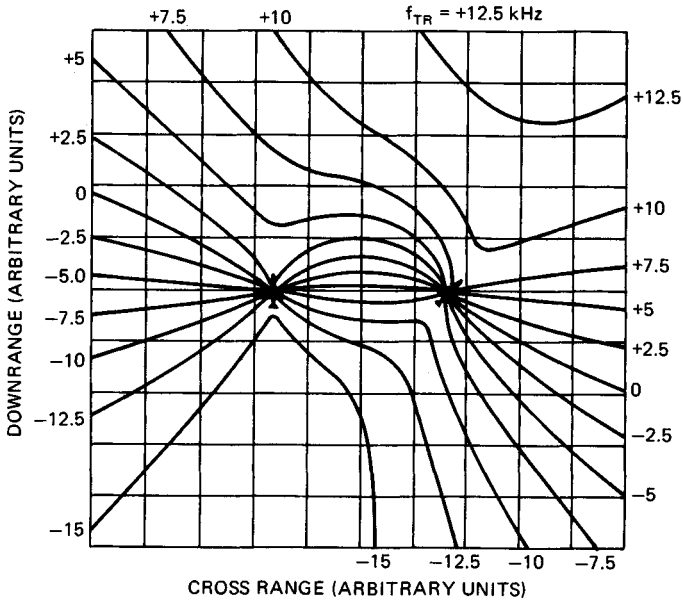
FIG. 25.6   Bistatic isodoppler contours for two dimensions and a flat earth.[90]

direction of the receiver. Bistatic cross sections are more complex than monostatic cross sections since σ_B is a function of aspect angle and bistatic angle.

Three regions of bistatic RCS are of interest: pseudo-monostatic, bistatic, and forward scatter (sometimes called near-forward scatter[99]). Each region is defined by the bistatic angle. The extent of each region is set primarily by physical characteristics of the target.

Pseudo-Monostatic RCS Region.   The Crispin and Siegal monostatic-bistatic equivalence theorem applies in the pseudo-monostatic region:[36] for vanishingly small wavelengths the bistatic RCS of a sufficiently smooth, perfectly conducting target is equal to the monostatic RCS measured on the bisector of the bistatic angle. Sufficiently smooth targets typically include spheres, elliptic cylinders, cones, and ogives. Figure 25.7 shows the theoretical bistatic RCS of two perfectly conducting spheres as a function of bistatic angle.[1,92–95] For the larger sphere (near the optics region) the pseudo-monostatic region extends to $\beta = \sim 100°$, with an error of 3 dB. And even for the smaller sphere (in the resonance region) the pseudo-monostatic region extends to $\beta = \sim 40°$. Measurements[38] of a sphere with $a = 0.42 \, \lambda$, where a is the sphere radius, match within 3 dB the values for the smaller sphere in Fig. 25.7.

For targets of more complex structure, the extent of the pseudo-monostatic region is considerably reduced. A variation of the equivalence theorem developed by Kell[41] applies to this case: for small bistatic angles, typically less than 5°, the bistatic RCS of a complex target is equal to the monostatic RCS measured on the bisector of the bistatic angle at a frequency lower by a factor of cos $(\beta/2)$.

Kell's complex targets are defined as an assembly of discrete scattering cen-
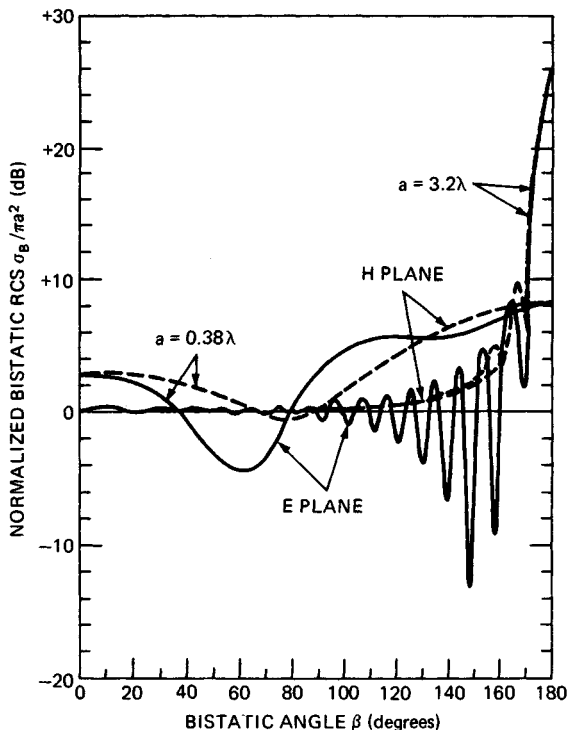
FIG. 25.7 Theoretical bistatic RCS for two perfectly conducting spheres, where a = sphere radius and λ = wavelength.[1,92–95]

ters (simple centers such as flat plates, reflex centers such as corner reflectors, skewed reflex centers such as a dihedral with corner $\neq$ 90° and stationary phase regions for creeping waves). When the wavelength is small compared with the target dimensions, these complex target models approximate many aircraft, ships, ground vehicles, and some missiles. The targets can be composed of conducting and dielectric materials.

The cos ($\beta/2$) frequency reduction term has little effect in Kell's pseudo-monostatic region, $0 < \beta < \sim 5°$, since a 5° bistatic angle corresponds to less than 0.1 percent shift in wavelength. At $\beta > 5°$ the change in radiation properties from discrete scattering centers is likely to dominate any cos ($\beta/2$) frequency reduction effect.[41] Thus the cos ($\beta/2$) term is often ignored.

Both versions of the equivalence theorem are valid when the positions of the transmitter and receiver are interchanged, given that the target-scattering media are reciprocal. Most media are reciprocal. Exceptions are gyrotropic media, such as ferrite materials and the ionosphere.[103]

Whenever the equivalence theorem is valid, Kell[41] provides a simple method for deriving bistatic RCS data from monostatic RCS data when plotted as a function of target aspect angle. Bistatic RCS data for the same polarization is obtained by translating along the target aspect angle axis by one-half of the desired bistatic angle. If monostatic RCS data is also available as a function of frequency, the monostatic curve for f sec ($\beta/2$), where f is the bistatic frequency, is used to

estimate the bistatic RCS at f. As outlined earlier, this correction is usually small.

Bistatic RCS Region. The bistatic angle at which the equivalence theorem fails to predict the bistatic RCS identifies the start of the second, bistatic region. In this region the bistatic RCS diverges from the monostatic RCS. Kell[41] identified three sources of this divergence for complex targets and for a target aspect angle fixed with respect to the bistatic bisector. These sources are (1) changes in relative phase between discrete scattering centers, (2) changes in radiation from discrete scattering centers, and (3) changes in the existence of centers—appearance of new centers or disappearance of those previously present.

The first source is analogous to fluctuations in monostatic RCS as the target aspect angle changes, but now the effect is caused by a change in bistatic angle.[104] The second source occurs when, for example, the discrete scattering center reradiates, i.e., retroreflects, energy toward the transmitter and the receiver is positioned on the edge of or outside the retroflected beamwidth; thus the received energy is reduced. The third source is typically caused by shadowing, for example, by an aircraft fuselage blocking one of the bistatic paths—transmitter or receiver LOS to a scattering center.

In general, this divergence results in a bistatic RCS lower than the monostatic RCS for complex targets. Exceptions include (1) some target aspect angles that generate a low monostatic RCS and a high bistatic specular RCS at specific bistatic angles, (2) targets that are designed for low monostatic RCS over a range of aspect angles, and (3) shadowing that sometimes occur in a monostatic geometry and not in a bistatic geometry.[92]

Ewell and Zehner[97] measured the monostatic and bistatic RCS of coastal freighters at X band when both the transmitter and the receiver were near grazing incidence. The data was plotted as a ratio of bistatic to monostatic RCS, σ_B/σ_M. The measurements match Kell's model: of the 27 data points, 24 show bistatic RCS lower than monostatic RCS. The bistatic RCS reduction starts at about $\beta = 5°$ and trends downward to $\sigma_B/\sigma_M \approx -15$ dB at $\beta \approx 50°$. Most of the data points are in the region $5° < \beta < 30°$ where -2 dB $> \sigma_B/\sigma_M > -12$ dB.

