chapter

7

Radar Clutter

7.1 INTRODUCTION TO RADAR CLUTTER

Clutter is the term used by radar engineers to denote *unwanted* echoes from the natural environment. It implies that these unwanted echoes "clutter" the radar and make difficult the detection of wanted targets. Clutter includes echoes from land, sea, weather (particularly rain), birds, and insects. At the lower radar frequencies, echoes from ionized meteor trails and aurora also can produce clutter. The electronic warfare technique known as *chaff*,* although not an example of the natural environment, is usually considered as clutter since it is unwanted and resembles clutter from rain. Clutter is generally distributed in spatial extent in that it is much larger in physical size than the radar resolution cell. There are also "point," or discrete, clutter echoes, such as TV and water towers, buildings, and other similar structures that produce large backscatter. Large clutter echoes can mask echoes from desired targets and limit radar capability. When clutter is much larger than receiver noise, the optimum radar waveform and signal processing can be quite different from that employed when only receiver noise is the dominant limitation on sensitivity.

Radar echoes from the environment are not always undesired. Reflections from storm clouds, for example, can be a nuisance to a radar that must detect aircraft; but storm clouds

*Chaff is an electronic countermeasure that consists of a large number of thin passive reflectors, often metallic foil strips. When released from an aircraft they are quickly dispersed by the wind to form a highly reflecting cloud. A relatively small bundle of chaff can form a cloud with a radar cross section comparable to that of a large aircraft.

containing rain are what the radar meteorologist wants to detect in order to measure rainfall rate over a large area. The backscatter echoes from land can interfere with many applications of radar, but they are the target of interest for ground-mapping radar, synthetic aperture radars, and radars that observe earth resources. Thus the same environmental echo might be the desired signal in one application and the undesired clutter echo in another. The observation of land, sea, weather and other natural phenomena by radar and other sensors for the purpose of determining something about the environment is known as *remote sensing of the environment*, or simply *remote sensing*. All radars, strictly speaking, are remote sensors; but the term is usually applied only to those radars whose major function is to observe the natural environment for the purpose of extracting information about the environment. A prime example of a radar used for remote sensing is the doppler weather radar.

Echoes from land or sea are examples of *surface clutter*. Echoes from rain and chaff are examples of *volume clutter*. The magnitude of the echo from distributed surface clutter is proportional to the area illuminated. In order to have a measure of the clutter echo that is independent of the illuminated area, the *clutter cross section per unit area*, denoted by the symbol σ^0, is commonly used to describe surface clutter. It is given as

$$\sigma^0 = \frac{\sigma_c}{A_c} \tag{7.1}$$

where σ_c is the radar cross section of the clutter occupying an area A_c. The symbol σ^0 is spoken, and sometimes written, as *sigma zero*. It has also been called the *scattering coefficient, differential scattering cross section, normalized radar reflectivity, backscattering coefficient*, and *normalized radar cross section* (*NRCS*). The zero is a superscript since the subscript is reserved for the polarization employed. Sigma zero is a dimensionless quantity, and is often expressed in decibels with a reference value of one m^2/m^2.

Similarly, a cross section per unit volume is used to characterize volume clutter. It is defined as

$$\eta = \frac{\sigma_c}{V_c} \tag{7.2}$$

where σ_c in this case is the radar cross section of the clutter that occupies a volume V_c. Clutter cross section per unit volume, η, is sometimes called the *reflectivity*.

In the next section, the radar range equation for targets in surface clutter is derived along with a brief review of the general character of scattering from surface clutter. This is followed by descriptions of radar echoes from land, sea, weather, and the atmosphere. The chapter concludes by describing methods that might be used to enhance the detection of targets in clutter.

7.2 SURFACE-CLUTTER RADAR EQUATION

The radar equation describing the detection of a target in surface clutter is different from the radar equation discussed in Chap. 2, where it was assumed that detection sensitivity

was limited by receiver noise. The radar equation for detection of a target in clutter leads to different design guidelines than the radar equation for detection of a target when limited by receiver noise.

Low Grazing Angle *Consider the geometry of Fig. 7.1 which depicts a radar illuminating* the surface at a grazing angle ψ. Assume the grazing angle is small. A small grazing angle usually implies that the extent of the resolution cell in the range dimension is determined by the radar pulse width τ rather than the elevation beamwidth. The width of the cell in the cross-range dimension is determined by the azimuth beamwidth θ_B and the range R. From the simple radar equation [such as Eq. (1.7) of Sec. 1.2] the received echo power P_r is

$$P_r = \frac{P_t G A_e \sigma}{(4\pi)^2 R^4} \qquad [7.3]$$

where

P_t = transmitter power, W
G = antenna gain
A_e = antenna effective aperture, m^2
R = range, m
σ = radar cross section of the scatterer, m^2

When the echo is from a target (rather than from clutter), we let $P_r = S$ (received target signal power) and $\sigma = \sigma_t$ (target cross section). The signal power returned from a target is then

$$S = \frac{P_t G A_e \sigma_t}{(4\pi)^2 R^4} \qquad [7.4]$$

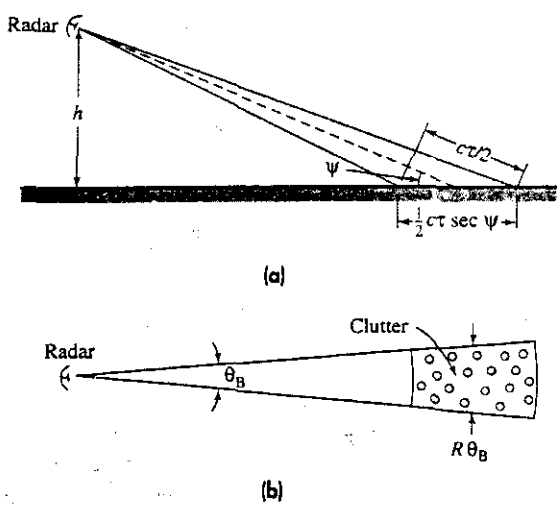


Figure 7.1 Geometry of radar surface clutter. (a) Elevation view showing the extent of the surface illuminated by the radar pulse, (b) plan view showing the illuminated clutter patch (or resolution cell) consisting of individual, independent scatterers.

When the echo is from clutter, the cross section σ becomes $\sigma_c = \sigma^0 A_c$, where the area A_c of the radar resolution cell is given from Fig. 7.1 as

$$A_c = R\theta_B (c\tau/2) \sec \psi \qquad [7.5]$$

θ_B = two-way azimuth beamwidth, c = velocity of propagation, τ = pulsewidth, and ψ = grazing angle (defined with respect to the tangent to the surface). The extent of the area A_c in the range coordinate (the range resolution) is $c\tau/2$, where the factor of 2 in the denominator accounts for the two-way propagation of radar. With these definitions, the radar equation for the surface-clutter echo-signal power C is

$$C = \frac{P_t G A_e \sigma^0 \theta_B (c\tau/2) \sec \psi}{(4\pi)^2 R^3} \qquad [7.6]$$

The received echo power C from surface clutter is seen to vary inversely as the cube of the range rather than the fourth power, as was the case for point targets in free space.

When the echo from surface clutter is large compared to receiver noise, the signal-to-clutter ratio is Eq. (7.4) divided by Eq. (7.6), or

$$\frac{S}{C} = \frac{\sigma_t}{\sigma^0 R \theta_B (c\tau/2) \sec \psi} \qquad [7.7]$$

If the maximum range $R_{\max}$ corresponds to the minimum discernible signal-to-clutter ratio $(S/C)_{\min}$, then the radar equation for the detection of a target in surface clutter at low grazing angle is

$$R_{\max} = \frac{\sigma_t}{(S/C)_{\min} \sigma^0 \theta_B (c\tau/2) \sec \psi} \qquad [7.8]$$

If pulse compression is used, the pulse width τ is that of the compressed pulse.

The azimuth beamwidth θ_B in the above equations is the two-way beamwidth. When the gaussian function can be used to approximate the beam shape (which is usually a good assumption), the two-way beamwidth is smaller than the one-way beamwidth by $\sqrt{2}$. (If the one-way beamwidth is used to calculate σ^0, the value of σ^0 will be lower by 1.5 dB than if the two-way beamwidth were used.)

The radar equation for surface clutter, Eq. (7.8), is quite different from the noise-dominated radar equation derived in Sec 1.2. The range appears as the first power rather than the fourth power as in the usual radar equation [Eq. (7.3)]. This results in greater variation of the maximum range of a clutter-dominated radar than a noise-dominated one when there is uncertainty or variability in the parameters of the radar equation. For example, if the target cross section in Eq. (7.8) were to change by a factor of two, the maximum range would also change by a factor of two. However, the same change of a factor of two in target cross section would only cause a variation in range of 1.2 (the fourth root of 2) when radar performance is determined by receiver noise alone.

There are other significant differences that affect radar design when clutter is the dominant limitation. The transmitter power does not appear explicitly in the surface-clutter radar equation. Increasing the transmitter power will increase the target signal, but the clutter echo will also increase by the same amount. Thus there is no net gain in the

detectability of desired targets. The only requirement on the transmitter power when Eq. (7.8) is used is that it be great enough to cause the clutter power at the radar receiver to be large compared to receiver noise.

Neither the antenna gain nor effective aperture enters explicitly, except as affected by the azimuth beamwidth θ_B. Equation (7.8) indicates that the narrower the azimuth beamwidth, the greater the range. Also, the narrower the pulse width, the greater the range, which is just the opposite of conventional radar detection of target echo signals in noise. A long pulse is desired when the radar is dominated by noise so as to increase the signal-to-noise energy ratio. A long pulse, on the other hand, decreases the signal-to-clutter ratio.

If the statistics of the clutter echoes are similar to the statistics of receiver noise (gaussian probability density function), the minimum signal-to-clutter ratio in Eq. (7.8) can be selected similarly to that for the signal-to-noise ratio as described in Chap. 2. Gaussian statistics, however, are rarely applicable to sea clutter and seldom for land clutter, unless the resolution cell is large. Therefore, the selection of the minimum signal-to-clutter ratio required for detection of a target in sea clutter can be difficult. When no other information is available concerning the statistics of clutter, many engineers will cautiously use the gaussian model that is used for receiver noise—and hope for the best. Even if the clutter can be described by gaussian statistics, clutter echoes do not vary with time in the same manner as receiver noise. The temporal correlation of successive clutter echoes has to be taken into account.

Radars integrate (add together) a number of echoes from a target to enhance detection, as was discussed in Sec. 2.6. Integration of pulses is generally much less effective when detection is limited by clutter than when limited by receiver noise. The statistics of receiver noise are independent (uncorrelated) in a time equal to $1/B$, where B = receiver bandwidth (usually measured in the IF portion of the radar receiver). When pulses are integrated, the sum of the noise-power fluctuations do not increase as rapidly as does the sum of the target echoes so that increased signal-to-noise ratio is obtained as the number of pulses integrated is increased. On the other hand, with perfectly stationary clutter (rocks or fence posts, for example), the clutter echo does not fluctuate from pulse to pulse. It builds up at the same rate as the target signal. There is no increase in signal-to-clutter ratio, so that integrating pulses provides no benefit. Sea clutter, however, changes with time, but slowly. At X band, for example, the decorrelation time of sea clutter is about several milliseconds. One might talk about an effective number, n_{eff}, of pulses integrated for incorporating in the numerator of Eq. (7.8), but it is likely to be much smaller than the n_{eff} for a noise-limited radar with the same number of received pulses. It is much more difficult to determine n_{eff} for clutter than for noise. Because of the uncertainties in determining n_{eff} for clutter, the conservative engineer might omit any gain due to pulse integration when detection is dominated by stationary or slowly changing clutter rather than receiver noise.

System losses also are not included explicitly in the radar equation of Eq. (7.8). Many of the system losses mentioned in Sec. 2.12 affect the target and clutter echo signals the same. Losses, therefore, have less effect on detection when signals are limited by clutter than when signals are limited by receiver noise. As long as losses do not make invalid the assumption that the clutter echo is large compared to receiver noise, losses have a lesser effect than in a noise-dominated radar.

High Grazing Angle Next consider the case where the radar observes surface clutter near perpendicular incidence. The clutter area viewed by the radar is determined by the antenna beamwidths θ_B and ϕ_B in the two principal planes. The clutter illuminated area A_c in Eq. (7.1) is $(\pi/4)R\theta_B R\phi_B/2 \sin \psi$, where ψ = grazing angle and R = range. The factor $\pi/4$ accounts for the elliptical shape of the illuminated area, and the factor of 2 in the denominator is necessary since in this case θ_B and ϕ_B are the one-way beamwidths. Substituting $\sigma = \sigma^0 A_c$ in Eq. (7.3), letting $P_r = C$ (the clutter echo power), and taking $G = \pi^2/\theta_B\phi_B$,[1] the clutter radar equation in this case is

$$C = \frac{\pi P_t A_e \sigma^0}{128R^2 \sin \psi} \qquad [7.9]$$

The clutter power is seen to vary inversely as the square of the range. This equation applies to the echo power received from the ground by a radar altimeter or the remote sensing radar known as a scatterometer. An equation for detecting a target at high grazing angles could be derived, but this represents a situation not often found in practice.

Grazing, Incidence, and Depression Angles In most of this chapter, the grazing angle is used to describe the aspect at which clutter is viewed. There are two other angles that are sometimes used instead of the grazing angle (Fig. 7.2). The *incidence angle* is defined with respect to the normal to the surface; the *grazing angle* is defined with respect to the tangent to the surface; and the *depression angle* is defined with respect to the local horizontal at the radar. The incidence angle is the complement of the grazing angle. When the earth's surface can be considered smooth and flat, the depression angle and the grazing angle are the same. When the earth's curvature must be taken into account, as in spaceborne radars, the depression angle can be quite different from the grazing angle. The incidence angle is usually used when considering earth backscatter at near perpendicular incidence, as in the altimeter and the scatterometer. The grazing angle is preferred in most other applications. Some engineers (as in references 3 and 4) prefer to use the depression angle when a rough or varying earth's surface is viewed at low grazing angles since it might be easier to determine than the grazing angle when the earth is not a flat surface.

Variation of Surface Clutter with Grazing Angle The general form of surface clutter as a function of grazing angle is shown in Fig. 7.3. There are three different scattering regions. At high grazing angles, the radar echo is due mainly to reflections from clutter that can be represented as a number of individual *planar facets* oriented so that the incident

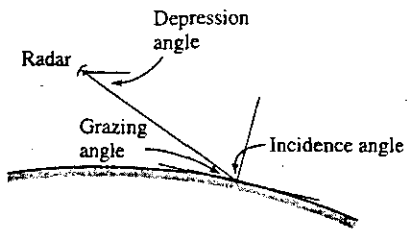


Figure 7.2 Angles used in describing geometry of the radar and surface clutter.

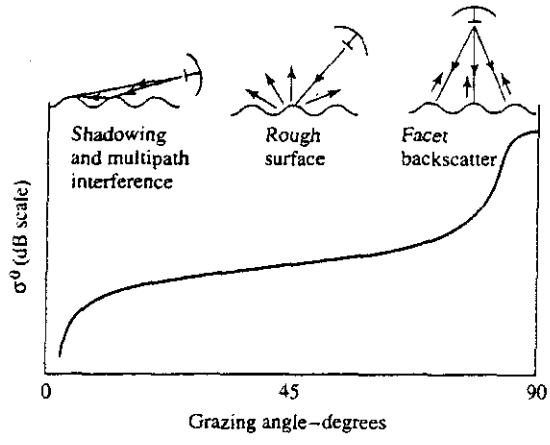


Figure 7.3 General nature of the variation of surface clutter as a function of grazing angle, showing the three major scattering regions.

energy is directed back to the radar. The backscatter can be quite large at high grazing angles. At the intermediate grazing angles, scattering is somewhat similar to that from a *rough surface*. At low grazing angles, back scattering is influenced by *shadowing* (masking) and by *multipath propagation*. Shadowing of the trough regions by the crests of waves prevents low-lying scatterers from being illuminated. Multipath reduces the energy propagating at low angles because of cancellation of the direct energy by the out-of-phase surface-reflected energy (similar to the multipath phenomenon described in Sec. 8.2). The curve drawn in Fig. 7.3 is descriptive of the general character of both land and sea scattering; but there are significant differences in the details depending on the particular type of clutter. The difference between the maximum clutter at perpendicular incidence and the minimum clutter at grazing incidence can be many tens of dB.

Mean and Median Values The characteristics of clutter echoes are usually given in statistical terms. It is often described either by the mean (average) value of σ^0 or the median (the value exceeded 50 percent of the time). For a Rayleigh probability density function, the difference between the mean and the median is small (a few percent); but for non-Rayleigh distributions, the mean and the median can be quite different. In some cases, the accuracy of clutter measurements might not be sufficient to make a distinction between the mean and the median. Nevertheless, when using clutter data, it is best to know whether it is the mean or the median that is being used. According to *Nathanson*,[2] the mean value is much better behaved than the median. The median can vary depending on the siting, pulse duration, masking, and other factors.

For proper radar design, the probability density function of the clutter echoes should be known so that the receiver detector can be designed appropriately. Unfortunately, when the statistics of the clutter cannot be described by the classical Rayleigh probability density function, it is often difficult to define a specific quantitative statistical description. For this reason, many radar designs have been based on the mean value of the clutter σ^0 rather than some statistical model. Further discussion of surface-clutter probability density functions is given later in Sec. 7.5.

7.3 LAND CLUTTER

The general nature of land clutter at low, medium, and high grazing angles is described in this section.

Land Clutter at Low Grazing Angles An extensive multiple-frequency database of land clutter at low angles was acquired by the MIT Lincoln Laboratory and reported by J. B. Billingsley and J. F. Larrabee.[3] This is one of the few collections of land clutter data that have been obtained over a wide variety of terrain, with good "ground truth," good calibration, and observations over a long period of time. The Lincoln Laboratory data provides much better understanding than previously, and it has caused some earlier notions about the nature of land clutter to be modified.