Glint Reduction in the Bistatic RCS Region. A second effect can occur in the bistatic region. When the bistatic RCS reduction is caused by a loss or attenuation of large discrete scattering centers, for example through shadowing, target glint is often reduced. Target glint is the angular displacement in apparent phase center of a target return and is caused by the phase interference between two or more dominant scatters within a radar resolution cell. As the target aspect angle changes, the apparent phase center shifts, often with excursions beyond the physical extent of the target. These excursions can significantly increase the errors in angle tracking or measurement systems. When the returns from dominant scatterers are reduced in the bistatic region, the source and hence the magnitude of glint excursions are reduced. Limited measurements for tactical aircraft show that, for a 30° bistatic angle, peak glint excursions can be reduced by a factor of 2 or more, with most of the excursions contained within the physical extent of the target.[54]

Forward-Scatter RCS Region. The third bistatic RCS region, forward scatter, occurs when the bistatic angle approaches 180°. When $\beta = 180°$, Siegel[33] showed, based on physical optics, that the forward-scatter RCS, σ_F, of

a target with silhouette (or shadow) area A is $\sigma_F = 4\pi A^2/\lambda^2$, where λ, the wavelength, is small compared with the target dimensions. The targets can be either smooth or complex structures and, from the application of Babinet's principle, can be totally absorbing.[37,91]

For $\beta < 180°$, the forward-scatter RCS rolls off from σ_F. The rolloff is approximated by treating the shadow area A as a uniformly illuminated antenna aperture. The radiation pattern of this *shadow aperture* is equal to the forward-scatter RCS rolloff when $(\pi - \beta)$ is substituted for the angle off the aperture normal. A sphere of radius a will roll off 3 dB at $(\pi - \beta) \simeq \lambda/\pi a$, when $a/\lambda \gg 1$.[15] Although the $a/\lambda \gg 1$ criterion is not satisfied in Fig. 25.7, the curve for $a = 3.2\lambda$ still exhibits this phenomenon: 3 dB reduction in σ_F at $\beta \simeq 174°$. (The value of σ_F at $\beta = 180°$ also matches $4\pi A^2/\lambda^2$ within 1 dB.) Figure 25.7 shows the rolloff approximating $J_0(x)/x$ down to $\beta \simeq 130°$, where J_0 is a Bessel function of zero order. A linear aperture of length D, with aspect angle orthogonal to the transmitter LOS, will roll off 3 dB at $(\pi - \beta) = \lambda/2D$, where $D/\lambda \gg 1$. The forward-scatter RCS rolloff continues, with sidelobes approximating $\sin x/x$ over the forward-scatter quadrant ($\beta > 90°$).[105] For other aspect angles and targets with complex shadow apertures, calculation of the forward-scatter RCS rolloff usually requires computer simulation.

The forward-scatter RCS of more complex bodies has been simulated and measured; the bodies were both reflecting and absorbing.[34,37,38,92,98,100–102] Paddison et al.[100] report both measurements and calculations via computer simulation of forward-scatter RCS for a right circular aluminum cylinder at 35 GHz and bistatic angles up to 175.4°. Calculations were made via the method of moments,[106] and measurements were made by Delco.[98] A good match between measurements and calculations was obtained for targets with dimensions of the order of several wavelengths. A similar match to Delco measurements was obtained by Cha et al., using physical-theory-of-diffraction methods for targets that are larger than several wavelengths and the method of moments otherwise.[102]

Figure 25.8 shows calculations of a 16- by 1.85-cm cylinder with 992 facets at 35 GHz, for three fixed transmitter-to-target geometries: (a) near end on, (b) 45° aspect angle, and (c) broadside.[100] The broadside geometry shows the classic forward-scattering lobe from a rectangular aperture, with approximate $\sin x/x$ sidelobe rolloff out to $\beta \simeq 110°$. The three bistatic RCS regions are quite distinct: pseudo-monostatic at $\beta < 20°$, bistatic at $20° < \beta < 140°$, and forward scatter at $\beta > 140°$. The other two geometries show a similar but broader forward-scatter lobe, as is expected since the silhouette area and hence the shadowing aperture are smaller. The 45° aspect geometry is of interest because the RCS in the bistatic region is larger than the monostatic RCS for most bistatic angles. The large spike at $\beta = 90°$ is the bistatic specular lobe, analogous to the monostatic specular lobe in the broadside geometry. While Fig. 25.8 shows the clear dependency of bistatic RCS on both aspect and bistatic angle, it also serves to caution attempts to use oversimplified bistatic RCS models, especially in the bistatic region.

25.8 CLUTTER

The bistatic radar cross section of clutter σ_c is a measure, as is the monostatic radar clutter cross section, of the energy scattered from a clutter cell area A_c in the direction of the receiver. It is defined as $\sigma_c = \sigma_B{}^0 A_c$, where $\sigma_B{}^0$ is the scattering coefficient, or the clutter cross section per unit area of the illuminated sur-
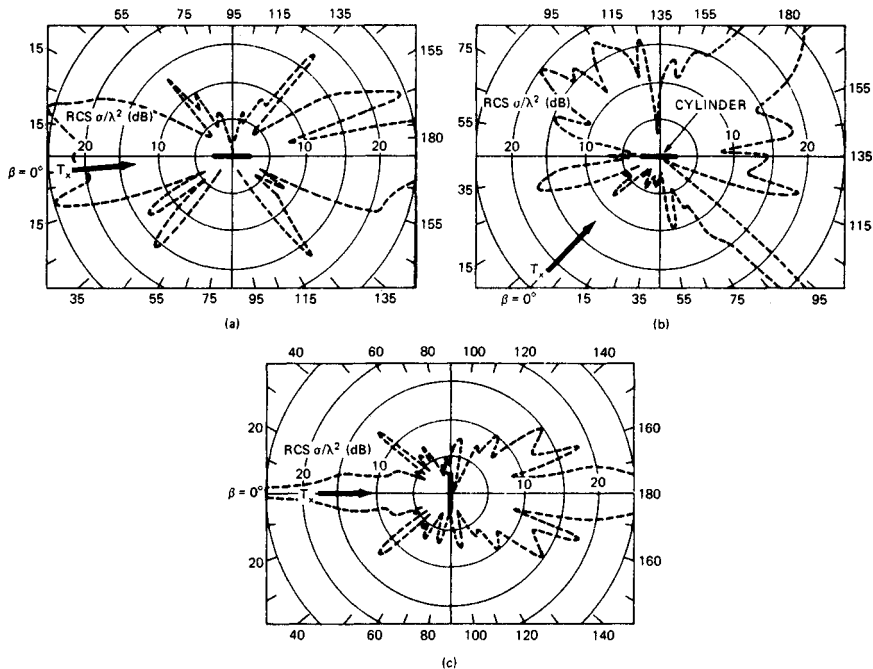
FIG. 25.8 Calculated bistatic RCS, replotted as a function of bistatic angle for a conducting cylinder, 16 by 1.85 cm at 35 GHz, *HH* polarization.[100] (*a*) Near end on. (*b*) 45° aspect angle. (*c*) Broadside.

face. The clutter cell area is given for beam- and range-limited cases in Sec. 25.5. This section considers measured and estimated values of $\sigma_B{}^0$, which vary as a function of the surface composition, frequency, and geometry. And, in contrast to the monostatic case, little measured data for $\sigma_B{}^0$ has been reported.[42,43,107-115]

The available database for terrain and sea clutter at microwave frequencies consists of six measurement programs, which are summarized in Table 25.2. The measurement angles shown in Table 25.2 are defined in Fig. 25.9, which is a clutter-centered coordinate system similar to those used in all the measurement programs. Because terrain and sea are reciprocal media, θ_i and θ_s are interchangeable in the subsequent data.[103] The Pidgeon data was analyzed by Domville[109] and Nathanson.[116] Vander Schurr and Tomlinson[117] analyzed the Larson and Cost data.

In addition to this database, bistatic reflectivity measurements have been made at optical[118] and sonic[119] wavelengths and of buildings,[120] airport structures,[121] and planetary surfaces.[66,122] In each of these measurements, the reflectivity data is expressed in terms of reflected power, not $\sigma_B{}^0$.[103]

The bistatic angle is calculated from the angles in Fig. 25.9 by the use of direction cosines:

$$\beta = \cos^{-1}(\sin\theta_i \sin\theta_s - \cos\theta_i \cos\theta_s \cos\phi) \qquad (25.19)$$

TABLE 25.2 Summary of Measurement Programs for Bistatic Scattering Coefficient, σ_B^0

Reference (date)	Organization	Author	Surface composition	Frequency	Polarization	Measurement angles (degrees) θ_i	θ_s	ϕ
42 (1965)	Ohio State University (Antenna Laboratory)	Cost, Peake	Smooth sand Loam Soybeans	10 GHz	VV, HH, HV	5–30, 10–70	5–30, 5–90	0–145, 0, 180
			Rough sand Loam with stubble Grass	10 GHz	VV, HH, HV	5–70	5–90	0–180
43 (1966)	Johns Hopkins University (APL)	Pidgeon	Sea (sea states 1, 2, 3)	C band	VV, VH	0.2–3	10–90	180
107 (1967)			Sea (Beaufort, wind 5)	X band	HH	1–8	12–45	180
108 (1967)	GEC (Electronics) Ltd., England	Domville	Rural land Urban land Sea (20-kn wind)	X band	VV, HH	6–90*	6–180*	180, 165
109 (1968)			Sea (20-kn wind)	X band	VV, HH	~ 0–90*	~ 0–180*	180, 165
110 (1969)			Semidesert	X band	VV, HH	~ 0	?	180, 165
111 (1977)	University of Michigan (ERIM)	Larson, Heimiller	Grass with cement taxiway	1.3 and 9.4 GHz	HH, HV	10, 40	5, 10, 20	0–180
112 (1978)			Weeds and scrub trees	GHz	HH, HV	10, 15, 20	5, 10, 20	0–105
113 (1982)	Georgia Institute of Technology (EES)	Ewell, Zehner	Sea (0.9-m, 1.2–1.8-m waveheights)	9.38 GHz	VV, HH	~ 0	~ 0	90–160
114 (1984)								
115 (1988)	University of Michigan (Department of Electrical Engineering and Computer Science)	Ulaby et al.	Visually smooth sand	35 GHz	VV, HH VH, HV	24	24	0–170
			Visually smooth sand	35 GHz	VV, HH	30	30	0–170
			Rough sand Gravel		VH, HV	30	10–90	0–90

*Measured and interpolated data ranges.
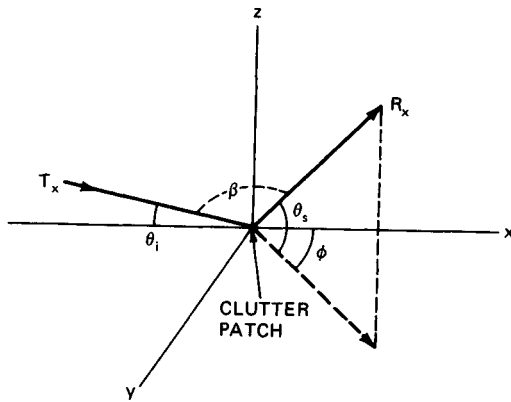
FIG. 25.9 Coordinate system for bistatic clutter measurements. θ_i = incident angle (in xz plane); θ_s = scattering angle (in plane containing z axis); ϕ = out-of-plane angle (in xy plane).

Two measurement sets are of interest: in plane, where $\phi = 180°$, and out of plane, where $\phi < 180°$. When $\phi = 180°$, $\beta = \theta_s - \theta_i$. In the monostatic case $\beta = 0$ and $\theta_s = \theta_i$ with $\phi = 180°$. Most of the data is taken at X band, with the most substantial in-plane database provided by Domville.[108–110] Because the database is sparse, mean values for $\sigma_B{}^0$ are usually given, with occasional standard deviations and probability distributions calculated.

In-Plane Land Clutter Scattering Coefficient.   Figure 25.10 is a plot of Domville's X-band, vertically polarized data summary for rural land, consisting of open grassland, trees, and buildings.[108] Domville reports that since the data was a composite of different sources and averaged over different terrain conditions, differences of 10 dB in the values sometimes occurred. The spread in raw data within any data set ranged from 1 dB to 4.5 dB, however. The measured database consists of points near the lines $\theta_i = \theta_s$, $\theta_i = 90°$, and $\theta_s = 90°$ and of points along the specular ridge near the forward-scatter region. The remaining data points are interpolations.

Domville also summarized in-plane data for forest and urban areas.[108] The shape of constant $\sigma_B{}^0$ contours for all Domville's terrain types are similar. For urban areas $\sigma_B{}^0$ is generally 3 to 6 dB higher. The extent of the specular ridge is smaller, however. Because forest terrain is a more uniform scatterer, the cones extend into the forward quadrant ($\theta_s > 90°$). The ridge extent is smaller and its magnitude is about 16 dB below that of rural land. Other values of $\sigma_B{}^0$ for forest terrain are similar to those of rural land for $\theta_s < 90°$.

Domville reports[109] that at low θ_i no significant variation in $\sigma_B{}^0$ was observed for rural and forest terrain when measured at a small out-of-plane angle, $\phi = 165°$. Also at low θ_i no significant variation between horizontal, vertical, and crossed polarizations was observed for rural and forest terrain.

For semidesert, $\sigma_B{}^0$ was measured[110] at -40 dB for both horizontal and vertical polarization at $\theta_i < \sim 1°$ and for all $\theta_s > \sim 1°$. Crossed-polarization measurements were 5 to 10 dB lower. Also, $\sigma_B{}^0$ is reduced by about 0.3 dB/° as ϕ moves from 180 to 165°.

Although terrain conditions are different, the Cost in-plane data[42] matches the
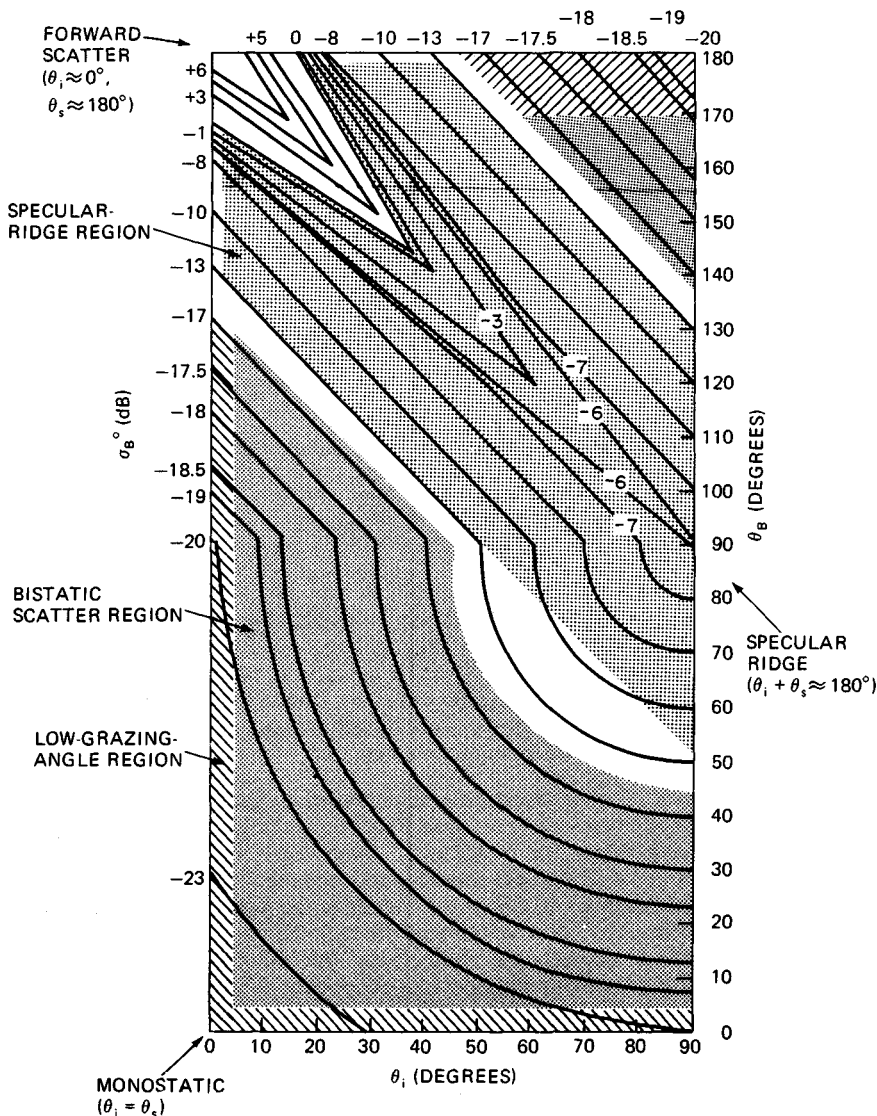
FIG. 25.10 X-band, vertically polarized, σ_B^0, in-plane ($\phi = 180°$) data summary for rural land.[108]

Domville data[108] within about 10 dB. The Cost data curves do not always approach the bistatic specular ridge monotonically even though the terrain conditions appear to be more uniform.

The in-plane Domville land clutter data can be divided into three regions: a low-grazing-angle region, where $\theta_i < \sim 3°$ or $\theta_s < \sim 3°$, the hatched area in Fig. 25.10; a specular-ridge region, where $140° \leq (\theta_i + \theta_s) \leq 220°$, the dotted area; and

a bistatic scatter region, where (θ_i, θ_s) assume values shown as the shaded areas in Fig. 25.10. Each region can be modeled, by a *semiempirical process* (containing arbitrary constants that are adjusted to fit empirical data), as follows.

The low-grazing-angle and bistatic scatter regions are based on the constant-γ monostatic clutter model:

$$\sigma_M^{\,0} = \gamma \sin \theta_i \qquad (25.20)$$

where $\sigma_M^{\,0}$ is the monostatic scattering coefficient, θ_i is the monostatic, or incident, angle on Fig. 25.9, and γ is a normalized reflectivity parameter. For farmland $\gamma \approx -15$ dB, and for wooded hills $\gamma \approx -10$ dB.[123]

The constant-γ bistatic-scatter-region model is developed by using a variation of the monostatic-bistatic equivalence theorem (Sec. 25.8), where $\sin \theta_i$ is replaced by the geometric mean of the sines of the incident and scattering angles, $(\sin \theta_i \sin \theta_s)^{1/2}$, in Eq. (25.20).[123] Hence

$$(\sigma_B^{\,0})_b = \gamma \, (\sin \theta_i \sin \theta_s)^{\frac{1}{2}} \qquad (25.21)$$

where $(\sigma_B^{\,0})_b$ is the scattering coefficient in the bistatic scatter region. Now γ can be estimated from Fig. 25.10 by using monostatic data, which is plotted along the line $\theta_i = \theta_s$. A value of $\gamma = -16$ dB in Eq. (25.20) fits the monostatic data within about 2 dB. Using $\gamma = -16$ dB in Eq. (25.21) yields a match within 3 dB to the bistatic data, including the small triangle in the forward quadrant.