Forty-two different sites widely dispersed geographically across North America were measured, with most of the sites in western Canada. Measurements were made over a period of almost 3 years with an average time of 17 days at a site. Some sites were visited more than once to determine seasonal variations. Measurements were made at five frequencies: VHF (167 MHz), UHF (435 MHz), *L* (1.23 GHz), *S* (3.24 GHz), and *X* (9.2 GHz) bands. The radars were mobile with antennas mounted on a tower that could be extended to heights of 30, 60, or 100 ft. Range resolution was either 150 m or 36 m at VHF and UHF. At the other three frequencies, the resolution was either 150 m or 15 m. Both vertical and horizontal polarizations were employed. The rms accuracy of the clutter echo measurements over all sites was said to be 2 dB, a very good value for field operations.

The radar measured $\sigma^0 F^4$ (called *clutter strength* in the Lincoln Laboratory Report), where σ^0 is the clutter cross-section per unit area and F is the *propagation factor* that sometimes appears in the radar equation to account for propagation effects such as multipath reflections, diffraction, and attenuation. The authors define the propagation factor F as the "ratio of the incident field that actually exists at the clutter cell being measured to the incident field that would exist there if the clutter cell existed by itself in free space." Even though one might wish to separate clutter backscatter (σ^0) from propagation effects (F), it is difficult to do so and it is not generally done. (Many of the measurements of σ^0 reported in the literature are measurements of $\sigma^0 F^4$ even though they might be said to be of σ^0.)

Clutter observations were made at low depression angles, at ranges from 1 to 25 or 50 km or more. The depression angle was used rather than the grazing angle since it was difficult to define the grazing angle over a non-flat surface such as natural terrain. The depression angle in the Lincoln Laboratory report is "the complement of the incidence angle at the backscattering terrain point under consideration." This definition of depression angle includes the effect of earth curvature on the angle of illumination but not the effect of the local terrain slope.

The results of the Lincoln Laboratory measurements are summarized in Fig. 7.4 and Table 7.1. Clutter strength is given as the median value of the measured means by terrain type and frequency. The values in the figure and table were averaged over both vertical and horizontal polarizations, and with both 150 m and 15 or 36 m range resolution. This

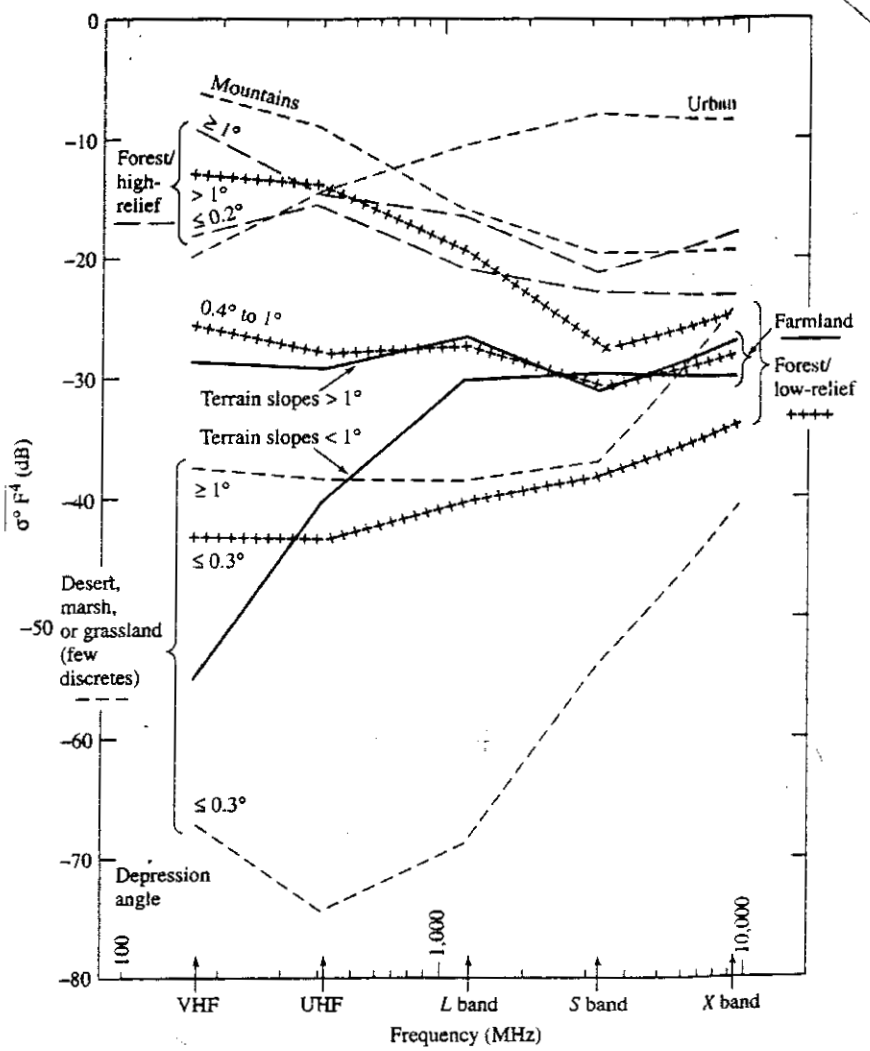


Figure 7.4 Mean clutter strength as a function of frequency for various terrain types. (From J. B. Billingsley and J. F. Larrabee.[3] Reprinted with permission of MIT Lincoln Laboratory, Lexington, Massachusetts.)

averaging was done since the variations of the mean clutter echo with both polarization and resolution were small, generally about 1 or 2 dB.

Figure 7.5 shows the mean value of rural clutter strength (no urban areas are included) as a function of frequency, measured over 36 sites. The mean values are almost independent of frequency. The average of the means is 29.2 dB, and all five values lie within 1.7 dB of this value. The standard deviation is indicated by the vertical bars, and the extreme horizontal bars in the figure represent the extremes of the measurements at each frequency. The values in Fig. 7.5 are *before* terrain classification. The corresponding one-sigma patch-to-patch variation in mean clutter strength by frequency band *after* terrain classification is 3.9, 3.8, 2.9, 2.7, and 2.3 dB at VHF, UHF, *L*, *S*, and *X* bands,

Table 7.1 Median value of mean land clutter strength over many measurements* by terrain type and frequency

Terrain Type	Median Value of $\sigma^0 F^4$ (dB)				
	Frequency Band				
	VHF	UHF	*L* Band	*S* Band	*X* Band
URBAN	−20.9	−16.0	−12.6	−10.1	−10.8
MOUNTAINS	−7.6	−10.6	−17.5	−21.4	−21.6
FOREST/HIGH-RELIEF (Terrain Slopes > 2°)					
High depression angle (>1°)	−10.5	−16.1	−18.2	−23.6	−19.9
Low depression angle (≤0.2°)	−19.5	−16.8	−22.6	−24.6	−25.0
FOREST/LOW-RELIEF (Terrain slopes < 2°)					
High depression angle (>1°)	−14.2	−15.7	−20.8	−29.3	−26.5
Intermediate depression angle (0.4° to 1°)	−26.2	−29.2	−28.6	−32.1	−29.7
Low depression angle (≤0.3°)	−43.6	−44.1	−41.4	−38.9	−35.4
AGRICULTURAL/HIGH-RELIEF (Terrain Slopes > 2°)	−32.4	−27.3	−26.9	−34.8	−28.8
AGRICULTURAL/LOW-RELIEF					
Moderately low-relief (1° < terrain slopes < 2°)	−27.5	−30.9	−28.1	−32.5	−28.4
Very low-relief (terrain slopes < 1°)	−56.0	−41.1	−31.6	−30.9	−31.5
DESERT, MARSH, OR GRASSLAND (Few discretes)					
High depression angle (≥1°)	−38.2	−39.4	−39.6	−37.9	−25.6
Low depression angle (≤0.3°)	−66.8	−74.0	−68.6	−54.4	−42.0

*Medianized (central) values within groups of like-classified measurements in a given frequency band, including both *V*- and *H*-polarizations and high (15 or 36 m) and low (150 m) range resolution.

respectively. If the terrain can be separated by class (as it is in Fig. 7.4 and Table 7.1), the variability in all bands is substantially reduced.

Some of the other findings of this comprehensive measurement program were:

- Most of the significant land clutter echoes at low angles come from spatially localized or discrete vertical features associated with the high regions of the visible landscape (such as trees, tree lines, buildings, fences, or high points of the terrain). Low regions of terrain are shadowed at low grazing angles so that the clutter is spatially patchy. Clutter occurs "within kilometer-sized macroregions of general geometric visibility, each containing hundreds or thousands of spatial resolution cells." Groups of cells that produce a strong return are often separated by cells with weak echoes or just receiver noise. (The patchiness of the clutter is what gives rise at low grazing angles to interclutter visibility, Sec. 7.8).

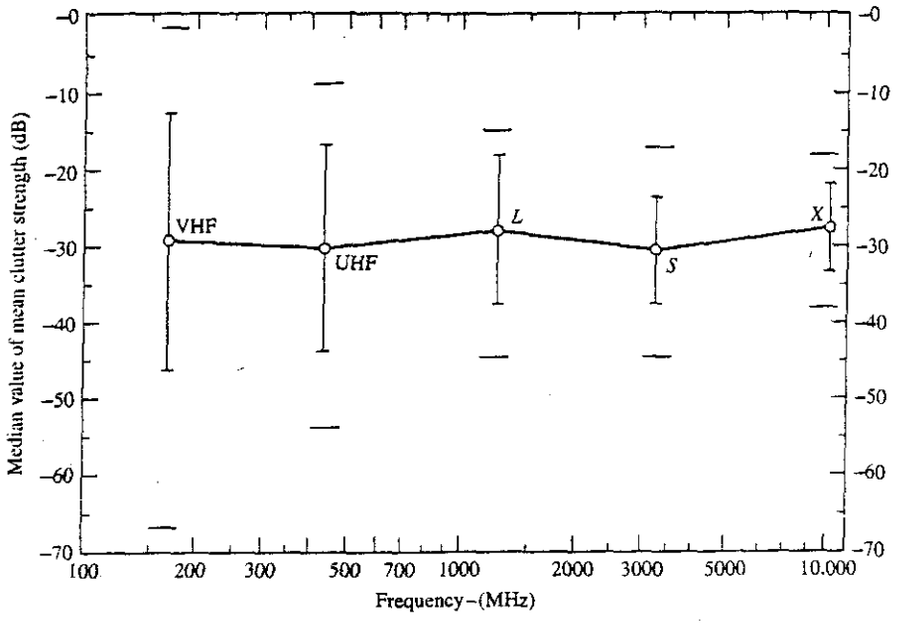


Figure 7.5 General dependence at low depression angles of the mean value of land clutter (circles), standard deviation (vertical lines), and the extreme measured values (horizontal bars) as a function of frequency at 36 rural sites.
(After J. B. Billingsley and J. F. Larrabee.[3])

- Over the range of depression angles employed in these measurements, the mean clutter strengths increase and the cell-to-cell fluctuations decrease with increasing angle. This is attributed to the reduction in masking with increasing angle. At the lower angles where masking occurs, the statistics can be represented by the Weibull probability density function. At depression angles of 6 to 8°, clutter is no longer spiky and is represented by the Rayleigh probability density.
- The difference between vertical and horizontal polarization is small, typically about 1 or 2 dB. The median difference between polarizations over all frequencies is 1.5 dB and the standard deviation is 2.8 dB, with little apparent dependence on frequency. (An exception is at VHF in steep mountainous terrain where the measurements show that the mean clutter strength is 7 to 8 dB greater with vertical polarization than with horizontal.) There was also little apparent dependence of the polarization on the range resolutions that were used. Even though the nominal calibration accuracy is 2 dB over all sites, the authors of reference 3 "accept the conclusion that, on average, the mean ground clutter strength is often 2 dB or so stronger at vertical polarization than at horizontal." They suggest that this may be due to the vertical orientation of many discrete clutter scatterers.
- The variation of clutter echo due to weather or season was found to be small, usually less than 1.5 dB for weather and 3 dB for changes in season. This was believed to be related to low-angle clutter being dominated by discrete scatterers.

- The effect of vertical discrete objects on the overall clutter strength is large even when they are relatively sparse. They contribute to clutter in a major way and are relatively unchanged with season. For example, the telephone lines, trees, and the fence around a wheat field are much more significant scatterers than the wheat itself. Even though the wheat field changes appearance with the seasons, it has little effect on the overall clutter strength. The echo from a tree line contributes strong clutter echoes no matter whether the trees are in leaf or bare, wet or dry.
- When a ground-based radar experiences strong clutter from visible terrain at ranges of 100 km or more, it is generally due to echoes from mountains that rise high enough to be within the line-of-sight of the radar. The echoes from mountains at long range are significantly greater when viewed with VHF than with microwave frequencies.
- Forests provide mean clutter echoes at VHF that are 10 to 15 dB stronger than at microwaves. This is attributed to the decreased loss at the lower frequencies when propagating radar energy through forests. Clutter from farmland, however, is 20 or 30 dB less at VHF than at microwaves since multipath (Sec. 8.2) is more of a factor with relatively flat, smooth surfaces. (Multipath, when it is present, reduces the energy at low angles.)
- Since forests tend to destroy the multipath, forest clutter echoes at high illumination angles are more likely to be the intrinsic value of σ^0 rather than $\sigma^0 F^4$.
- The median difference of the mean ground clutter strength between the low (150 m) and high resolutions (15 or 36 m) used in these measurements is less than 2 dB. Thus, on average, over many measurements, there is no significant difference expected between low and high range-resolution.
- Although the mean value of the clutter echo signal does not depend on resolution, the *variation* of the clutter echo amplitudes is a function of resolution. The smaller the resolution cell, the less the averaging within the cell and the greater is the variability from cell to cell. Reference 3 provides measurements of the ratio of the standard deviation to the mean, the skewness (third central moment), and kurtosis (fourth central moment) for the seven major terrain types at the five frequencies. Tables of the 50-, 70-, and 90-percentile levels are also given.
- Many past radar designs were based on only the mean value of the expected clutter. Different types of clutter, however, might have the same mean value, but quite different statistical variations. Figure. 7.6a is a histogram of the clutter for a particular region of forest at X band. Its mean is about the same as the histogram from farmlands shown in Fig. 7.6b. The distributions, however, are seen to be quite different. It should not be expected that the performance of a radar design based only on the mean value of clutter will be the same in these two regions. The mean-to-median ratio for forest in this case was 8 dB and for level farmland it was a very large value of 33 dB. The histograms for VHF over these same two sites are shown in Figs. 7.6c and d. The means are quite different in these two cases. The mean-to-median ratio was 4 dB for forest and 15 dB for farmland.
- The most likely day-to-day difference in the mean clutter strength was 0.2 dB, the average difference was about 1 dB, and the one-sigma range of variability beyond

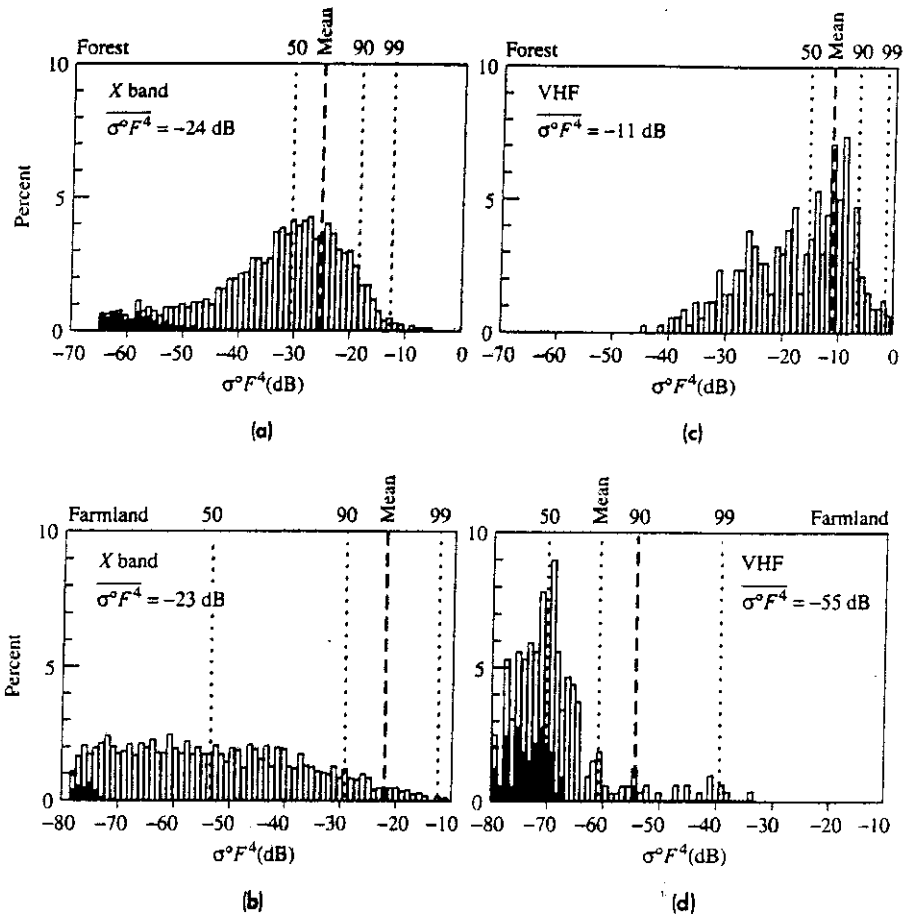


Figure 7.6 Examples of histograms of *X*-band land clutter at low depression angles in (a) a particular region of forest and (b) a particular level-farmland; both with the same approximate mean value of clutter strength, but with different statistical distributions. Histograms for the same terrain, but at VHF, are shown in (c) and (d). Black values are receiver noise.