The low-grazing-angle region is modeled by the sine of the arithmetic mean of the incident and scattering angles, $\sin [(\theta_i + \theta_s)/2]$. Hence

$$(\sigma_B^{\,0})_1 = \gamma \sin [(\theta_i + \theta_s)/2] \qquad (25.22)$$

where $(\sigma_B^{\,0})_1$ is the scattering coefficient in the low-grazing-angle region. The data match is again ~3 dB for $\gamma = -16$ dB, including the small quadrilateral in the upper right corner of Fig. 25.10. Since $(\theta_i + \theta_s)/2 = \theta_i + \beta/2$, Eq. (25.22) is an exact application of the monostatic-bistatic equivalence theorem. For very low grazing angles (θ_i or $\theta_s << \sim 1°$), but excluding the specular-ridge region, the calculations for $(\sigma_B^{\,0})_1$ must be multiplied by the pattern propagation factors F_T^2 and F_R^2 and the loss terms L_T and L_R.[123]

The specular-ridge region is modeled for values of $(\sigma_B^{\,0})_s \le 1$ by a variation of the Beckman and Spizzichino theory of forward scattering from rough surfaces:[124,125]

$$(\sigma_B^{\,0})_s = \exp [- (\beta_c/\sigma_s)^2] \qquad (25.23)$$

where $(\sigma_B^{\,0})_s$ = scattering coefficient in the specular-ridge region
$\qquad \sigma_s$ = rms surface slope
$\qquad \beta_c$ = angle between vertical and the bistatic bisector of θ_i and θ_s
$\qquad\quad$ = $|90 - (\theta_i + \theta_s)/2|$

For flat terrain $\sigma_s \approx 0.1$ rad. With a value of $\sigma_s = 0.17$ rad, Eq. (25.23) matches the specular ridge in Fig. 25.10 within 5 dB, for $(\sigma_B^{\,0})_s \le 1$.

In-Plane Sea Clutter Scattering Coefficient. Limited in-plane sea clutter measurements have been taken.[43,107,109] The Domville data[109] contains a broad range of θ_i, θ_s measurement conditions but unfortunately estimates only wind conditions and not sea state. For vertical polarization, the Domville X-band data[109] and the Pidgeon C-band data[43] show spreads of about 10 dB, and their averages match within ±5 dB. For horizontal polarization, the Domville X-band data[109] and the Pidgeon X-band data[107] again show spreads of about 10 dB, but the match is only about ±10 dB.

In view of the limited database and the uncertainties in some of the measurement conditions, caution must be exercised in modeling this data. An approximate model is the direct application of the constant-γ monostatic clutter model, Eq. (25.20), when either θ_i or θ_s is held constant. Then for the region θ_i, $\theta_s > \sim 2°$ and $\theta_i + \theta_s < \sim 100°$, $\gamma = -20$ dB matches the available vertically polarized data within about 5 dB for a 20-kn wind ($\approx$ sea state 3 when fully developed).

Below about 2°, pattern propagation factors and losses affect the measurements. Values for $\sigma_B{}^0$ of -50 dB ±5 dB have been measured.[43] When the pattern propagation factors and losses are included in measurements, the data is sometimes called effective $\sigma_B{}^0$.[114] For $\theta_i + \theta_s > \sim 100°$, $\sigma_B{}^0 > 0$ dB, reaching $+10$ dB in the specular-ridge region. For horizontal polarization $\sigma_B{}^0$ is typically 1 to 5 dB lower,[109] but this difference is not significant compared with the data spread. Measured cross-polarized (VH) values for $\sigma_B{}^0$ are 10 to 15 dB lower than those for copolarized (VV) values at $\theta_i < 1°$ but only 5 to 8 dB lower at $\theta_i \approx 3°$.[43]

Out-of-Plane Scattering Coefficient. Limited out-of-plane land clutter measurements have been taken.[42,111,112,115] The Cost[42] and Ulaby[115] data shows reasonable correlation but only limited correlation with the Larson[111,112] data. There does not appear to be a satisfactory model of the available data.

However, general trends are apparent for all polarizations. First, $\sigma_B{}^0$ usually approaches a minimum as ϕ approaches 90°, with values 10 to 20 dB below the monostatic value ($\theta_i = \theta_s$, $\phi = 180°$). Second, out-of-plane $\sigma_B{}^0$ values are not significantly different (within ~ 5 dB) from in-plane $\sigma_B{}^0$ values for $\phi < \sim 10°$ and $\phi > \sim 140°$, i.e., angles close to in-plane conditions. The $\phi < \sim 10°$ limit is based on Cost, Ulaby, and Domville data; the $\phi > \sim 140°$ limit, on Ulaby and Larson data.

Ewell[113,114] measured horizontally and vertically polarized out-of-plane $\sigma_B{}^0$ for sea clutter at θ_i and θ_s near grazing incidence (θ_i, $\theta_s << 1°$). Visual estimates of sea conditions ranged from 0.9- to 1.8-m waveheight. Ratios of bistatic to monostatic scattering coefficients (median values) were calculated, with bistatic angles, $\beta \approx 180° - \phi$, ranging from 23° to 85°. The data implicitly included pattern propagation factors and losses. Since antenna heights were different, F_T, F_R, L_T, and L_R are expected to be different but were not measured. In all cases the measured bistatic to monostatic ratios were less than unity. In two cases they ranged from -2 dB to -12 dB, and in the third case they dropped from ~ -5 dB at $\beta = 23°$ to -20 to -25 dB at $\beta = 60°$. The trend was generally downward as β increased. Values for horizontal and vertical polarization showed no significant differences. For the most part both monostatic and bistatic data exhibited nearly log-normal amplitude distributions.

25.9 SPECIAL TECHNIQUES, PROBLEMS, AND REQUIREMENTS

Pulse Chasing.[49,73,126,127,129] The concept of pulse chasing has been proposed as a means to reduce the complexity and cost of multibeam bistatic

receivers, which are one solution to the beam scan-on-scan problem. The simplest pulse-chasing concept replaces the multibeam receive system (*n* beams, receivers, and signal processors) with a single beam, receiver, and signal processor. As shown in Fig. 25.11, the single receive beam rapidly scans the volume covered by the transmit beam, essentially chasing the pulse as it propagates from the transmitter: hence the term *pulse chasing*. In addition to the usual requirements for solving the bistatic triangle, pulse chasing requires knowledge of θ_T and pulse transmission time,[126] which can be provided to the receive site by a data link. Alternatively, if the transmit beam scan rate and the pulse repetition frequency (PRF) are uniform, the receive site can estimate these parameters as the transmit beam passes by the receive site.[127]
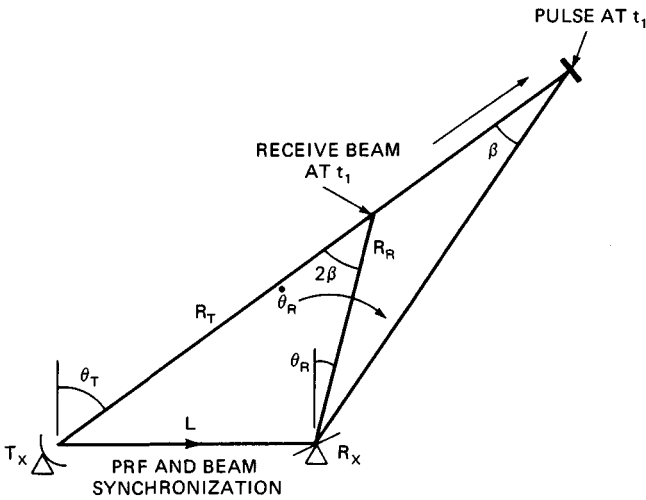
FIG. 25.11 Pulse chasing for the single-beam, continuous-scan case.

The receive beam-scanning rate must be at the transmitter's pulse propagation rate, modified by the usual geometric conditions. This rate, $\dot{\theta}_R$, is given by[73]

$$\dot{\theta}_R = c \tan (\beta/2)/R_R \qquad (25.24)$$

For typical geometries $\dot{\theta}_R$ can vary from $1°/\mu s$ to $0.01°/\mu s$. These rates and rate changes require an inertialess antenna such as a phased array and fast diode phase shifters. Normally a phased array antenna used for surveillance is programmed to switch beams in increments of a beamwidth. Fractional shifts of a beamwidth can be achieved by changing the phase of a few (symmetric) pairs of phase shifters in the array. In this way a pseudo-continuous beam scan can be generated, with the required rates and rate changes.[128]

Because of pulse propagation delays from the target to the receiver, the pointing angle of the receive beam θ_R must lag the actual pulse position. For an instantaneous pulse position that generates a bistatic angle β, $\theta_R = \theta_T - 2\beta$. In terms of the bistatic triangle, the required receive beam-pointing angle is[73]

$$\theta_R = \theta_T - 2 \tan^{-1}\left(\frac{L \cos \theta_T}{R_T + R_R - L \sin \theta_T}\right) \qquad (25.25)$$

The minimum receive beamwidth $(\Delta\theta_R)_m$ required to capture all returns from a range cell intersecting the common beam area is approximated by[73]

$$(\Delta\theta_R)_m \simeq (c\tau_u \tan(\beta/2) + \Delta\theta_T R_T)/R_R \qquad (25.26)$$

where τ_u is the uncompressed pulse width. The approximation assumes that respective rays from the transmit and receive beams are parallel. The approximation is reasonable when $(R_T + R_R) >> L$ or when $L >> c\tau_u$.

Other implementations of pulse chasing are possible. In one concept the n-beam receive antenna is retained and two receiver–signal processors (RSPs) are time-multiplexed across the n beams. One RSP steps across the even-numbered beams, and the other RSP steps across the odd-numbered beams, so that returns in beam pairs are processed simultaneously: (1,2), (2,3), (3,4), etc. This *leapfrog* sequence is required to capture all returns in the common-beam area.

A second concept uses two beams and two RSPs step-scanning over the volume covered by the n-beam antenna. It uses an identical leapfrog sequence. Both concepts relax the fractional beam scan requirements by either sampling or stepping the beams in units of a beamwidth. Since they both process returns across two beamwidths before switching, the beam dwell time T_b is approximately $2(\Delta\theta_R)_m R_R/c$ and the stepping rate is T_b^{-1}. The approximation assumes negligible phase-shift delays and settling times.

Beam Scan on Scan. If high-gain scanning antennas are used by both the transmitter and the receiver in a bistatic surveillance radar, inefficient use is made of the radar energy, since only the volume common to both beams (the bistatic footprint) can be observed at any given time. Targets illuminated by the transmit beam outside the footprint are lost to the receiver. Four remedies to the beam scan-on-scan problem are possible: (1) fix the transmit beam for the time required to complete a surveillance frame by the receive beam, step the transmit beam one beamwidth and complete a second surveillance frame, and so forth until the transmit beam has stepped across the surveillance sector; (2) scan the transmit beam and use multiple simultaneous receive beams to cover the surveillance sector; (3) scan the transmit beam and chase the transmitted pulse with the receiver beam; and (4) broaden the transmit beamwidth to floodlight the surveillance sector, and scan the receive beam across the surveillance sector. The first and fourth remedies require a dedicated transmitter; the second and third do not.

The step-scan transmitter remedy increases the surveillance frame time by the number of required transmit beam steps. This increase is usually not acceptable for surveillance operations, and as a consequence the remedy is seldom considered. The multibeam-receiver remedy restores the frame time but increases receiver cost and complexity, since a multiple-beam antenna is required and an RSP must be used for each beam. The pulse-chasing remedy can remove requirements for multiple beams and RSPs, but at the penalty of using an inertialess (phased array) antenna with both complex and precise beam scheduling and/or multiplexing requirements, depending upon the configuration. The floodlight transmitter remedy removes all these complexities. It has the benefits of increasing data rates and simultaneously servicing multiple receivers. It incurs the pen-

alty of a reduced S/N, directly as G_T is reduced; it also suffers increased sidelobe clutter levels. In short, there are no simple and inexpensive remedies to the beam scan-on-scan problem without suffering a penalty in surveillance performance.

Sidelobe Clutter. As with a monostatic radar, a bistatic radar must contend with sidelobe clutter. When both transmitter and receiver are ground-based and separated by a baseline range L, only ground clutter from regions having an adequate line of sight (LOS) to both the transmitter and the receiver will enter the receiver antenna sidelobes. That region is defined for a smooth earth as the common-coverage area A_C. From Eqs. (25.9), (25.10), and (25.12), when $h_t = 0$ and $L \geq r_R + r_T = 130 \ (\sqrt{h_R} + \sqrt{h_T})$, A_C for ground clutter is zero, and no sidelobe (or main-lobe) clutter enters the receive antenna. Targets with adequate LOS to both transmitter and receiver can be detected in a thermal-noise-limited background. This situation is analogous to a monostatic radar detecting targets at ranges greater than r_T.

This development applies to ocean scenarios but seldom is valid for land scenarios. In land scenarios variable terrain can decrease clutter levels by masking a clutter LOS when $L < r_R + r_T$ or increase clutter levels by generating a clutter LOS when $L > r_R + r_T$.

When the transmitter and/or the receiver is elevated or airborne, LOS restrictions are greatly reduced but not necessarily eliminated. Two clutter problems unique to bistatic radars are encountered in this situation. The first occurs when a floodlight transmit beam is used. To a first order, sidelobe clutter levels are reduced only by the one-way receive antenna sidelobes, in contrast to two-way sidelobe clutter reduction for a monostatic radar.

The second problem occurs when the transmitter and/or the receiver are moving, e.g., when airborne. Now the bistatic clutter doppler returns skew and spread, depending upon the geometry for each clutter patch and the kinematics of the transmit and receive platforms. Doppler skew is defined in terms of isodoppler contours, or isodops, given by Eq. (25.18) for two dimensions and a flat earth. The skew is range- and angle-dependent. The range-dependent skewing effect is not present in an airborne monostatic radar. Clutter spread in a particular sidelobe range cell is centered on the doppler skew present in the range cell.

These skewing and spreading effects, along with increased clutter levels, can greatly complicate the ability of a bistatic radar to detect targets in clutter. Remedies include conventional doppler filtering and high time-bandwidth waveforms; the judicious use of masking when available; control of the geometry, especially when a dedicated or cooperative transmitter is available; design of very low receive (and transmit when possible) antenna sidelobe levels; sidelobe blanking of discrete clutter returns; range or range-doppler averaging in the constant false-alarm rate (CFAR) unit for homogeneous clutter; and spatial excision of clutter returns. One implementation of this last technique relies on knowledge of the geometry and kinematics to predict the clutter doppler and spread in a given area. Then a filter or gate is set to excise main-beam clutter returns in that area. The amount of range-doppler space excised by this procedure can be as high as 8 percent.[45]

Time Synchronization. Time synchronization is required between the bistatic transmitter and receiver for range measurement. Timing accuracies on the order of a fraction of the transmitter's (compressed) pulse width are typically desired over the duration of an operation. Time synchronization can be accomplished directly by receiving a signal from the transmitter, demodulating the signal if necessary, and using the demodulated signal to

synchronize a clock in the receiver. The transmitter signal can be sent via landline, via a communication link, or directly at the transmitter's RF if an adequate line of sight (LOS) exists between transmitter and receiver. If an adequate LOS is not available, it can be sent via a scatter path, where the scatterer has adequate LOS to both the transmitter and the receiver.[49] In this case, the scatterer must lie in the common-coverage area, as defined by Eq. (25.12). Transmission via tropospheric scatter can also be used in special cases.[50] In all these direct time synchronization schemes, implementation is straightforward, much like the initial synchronization process in communication systems. They can also be used for any type of transmitter pulse repetition interval (PRI) modulation: stable, staggered, jittered, and random. With time synchronization established, target range is calculated via Eq. (25.7) or similar methods.

For stable PRIs, time synchronization can be accomplished indirectly by using identical stable clocks at the transmitter and receiver sites. The clocks can be synchronized periodically, for example, whenever the transmitter and receiver are within LOS or located together if one or both are mobile. Direct time synchronization methods can be used for this task. Alternatively, the stable clocks can be slaved to a second source, such as Navstar GPS or loran C.[50,82,130] Indirect time synchronization can also be employed with a dedicated or cooperative transmitter using random PRIs if a random code sequence is established a priori and is known by the receive site.

In direct time synchronization, the required clock stability between updates is, to a first order, $\Delta\tau/T_u$, where $\Delta\tau$ is the required timing accuracy and T_u is the clock update interval. The update interval typically ranges from a minimum of the transmitter's interpulse period to a maximum of the transmitter's antenna scan period. The former usually requires a dedicated link between transmitter and receiver; the latter can be implemented whenever the transmit beam scans past the receive site, given an adequate LOS, and is sometimes called *direct breakthrough*.[130] Temperature-controlled crystal oscillators can often satisfy these requirements.

However, when direct-breakthrough time synchronization is used, multipath and other propagation anomalies, as well as radio-frequency interference (RFI), will degrade the accuracy of updating. Errors of ± 1 μs have been measured when a direct LOS is available.[131] They increase to ± 5 μs over a tropospheric propagation path.[130]

Since two clocks are used in indirect time synchronization, clock stability is, to a first order, $\Delta\tau/2T_u$. For T_u on the order of hours, atomic clocks are usually required to satisfy this requirement. Temperature-controlled crystal oscillators, an integral part of atomic clocks, will usually satisfy short-term (< 1 s) stability requirements. If the stable clocks are slaved to a second source, estimated timing accuracies of 0.5 μs for loran-C and <0.1 μs for Navstar GPS are reported.[82]

Phase Synchronization and Stability. As with monostatic radars, doppler or MTI processing can be used by the bistatic receiver to reject clutter or chaff. If noncoherent MTI is acceptable for clutter rejection, the bistatic receiver can use a clutter reference, exactly as a monostatic radar would, given that clutter patches are illuminated by the transmitter.

In one bistatic noncoherent MTI implementation, called *phase priming*, an oscillator at the receiver was phase-synchronized at the PRF rate with a small sample of close-in clutter returns.[132] Phase coherence was obtained within about 10 μs and extended over 1 ms. The process was found insensitive to the clutter signal level but quite sensitive to pulse-to-pulse phase fluctuations in the clutter signal.

If coherent processing is required, phase synchronization can be established with methods similar to those used for time synchronization: directly by phase-locking the receiver to the transmit signal or indirectly by using identical stable clocks in the transmitter and the receiver. Phase accuracy, or stability, requirements are the same as those for coherent processing by a monostatic radar: from 0.01λ to 0.1λ, or 3.6° to 36° of RF phase over a coherent processing interval,[133] with 0.01λ representing more typical design requirements.

Direct phase locking can be implemented as in direct time synchronization: via landline, communication link, or at the transmitter's RF. If a direct RF link is used, adequate transmitter-to-receiver LOS is again required. It is also subject to multipath and to phase reversals if coherent operation is required across transmitter antenna sidelobes. However, this latter problem can be overcome by a Costas loop for phase reversals near 180°.[134] An extension of direct-path phase locking is the use of the direct-path signal as a reference signal in a correlation processor.[135]

For direct-path phase locking, clock stability is $\Delta\phi/2\pi f\Delta T$, where $\Delta\phi$ is the allowable rms sinusoidal phase error, f is the transmitter frequency, and ΔT is the difference in propagation time between the transmitter-target-receiver path and the transmitter-receiver (direct) path.[133] As with time synchronization, this requirement can usually be satisfied by a temperature-controlled crystal oscillator.