(From J. B. Billingsley and J. F. Larrabee.[3] Reprinted with permission of MIT Lincoln Laboratory, Lexington, Massachusetts.)

the mean was 2.5 dB. Occasionally when the day-to-day difference was large (near 10 dB), it was attributed to measurement equipment malfunctions or singular events such as a train passing through the scene.

It is said that land clutter's most salient attribute is its variability. This needs to be kept in mind when designing a radar to detect targets in clutter. The design should be conservative in order to provide reliable detection over a wide range of possible clutter values.

Figure 7.7 is an X-band PPI display of the clutter seen at Gull Lake West, in Manitoba, Canada, one of the sites measured by Lincoln Laboratory. The maximum range is 7 km, range resolution is 15 m, and the polarization is horizontal. Cells with $\sigma^0 F^4 \geq -40$ dB are shown in white. The area outlined in black lines in the northwest is a sector in which much of the data was recorded. This particular area is a forested wetland out to 3.5 km range, followed by a swampy open pond from 3.5 to 5 km, then to a higher sand dune along the shore of Lake Winnipeg between 5 and 6 km. Beyond that is the water of Lake Winnipeg. This PPI display illustrates that land clutter consists of many different kinds of textures and degrees of spatial correlation. It shows that clutter does not occur as "random salt and pepper."

The cumulative amplitude distributions of the X-band echo from rural low-relief and rural high-relief terrain are shown in Fig. 7.8 as a function of depression angle.[4] By low relief is meant slopes $<2°$ and variations in height <100 ft; and by high relief, slopes $>2°$ and variations in height >100 ft. At the higher depression angles, the slope of the distribution approaches that of the Rayleigh, which indicates that microshadowing of the clutter is small at higher angles.

Figure 7.9, derived from Billingsley,[4] shows the variation of the mean and median values for low- and high-relief rural land clutter. At the top of this figure are values of the Weibull skewness parameter, which is mentioned later in Sec. 7.5 and Eq. (7.18).

Further information about the Lincoln Laboratory ground-clutter measurements can be found in the detailed reports by Billingsley.[3–5]

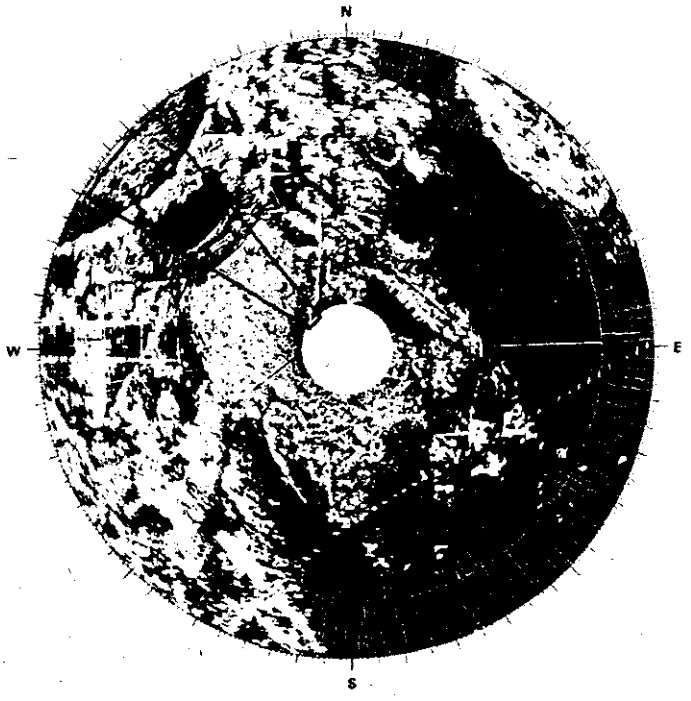


Figure 7.7 PPI clutter map at Gull Lake West, Manitoba, Canada. Cells with $\sigma^0 F^4 \geq -40$ dB are shown in white. (From J. B. Billingsley and J. F. Larrabee.[3] Reprinted with permission of MIT Lincoln Laboratory, Lexington, Massachusetts.)

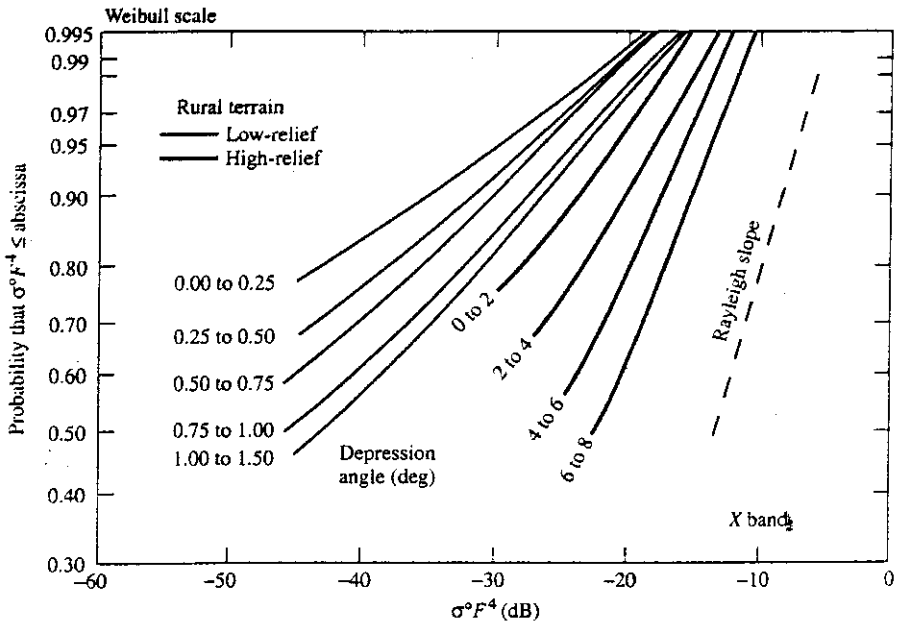


Figure 7.8 Cumulative amplitude probability distributions as a function of depression angle for X-band clutter from rural terrain of low and high relief. A straight line on this graph indicates Weibull clutter. The Rayleigh slope line is Rayleigh distribution (Weibull with $\alpha = 2$) is shown for comparison.
(From J. B. Billingsley.[4] Reprinted with permission of MIT Lincoln Laboratory, Lexington, Massachusetts.)

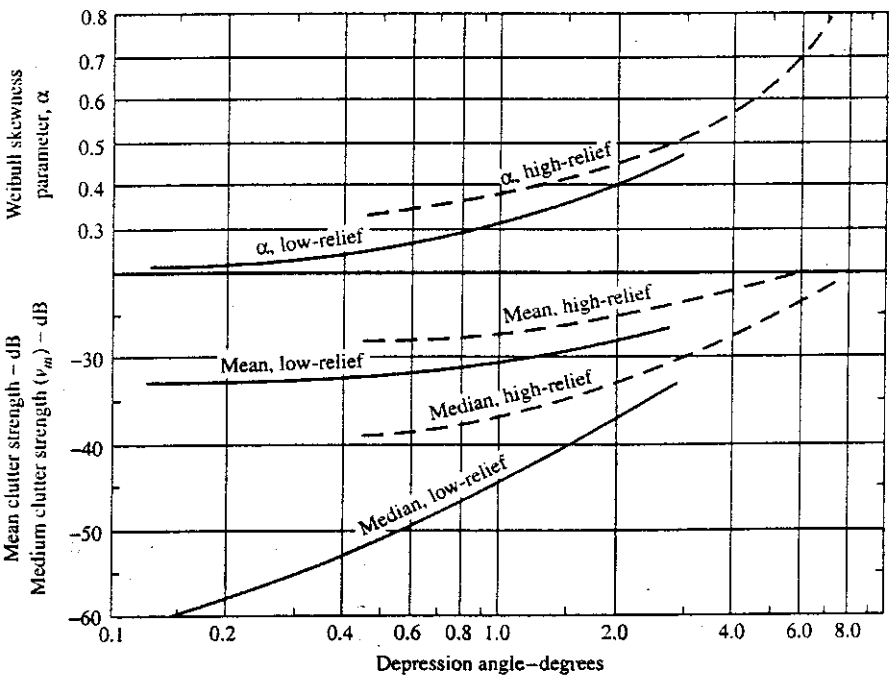


Figure 7.9 Mean and median values of low- and high-relief rural terrain as a function of depression angle. The upper curves give the Weibull skewness parameter.
(After J. Billingsley.[4])

Land Clutter at Medium Grazing Angles There exist many measurements of land clutter at medium grazing angles, from a few degrees to about 70° (see, for example, references 2,6,9,10,12,15,17–19). Much of the information in this angle regime has been obtained from aircraft or spacecraft, where there is poor ground truth about the nature of the clutter being viewed by the radar. Radars mounted on portable "cherry pickers" also have been used to acquire data at medium grazing angles, especially by those interested in remote sensing. Ground truth is easier to obtain with this type of instrumentation since the radar is a short distance from the clutter and one can readily discern its nature. Radars mounted on cherry pickers, however, do not obtain the highly averaged values possible with airborne or spaceborne sensors.

An example of the average value of clutter that might be typical of North America in the summer is shown in Fig. 7.10. This figure was derived from data presented by Moore et al.[6] It represents a combination of measurements made by the Skylab S-193 scatterometer (at 13.9 GHz) and the University of Kansas "microwave-active-spectrometer system"* mounted on a cherry picker. The ground-based cherry-picker radars could operate from 1 to 18 GHz. The data from the cherry-picker radars were combined by Moore et al. by comparing the 13.8-GHz ground-based data with the 13.9-GHz Skylab measurements. The absolute level of the model was determined by the Skylab measurements, but the relationships among the other frequencies were set by the ground-based measurements. The curve in Fig. 7.10 for L band combines the horizontal and vertical polarizations, and represents an average over the two years for which measurements were reported. The X-band values at low grazing angles were taken from the means averaged over all terrain types, as presented by Billingsley.[4] The dashed portion of the curve is a bold interpolation.

Many experimental measurements of land clutter seem to indicate that the value of sigma zero at grazing angles from a few degrees to perhaps 70° is approximately

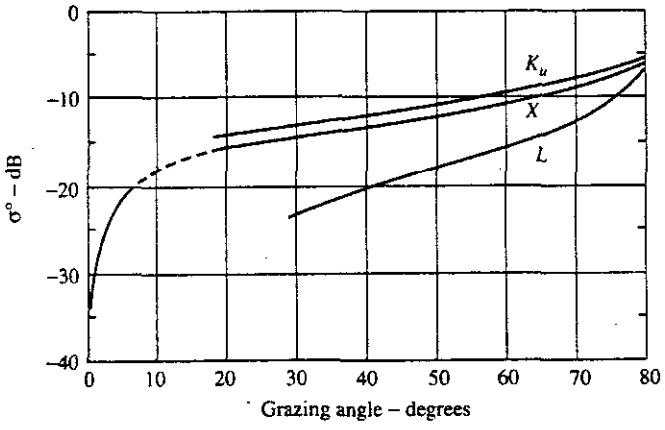


Figure 7.10 Clutter sigma zero that might be typical of North America in the summer.
(After Moore et al.[6] Courtesy of the IEEE. The X-band values at low angles are taken from Billingsley.[4])

* "A radar by any other name is still a radar.

proportional to the sine of the grazing angle ψ. For this reason, land clutter is sometimes described by the parameter γ, defined as

$$\gamma = \frac{\sigma^0}{\sin \psi} \qquad [7.10]$$

The parameter γ is said to be almost independent of the grazing angle in this angular regime. (Note that a perfectly rough surface that reradiates in the direction of the source a power per unit solid angle independent of the direction of the incident energy will result in $\sigma^0 = 2 \sin \psi$.[7] Substituting in Eq. (7.10), γ in this case is 2, or 3 dB.) Nathanson[2] states that the maximum value of γ for grazing angles from 6 to 70°, all frequencies from 0.4 to 35 GHz, and for all polarizations is -3 dB. The median is -14 dB, and the minimum is -29 dB. In this oversimplification, these values are a maximum of various experiments rather than a peak value in time. Barton[8] adds that for rural terrain covered by crops, bushes, and trees, the value of γ lies between -10 and -15 dB, and that urban and mountain clutter might have a value of γ that approaches -5 dB. Billingsley's X-band data for all terrain types at depression angles from 1 to 8° can be approximated by $\gamma = -11$ dB.[4]

These values of γ are said to apply over a wide range of frequencies; but it is known that they are frequency dependent, as was indicated in Fig. 7.10. Nathanson suggests that since these values are based to large extent on X-band data, they might be considered as applicable for X band and that the frequency dependence can be represented by

$$\gamma = \gamma_{10} + 5 \log \left(\frac{f}{10} \right) \qquad [7.11]$$

where f is the frequency in GHz and γ_{10} is the value of γ at 10 GHz. For example, if $\gamma_{10} = -14$ dB at X band, γ will be -18.5 dB at L band. Equation (7.11) is shown by the dashed curve in Fig. 7.11.

The solid curves of Fig. 7.11 are derived from Table 7.16 of Ref 2. They plot the maximum value of γ for various terrains. These are an indication of the decibel average

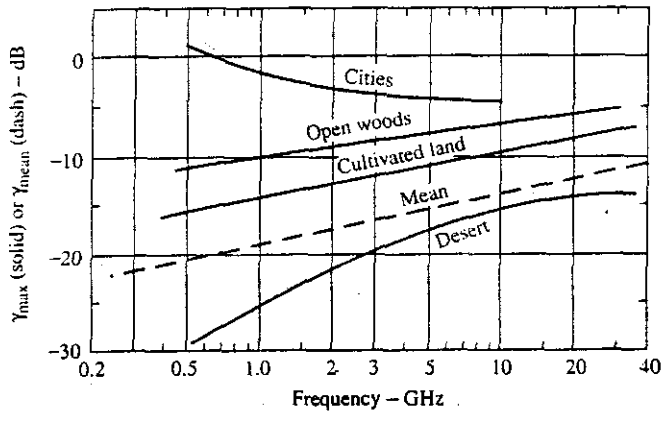


Figure 7.11 Dashed curve is the mean value of γ as given by Eq. (7.11) with $\gamma_{10} = -14$ dB at X band. The solid curves are the maximum values of γ as given by Table 2.16 of Ref. 2.

of the maximum values reported by various experimenters rather than an indication of the maximum expected return. (Because of the wide variation in the data, some liberties have been taken in smoothing the values from which Fig. 7.11 was derived.)

Land Clutter at High Grazing Angles (Near Vertical Incidence) The values of σ^0 can be large near vertical incidence. As mentioned previously, scattering in this regime is due to those facets (surfaces that are "flat" relative to a wavelength) that reflect incident energy back to the radar. The values of σ^0 near vertical incidence can also be affected by the antenna pattern and the antenna gain.

When determining the values of σ^0 at perpendicular incidence, the effect of the antenna pattern shape must be taken into account. Because of the finite antenna beamwidth, measured values of σ^0 at vertical incidence are sometimes lower than actual. This results from the averaging done over a finite beamwidth where the clutter echo is changing rapidly with angle. With a finite beamwidth antenna, the large value at the peak (90° grazing angle) is averaged with lower values at angles off normal so that the value obtained in this manner is less than it really is at 90°.

The antenna gain can influence, in some cases, the value of σ^0 near vertical incidence. Schooley[7] has shown that the value of σ^0 is $4/(\theta_{2B})^2$ at normal incidence (grazing angle = 90°) for an infinite perfectly smooth reflecting surface when the antenna is a pencil beam of two-way beamwidth θ_{2B}. The one-way beamwidth $\theta_B = \sqrt{2}\ \theta_{2B}$. Since the maximum gain G of an antenna is approximately equal to $\pi^2/(\theta_B)^2$, the value of sigma zero in this case is

$$\sigma^0 = \frac{8}{\theta_B^2} = \frac{8G}{\pi^2} \approx G \qquad [7.12]$$

Thus the antenna gain G can have a significant effect on the value of σ^0 at normal incidence and in the vicinity of normal incidence for a perfectly smooth surface. This probably applies, however, more to a perfectly smooth sea than to land. Over a perfectly flat earth with a reflection coefficient less than unity, the value of σ^0 will be reduced by the (power) reflection coefficient. If the surface is slightly rough (roughness small compared to a wavelength), the value of σ^0 will also be reduced similar to that for an antenna with errors in the aperture distribution.

Examples of the values of σ^0 near vertical incidence may be found in Chapter 12, Ground Echo, in the *Radar Handbook.*[9]

The value of σ^0 at normal incidence is generally larger than unity. That is, the radar cross section is greater than the area of the radar resolution cell.

Other Land Clutter Topics Other subjects of interest in the radar scattering from land are briefly summarized below.

Discrete Echoes Buildings and other constructed objects can result in large echoes, known as *discretes* or *point clutter.* They can range from 10^4 to 10^5 m^2 at *S* band.[10] They can be even larger at the lower frequencies. W. H. Long et al.[11] state that at the higher radar frequencies (assumed to be *S* band or above) discrete clutter echoes of 10^4 m^2 cross section might have a density of 1 per mi^2; 10^5 m^2 discretes, a density of 0.1 per mi^2; and

10^6 m^2 discretes, only 0.01 per mi^2. It takes a good MTI radar to eliminate large discrete echoes. Other techniques (such as blanking based on a clutter map) might be used to reduce the effects of discrete echoes when the MTI cannot eliminate them.

Snow The effect of snow covering the ground depends on its thickness, water content, and frequency. Dry snow (no liquid water) has a dielectric constant ranging from 1.4 to 2, so that significant transmission of energy can take place across the air-snow boundary over a wide range of angles.[12] Dry snow is of low loss so that radar energy is not highly attenuated when propagating through the snow. The ground beneath the snow, therefore, can have a major effect on σ^0, especially at low frequencies and shallow snow cover where the attenuation is quite low. It was found, for example, that 15 cm of dry powder-like snow had no effect on the measured backscatter over the frequency range from 1 to 8 GHz.[13] Any backscatter from snow under these conditions was completely dominated by the contribution from the underlying surface. With 12 cm of wet snow, however, σ^0 decreased approximately 5 to 10 dB at grazing angles between 30 and 80°.