For matched stable clocks in the transmitter and receiver, phase stability is usually required over a coherent processing time T. Thus clock stability is $\Delta\phi/2\pi fT$. Again, atomic clocks are usually required, with crystal oscillators used for short-term stability. However, when $T < \sim 1$ s, integral crystal oscillators are usually acceptable. Quadratic phase errors caused by long-term drift in the stable clocks are usually smaller than allowable short-term sinusodial phase errors and can often be ignored.

In most types of SAR images, the integrated sidelobe ratio (ISLR) is an important criterion for image quality. It is a measure of the energy from a particular target that appears at image locations other than that corresponding to the target. Typically a -30- to -40-dB ISLR allocation for clock, or stable local oscillator (stalo), phase noise is desired.[53] When a single stalo is used, as in the monostatic case, these levels can be achieved for long coherent integration terms ($T > 10$ s) since low-frequency components of the phase noise are partially canceled in the demodulation process. However, since both bistatic phase synchronization techniques use two stalos, these low-frequency components do not cancel, resulting in higher ISLRs. Thus the bistatic SAR image quality, in terms of doppler or azimuth sidelobes, is degraded for coherent integration times greater than about 1 s at X band[53] unless very-high-quality clocks are used.

When direct-path phase locking is used by a bistatic SAR, the required motion-compensation phase shift (to track the target phase) must correct for relative motion between transmitter and receiver. When matched stable clocks are used, this correction is not required.[133]

Either time or phase errors can dominate synchronization requirements, depending upon the range and doppler accuracies needed. While all these requirements usually can be met, implementation is more complicated, time-consuming, and costly when compared with a monostatic system, which uses one clock for both time and phase synchronization.

REFERENCES

1. Skolnik, M. I.: "Introduction to Radar Systems," McGraw-Hill Book Company, New York, 1980.
2. *Microwave Syst. News Commun. Technol.*, vol. 18, p. 60, February 1988.
3. Heimiller, R. C., J. E. Belyea, and P. G. Tomlinson: Distributed Array Radar, *IEEE Trans.*, vol. AES-19, pp. 831–839, 1983.
4. Steinberg, B. D.: "Principles of Aperture and Array System Design—Including Random and Adaptive Arrays," John Wiley & Sons, New York, 1976.
5. Steinberg, B. D., and E. Yadin: Distributed Airborne Array Concepts, *IEEE Trans.*, vol. AES-18, pp. 219–226, 1982.
6. Steinberg, B. D.: High Angular Microwave Resolution from Distorted Arrays, *Proc. Int. Comput. Conf.*, vol. 23, 1980.
7. Easton, R. L., and J. J. Fleming: The Navy Space Surveillance System, *Proc. IRE*, vol. 48, pp. 663–669, 1960.
8. Mengel, J. T.: Tracking the Earth Satellite, and Data Transmission by Radio, *Proc. IRE*, vol. 44, pp. 755–760, June 1956.
9. Merters, L. E., and R. H. Tabeling: Tracking Instrumentation and Accuracy on the Eastern Test Range, *IEEE Trans.*, vol. SET-11, pp. 14–23, March 1965.
10. Scavullo, J. J., and F. J. Paul, "Aerospace Ranges: Instrumentation," D. Van Nostrand Company, Princeton, N.J., 1965.
11. Steinberg, B. D., et al.: First Experimental Results for the Valley Forge Radio Camera Program, *Proc. IEEE*, vol. 67, pp. 1370–1371, September 1979.
12. Steinberg, B. D.: Radar Imaging from a Distributed Array: The Radio Camera Algorithm and Experiments, *IEEE Trans.*, vol. AP-29, pp. 740–748, September 1981.
13. Salah, J. E., and J. E. Morriello: Development of a Multistatic Measurement System, *IEEE Int. Radar Conf.*, pp. 88–93, 1980.
14. Multistatic Mode Raises Radar Accuracy, *Aviat. Week Space Technol.*, pp. 62–69, July 14, 1980.
15. Skolnik, M. I.: An Analysis of Bistatic Radar, *IRE Trans.*, vol. ANE-8, pp. 19–27, March 1961.
16. Caspers, J. M.: Bistatic and Multistatic Radar, chap. 36 in Skolnik, M. I. (ed.): "Radar Handbook," McGraw-Hill Book Company, New York, 1970.
17. Ewing, E. F.: The Applicability of Bistatic Radar to Short Range Surveillance, *IEE Conf. Radar 77, Publ. 155*, pp. 53–58, London, 1977.
18. Ewing, E. F., and L. W. Dicken: Some Applications of Bistatic and Multi-Bistatic Radars, *Int. Radar Conf.*, pp. 222–231, Paris, 1978.
19. Farina, A., and E. Hanle: Position Accuracy in Netted Monostatic and Bistatic Radar, *IEEE Trans.*, vol. AES-19, pp. 513–520, July 1983.
20. Hanle, E.: Survey of Bistatic and Multistatic Radar, *Proc. IEE*, vol. 133, pt. F, pp. 587–595, December 1986.
21. Taylor, A. H., L. C. Young, and L. A. Hyland: U.S. Patent 1,981,884, System for Detecting Objects by Radio, Nov. 27, 1934.
22. Williams, A. F.: The Study of Radar, "Research Science and Its Application in Industry," vol. 6, Butterworth Scientific Publications, London, 1953, pp. 434–440.
23. Watson-Watt, Sir R.: "The Pulse of Radar," Dial Press, New York, 1959.
24. Skolnik, M. I.: Fifty Years of Radar, *Proc. IEEE*, vol. 73, pp. 182–197, February 1985.

25. Guerlac, H. E.: "Radar in World War II," vols. I and II, Tomask/American Institute of Physics, New York, 1987.

26. Price, A.: "The History of US Electronic Warfare," vol. 1, The Association of Old Crows, 1984.

27. Barton, D. K.: Historical Perspective on Radar, *Microwave J.*, vol. 23, p. 21, August 1980.

28. Summers, J. E., and D. J. Browning: An Introduction to Airborne Bistatic Radar, *IEE Colloq. Ground Airborne Multistatic Radar*, pp. 2/1–2/5, London, 1981.

29. Eon, L. G.: An Investigation of the Techniques Designed to Provide Early Warning Radar Fence for the Air Defense of Canada, *Defense Research Board (Canada), Rept. TELS* 100, Dec. 1, 1952.

30. Sloane, E. A., J. Salerno, E. S. Candidas, and M. I. Skolnik: A Bistatic CW Radar, *MIT Lincoln Laboratory Tech. Rept.* 82, AD 76454, Lexington, Mass., June 6, 1955.

31. Skolnik, M. I., J. Salerno, and E. S. Candidas: Prediction of Bistatic CW Radar Performance, *Symp. Radar Detection Theory, ONR Symp. Rept. ACR*-10, pp. 267–278, Washington, Mar. 1–2, 1956.

32. Skolnik, M. I.: private communication, September 1986.

33. Siegel, K. M., et al.: Bistatic Radar Cross Sections of Surfaces of Revolution, *J. Appl. Phys.*, vol. 26, pp. 297–305, March 1955.

34. Siegel, K. M.: Bistatic Radars and Forward Scattering, *Proc. Nat. Conf. Aeronaut. Electron.*, pp. 286–290, May 12–14, 1958.

35. Schultz, F. V., et al.: Measurement of the Radar Cross-Section of a Man, *Proc. IRE*, vol. 46, pp. 476–481, February 1958.

36. Crispin, J. W., Jr., et al.: "A Theoretical Method for the Calculation of Radar Cross Section of Aircraft and Missiles," *University of Michigan, Radiation Lab. Rept.* 2591-1-*H*, July 1959.

37. Hiatt, R. E., et al.: Forward Scattering by Coated Objects Illuminated by Short Wavelength Radar, *Proc. IRE*, vol. 48, pp. 1630–1635, September 1960.

38. Garbacz, R. J., and D. L. Moffett: An Experimental Study of Bistatic Scattering from Some Small, Absorber-Coated, Metal Shapes, *Proc. IRE*, vol. 49, pp. 1184–1192, July 1961.

39. Andreasen, M. G.: Scattering from Bodies of Revolution, *IEEE Trans.*, vol. AP-13, pp. 303–310, March 1965.

40. Mullin, C. R., et al.: A Numerical Technique for the Determination of the Scattering Cross Sections of Infinite Cylinders of Arbitrary Geometric Cross Section, *IEEE Trans.*, vol. AP-13, pp. 141–149, January 1965.

41. Kell, R. E.: On the Derivation of Bistatic RCS from Monostatic Measurements, *Proc. IEEE*, vol. 53, pp. 983–988, August 1965.

42. Cost, S. T.: "Measurements of the Bistatic Echo Area of Terrain of X-Band," *Ohio State University, Antenna Lab. Rept.* 1822-2, May 1965.

43. Pidgeon, V. W.: Bistatic Cross Section of the Sea, *IEEE Trans.*, vol. AP-14, pp. 405–406, May 1966.

44. Lefevre, R. J.: Bistatic Radar: New Application for an Old Technique, *WESCON Conf. Rec.*, pp. 1–20, San Francisco, 1979.

45. Fleming, F. L., and N. J. Willis: Sanctuary Radar, *Proc. Mil. Microwaves Conf.*, pp. 103–108, London, Oct. 22–24, 1980.

46. Forrest, J. R., and J. G. Schoenenberger: Totally Independent Bistatic Radar Receiver with Real-Time Microprocessor Scan Correction, *IEEE Int. Radar Conf.*, pp. 380–386, 1980.