The attenuation in propagating through snow depends on the amount of water in the snow. When the sun shines, it melts the snow and produces water, which increases attenuation. At night this water freezes. Thus there can be large variations in radar scattering, depending on whether it is night or day.[6] For example, it was reported that in the morning as the snow began to melt and produce water, the radar return decreased as much as 10 dB within half an hour (at the higher microwave frequencies, 40° grazing angle, and 48 cm snow depth). Figure 7.12 shows the effect of dry and wet snow as a function of frequency at a grazing angle of 40°.[14] At L band, the snow is essentially transparent and the echo is primarily from the underlying ground. At the upper end of the microwave frequency spectrum, at K_a band, the snow is likely to have a dominant effect on the backscatter.

Other examples of radar measurements of snow can be found in *Microwave Remote Sensing*, vol III.[15]

Snow and Altimeters The effect of snow (or ice) over the ground is dramatically illustrated by the problem it can introduce in some radar altimeters. Serious, life-threatening

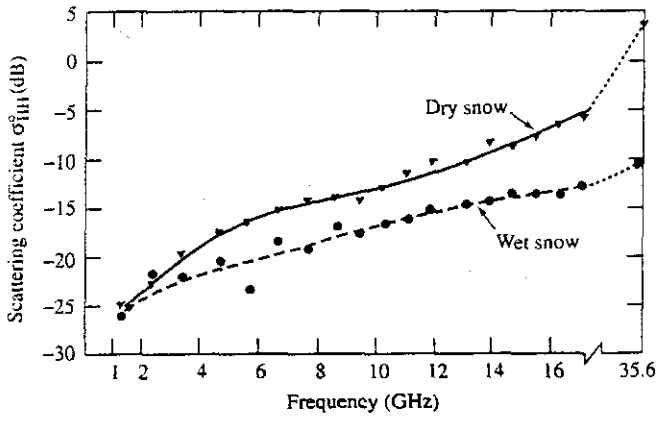


Figure 7.12 Effect of wet and dry snow as a function of frequency for a grazing angle of 40°.
(Reprinted from Ulaby and Stiles, Ref. 14, Copyright 1981, with permission from Elsevier Science.)

errors in the measurement of the aircraft altitude can result, especially at UHF or lower frequencies where the propagation loss in snow is low.[16] In some cases, the echo from the air-snow interface can be so small that it might be missed and the altitude measurement is falsely determined by the reflection from the ground underneath the snow. To make matters worse, the dielectric constant of snow compared to air results in a greater altitude measurement for a given time delay if the velocity of propagation in air rather than snow is assumed. (The velocity of propagation in snow is about 0.8 that in air; in ice it is 0.55 times that in air.) It can result in a pilot being misled as to the altitude of the aircraft above the snow and might crash into the snow, thinking that the plane's altitude was safely above the surface. This can be a serious problem with UHF altimeters when flying over a region such as the Greenland ice cap. Altimeters are now operated at the higher microwave frequencies (4.2–4.4 GHz) where the propagation loss is higher than that at UHF so that altimeter measurements are more likely to indicate the distance to the air-snow interface rather than to the ground beneath the snow.

Remote Sensing Much information on land backscatter reported in the literature has been obtained for purposes of developing radar for the remote sensing of the environment. The goal of the remote sensing radar designer is different from that of other radar designers. In remote sensing the interest is in acquiring information about the environment (such as soil moisture or crop census) while other radar engineers are interested in detecting targets in unwanted clutter. The data of interest to the remote sensing designer, therefore, might not always be useful to the radar designer, and vice versa. Examples of the backscatter from land obtained mainly for purposes of remote sensing design are given by Ulaby and Dobson[17] as well as in a computer database at the University of Massachusetts.[18]

Variability of Land Clutter As mentioned earlier in this chapter, the chief characteristic of clutter is its variability. There is seldom precise agreement among similar data taken by different investigators. Long[19] points out that "two flights over apparently the same type of terrain may at times differ by as much as 10 dB in σ^0." Another analysis of the problem of the variability of the value of σ^0 from eleven sets of measurements of agricultural crops under "similar" conditions and for approximately the same radar parameters (frequency, polarization, and grazing angle) found maximum differences of approximately 17 dB.[20] Some of these differences are due to the absence of good ground truth (that is, the measurements are not from similar terrain), accurate system calibration, or a lack of accurate data processing (including accounting for the effect of range variation of clutter as well as the effect of the antenna pattern shape). The variation and uncertainty in the value of land clutter is something the radar designer has learned to accept and compensate for by conservative system design.

Theory of Land Clutter There have been many attempts to describe scattering from land using empirical models; but these attempts have not been completely successful. In principle, if the shape and the dielectric properties of the scattering surface are known, Maxwell's equations can be solved numerically. In practice, however, the land features that cause scattering are generally too complex to describe simply.

Radar and Land Clutter Information about radar backscatter from land is required for several diverse applications, each with its own special needs that differ from the others. These applications include:

- *Detection of aircraft over land.* Clutter echoes might be 50 to 60 dB, or more, greater than aircraft echoes. MTI or pulse doppler radar (Chap. 3) is commonly used for this application to remove the large unwanted clutter.
- *Detection of moving surface-targets over land.* Vehicles or personnel can be separated from clutter by means of properly designed doppler processing radar.
- *Altimeters.* A large echo from the ground or the sea is desired for the measurement of the height of an aircraft or spacecraft since "clutter" is the target. The altimeter has also been used in "map matching" for missile guidance, as well as for remote sensing.
- *Detection and height measurement of terrain features.* This provides warning to an aircraft that it is approaching high ground so that the aircraft can fly around it (terrain avoidance) or follow the contour of the land (terrain following).
- *Mapping, or imaging, radars.* These utilize high resolution to recognize ground objects by their shape and by contrast with their surroundings. Such radars might be used by the military for navigation or target recognition. The synthetic aperture radar (SAR), sidelooking airborne radar (SLAR), and military air-to-surface radar are examples.
- *Remote sensing.* Imaging radars (such as SAR), altimeters, and scatterometers (a radar that measures sigma zero as a function of elevation angle) are used to obtain specific information about the characteristics of the earth's surface.

7.4 SEA CLUTTER

The radar echo from the sea when viewed at low grazing angles is generally smaller than the echo from land. It usually does not extend as far in range as land clutter and is more uniform over the oceans of the worlds than typical land clutter. It has been difficult, however, to establish reliable quantitative relationships between sea echo measurements and the environmental factors that determine the sea conditions. Another difficulty in dealing with sea echo is that the surface of the sea continually changes with time. Nevertheless, there does exist a large body of information regarding the radar echo from the sea that can be used for radar design and provide a general understanding of its effect on radar performance.

The nature of the radar echo (clutter) from the sea depends upon the shape of the sea surface. Echoes are obtained from those parts of the sea whose scale sizes (roughness) are comparable in dimension to the radar wavelength. The shape, or roughness, of the sea depends on the wind. Sea clutter also depends on the pointing direction of the radar antenna beam relative to the direction of the wind. Sea clutter can be affected by

contaminants that change the water surface-tension. The temperature of the water relative to that of the air is also thought to have an effect on sea clutter.

The *sea* generally consists of waves that result from the action of the wind blowing on the water surface. Such waves, called *wind waves,* cause a random-appearing ocean-height profile. *Swell waves* occur when wind waves move out of the region where they were originally excited by the wind or when the wind ceases to blow. Swell waves are less random and sometimes appear to be somewhat sinusoidal. They can travel great distances (sometimes thousands of miles) from the place where they originated. The echoes from an *X*-band radar viewing swell at low grazing angles will be small if there is no wind blowing, even if the swell waves are large. If a wind occurs, the surface will roughen and radar echoes will appear.

Sea state is a term used by mariners as a measure of wave height, as shown in Table 7.2. The sea state description shown in this table is that of the World Meteorological Organization. Sea state conditions can also be described by the Douglas scale, the Hydrographic Office scale, and the Beaufort scale. The Beaufort is actually a wind-speed scale.[21] Each gives slightly different values, so when a sea state is mentioned one should check which scale is being used.

Although sea state is commonly used to describe the roughness of the sea, it is not a complete indicator of the strength of sea clutter. Wind speed is often considered a better measure of sea clutter, but it is also limited since the effect of the wind on the sea depends on how long a time it has been blowing (called the *duration*) and over how great a distance (called the *fetch*). Once the wind starts to blow, the sea takes a finite time to grow and reach equilibrium conditions. When equilibrium is reached it is known as a *fully developed sea.* For example, a wind speed of 10 kt with a duration of 2.4 h and a fetch of at least 10 nmi produces a fully developed sea with a significant wave height (average height of the one-third highest waves) of 1.4 ft.[22] It corresponds to sea state 2. A 20 kt

Table 7.2 World Meteorological Organization sea state

	Wave Height		
Sea State	**Feet**	**Meters**	**Descriptive Term**
0	0	0	Calm, glassy
1	0–$\frac{1}{3}$	0–0.1	Calm, rippled
2	$\frac{1}{3}$–1$\frac{2}{3}$	0.1–0.5	Smooth, wavelets
3	2–4	0.6–1.2	Slight
4	4–8	1.2–2.4	Moderate
5	8–13	2.4–4.0	Rough
6	13–20	4.0–6.0	Very rough
7	20–30	6.0–9.0	High
8	30–45	9.0–14	Very high
9	over 45	over 14	Phenomenal

wind blowing for 10 h over a fetch of 75 nmi results in a significant wave height of about 8 ft and corresponds to sea state 4.

Average Value of σ^0 as a Function of Grazing Angle A composite of sea clutter data from many sources is shown in Fig. 7.13. This figure was derived from data for winds varying from approximately 10 to 20 kt, and can be considered representative of sea state 3. (Sea state 3 is roughly the medium value of sea state; i.e., about half the time over the oceans of the world, the sea state is 3 or less.) It is believed that Fig. 7.13 is representative of the average behavior of sea clutter, but there is more uncertainty in the data than is indicated by the thin lines with which the curves were drawn.

The curves of Fig. 7.13 provide the following conclusions for sea clutter with winds from 10 to 20 kt:

- At high grazing angles, above about 45°, sea clutter is independent of polarization and frequency.
- Sea clutter with vertical polarization is larger than with horizontal polarization. (At higher wind speeds the differences between the two polarizations might be less.)
- Sea clutter with vertical polarization is approximately independent of frequency. (This seems to hold at low grazing angles even down to frequencies in the HF band.)
- At low grazing angles, sea clutter with horizontal polarization decreases with decreasing frequency. This is apparently due to the interference effect at low angles between the direct radar signal and the multipath signal reflected from the sea surface.

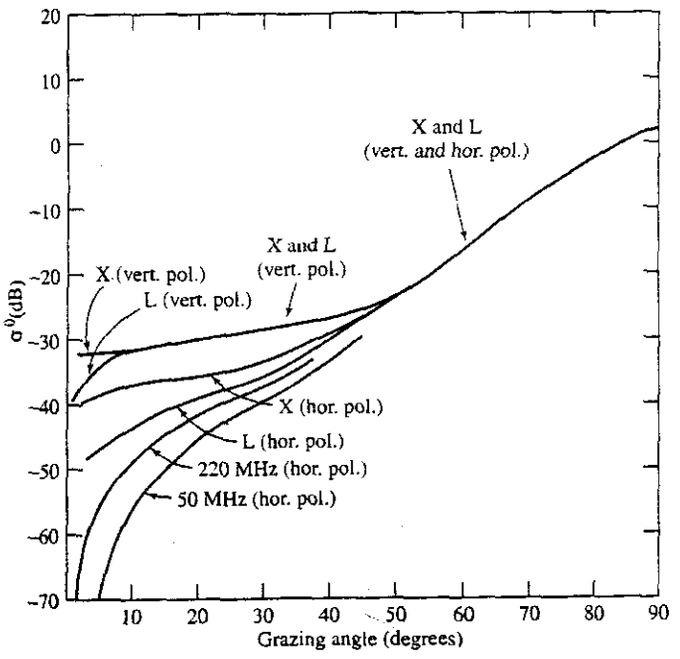


Figure 7.13 Composite of averaged sea clutter σ^0 data from various sources for wind speeds ranging from 10 to 20 kt.

There is no simple law that describes the frequency dependence of sea clutter with horizontal polarization.

Effect of Wind The wind is the most important environmental factor that determines the magnitude of the sea clutter. At low grazing angles and at microwave frequencies, backscatter from the sea is quite low when the wind speed is less than about 5 kt. It increases rapidly with increasing wind from about 5 to 20 kt, and increases more slowly at higher wind speeds. At very high winds, the increase is small with increasing wind.

Figure 7.14 was derived from experimental data of John Daley et al. using the Naval Research Laboratory Four-Frequency Airborne Radar.[23–25] As mentioned, sea clutter at the higher microwave frequencies and low grazing angles increases with increasing wind speed, but begins to level off at winds of about 15 to 25 kt. When viewed at vertical incidence (grazing angle of 90°) and with zero or low wind speed, the sea surface is flat and a large echo is directed back to the radar. As the wind speed increases and the sea surface roughens, some of the incident radar energy is scattered in directions other than back to the radar, so that σ^0 will decrease. According to Daley et al., the value of σ^0 at vertical incidence in Fig. 7.14 can be represented by $25\ w^{-0.6}$, where w is the wind speed in knots.

At low grazing angles, less than about one degree, it is difficult to provide a quantitative measure of the effect of the wind on sea clutter. This is due to the many factors that influence σ^0 at low angles, such as shadowing of parts of the sea by waves, multipath interference, diffraction, surface traveling (electromagnetic) waves, and ducted propagation.

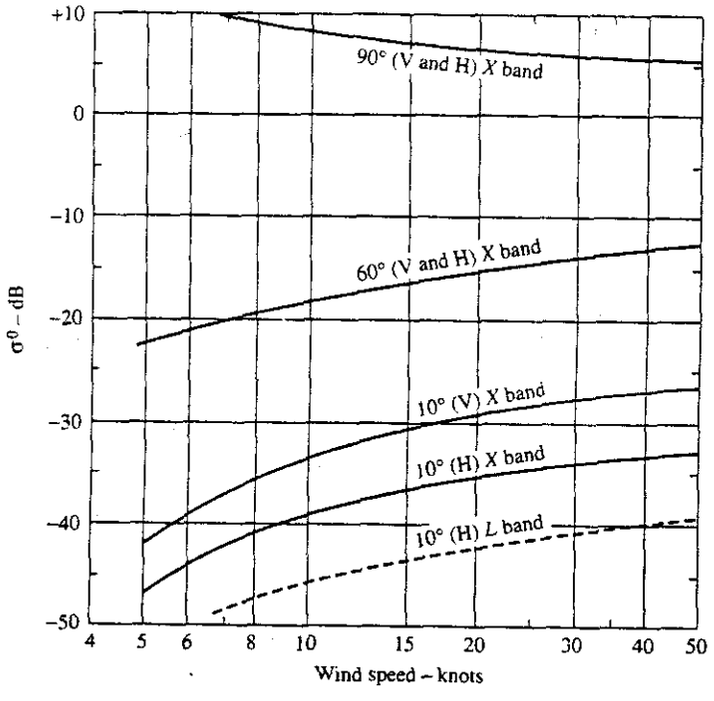


Figure 7.14 Effect of wind speed on sea clutter at several grazing angles. Radar looking upwind. Solid curves apply to *X* band, dashed curve applies to *L* band. [After J. C. Daley, et al.[23–25]]

Another factor affecting the ability to obtain reproducible measurements of sea clutter is that a finite time and a finite fetch are required for the sea to become fully developed. Unfortunately, few measurements of sea clutter mention the fetch and duration of the wind.

Sea clutter is largest when the radar looks into the wind (upwind), smallest when looking with the wind (downwind), and intermediate when looking perpendicular to the wind (crosswind). There might be as much as 5 to 10 dB variation in σ^0 as the antenna rotates 360° in azimuth.[26] Backscatter is more sensitive to wind direction at the higher frequencies than at lower frequencies; horizontal polarization is more sensitive to wind direction than vertical polarization; the ratio of σ^0 measured upwind to that measured downwind decreases with increasing grazing angle and sea state; and at UHF the backscatter is practically insensitive to wind direction at grazing angles greater than 10°.

The orthogonal component of polarization from sea clutter (cross polarization response) at grazing angles from 5 to 60° appears to be about 5 to 15 dB less than the echo from the same polarization (co-pol) as transmitted.[25]

Sea Clutter with High-Resolution Radar (Sea Spikes) The use of a clutter density, σ^0 (clutter cross section per unit area), to describe sea clutter implies that the clutter echo is independent of the illuminated area. When sea clutter is viewed by a high-resolution radar, especially at the higher microwave frequencies (such as *X* band), sea clutter is not uniform and cannot be characterized by σ^0 alone. High-resolution sea clutter is spiky. The individual echoes seen with high-resolution radar are called *sea spikes*. They are sporadic and have durations of the order of seconds. They are nonstationary in time, spatially nonhomogeneous, and have a probability density function that is non-Rayleigh. Sea spikes are important since they are the major cause of sea clutter at the higher microwave frequencies at low grazing angles, with any radar resolution.

Fig. 7.15 is an example of the time history of sea spikes in sea state 3 for pulse widths varying from 400 to 40 ns and with vertical polarization.[27,28] With 40-ns pulse width (6-m range resolution), Fig. 7.15 shows that the time between sea spikes can be several tens of seconds and the duration of each spike is of the order of one or a few seconds. As the pulse width is increased, more sea spikes appear within the larger resolution cell of the radar and the time between spikes decreases. The peak radar cross section of sea spikes in this example is almost 10 m^2. (At times they have been observed to be of even higher cross section.) The relatively large cross section and time duration of sea spikes can result in their being mistaken for small radar targets. This is a major problem with sea spikes; they can cause false alarms when a conventional detector based on gaussian receiver noise is used.