47. Pell, C., et al.: An Experimental Bistatic Radar Trials System, *IEE Colloq. Ground Airborne Multistatic Radar*, pp. 6/1–6/12, London, 1981.

48. Schoenenberger, J. G., and J. R. Forrest: Principles of Independent Receivers for Use with Co-operative Radar Transmitters, *Radio Electron. Eng.*, vol. 52, pp. 93–101, February 1982.

49. Soame, T. A., and D. M. Gould: Description of an Experimental Bistatic Radar System, *IEE Int. Radar Conf. Publ.* 281, pp. 12–16, 1987.

50. Dunsmore, M. R. B.: Bistatic Radars for Air Defense, *IEE Int. Radar Conf. Publ.* 281, pp. 7–11, 1987.

51. Lorti, D. C., and M. Balser, Simulated Performance of a Tactical Bistatic Radar System, *IEEE EASCON 77 Rec. Publ.* 77 CH1255-9, pp. 4-4A–4-40, Arlington, Va., 1977.

52. Tactical Bistatic Radar Demonstrated, *Def. Electron.*, no. 12, pp. 78–82, 1980.

53. Auterman, J. L.: Phase Stability Requirements for a Bistatic SAR, *Proc. IEEE Nat. Radar Conf.*, pp. 48–52, Atlanta, March 1984.

54. Bistatic Radars Hold Promise for Future Systems, *Microwave Syst. News*, pp. 119–136, October 1984.

55. Griffiths, H. D., et al.: Television-Based Bistatic Radar, *Proc. IEE*, vol. 133, pt. F, pp. 649–657, December 1986.

56. Tomiyasu, K., Bistatic Synthetic Aperture Radar Using Two Satellites, *IEEE EASCON Rec.*, pp. 106–110, Arlington, Va., 1978.

57. Lee, P. K., and T. F. Coffey: Space-Based Bistatic Radar: Opportunity for Future Tactical Air Surveillance, *IEEE Int. Radar Conf.*, pp. 322–329, Washington, 1985.

58. Hsu, Y. S., and D. C. Lorti: Spaceborne Bistatic Radar—An Overview, *Proc. IEE*, vol. 133, pt. F, pp. 642–648, December 1986.

59. Anthony, S., et al.: Calibration Considerations in a Large Bistatic Angle Airborne Radar System for Ground Clutter Measurements, *Proc. IEEE Nat. Radar Conf.*, pp. 230–234, Ann Arbor, Mich., Apr. 20, 1988.

60. Walker, B. C., and M. W. Callahan: A Bistatic Pulse-Doppler Intruder-Detection Radar, *IEEE Int. Radar Conf.*, pp. 130–134, 1985.

61. Dawson, C. H.: Inactive Doppler Acquisition Systems, *Trans. AIEE*, vol. 81, pp. 568–571, January 1963.

62. Detlefsen, J.: Application of Multistatic Radar Principles to Short Range Imaging, *Proc. IEE*, vol. 133, pt. F, December 1986.

63. Nicholson, A. M., and G. F. Ross: A New Radar Concept for Short Range Application, *IEEE Int. Radar Conf.*, 1975.

64. Tyler, G. L.: The Bistatic Continuous-Wave Radar Method for the Study of Planetary Surfaces, *J. Geophys. Res.*, vol. 71, pp. 1559–1567, Mar. 15, 1966.

65. Tyler, G. L., et al.: Bistatic Radar Detection of Lunar Scattering Centers with Lunar Orbiter 1, *Science*, vol. 157, pp. 193–195, July 1967.

66. Pavel'yev, A. G., et al.: The Study of Venus by Means of the Bistatic Radar Method, *Radio Eng. Electron. Phys. (U.S.S.R.)*, vol. 23, October 1978.

67. Zebker, H. Z., and G. L. Tyler: Thickness of Saturn's Rings Inferred from Voyager 1 Observations of Microwave Scatter, *Science*, vol. 113, pp. 396–398, January 1984.

68. Tang, C. H., et al.: Measurements of Electrical Properties of the Martian Surface, *J. Geophys. Res.*, vol. 82, pp. 4305–4315, September 1977.

69. Zhou Zheng-Ou, et al.: A Bistatic Radar for Geological Probing, *Microwave J.*, pp. 257–263, May 1984.

70. Peterson, A. M., et al.: Bistatic Radar Observation of Long Period, Directional Ocean-Wave Spectra with Loran-A, *Science*, vol. 170, pp. 158–161, October 1970.

71. Doviak, R. J., et al.: Bistatic Radar Detection of High Altitude Clear Air Atmospheric Targets, *Radio Sci.*, vol. 7, pp. 993–1003, November 1972.

72. Wright, J. W., and R. I. Kressman: First Bistatic Oblique Incidence Ionograms Between Digital Ionosondes, *Radio Sci.*, vol. 18, pp. 608–614, July–August 1983.

73. Jackson, M. C.: The Geometry of Bistatic Radar Systems, *IEE Proc.*, vol. 133, pt. F, pp. 604–612, December 1986.

74. Davies, D. E. N.: Use of Bistatic Radar Techniques to Improve Resolution in the Vertical Plane, *IEE Electron. Lett.*, vol. 4, pp. 170–171, May 3, 1968.

75. McCall, E. G.: Bistatic Clutter in a Moving Receiver System, *RCA Rev.*, pp. 518–540, September 1969.

76. Crowder, H. A.: Ground Clutter Isodops for Coherent Bistatic Radar, *IRE Nat. Conv. Rec.*, pt. 5, pp. 88–94, New York, 1959.

77. Dana, R. A., and D. L. Knapp: The Impact of Strong Scintillation on Space Based Radar Design, I: Coherent Detection, *IEEE Trans.*, vol. AES-19, July 1983.

78. Pyati, V. P.: The Role of Circular Polarization in Bistatic Radars for Mitigation of Interference Due to Rain, *IEEE Trans.*, vol. AP-32, pp. 295–296, March 1984.

79. McCue, J. J. G.: Suppression of Range Sidelobes in Bistatic Radars, *Proc. IEEE*, vol. 68, pp. 422–423, March 1980.

80. Buchner, M. R.: A Multistatic Track Filter with Optimal Measurement Selection, *IEE Radar Conf.*, pp. 72–75, London, 1977.

81. Farina, A.: Tracking Function in Bistatic and Multistatic Radar Systems, *Proc. IEE*, vol. 133, pt. F, pp. 630–637, December 1986.

82. Retzer, G.: Some Basic Comments on Multistatic Radar Concepts and Techniques, *IEE Colloq. Ground Airborne Multistatic Radar*, pp. 3/1–3/3, London, 1981.

83. Hoisington, D. B., and C. E. Carroll: "Improved Sweep Waveform Generator for Bistatic Radar," U.S. Naval Postgraduate School, Monterey, Calif., August 1975.

84. Kuschel, H.: Bistatic Radar Coverage—A Quantification of System and Environmental Interferences, *IEE Int. Radar Conf. Publ.* 281, pp. 17–21, 1987.

85. Barrick, D. E. Normalization of Bistatic Radar Return, *Ohio State University, Res. Found. Rept.* 1388-13, Jan. 15, 1964.

86. Peake, W. H., and S. T. Cost: The Bistatic Echo Area of Terrain of 10 GHz, *WESCON 1968*, sess. 22/2, pp. 1–10.

87. Weiner, M. M., and P. D. Kaplan: Bistatic Surface Clutter Resolution Area at Small Grazing Angles, *MITRE Corporation, RADC-TR*-82-289, AD A123660, Bedford, Mass., November 1982.

88. Moyer, L. R., C. J. Morgan, and D. A. Rugger: An Exact Expression for the Resolution Cell Area in a Special Case of Bistatic Radar Systems, *Trans. IEEE*, vol. AES-25, July 1989.

89. Lorti, D. C., and J. J. Bowman: Will Tactical Aircraft Use Bistatic Radar?, *Microwave Syst. News*, vol. 8, pp. 49–54, September 1978.

90. Moyer, L. R. (TSC): private communication, February 1988.

91. Kock, W. I.: Related Experiments with Sound Waves and Electromagnetic Waves, *Proc. IRE*, vol. 47, pp. 1200–1201, July 1959.

92. Siegel, K. M., et al.: RCS Calculation of Simple Shapes—Bistatic, chap. 5, "Methods of Radar Cross-Section Analysis," Academic Press, New York, 1968.

93. Weil, H., et al.: Scattering of Electromagnetic Waves by Spheres, *University of Michigan, Radiat. Lab. Stud. Radar Cross Sections X, Rept.* 2255-20-T, contract AF 30(602)-1070, July 1956.

94. King, R. W. P., and T. T. Wu: "The Scattering and Diffraction of Waves," Harvard University Press, Cambridge, Mass., 1959.

95. Goodrich, R. F., et al.: Diffraction and Scattering by Regular Bodies—I: The Sphere, *University of Michigan, Dept. Electr. Eng. Rept.* 3648-1-T, 1961.

96. Matsuo, M., et al.: Bistatic Radar Cross Section Measurements by Pendulum Method, *IEEE Trans.*, vol. AP-18, pp. 83–88, January 1970.

97. Ewell, G. W., and S. P. Zehner: Bistatic Radar Cross Section of Ship Targets, *IEEE J. Ocean. Eng.*, vol. OE-5, pp. 211–215, October 1980.

98. Radar Cross-Section Measurements, *General Motors Corporation, Delco Electron. Div. Rept. R*81-152, Santa Barbara, Calif., 1981.