Under calm conditions, sea state 1 or less, the echo signals still have the same spiky appearance as in Fig. 7.15, but are as much as 40 dB lower in cross section. A time history similar to that of Fig. 7.15, but for horizontal polarization, would show that sea spikes with this polarization occur slightly less frequently and are sharper (of briefer duration) than with vertical polarization.

Based on airborne radar measurements made at *L*, *S*, and *X* bands with pulse widths ranging from 0.5 to 5 μs, clutter is more spiky when the radar looks upwind or downwind rather than crosswind and with low rather than high grazing angles.[29]

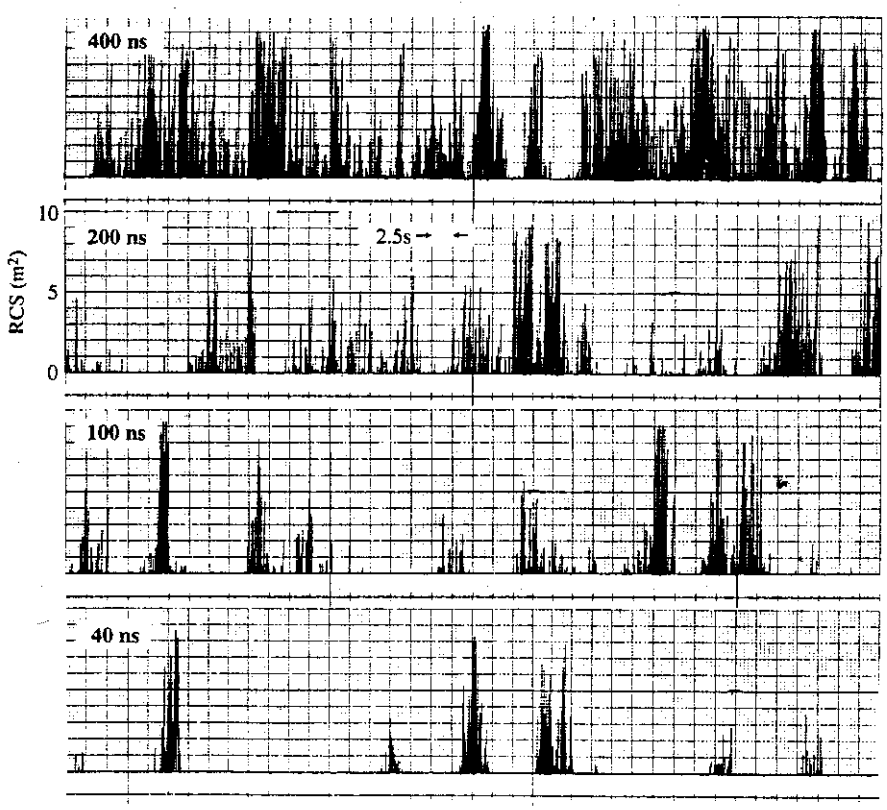


Figure 7.15 Amplitude as a function of time at a fixed range-resolution cell for low grazing angle *X*-band (9.2 GHz) sea clutter obtained off Boca Raton, Florida, with pulse widths ranging from 400 ns to 40 ns in a windblown sea with many white caps (sea state 3). Cross-range resolution is approximately 9 m, grazing angle of 1.4°, and vertical polarization.
(From J. P. Hansen and V. F. Cavaleri.[27])

Sea spikes are evident when the radar resolution is less than the water wavelength. The physical size of sea spikes usually is less than the resolution of the radar which observes them, and it is said they appear to move at approximately the surface wave velocity.[30] Sea spikes obviously are present with low as well as high resolution. At low resolution, the sum of many individual sea spikes within the resolution cell produces an almost continuous noiselike echo. At the lower frequencies where the radar wavelength is large compared to the sea-surface features that give rise to sea spikes, it would be expected that the backscatter is no longer characterized by sea spikes.

Another characteristic of sea spikes is a relatively rapid and high-percentage pulse-to-pulse amplitude modulation. According to Hansen and Cavaleri,[27] measured modulation frequencies at *X* band vary from 20 to 500 Hz. The modulating frequency seems to be affected by the type of physical surface (breaking water, sharp wave crests, ripples, etc.), the relative wind speed and direction, and polarization of the radar. The

characteristic amplitude modulations can be used for recognizing and rejecting sea spike echoes from true target echoes.[31]

In addition to changes in clutter characteristics with narrow pulse widths, as illustrated in Fig. 7.15, similar effects have been noted with narrow antenna beamwidths.[32] In some cases, target detection in clutter can actually be enhanced when a smaller antenna with a lower resolution is used rather than one with a high resolution that is bothered by sea spikes.[32] On the other hand, with very high resolution, targets can be seen in the clear regions between the sea spikes, so that subclutter visibility in the traditional sense might not be required.

Origin of Sea Spikes Sea spikes are associated with breaking waves or waves about to break. Visible whitecaps are also associated with breaking waves, but whitecaps themselves do not appear to be the cause of sea spikes since whitecaps are mainly foam with entrapped air (which does not result in significant backscatter). It has been reported[33] that about 50 percent of the time the whitecap is seen visually either simultaneously or a fraction of a second later than the appearance of a sea spike on the radar display. About 35 to 40 percent of the time a spike is observed when the waves have a very peaked crest, but with no whitecap developed. A sea spike echo can appear without the presence of a whitecap, but no whitecap is seen in the absence of a radar observation of a spike. Thus it can be concluded that the whitecap is not the cause of the sea spike echo.

Wetzel[34] has offered an explanation for the origin of sea spikes based on the entraining plume model of a spilling breaker.[35] In this model "a turbulent plume emerges from the unstable wave crest and accelerates down the forward face of the breaking wave, entraining air as it goes." Wetzel further assumes "that the breaking 'event' involves a cascade of discrete plumes emitted along the wave crest at closely spaced times." Based on this model and some assumptions about the characteristics of the plume, Wetzel was able to account for the peak radar cross section, frequency dependence, the spikier nature of horizontal polarization, similarity of the appearance of sea spikes with both horizontal and vertical polarizations at very low grazing angles, appearance of the characteristic modulation, and other properties. He has pointed out that this model is based on a simplistic hypothesis that requires further elaboration. It does not account, however, for the scattering when the radar views the sea downwind and it does not adequately explain the internal amplitude modulations of the sea spikes.

Detection of Signals in High-Resolution Sea Clutter The nature of sea spikes as seen by high-resolution radar results in a probability density function (pdf) that is not Rayleigh. Therefore, conventional methods for detection of signals in gaussian noise do not apply. (A Rayleigh pdf for clutter power is equivalent to a gaussian pdf for receiver noise voltage.) Non-Rayleigh pdfs have "high tails"; that is, there is a higher probability of obtaining a large value of clutter than when it is Rayleigh. A receiver detector designed as in Chap. 2 on the basis of gaussian noise, or Rayleigh clutter, will result in a high false-alarm rate when confronted with sea spikes. The statistics of non-Rayleigh sea clutter are not easily quantified and can vary with resolution and sea state. Thus conventional receiver detector-design based on the assumption of gaussian noise cannot be applied. To avoid excessive false alarms with non-Rayleigh sea clutter, the detection-decision

threshold might have to be increased (perhaps 20 to 30 dB). The high thresholds necessary to avoid false alarms reduce the probability of detecting desired targets. When sea spikes are a concern, detection criteria other than those based on gaussian or Rayleigh statistics must be used if a severe penalty in detection capability is to be avoided.

One method for dealing with a sea spike is to recognize its characteristic amplitude modulations and delete the sea spike from the receiver.[31] Another method is to employ a receiver with a log-log output-input characteristic whose function is to provide greater suppression of the higher values of clutter than a conventional logarithmic receiver. In a log-log receiver, the logarithmic characteristic progressively declines faster than the usual logarithmic response by a factor of 2 to 1 over the range from noise level to +80 dB above noise.[36]

Conventional pulse-to-pulse integration does not improve the detection of targets in spiky clutter because of the long correlation time of sea spikes. However, a high antenna scan rate (several hundred rpm) allows independent observations of the clutter to be made, so that scan-to-scan integration can be performed.[37] Similarly, if the target is viewed over a long period of time before a detection decision is made, the target can be recognized by its being continuously present on the display while sea spikes come and go. One method for achieving this is *time compression,* as mentioned later in Sec. 7.8.

The effect of sea spikes is less important for radars that are to detect ships since ship cross sections are large compared to the cross sections of sea spikes. Sea spikes might interfere, however, with the detection of small targets such as buoys, swimmers, submarine periscopes, debris, and small boats. With ultrahigh resolution (ultrawideband radar) where the range resolution might be of the order of centimeters, sea spikes are relatively sparse in both time and space so that it should be possible to see physically small targets when they are located in between the spiky clutter.

Sea Clutter at Very Low Grazing Angles[38] Clutter at very low grazing angles differs from that at higher angles because of shadowing, ducted propagation (Sec. 8.5), and the changing angle at which the radar ray strikes the fluctuating sea surface.

The surface of the sea is seldom perfectly flat. It usually has a time-varying angle with respect to the radar. A grazing angle might be defined with respect to the horizontal, but it is difficult to determine the angle made with respect to the dynamic sea surface. Refraction by the atmosphere can also change the angle the radar ray makes with the surface.

Shadowing of wave troughs by wave crests can occur. Scatterers as seen by the radar are thought to be mainly those from the crests, especially for horizontal polarization. Attempts to compute the effect of shadowing by simple geometrical considerations (to determine how much of the sea surface is masked) have not proven successful. One reason for failure of geometrical shadowing is that scattering features (such as sloshes, plumes, and other effects of breaking waves) are not uniformly distributed and are more likely to be near the crests of the waves. Diffraction effects can occur which complicate shadowing calculations. Thus the effect of shadowing is more complicated than just simple masking.[38,39]

Another factor to consider when searching for a theoretical understanding of sea spikes (and microwave sea clutter in general) is the effect of a surface traveling radar wave that

is launched when the incident wave has a component of electric field in the plane of incidence.[40] The effect was mentioned in Sec. 2.7 and shown in Fig. 2.10 for scattering from a long thin road. The surface traveling wave and its reflection from a discontinuity can be a reason why microwave sea clutter seen with vertical polarization usually is larger than sea clutter seen with horizontal polarization.

At very low grazing angles (less than about one degree), sea clutter should decrease rapidly with decreasing grazing angle, especially for horizontal polarization. Figure 7.3, which is a generic plot of clutter echo strength as a function of grazing angle, attempts to show this behavior. The decrease of clutter at low angles results from the cancellation of the direct and surface scattered waves as illustrated in the ideal representation given in Sec. 8.2. Measurements[41] have demonstrated this rapid decrease in sea clutter below some *critical angle* (the angle at which clutter changes from a R^{-3} dependence at short range to an R^{-7} at longer range and low grazing angle, where R = range). Not all low-angle sea clutter measurements, however, show this effect. For example, the curve for X-band sea clutter in Fig. 7.13 does not show the presence of a critical angle below which the clutter decreases rapidly.

One reason for the lack of a critical angle in some cases is that at very low grazing angles ducted propagation can occur.[42] Erroneous measurements of σ^0 can be made unless propagation effects are separated from sea-surface scattering. This requires that ducting propagation be accounted for using the proper propagation models.[43] The commonly encountered evaporation duct, which occurs a large portion of the time over most of the oceans of the world, probably is the dominant factor that gives rise to larger values of sea clutter than expected with normal (nonducting) refractive conditions.

It is difficult to obtain reliable empirical information on sea clutter at very low grazing angles that is correlated with environmental conditions. It is also difficult to obtain a theoretical understanding of the nature of the scatterers and the propagation medium at low angles. Fortunately, sea clutter is quite low at very low grazing angles and might not be a serious factor in most radar applications that have to detect targets in a sea clutter background. For example, at X band and sea state 3, σ^0 is less than -40 dB at 1.0°, less than -45 dB at 0.3°, and less than -50 dB at 0.1° grazing angle.[44] At lower frequencies, sea clutter is even less than at X band when the polarization is horizontal.

Sea Clutter at Vertical (Normal) Incidence Although sea clutter usually is much lower than land clutter at low grazing angles, it is higher than land clutter at perpendicular incidence. In the discussion of clutter at vertical (normal) incidence over a flat, perfectly reflecting land surface, it was mentioned that the value of σ^0 is approximately equal to the antenna gain G. The same is true over the sea. Thus when examining measured values of sea (or land) clutter at or near normal incidence, it needs to be kept in mind that the antenna has a significant effect on the value of the backscatter clutter power per unit area when the surface is flat or even when it is slightly rough.

Theory of Sea Clutter Many theoretical models have been proposed over the years to explain sea clutter. Most of the discussion here applies to moderate or low grazing angles. Scattering near vertical incidence generally requires a different theoretical model than at the lower grazing angles.

Past attempts to explain sea echo have been based on two different approaches. In one, the clutter is assumed to originate from scattering features at or near the sea surface. Examples include reflection from a corrugated surface,[45] backscatter from droplets of spray thrown into the air above the sea surface,[46] and backscatter from small facets, or patches, that lie on the sea surface.[41] Each of these can be applied to some limited aspect of sea echo, but none has provided an adequate explanation that accounts for the experimental evidence.

The other approach is to derive the scattering field as a boundary-value problem in which the sea surface is described by some kind of statistical process. One of the first attempts assumed that the surface disturbances could be described by a gaussian probability density function. It has been said[47] that gross observation of the sea characterized by large water wavelengths shows that the surface can be considered to be approximately gaussian, but observation of the sea's fine structure (which is what is of interest for radar backscatter) shows it is not so. Calculations of the scattering of the sea based on a gaussian surface[48] produce results that appear at first glance to be reasonable, but on close examination do not match experimental data. The conclusion that the sea cannot be represented as a gaussian statistical surface was also found by application of the theory of chaos in nonlinear dynamical systems to experimental sea clutter data.[49]

Bragg Scatter A model that has been successfully used to describe sea echo at long radar wavelengths (HF and VHF) is Bragg resonance, or Bragg scatter. This type of backscatter is so named because of its similarity to X-ray diffraction in crystals as originally put forward by Sir Lawrence Bragg and his father, Sir William Bragg, for which they jointly received the 1915 Nobel Prize in Physics. Bragg scatter is based on the coherent reinforcement of scattering from a periodic scatterer. Although real seas driven by the wind do not appear periodic, the sea can be considered as made up of a large number of individual sinewaves of different wavelengths and directions, similar to the manner in which an electrical engineer describes a noiselike voltage waveform by its frequency spectrum (also a collection of individual sinewaves of different frequencies). The Fourier transform is the means by which one can convert the voltage waveform to the frequency spectrum, and vice versa. The sea spectrum is two-dimensional (wave height as a function of frequency and direction), which differs from the one-dimensional frequency spectrum usually considered by the electrical engineer.

A rough sea surface can be described by its vertical displacement from the mean. A Fourier transform of this surface displacement gives a spectrum. Scattering from such a surface can be characterized as scattering from that particular (sinewave) frequency component of the surface spectrum that is *resonant* with the radar frequency. By resonant is meant backscattering such that the echo energy from each cycle of the sinewave reinforces coherently, as depicted in Fig. 7.16. In this figure, the radar wavelength λ_r is twice that of the spectral component λ_w (water wavelength) so that coherent addition takes place. Echoes, from any other spectral component that is not at the resonant frequency, add noncoherently and result in much smaller amplitude than the echo that experiences coherent addition from each cycle. Hence, the major scattering effect is from the resonant component and not from the other components of the sea-surface spectrum.

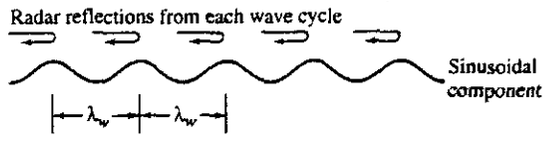


Figure 7.16 Representation of "resonant," or Bragg, scatter from a sinusoidal component of the sea-surface spectrum. Grazing angle = 0°. When the radar wavelength λ_r equals twice the wavelength λ_w of the resonant spectral component of the sea spectrum, then in-phase (coherent) addition of scattering takes place.

In general, the grazing angle is not zero, as was shown in Fig. 7.16. The Bragg resonant condition relating the radar wavelength λ_r and the water wavelength λ_w for a grazing angle ψ is given as

$$\lambda_r = 2\lambda_w \cos \psi \qquad [7.13]$$

The first to successfully apply the Bragg scattering model to sea clutter was Douglas Crombie,[50] who used it to explain the distinctive nature of the sea clutter doppler spectrum from HF radar. Bragg scatter forms the theoretical basis for remote sensing of sea and wind conditions over the sea with HF radar.[51,52] The Bragg model, however, has not been successful in explaining experimental observations of sea echo at the higher microwave frequencies.

Composite Surface Model At the higher microwave frequencies, the resonant water waves of the classical Bragg model that might contribute to backscatter have wavelengths of the order of centimeters. These short water waves are said to ride on the higher-amplitude long waves. The short-wavelength water waves are primarily responsible for the backscattering, and the long-wavelength water waves act to tilt the shorter waves. The effect of tilt can be taken into account by averaging over the tilt angles. The sea surface is thus modeled as a composite surface with two scales of roughness.[53] It has been called the *two-roughness* model or the *composite surface* model.

Long-wavelength water waves are called *gravity waves* because their velocity of propagation is determined primarily by gravity. *Capillary waves*, also called *ripples*, are small waves whose velocity is determined mainly by the surface tension of the water. Waves less than about 1.73 cm are capillary waves; waves of longer wavelength are considered gravity waves. At *X* band, the composite surface model is said to consist of short capillary waves riding on longer gravity waves. (There are, however, other surface effects in addition to capillary waves that can give rise to scattering at the higher frequencies; and these surface effects can produce greater backscatter than capillaries.)

The combination of the Bragg theory of sea echo and the composite-surface model has had considerable acceptance by some theorists since it is one of the few theories that provides an analytical basis for computation. Bragg theory, however, has serious limitations at microwave frequencies, which make its validity suspect. These include:

- The composite-surface model for microwave sea echo does not explain many of the important experimental observations of sea clutter, especially at low grazing angles.

- The theoretical formulation of Bragg scattering as a global boundary-value problem is based on the assumption that the sea-surface displacements are small compared to the radar wavelength, an assumption usually not satisfied at microwave frequencies.[54]
- Examination of sea echo obtained with high-resolution radar indicates that the dominant scattering features (sea spikes) are of relatively large amplitude (order of a few square meters radar cross section) and have a time duration of seconds. When observed with sufficient resolution, they are seen by the radar intermittently. This is far different from the view that small waves ride on top of big waves as in the composite model. Most proponents of Bragg scatter tend to consider the capillaries or ripples as the small waves. In photographs of rough sea that show capillaries or ripples, however, there are almost always present other, more significant, sea-surface features. These are generally larger than capillaries and have sharp features that can give rise to larger echoes than the capillaries. Any theory of microwave sea clutter has to be based on the observed nature of the sea and what the radar sees, which at low grazing angle is predominantly (maybe even exclusively) sea spikes.

Scattering by Surface Features One of the limitations to obtaining a satisfactory theory of sea echo is the lack of a quantitative description of the sea surface. If the exact nature of the sea surface is described, the calculation of electromagnetic scattering is achieved by solution of Maxwell's equations using more or less standard methods. Oceanographers and hydrodynamicists, however, have not been very interested in the small-scale features of the sea surface that give rise to radar scattering at microwave frequencies. The lack of a full understanding of these small-scale features has been a limitation in formulating a suitable model.

At the beginning of this discussion on sea echo theory, it was mentioned that early theorists tried to explain sea echo by postulating rather simple surface features (sinewaves, facets, and spray). Wetzel[55] has enumerated several other sea-surface features that could account for sea scatter. These are:

- *Plumes:* which slide down the front faces of breaking waves
- *Sloshes or hydraulic shocks:* structures associated with localized wind-puffs or the passage of a steep wave
- *Pools of surface roughness:* associated with breaking waves
- *Wedgelike structures*[56]
- *Pyramidal cusplike structures*

In addition, scattering from the tops of the disturbed crests might be included in the list. If the size, shape, distribution, and transient behavior of these surface features were known, one might begin to formulate a theory of sea echo.

Sea Spikes Whatever its cause and nature, the scattering feature known as the sea spike has to be a major part of any theory of sea clutter at low grazing angles at the higher microwave frequencies. Experiments show that the sea spike is all that a high-resolution radar sees. There are no other major scatterers present in the echo. Thus the sea spike should be the basis for any radar sea scatter theory at microwave frequencies (*L* band and

higher), and Bragg scatter is the basis for sea scatter theory at the lower radar frequencies (HF and VHF).

Sea Clutter Models and Radar Design As indicated, it has been difficult to formulate a satisfactory theory of microwave sea clutter that starts from fundamental principles and fully describes the experimental observations of electromagnetic scatter from the sea. A satisfactory theory should allow the radar engineer to determine optimum waveforms and related signal processing to maximize radar performance when sea clutter limits detection. Lacking a suitable theory, radar engineers have been able to design radars that operate satisfactorily in the presence of sea clutter by using simplified models and conservative design. Empirical observations are still a widely used basis for specifying a sea clutter model for engineering design or procurement specification. This is an engineering expedient that will be used until a theory suitable for reliable design is formulated.

There exist several collections of averaged values of σ^0 used to describe sea clutter. Two of the most popular are the tables given in Nathanson's book[57] and the formulas of Georgia Tech.[58] Other models have been described by Morchin.[59]

The values of σ^0 in the literature are usually the mean or the median. There is less information on the statistical properties of sea clutter. Design of the radar signal processor and detector by the methods of classical detection theory requires knowledge of the probability density function of the clutter. The spatial as well as temporal variations of clutter are also necessary, especially with high radar-resolution. These descriptions have been difficult to obtain. It has been even more difficult to relate them to the environmental parameters that affect sea clutter. There has been only limited success in devising a theory of signal processing and detector design that can be used by the radar designer when the clutter with which the radar echo must compete is time varying (nonstationary) and spatially nonhomogeneous.

If there were a satisfactory model for sea clutter, it would also provide information about the physical mechanisms occurring at the sea surface. In remote sensing, the radar is used as an instrument to measure some characteristic of the sea, such as the sea state, surface currents, and the winds that drive the sea. One can obtain empirical relations between the radar echo and some environmental parameters of the sea to be measured; but a satisfactory theory of sea clutter could advance one's ability to perform remote sensing of the environment, as well as improved detection of aircraft, missiles, ships, and other targets.

Backscatter from Sea Ice[60–62] At low grazing angles, as might occur with a shipboard radar, there is little backscattered energy from smooth, flat ice. On a PPI or similar display the areas of ice will be dark except perhaps at the edges. If areas of water are present along with the ice, the sea clutter seen on the display will be relatively bright in contrast with the ice echoes. Backscatter from rough ice, such as floes and pack ice, can produce an effect on the radar display similar to sea clutter. Clutter echoes from rough ice, however, can be distinguished from sea clutter since its pattern will remain stationary from scan to scan but the sea clutter echo pattern will change with time.[63] Measurements of ice fields near Thule, Greenland indicated that the radar backscatter (σ^0) varied linearly with frequency over a range of grazing angles from 1° to 10°, was proportional

to the grazing angle from 2° to 10°, but was inversely proportional to grazing angle from 1° to 2°.[64] The scatterometer, a radar that measures σ^0 as a function of elevation angle in the region near vertical incidence, has been used as an ice sensor to differentiate between first-year ice and multiyear ice (ice that doesn't completely melt in the summer). Shipboard radars with multiple polarizations are also able to distinguish between first-year ice, multiyear ice, and icebergs.[65] Multiyear ice is harder, and more difficult for a ship to penetrate, than first-year ice. Imaging radars, such as SAR and SLAR, have been used to image ice fields to determine the nature of the ice and inform shipping of the best routes to travel through the ice.[66]

Radar can also detect icebergs, especially if they have faces that are nearly perpendicular to the radar direction of propagation.[63] Icebergs with sloping surfaces can have a small echo even though they may be large in size. Icebergs can be readily identified in radar images by the characteristic shadows (absence of echo) they produce when viewed at low grazing angles.[67] Growlers, which are small icebergs that are large enough to be of danger to ships, are poor radar targets because of their small size and shape.

Oil Slicks[68] Oil on the surface of the sea has a smoothing effect on breaking waves. Oil slicks, therefore, are readily detectable since they appear dark on a radar PPI compared to the surrounding sea. Vertical polarization produces greater contrast in radar images of oil slicks than horizontal polarization.

7.5 STATISTICAL MODELS FOR SURFACE CLUTTER

Because of the highly variable nature of clutter echoes it is often described by a probability density function (pdf) or a probability distribution (Sec. 2.4). This section describes several statistical models that have been suggested for characterizing the fluctuations of the surface-clutter cross section per unit area, or σ^0. They can apply to both sea and land clutter. The term *distribution*, as in *Rayleigh distribution*, is used here to indicate the statistical nature of the phenomenon and applies to either the pdf or the probability distribution function. In this chapter, however, the pdf rather than the probability distribution is usually used to describe clutter statistics.

Rayleigh Distribution[69] This popular model is based on the assumption that there are a large number of randomly located independent scatterers within the clutter surface area illuminated by the radar. (The assumption of a large number is usually satisfied with as few as ten scatterers.) It is further assumed that none of the individual scatterers is significantly larger than the others.

If the radar receiver uses a linear detector, the probability density function of the voltage envelope of the Rayleigh distributed clutter at the receiver output is[17]

$$p(v) = \frac{2v}{m_2} \exp\left(-\frac{v^2}{m_2}\right) \qquad v \geq 0 \tag{7.14a}$$

where m_2 is the mean-square value (second moment) of the envelope v. (This equation differs slightly from the form given in Ref. 17.) The mean m_1 (first moment) is $(m_2\pi/4)^{1/2}$, and

the median is $(m_2 \ln 2)^{1/2}$ so that the mean-to-median ratio is $[\pi/(4 \ln 2)]^{1/2} = 1.06$, or 0.27 dB. The mean of the Rayleigh distribution is proportional to the standard deviation, or

$$\text{standard deviation} = \sqrt{\frac{4}{\pi} - 1} \times \text{mean} = 0.523 \times \text{mean}$$

The Rayleigh pdf of v normalized to its median v_m, rather than the mean-square value, is

$$p(v_n) = 2(\ln 2) v_n \exp[-(\ln 2) v_n^2] \qquad v_n \geq 0 \qquad [7.14b]$$

where $v_n = v/v_m$.

The Rayleigh pdf also describes the envelope of the output of the receiver when the input is gaussian noise, as was discussed in Chap. 2 for the detection of signals in noise. The theory given in Chap. 2 for signals in noise applies to the detection of signals in clutter when the clutter statistics are Rayleigh and the clutter echoes are independent pulse to pulse (as is receiver noise). However, this is not often true with clutter. The noise voltage from a receiver is independent from pulse to pulse since it is decorrelated in a time $1/B$, where B = bandwidth. The decorrelation time of clutter can be much longer. This is important when attempting to integrate pulses for improving detection since clutter echoes are generally not decorrelated pulse to pulse. (Correlation of clutter echoes was discussed in relation to the surface-clutter radar equation in Sec. 7.2.) To apply the methods of Chap. 2 based on Rayleigh statistics, the clutter might be decorrelated by changing the frequency pulse to pulse or by waiting a sufficiently long time between pulses to allow the temporal decorrelation of the clutter to occur (if there is clutter motion). However, these methods for decorrelating clutter are not practical in all applications. For example, frequency agility cannot be used with doppler processing. Waiting for the clutter to decorrelate due to its own internal motion might take too long.

If the receiver uses a square-law detector, the output voltage is proportional to the input signal power P. The probability density function in this case is the exponential, or[17]

$$p(P) = \frac{1}{\overline{P}} \exp\left(-\frac{P}{\overline{P}}\right) \qquad P \geq 0 \qquad [7.15]$$

where $\overline{P}$ is the mean value of the power. The standard deviation of the exponential pdf is equal to the mean.

Although the output of a linear receiver is described by the Rayleigh pdf, as in Eq. (7.14), and the output of the square-law receiver is given by the exponential pdf of Eq. (7.15), they both are considered to belong to the Rayleigh model. The term *Rayleigh clutter* derives from the basic nature of the clutter model rather than whether a linear or a square-law receiver is used.

Log-Normal Distribution As mentioned, the Rayleigh clutter model usually applies when the radar resolution cell is large so that it contains many scatterers, with no one scatterer dominant. It has been used to characterize relatively uniform clutter. However, it is not a good representation of clutter when the resolution cell size and the grazing angle are small. Under these conditions, there is a higher probability of getting large values of clutter (higher "tails") than is given by the Rayleigh model.

One of the first models suggested to describe non-Rayleigh clutter was the log-normal pdf since it has a long tail (when compared to the Rayleigh).[70] In the log-normal pdf the clutter echo power expressed in dB is gaussian. The log-normal pdf for the echo power when the receiver has a square-law detector is[17]

$$p(P) = \frac{1}{\sqrt{2\pi}\, sP} \exp\left[-\frac{1}{2s^2}\left(ln\frac{P}{P_m}\right)^2\right] \qquad P \geq 0 \qquad [7.16]$$

where s = standard deviation of $ln\ P$, and P_m = median value of P. The ratio of the mean to the median is $\exp(s^2/2)$. With a linear receiver, the pdf for the normalized output voltage amplitude $v_n = v/v_m$, with v_m = median value of v, is

$$p(v_n) = \frac{2}{\sqrt{2\pi}\, sv_n} \exp\left[-\frac{2}{s^2}(ln\ v_n)^2\right] \qquad v_n \geq 0 \qquad [7.17]$$

where s remains the standard deviation of $ln\ P$.

The log-normal distribution is specified by two parameters (the standard deviation and the median), whereas the Rayleigh is specified by only one parameter (the mean square value). Log-normal clutter is often characterized by the ratio of its mean to median.[71] Based on measurements by different experimenters, as reported in Ref. 71, the mean-to-median ratio for sea clutter varies from about 0.6 dB for a grazing angle of 4.7° and low sea state, to 5.75 dB for a high sea state and low grazing angle of 0.5°. For a particular ground clutter measurement at low grazing angles, it had a value of about 2.6 dB.

It should be expected that because of its two parameters the log-normal pdf can be made to fit experimental data better than can the one-parameter Rayleigh. However, it is sometimes said that the log-normal tends to predict higher tails for the pdf than normally experienced with most non-Rayleigh clutter, just as the Rayleigh model tends to predict lower values.

Weibull Distribution[17,72] The Weibull distribution is a two-parameter family that can be made to fit clutter measurements that lie between the Rayleigh and the log-normal. The Rayleigh is actually a special case of the Weibull; and with appropriate selection of the distribution's parameters, the Weibull can be made to approach the log-normal.

If v is the amplitude of the voltage out of a linear detector, the Weibull pdf for the normalized amplitude $v_n = v/v_m$ is

$$p(v_n) = \alpha(ln\ 2)v_n^{\alpha-1} \exp\left[-(ln\ 2)v_n^{\alpha}\right] \qquad v_n \geq 0 \qquad [7.18]$$

where α is a parameter that relates to the skewness of the distribution (sometimes called the *Weibull skewness parameter*), and v_m is the median value of the distribution. When $\alpha = 2$, the Weibull takes the form of the Rayleigh; and when $\alpha = 1$, it is the exponential pdf. The mean-to-median ratio is $(ln\ 2)^{-1/\alpha}\,\Gamma(1 + 1/\alpha)$, where $\Gamma(\cdot)$ is the gamma function. With a square-law detector, the Weibull pdf for the power $P = v^2$ is

$$p(P_n) = \beta(ln\ 2)P_n^{\beta-1} \exp\left[-(ln\ 2)P_n^{\beta}\right] \qquad P_n \geq 0 \qquad [7.19]$$

where $\beta = \alpha/2$, $P_n = P/P_m$, and $P_m = v_m^2$ is the median value of P.

The Weibull distribution has been applied to land clutter,[73,74] sea clutter,[75,76] weather clutter,[77] and sea ice.[78] Table 7.3 lists examples of the Weibull skewness parameter α for several types of land and sea clutter.[79]

Table 7.3 Weibull clutter parameters.[79]

Terrain/Sea State	Frequency	Beamwidth (deg)	Pulse Width (μs)	Grazing Angle	Weibull Parameter
Rocky mountains	S	1.5	2	.....	0.52
Wooded hills	L	1.7	3	~0.5	0.63
Forest	X	1.4	0.17	0.7	0.51–0.53
Cultivated land	X	1.4	0.17	0.7–5.0	0.61–2.0
Sea state 1	X	0.5	0.02	4.7	1.45
Sea state 3	K_u	5	0.1	1.0–30.0	1.16–1.78

With regard to land clutter at small depression angles, Billingsley[4] states: "Weibull formulations of clutter amplitude statistics represent better engineering approximations to clutter spatial amplitude distributions than do log normal formulations. Log normal formulations of clutter amplitude statistics tend to provide too much spread in the statistics."

K-Distribution This two-parameter probability distribution for modeling the statistics of clutter, proposed by Jakeman and Pusey,[80] is described as a compound distribution that consists of two components. The K-distribution probability density function describing the voltage amplitude x is

$$p(x) = \frac{2b}{\Gamma(\nu)}\left(\frac{bx}{2}\right)^{\nu} K_{\nu-1}(bx) \qquad [7.20]$$

where b is a scale parameter[81] that relates only to the mean of the clutter or σ^0, $\Gamma(\cdot)$ is the gamma function, ν is the shape parameter which depends on the higher moments in relation to the mean, and K_ν is the modified Bessel function of the second kind. (The parameter x has been used in the above instead of the more usual v for voltage so as to not confuse it with the Greek letter ν which is the shape parameter.) The statistical moments of the K-distribution lie between those of the Rayleigh and the log-normal distributions, which is sometimes used as a justification for the K-distribution since experimental sea clutter measurements seem to have the same property. Although it can be applied to land,[82] there appears to have been more interest in its application to the sea.

For sea clutter, the K-distribution is said to be made up of two components that might be associated with experimental observations.[83] There is a fast varying component, with a correlation time on the order of 5 to 10 ms. (This applies for a particular set of data obtained at X band, with a 30 ns pulse duration and a two-way 3-dB beamwidth of 1.2°.) The fast component can be decorrelated pulse to pulse by means of frequency agility. It is sometimes called the *speckle component* and its statistics can be represented by a Rayleigh distribution. The other component has a longer decorrelation time, of the order of seconds, and is unaffected by frequency agility. The slowly varying component can be represented by a gamma distribution.[84] Thus the model assumes a Rayleigh-distributed

rapidly fluctuating component modulated by a slowly fluctuating gamma-distributed component which results in the compound K-distribution of Eq. (7.20).

In the literature it has been said that the rapid component "occurs because of the multiple nature of the clutter in any illuminated patch."[85] The slowly varying component is "thought to be associated with the sea swell structure." There is not universal agreement about this description. (It is also possible to describe sea clutter as consisting of two components with different correlation times because of the observed behavior of sea spikes, as was discussed in Sec. 7.4.)