99. Bachman, C. G.: "Radar Targets," Lexington Books, Lexington, Mass., 1982, p. 29.

100. Paddison, F. C., et al.: Large Bistatic Angle Radar Cross Section of a Right Circular Cylinder, *Electromagnetics*, vol. 5, pp. 63–77, 1985.

101. Glaser, J. I.: Bistatic RCS of Complex Objects Near Forward Scatter, *IEEE Trans.*, vol. AES-21, pp. 70–78, January 1985.

102. Cha, Chung-Chi, et al.: An RCS Analysis of Generic Airborne Vehicles' Dependence on Frequency and Bistatic Angle, *IEEE Nat. Radar Conf.*, pp. 214–219, Ann Arbor, Mich., Apr. 20, 1988.

103. Weiner, M. M. (MITRE Corporation): private communication, April 1988.

104. Pierson, W. A., et al.: The Effect of Coupling on Monostatic-Bistatic Equivalence, *Proc. IEEE*, pp. 84–86, January 1971.

105. Barton, D. K.: "Modern Radar System Analysis," Artech House, Norwood, Mass., 1988, pp. 121–123.

106. Burk, G. J., and A. J. Foggio: "Numerical Electromagnetic Code (NEC)—Method of Moments," Naval Ocean Systems Center, San Diego, 1981.

107. Pidgeon, V. W.: Bistatic Cross Section of the Sea for Beaufort 5 Sea, *Science Technol.*, vol. 17, American Astronautical Society, San Diego, 1968, pp. 447–448.

108. Domville, A. R.: The Bistatic Reflection from Land and Sea of X-Band Radio Waves, pt. I, *GEC (Electronics) Ltd., Memo. SLM* 1802, Stanmore, England, July 1967.

109. Domville, A. R.: The Bistatic Reflection from Land and Sea of X-Band Radio Waves, pt. II, *GEC (Electronics) Ltd., Memo SLM* 2116, Stanmore, England, July 1968.

110. Domville, A. R.: The Bistatic Reflection from Land and Sea of X-Band Radio Waves, pt. II—Suppl., *GEC-AEI (Electronics) Ltd., Memo. SLM* 2116 (*Suppl.*), Stanmore, England, July 1969.

111. Larson, R. W., et al.: Bistatic Clutter Data Measurements Program, *Environmental Research Institute of Michigan, RADC-TR-*77-389, AD-A049037, November 1977.

112. Larson, R. W., et al.: Bistatic Clutter Measurements, *IEEE Trans.*, vol. AP-26, pp. 801–804, November 1978.

113. Ewell, G. W., and S. P. Zehner: Bistatic Sea Clutter Return Near Grazing Incidence, *IEE Radar Conf. Publ.* 216, pp. 188–192, October 1982.

114. Ewell, G. W., Bistatic Radar Cross Section Measurements, chap. 7 in Currie, N. C. (ed.): "Technology of Radar Reflectivity Measurement," Artech House, Norwood, Mass., 1984.

115. Ulaby, F. T., et al.: Millimeter-Wave Bistatic Scattering from Ground and Vegetation Targets, *IEEE Trans.*, vol. GE-26, no. 3, May 1988.

116. Nathanson, F. E., "Radar Design Principles," McGraw-Hill Book Company, New York, 1969.

117. Vander Schurr, R. E., and P. G. Tomlinson: Bistatic Clutter Analysis, *Decision-Science Applications, Inc., RADC-TR-*79-70, April 1979.

118. Sauermann, G. O., and P. C. Waterman: Scattering Modeling: Investigation of Scattering by Rough Surfaces, *MITRE Corporation, Rept. MTR-*2762, *AFAL-TR-*73-334, January 1974.

119. Zornig, J. G., et al.: Bistatic Surface Scattering Strength at Short Wavelengths, *Yale University, Dept. Eng. Appl. Sci. Rept. CS-*9, AD-A041316, June 1977.

120. Bramley, E. N., and S. M. Cherry: Investigation of Microwave Scattering by Tall Buildings, *Proc. IEE*, vol. 120, pp. 833–842, August 1973.

121. Brindly, A. E., et al.: A Joint Army/Air Force Investigation of Reflection Coefficient at C and K_u Bands for Vertical, Horizontal and Circular System Polarizations, *IIT Research Institute, Final Rept., TR-76-67, AD-A031403*, Chicago, July 1976.

122. Tang, C. H., et al.: Bistatic Radar Measurements of Electrical Properties of the Martian Surface, *J. Geophys. Res.*, vol. 82, pp. 4305–4315, September 1977.

123. Barton, D. K.: Land Clutter Models for Radar Design and Analysis, *Proc. IEEE*, vol. 73, pp. 198–204, February 1985.

124. Beckman, P., and A. Spizzichino: "The Scattering of EM Waves from Rough Surfaces," Pergamon Press, New York, 1963.

125. Nathanson, F., Technology Service Corporation: private communication, May 1988.

126. Hanle, E.: Pulse Chasing with Bistatic Radar-Combined Space-Time Filtering, in Schussler, H. W. (ed.): "Signal Processing II: Theories and Applications," Elsevier Science Publishers B.V., North Holland, pp. 665–668.

127. Schoenenberger, J. G., and J. R. Forrest: Principles of Independent Receivers for Use with Co-operative Radar Transmitters, *Radio Electron. Eng.*, vol. 52, pp. 93–101, February 1982.

128. Frank, J., and J. Ruze, Beam Steering Increments for a Phased Array, *IEEE Trans.*, vol. AP-15, pp. 820–821, November 1967.

129. Freedman, N.: Bistatic Radar System Configuration and Evaluation, *Raytheon Company, Independ. Dev. Proj. 76D-220, Final Rept. ER76-4414*, Dec. 30, 1976.

130. Bovey, C. K., and C. P. Horne, Synchronization Aspects for Bistatic Radars, *IEE Int. Radar Conf. Publ. 281*, pp. 22–25, 1987.

131. Schoenenberger, J. G., et al.: Design and Implementation of a UHF Band Bistatic Radar Receiver, *IEE Colloq. Ground Airborne Multistatic Radar*, pp. 7/1–7/3, London, 1981.

132. Griffiths, H. D., and S. M. Carter: Provision of Moving Target Indication in an Independent Bistatic Radar Receiver, *Radio Electron. Eng.*, vol. 54, pp. 336–342, July–August 1984.

133. Kirk, Jr., J. C.: Bistatic SAR Motion Compensation, *IEEE Int. Radar Conf.*, pp. 360–365, 1985.

134. Costas, J. P.: Synchronous Communications, *Proc. IRE*, vol. 44, pp. 1713–1718, December 1956.

135. Retzer, G.: A Concept for Signal Processing in Bistatic Radar, *IEEE Int. Radar Conf.*, pp. 288–293, 1980.

Index

I.4

Index terms	Links		
Antennas: *(Continued)*			
design considerations for	**6**.34		
directivity of	**6**.3		
and ECCM	**9**.7		
effective aperture of	**6**.4		
errors in	**6**.41		
feed blockage in	**6**.34		
feed spillover in	**6**.37		
feeds for	**6**.21	**6**.26	**18**.11
gain of	**6**.3		
for HF OTH radar	**24**.5		
introduction to	**1**.3		
leakage through surfaces of	**6**.40		
lobe pattern of	**2**.34	**2**.62	
loss of gain in, due to errors	**22**.6	**22**.9	
low-sidelobe	**7**.37		
and ECCM	**9**.24		
mechanical considerations for	**6**.40		
mirror	**18**.32	**19**.34	
for missile guidance	**19**.33		
monopulse feeds for	**6**.21	**18**.11	
multiple beams for	**6**.19		
multiple reflector	**6**.23		
noise temperature of	**2**.28		
parabolic cylinder	**6**.10	**6**.11	**6**.15
paraboloidal reflector	**6**.10		
pattern ranges for	**6**.52		
polarization of	**6**.10		
radiation patterns of	**6**.4		
radomes for	**6**.44		
reciprocity in	**6**.2		
reflector analysis of	**6**.27		

This page has been reformatted by Knovel to provide easier navigation.

Index terms		Links

Index terms	Links	

Index terms	Links	

Index terms	Links		
Lambert's law	**12**.7	**13**.16	**13**.34
Laser frequencies	**1**.18		
Lens-effect loss	**2**.52		
Letter-bands for radar	**1**.14		
Likelihood ratio	**8**.2		
Limiter	**3**.30		
Limiting in MTI radar	**15**.41		
Line-type modulators	**4**.33		
Linear array	**7**.10		
Linear-beam tubes	**4**.12		
Linear-FM pulse compression	**10**.4		
delay lines for	**10**.12		
Stretch processor	**10**.8		
Taylor weighting in	**10**.30		
Lobe switching	**18**.3		
in height finding	**20**.2		
Local oscillator	**3**.11		
Log-CFAR	**8**.18		
Logarithmic amplifier	**3**.28		
Logarithmic detector	**3**.26		
Logarithmic devices	**3**.25		
Logarithmic receiver for ECCM	**9**.18		
Loss:			
in CFAR	**3**.47	**3**.53	**8**.13
filter mismatch	**15**.14		
Loss factors in radar equation	**2**.46		
Low-angle height finding	**20**.36		
Low-angle tracking	**18**.46	**18**.54	
Low-sidelobe antennas	**7**.37	**9**.24	

Index terms		**Links**

Index terms Links

Index terms	Links

Index terms	Links	

Index terms	Links	