Based on measurements with an X-band radar having a 30-ns pulse width, and a 1.2° (two-way) beamwidth (resulting in a cross-range resolution of 100 to 800 m), the shape parameter ν in the K-distribution generally falls within the range 0.1 and ∞.[81] The value of ν increases with increasing cell size; and when $\nu = \infty$, the K-distribution is the same as the Rayleigh. An empirical estimate of the shape parameter ν, based on the above radar characteristics, is

$$\log \nu = \frac{2}{3} \log \psi + \frac{5}{8} \log \Delta + \zeta - k \qquad [7.21]$$

where ψ = grazing angle in degrees (from 0.1 to 10°); Δ = cross-range resolution in meters; $\zeta = -1/3$ for up or down swell directions, $+1/3$ for across swell directions, and 0 for intermediate directions or when no swell waves exist; and k gives the effect of polarization with $k = 1$ for vertical and $k = 1.7$ for horizontal polarization. No significant statistical trend was noted for variations in the sea state, wind speed, or aspect angle relative to the wind direction. It has been suggested[86] that changes in the range resolution might have a similar effect on ν as the cross-range resolution Δ, and that the parameter ζ can be expressed as $-1/3 \cos 2\theta$, where the angle θ is zero when the radar antenna points in the direction of the swell. [There is also a slightly different version of Eq. (7.21).[86]] As the range-resolution cell becomes smaller, the clutter distribution deviates from Rayleigh and ν becomes smaller. The smaller the ν, the higher the threshold for a given false-alarm probability. However, the higher the resolution, the smaller will be the clutter echo. In one analysis of range resolution,[87] it was concluded that better performance will generally be achieved with the higher resolution radar, provided the K-distribution model is maintained.

It should be cautioned that Eq. (7.21) was obtained with measurements from only a single radar and might not be universally applicable to other situations. (It would seem that ν should vary to some extent with sea state, since experiments show that the probability density function of sea clutter changes with the sea state.[88]) Although most of the above information is based on X-band data, similar behavior of the parameter ν was also observed at S band and K_u band.[89]

The above discussion has applied mainly to the noncoherent detection of targets in K-distributed clutter. The model has been extended to coherent detection,[90,91] the addition of thermal noise,[92] and the performance of CFAR.[93]

Other Statistical Distributions Other distributions that have been proposed for describing the statistics of clutter include contaminated normal,[88] Ricean,[94] Rice squared,[95] gamma,[96] and log-Weibull.[97] Shlyakhin[98] provides an extensive listing of probability density

functions for describing non-Rayleigh fluctuating radar signals. He lists twelve two-parameter, six three-parameter, and seven multiparameter distributions.

Application of the Statistical Descriptions of Clutter Probability distributions such as those given in this section have been of interest for providing a concise description of the amplitude statistics of clutter. As mentioned previously, except for the Rayleigh, there is no physical basis for the use of these distributions. They are curve-fitting models used to describe experimental data. Analyses in the literature of the detection of targets in non-Rayleigh clutter can be found that employ log-normal clutter,[99,100] Weibull clutter,[88,101] and K-distribution clutter.[102,91]

The probability distributions that contain two parameters, such as the Weibull and the K, can be made to fit experimental data better than can a single-parameter distribution. Generally, the Weibull and K-distributions have similar curve-fitting capability; but it has been said that the Weibull is easier to use. However, curve fitting of a particular distribution is not always of sufficient help to the radar engineer trying to determine how to design a signal processor to detect signals in non-Rayleigh clutter since a single statistical expression does not do well in fitting the wide variety of real-world clutter.

A comparison of the Rayleigh pdf with examples of log-normal and Weibull is shown in Fig. 7.17. (The equivalence of the Rayleigh and the Weibull with $\alpha = 2$ is noted in the figure.) The ability of a two-parameter distribution like the Weibull to fit experimental data is illustrated by the quite different shapes for the $\alpha = 0.5$ and $\alpha = 2$ curves.

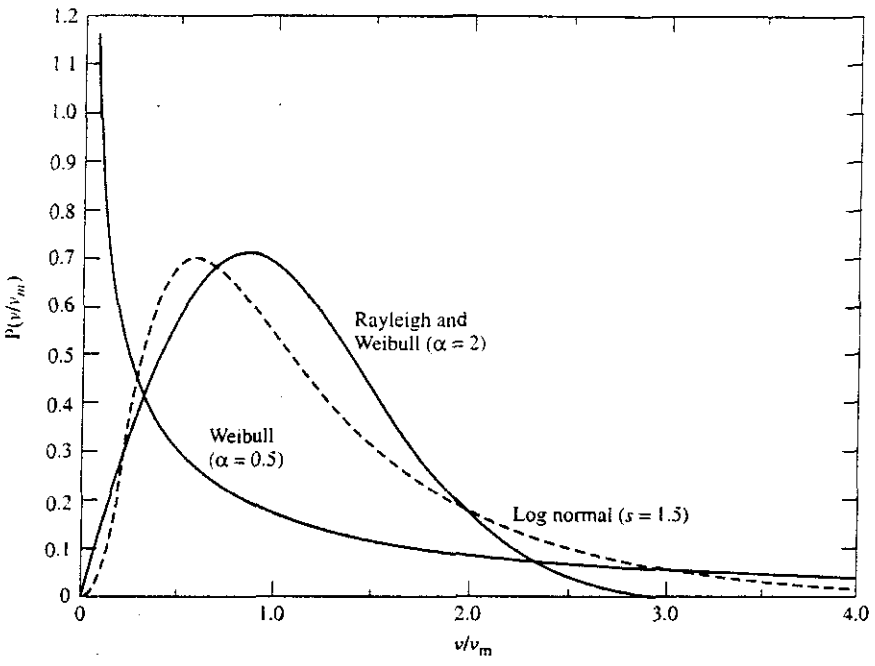


Figure 7.17 Probability density functions of the envelope v of the output of a linear detector, normalized by the median value v_m, for Rayleigh [Eq. (7.14b)], log-normal [Eq. (7.17), with a mean-to-median ratio of 5 dB], and Weibull [Eq. (7.18) with a skewness parameter of 0.5].

The theory for the detection of signals in clutter that is non-gaussian, partially correlated in time, nonhomogeneous, and nonstationary (which is the nature of real-world clutter) is a difficult problem. Classical detection-theory methods have generally been inadequate. They either contain restricting assumptions that might not be applicable or are difficult to apply to the varying nature of real-world clutter. The radar designer, when confronted with such a problem, has to design conservatively and from experience.*

7.6 WEATHER CLUTTER

Radar is much less affected by weather than are optical and infrared sensors; but weather can seriously degrade radar performance at the higher microwave and millimeter wave frequencies. Backscatter from precipitation that masks the desired target echo is the chief problem. Attenuation of the electromagnetic signal in propagating through precipitation also occurs; but at most microwave frequencies it is seldom strong enough to be of serious concern for the detection of targets.

Radar Equation for Detection of Rain[1,103–105] The radar equation describing the echo from meteorological particles such as rain, hail, sleet, and snow is of interest to the meteorologist as well as the radar engineer. The engineer needs to understand the effect weather has on the detection of targets. The meteorologist, on the other hand, is interested in using radar to determine the rate of fall of precipitation and the direction and speed of the wind. The radar equation derived below is based on the detection of rain, but it can be extended to other forms of precipitation.

Within the radar resolution cell there are many individual rain drops, each with some radar cross section σ_i. The total radar cross section, σ_c, from volume-distributed rain is the summation of the echoes from all the individual scatterers, and can be represented as

$$\sigma_c = V_c \eta = V_c \sum_i \sigma_i \qquad [7.22]$$

where η = radar cross section per unit volume, and V_c = volume of the radar resolution cell. The summation is taken over the unit volume. The volume V_c is

$$V_c = \frac{\pi}{4}(R\theta_B)(R\phi_B)\left(\frac{c\tau}{2}\right)\frac{1}{2\ ln\ 2} \qquad [7.23]$$

where R = range, θ_B = horizontal half-power beamwidth of the radar antenna, ϕ_B = vertical half-power beamwidth, τ = pulse duration, c = velocity of propagation, and ln is the natural logarithm. The factor $\pi/4$ is included to account for the elliptical shape of the antenna-beam projected area. In the interest of accuracy, the radar meteorologist reduces the

*A similar statement has been made before in this text, referring to those times when *full information is not available* to the radar designer. *Conservative* means being cautious and the inclusion of "safety factors" where appropriate. *Experience* implies that the designer learns from his or her own experience *and the experience of others*. The role of experience in engineering design, especially that which comes from failure, has been described by Duke University Civil Engineering Professor Henry Petroski in *To Engineer is Human—The Role of Failure in Successful Design*. New York: Vintage, 1992.

volume of Eq. (7.23) by a factor of 2 *ln* 2 to account for the fact that the effective volume of uniform rain illuminated by the *two-way* radar antenna pattern (described by a gaussian beamshape) is less than that indicated if the half-power (one-way) beamwidths are used in Eq. (7.23).

The echo power P_r, received from rain of radar cross section σ_c, is given by the simple radar equation of Sec. 1.2:

$$P_r = \frac{P_t G^2 \lambda^2 \sigma_c}{(4\pi)^3 R^4} \tag{7.24}$$

where P_t = transmitted power, G = antenna gain, λ = wavelength, and R = range. Using Eqs. (7.22) and (7.23), the radar equation becomes

$$P_r = \frac{P_t G^2 \lambda^2 \theta_B \phi_B c\tau}{1024(ln\ 2)\pi^2 R^2} \sum_i \sigma_i \tag{7.25}$$

The antenna gain is approximated by $G = \pi^2/\theta_B\phi_B$, if the antenna pattern can be described by a gaussian function.[1] With this substitution, the radar equation becomes

$$\overline{P}_r = \frac{P_t G \lambda^2 c\tau}{1024(ln\ 2)R^2} \sum_i \sigma_i \tag{7.26}$$

The bar over P_r denotes that the received power is averaged over many independent samples to smooth the fluctuations of the rain echo. (The radar meteorologist might average 30 or 40 samples to make the standard deviation of the received signal power less than one dB.)

The above radar equation assumes that the volume of the radar resolution cell is completely filled with uniform rain. If not, a correction must be made by introducing a dimensionless beam-filling factor, which is the fraction of the cell filled by rain. It is difficult to accurately estimate this correction. The resolution cell is not likely to be completely filled at long range or when the beam is viewing the edge of a precipitation cell.

If a single drop of rain can be considered as a sphere of diameter D_i, with circumference small compared to the radar wavelength (Rayleigh scattering region), the radar cross section is given by

$$\sigma_i = \frac{\pi^5 D_i^5}{\lambda^4} |K|^2 \tag{7.27}$$

where $|K|^2 = (\epsilon - 1)/(\epsilon + 2)$, and ϵ = dielectric constant of water. The value of $|K|^2$ for water varies with temperature and wavelength. At 10°C and 10 cm wavelength, it is approximately 0.93. Equation (7.27) also applies for ice. The value of $|K|^2$ for ice is about 0.197 and is independent of frequency in the centimeter-wavelength region. Substituting Eq. (7.27) into (7.26) yields

$$\overline{P}_r = \frac{\pi^5 P_t G c\tau}{1024(ln\ 2)R^2\lambda^2} |K|^2 \sum_i D_i^6 \tag{7.28}$$

Since the diameter D_i appears as the sixth power, in any distribution of drop sizes the small number of large drops will dominate the contribution to the radar echo.

Equation (7.28) does not include the attenuation of radar energy by rain. Attenuation usually is not of concern at microwave frequencies, except when accurate measurements of rainfall rate are required. (It is for this reason that the meteorological radar known as Nexrad is at S band, where the effects of attenuation are small.) The two-way attenuation of the radar signal in traversing the range R and back through uniform rain is $\exp(-2\alpha R)$, where α is the one-way attenuation coefficient. If rain is not uniform, the total attenuation must be expressed as the integrated value of $\exp(-2\alpha R)$ over the two-way path. When the radar equation has to include the effect of attenuation, it cannot be readily solved for the range R. (In practice, it can be solved by trial and error.)

The parameter $\sum D_i^6$ in Eq. (7.28) is called the *radar reflectivity factor* by the radar meteorologist and is denoted by Z. Experimentally, Z can be related to the rainfall rate r by

$$Z = \sum_i D_i^6 = ar^b \qquad [7.29]$$

where a and b are empirically determined constants. Equation (7.29) allows the echo signal power P_r to be related to the rainfall rate r. A number of experimenters have attempted to determine the constants in Eq. (7.29), but considerable variability exists among the reported results.[103] One form of Eq. (7.29) that has had wide acceptance is

$$Z = 200r^{1.6} \qquad [7.30]$$

where Z is in units of mm^6/m^3 and the rainfall rate r is in mm/h. (Sometimes the expression $Z = 300\,R^{1.4}$ has been used for the Nexrad WSR-88D weather radar system for rainfall estimation.) The units of Z must be properly accounted for when substituting into the radar equation. Equation (7.30) applies to stratiform rain which, in temperate climates, is relatively uniform over distances of 100 km or more (except for embedded showers that might have dimensions of the order of 10 km).[106] Intensities rarely exceed 3 mm/h except in the imbedded showers that can have rates up to 20 mm/h or more. Their vertical extent is from 4 to 6 km, and durations are of several hours. For orographic rain $Z = 31r^{1.71}$ and for thunderstorm rain $Z = 48r^{1.37}$. (Orographic rain is the rain produced when a mountain deflects moisture-laden wind upward.)

When Eq. (7.30) is substituted into Eq. (7.28) the radar equation becomes

$$\overline{P}_r = \frac{2.4P_tG\tau r^{1.6}}{R^2\lambda^2} \times 10^{-8} \qquad [7.31]$$

where r is in mm/h. R and λ in m, τ in seconds, and P_t in watts. This is the form of the equation used to measure rainfall. The received power varies as $1/R^2$ rather than the $1/R^4$ found in the usual radar equation for conventional "point" targets.

The backscatter cross section of rain per unit volume, η, as a function of wavelength and rainfall rate is shown plotted in Fig. 7.18. The dashed lines apply to Rayleigh scattering when $D_i << \lambda$, and were obtained by summing the values of σ_i given by Eq. (7.27) over the unit volume and using Eqs. (7.29) and (7.30) to give

$$\eta = \sum_i \sigma_i = 7f^4r^{1.6} \times 10^{-12} \quad \text{m}^2/\text{m}^3 \qquad [7.32]$$

where f is the radar frequency in GHz and r the rainfall rate in mm/h. The solid curves are exact values computed by Haddock.[107] Rayleigh scattering is seen to be a good approximation over most of the frequency range of interest to radar.

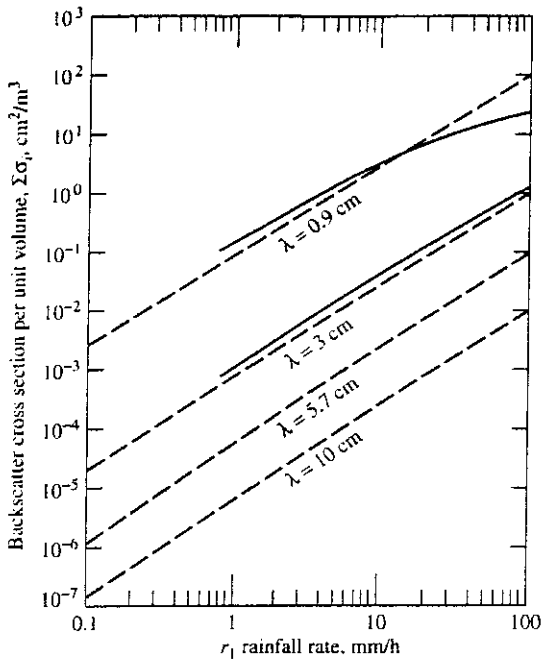


Figure 7.18 Exact (solid curves) and approximate (dashed curves) backscattering cross section per unit volume of rain at a temperature of 18°C. (From Gunn and East,[104] Quart. J. Roy. Meteor. Soc.)

The radar meteorologist prefers to describe the output of a weather radar by the parameter Z [Eq. (7.29)] rather than rainfall rate r. It would not be unusual for a weather radar to observe a rain cloud filled with rain that had not yet reached the ground or which evaporated before hitting the ground. Therefore a radar measurement of r in mm/h from such a cloud would have no meaning to an observer on the ground with a conventional rain gauge. By using Z rather than r, the meteorologist avoids this problem. The parameter Z is usually given in dB and denoted as dBZ. Using the relationship of Eq. (7.30), a rainfall rate of 1 mm/h equals 23 dBZ, 4 mm/h equals 33 dBZ, and 16 mm/h equals 42 dBZ. (This may be an incorrect usage of the precise definition of decibels as a power ratio, but it is the jargon of the radar meteorologist.) With the parameter Z instead of rainfall rate, Eq. (7.31) becomes (ignoring attenuation)

$$\bar{P}_r = \frac{1.2P_t G\tau Z}{R^2\lambda^2} \times 10^{-10} \qquad [7.33]$$

Sometimes the radar meteorologist uses, instead of Z, an equivalent radar reflectively factor Z_e when the assumption of Rayleigh scattering does not apply. It is defined as[108]

$$Z_e = \lambda^4\eta/(\pi^5|K|^2) \qquad [7.34]$$

The measurement of rainfall based on the Z-r relationship discussed in this section is not always sufficiently accurate for the meteorologist. Other microwave methods for obtaining potentially more accurate measurements are based on attenuation measured at one or two frequencies, or at two polarizations; differential reflectivity (comparison of

reflectivity measured with both horizontal and vertical polarizations); polarization differential propagation phase shift (difference in rates of phase change with increasing distance between vertically and horizontally polarized waves); and others.[109]

Radar Equation for Detection of Targets in Rain Rain can be a serious limitation to the detection of targets, especially at *L*-band and higher frequencies. The radar equation derived here indicates the important parameters that affect detection of targets in rain.

We derive the radar equation for detection of a target in rain by taking the ratio of the echo from the target and the echo from rain. It is assumed that the rain echo is much larger than receiver noise. The received signal power S from the target is (Sec. 1.2)

$$S = P_r = \frac{P_t G^2 \lambda^2 \sigma_t}{(4\pi)^3 R^4} \qquad [7.35]$$

where σ_t = target radar cross section, and the other parameters are defined as in Eq. (7.24). The rain clutter C is similar to that of Eq. (7.33), and is written

$$C = \frac{K_1 P_t G \tau Z}{R^2 \lambda^2} \qquad [7.36]$$

where K_1 is a constant = 1.2×10^{-10}. (The constant K_1 includes the velocity of propagation which has units of m/s.) The ratio of Eqs. (7.35) and (7.36) gives the signal-to-clutter ratio for a *single* pulse. If the maximum range R_{max} corresponds to the minimum detectable signal-to-clutter ratio $(S/C)_{min}$, then

$$R_{max}^2 = \frac{K_2 G \lambda^4 \sigma_t}{\tau Z (S/C)_{min}} \qquad [7.37]$$

where $K_2 = 4.2 \times 10^6$. Attenuation has not been included. (For Z, one can substitute $200r^{1.6}$, or any other suitable ar^b relationship.) It should be noted that the statistics of rain echo can be different from those of receiver noise, and the value of $(S/C)_{min}$ might not be easy to determine. We see from Eq. (7.37) that for long-range detection of targets in rain, the radar wavelength should be large (low frequency), the pulse width small, and the beamwidths small (high antenna gain).

The radar equation derived above for detection of targets in rain applies for one pulse. When a number of pulses are available from a target, they may be added together (integrated) to get larger signal-to-clutter ratio if the echoes are not correlated. It was mentioned that for land and sea clutter the pulses might not be decorrelated pulse to pulse, and the use of an effective number of pulses n_{eff} must be done with caution. Rain clutter, however, is likely to be decorrelated quicker than other clutter echoes and have the statistics of the Rayleigh pdf if the radar resolution cell is not too small and the prf is not too high. Therefore, it might be appropriate to include an effective number of pulses in the numerator of Eq. (7.37) when the conditions for independent pulses apply. The decorrelation (or independence) time of rain in seconds has been said to be

$$T_I = 0.2\lambda/\sigma_v \qquad [7.38]$$

where λ = radar wavelength in meters and σ_v = standard deviation of the velocity spectrum of the rain echo in m/s.[110] For example, at *S* band (λ = 10 cm) and σ_v = 1 m/s, the

decorrelation time is 0.02 s, which means there are only 50 independent samples of rain echo available per second. (The value of σ_v depends on wind shear, turbulence, and the terminal fall velocities of the precipitation.[111] It varies from 0.5 m/s for snow to 1 m/s for rain. In convective storms it might reach 5 m/s.)

Scattering from Snowfall Dry snow particles are essentially ice crystals, either single or aggregated. The relationship between Z and snowfall rate r is of the same form as that given by Eq. (7.29) for rain, but with different values of a,b. This allows the radar equation derived for rain, Eq. (7.28) or (7.31), to be applied to snow with the proper value of $|K|^2$ inserted (typically 0.197) and the snowfall rate r at the ground in millimeters of water measured when the snow is liquid (melted).

There have been fewer measurements of the Z-r relationship for snow than rain. The following two expressions have been suggested[112,113]

$$Z = 2000\, r^2 \qquad [7.39]$$

$$Z = 1780\, r^{2.21} \qquad [7.40]$$

Other values have been proposed, but there does not seem to be universal agreement.[114,115]

A radar is less affected by snow and ice than by rain since $|K|^2$ is smaller for ice than rain, and the snowfall rates (equivalent water content) are generally smaller than rainfall rates.

Scattering from Water-Coated Ice Spheres Moisture at altitudes where the temperature is below freezing takes the form of ice crystals, snow, or hail. As frozen particles fall to the warmer lower altitudes, they melt and form water-coated ice before becoming rain. Radar scattering by water-coated ice or snow is similar in magnitude to that of water drops of the same size and shape. Therefore, when an ice particle begins to change to liquid, radar backscatter increases since water-coated particles reflect more strongly than ice. Even for comparatively thin coatings of water, the composite particle scatters nearly as well as an all-water particle. One effect of this phenomenon is to give rise to what is called the *bright band.*

Radar observations of light precipitation show a horizontal bright band at an altitude at which the temperature is just above 0°C. It can be seen with a RHI display (range-height indicator). The center of the bright band might be 100 to 400 m below the 0°C isotherm. The echo in the center of the bright band is typically about 12 to 15 dB greater than the echo from the snow above it and about 6 to 10 dB greater than the rain beneath it.[103]

The bright band is due to changes in snow falling through the freezing level.[104] As snow just begins to melt, it changes from flat or needle-shaped particles which scatter feebly to similarly shaped particles which scatter relatively strongly due to their water coating. As melting progresses, the particles lose their extreme shapes and their velocity of fall increases. This results in a decrease in the number of particles per unit volume, and a reduction in the backscatter.

Scattering from Clouds and Fog Most water droplets in fair-weather (cumulus) clouds do not exceed 100 μm in diameter (1 μm = 10^{-6} m); consequently Rayleigh scattering

applies. Since the diameter of cloud droplets is small compared to the radar wavelength, echoes from fair-weather clouds are not of concern.

It might be possible to obtain weak echoes from a thick, intense fog at millimeter wavelengths; but at most radar frequencies, echoes due to fog may generally be regarded as insignificant.

Attenuation in Precipitation When precipitation particles are small compared to the radar wavelength (Rayleigh region), the attenuation due to absorption is small. This is the case for frequencies below *S* band. Since rain attenuation is usually small and unimportant at the longer wavelengths, the relative simplicity of the Rayleigh scattering approximation is of limited use for predicting attenuation through rain. The computation of rain attenuation must therefore be based on a more exact formulation, the results of which are shown in Fig. 7.19 as a function of wavelength and rainfall rate.[104] The attenuation produced by ice particles in the atmosphere, whether occurring as hail, snow, or ice-crystal clouds, is less than that caused by rain of an equivalent rate of precipitation.[116]

Effect of Weather on Radar Because the echo from precipitation varies as f^4, where f = frequency, UHF radars (420–450 MHz) are seldom bothered by weather effects. At *L* band weather echoes can be a problem, and some method for seeing aircraft targets in weather is usually needed. A radar at *S* band will have its range considerably reduced in modest rainfall if nothing is done to reduce the effect of rain. Radars at higher frequencies are even further degraded by rain. (Airborne weather-avoidance radars at *X* band, for example, can be severely degraded by heavy rain and prevent the radar from seeing hazardous weather.)

A typical specification for an air-surveillance radar might be that it has to detect its target when rainfall in the vicinity of the target is at the rate of 4 mm/h. This is called a

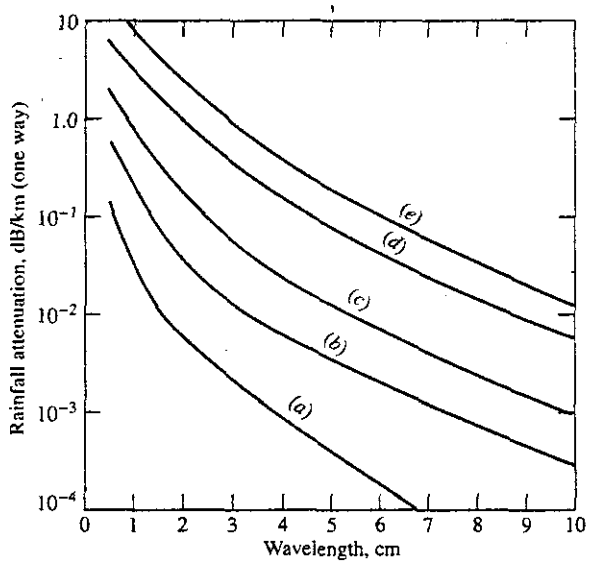


Figure 7.19 *One-way attenuation* in rain at a temperature of 18°C. (a) Drizzle—0.25 mm/h; (b) light rain—1 mm/h; (c) moderate rain—4 mm/h; (d) heavy rain—16 mm/h; (e) excessive rain—40 mm/h. In Washington, D.C., a rainfall rate of 0.25 mm/h is exceeded 450 h/yr, 1 mm/h is exceeded 200 h/yr, 4 mm/h is exceeded 60 h/yr, 16 mm/h is exceeded 8 h/yr, and 40 mm/h is exceeded 2.2 h/yr.

moderate rain. At Washington, D.C., this rainfall rate is equaled or exceeded approximately 60 h/yr.[117] As has been mentioned, attenuation is generally not a problem at frequencies below *X* band, unless the precipitation is very heavy.

7.7 OTHER SOURCES OF ATMOSPHERIC ECHOES

Radar echoes are sometimes obtained from regions of the atmosphere where no apparent reflecting sources seem to exist. Before the causes of these echoes were understood they were called *angels*. Today we know that discrete or "point" angel echoes are usually due to birds or insects. Even though the echo from a single bird or insect is small, it can be detected by radar if the range is short enough. For example, a radar that can detect a target with a radar cross section of 1 m^2 at a range of 200 nmi can detect a 10^{-2} m^2 target (a large bird) at a range greater than 60 nmi, and a 10^{-4} m^2 target (a large insect) at a range of 20 nmi. Angel echoes also can be caused by large flocks of birds, especially during migration season, or by large numbers of insects.

Distributed angel echoes can be due to inhomogeneities in the atmosphere such as clear-air convective cells and other turbulent atmospheric effects. Echoes from atmospheric turbulence are weak and can only be detected by powerful radars or at short range. Echoes from birds and insects, on the other hand, can be quite substantial in numbers and can be seen within a large area of the radar coverage. They can present a problem to an unprepared radar designed to detect aircraft. Operation at VHF or UHF can reduce the backscatter of insects and, to some extent, the echo from small birds when their cross sections are within the Rayleigh scattering region where the cross section is proportional to the fourth power of the frequency.

Knowledge of the scattering characteristics of birds and insects is important for understanding their effect on radar performance. In addition, radar has proven to be an important instrument for the study of the flight characteristics of insects (entomology) and birds (ornithology). It can obtain information not possible by other sensing means.

Birds Table 7.4 gives examples of the radar cross sections of birds taken at three frequencies with vertical polarization.[118] The largest values occur at *S* band. It is difficult, however, to be precise about the radar cross section of birds since echoes can fluctuate over quite large values, with the maximum and minimum differing at times by more than two orders of magnitude.[118]

The effect of birds on radar depends on the total cross sections that can be expected from large flocks. Table 7.5 gives representative estimates of the area over which bird echoes might be expected, the number of birds within the area, their density, individual cross section, and the volume reflectivity η (or radar cross section per unit volume). The value of η can be used in Eq. (7.2) to find the clutter cross section σ_c for the radar equation. This table is taken from an excellent review paper[119] published in the *Proceedings of the IEEE* by Charles Vaughn of NASA, Wallops Island, VA. (References to the sources of this information can be found in Vaughn's paper.) According to Table 7.5, most birds fly at altitudes less than a few hundred meters; but at times they have been observed much

Table 7.4 Radar cross sections of birds.[118]

Bird	Frequency Band	Mean Radar Cross Section (cm^2)	Median Radar Cross Section (cm^2)
Grackle	*X*	16	6.9
	S	25	12
	UHF	0.57	0.45
Sparrow	*X*	1.6	0.8
	S	14	11
	UHF	0.02	0.02
Pigeon	*X*	15	6.4
	S	80	32
	UHF	11	8.0

higher.[120] Over the ocean, the altitudes of migrating land birds have been reported as high as 4000 m.[121] Birds generally fly at speeds of from 15 to 40 kt,[122] and some have been observed at greater than 50 kt. These speeds are too high to be rejected by conventional microwave MTI radars without special design consideration.

The radar echo from a bird exhibits a periodic modulation that is caused by the beating of the bird's wings. A wing-beat fluctuation in radar cross section of 10 dB is common. A single tracked bird has sometimes been observed to fluctuate as much as 30 to 40 dB at the wing-beat rate.[119] The wing-beat frequency might vary from a few Hz to 10 to 20 Hz. The nonsinusoidal nature of the wing-beat modulation results in a spectrum with significantly higher harmonics. One estimate of the wing-beat frequency is given by $fL^{0.827} = 572$, where f is the wing-beat frequency in hertz and L is the length of the bird's wing in millimeters.[123] The spectral components of the wing-beat modulation has been said to be remarkably stable and suited for recognition.[124]

Williams and Williams[121] state that the great majority of radar observations of birds over the ocean are straight tracks at altitudes of more than 100 m, and are found at the season of peak land-bird migration on nearby continents. Migrating land and sea birds move over the oceans in massive waves that can extend up to a thousand kilometers wide and a hundred kilometers deep. The maximum density of radar echoes over the ocean is said to be significantly less than that detected with similar radars at continental sites. Williams and Williams also report that radar will often detect no bird echoes over the ocean for several days and then suddenly the display will be filled with small echoes, all moving at about the same direction and speed. Sea birds are usually seen close to the surface of the sea and move in curving paths rather than in the large number of parallel tracks typical of migrating land birds. Land birds cannot feed or rest over the sea and must make nonstop flights. Sea birds, on the other hand, feed from the sea and make use of dynamic soaring.

A statistical prediction of radar bird clutter was reported for the Distant Early Warning (DEW) line of aircraft surveillance radars that were located in northern Canada and

Table 7.5 Representative order of magnitude values for some high average areal number densities of birds (integrated over altitude), most of which are averaged over wide geographical areas. The cross section values apply for frequencies from *S* to *X* bands.

Type of Concentration	Area Affected (km^2)	No. of Birds	n ($Birds/m^3$)	σ (cm^2)	η (cm^{-1})
Blackbird roosts lower Miss. Valley (winter)	10^5	10^8	$10^{-6(1)}$	5–50	10^{-11}–10^{-10}
Single blackbird roost (Winter feeding area)	10^3–10^4	10^7	10^{-6}–$10^{-5(1)}$	5–50	10^{-11}–10^{-9}
Crows, gulls, geese, and ducks costal areas, wildlife refuges (winter)	10^3	10^4–10^6	10^{-9}–$10^{-6(2)}$	10–500	10^{-13}–10^{-9}
Shearwater migration, California coast	10^3	$> 10^6$	$10^{-5(3)}$	50–500	10^{-9}–10^{-8}
Greater Shearwater (*Puffinus gravis*) Georges Banks, non-breeding	10	10^5	$10^{-4(3)}$	50–500	10^{-9}–10^{-8}
Wading and seabird (breeding) U.S. Gulf Coast	10^5	10^6	$10^{-9(2)}$	50–500	10^{-13}–10^{-12}
Nocturnal fall migration U.S. east of Rocky Mts.	10^7	10^9	$10^{-7(4)}$	1–50	10^{-13}–10^{-12}
Nocturnal fall migration, one location, eastern U.S.	$> 10^3$	?	$10^{-6(4)}$	1–50	10^{-12}–10^{-10}
Quelea (*Quelea quelae*) breeding colony, semi-arid African savannahs	2	$> 10^7$	$10^{-2(1)}$	5–50	10^{-7}–10^{-6}
Quelea, semi-arid African Savannahs	10^9	10^9–10^{11}	10^{-9}–$10^{-7(1)}$	5–50	10^{-14}–10^{-11}

[1] [(1)]10 percent within 100 m of the ground.

[2] [(2)]50 percent within 200 m of the ground.

[3] [(3)]50 percent within 20 m of the surface.

[4] [(4)]50 percent within a 500 m interval.

[5] Sources of this information can be found in Vaughn,[119] from which this table was taken. *(Copyright 1985 IEEE)*

Alaska.[125] Birds in this region can be relatively large, and can have mean cross sections of about 0.02 m^2 at *L* band and 0.05 m^2 at UHF. A migrating flock might contain between 25 and 100 birds, but at times can be as large as 1000. The probability that a flock of Elders, a common northern bird, will have a radar cross section greater than 0.1 m^2 was said to be 90 percent, and the probability that the cross section will be greater than 1.0 m^2 is 50 percent. In northern Alaska, for example, from July to September, approximately 1.5 million large size birds such as ducks, swans, and geese can be expected to migrate across the length of the DEW line. It was estimated that the maximum number of birds

that might be seen on a particular radar screen at various locations along the DEW line during the peak of the migratory season could vary from 8,000 to 30,000.

There are times when birds or insects can cause serious degradation of radar performance if the radar operator and/or designer are not aware of the effects that birds can have as moving clutter echoes. Radars that have to operate reliably under all conditions need to account for clutter from birds and insects. Most civil air-traffic control air-surveillance radars employ sensitivity time control (to be described later in this chapter) to turn down the receiver gain at short ranges so as to reduce the clutter from nearby birds and insects. Military air-surveillance radars cannot reduce their gain by means of sensitivity time control in order to avoid low cross section birds and insects, since a military radar's target of interest can also have low cross section. During migratory seasons, bird echoes can extend over a wide area and can cause great concern to those who are not aware of what is happening. In addition to times of migration, Vaughn[119] states that mass bird concentrations can occur in the vicinity of breeding colonies, nocturnal roosts, premigratory staging areas, and concentrated feeding areas.

Vaughn summarizes the effect of birds on radar as follows:

> "From spring through fall, birds and/or insects are generally common to abundant in the atmosphere to an altitude of 1 or 2 km over most land areas of the world—especially at night. Some migrant bird species regularly fly above 4-km altitude and more than 1000 km from the nearest land. From the surface to 2-km altitude a density of 10^{-7} to 10^{-6} m^{-3} birds is not uncommon; in areas with social gatherings, 10^{-5} m^{-3} birds can be encountered during the day. A radar resolution cell of 10^6 to 10^7 m^3 (which will occur within 20 km of many radars) will result in a significant fraction of these cells occupied by at least one bird. Single birds typically have mean σ [radar cross section] from 10^0 to 10^2 cm^2, while migrant flocks, which often occur during the day, can have a mean σ greater than 10^4 cm^2."

Insects Despite their small size, insects can be detected by radar, especially at short range. In sufficient numbers, insects can *clutter* the display and prevent the detection of desired targets. It has been said that the density of airborne insects can be many orders of magnitude greater than the density of birds.[119]

Figure 7.20 is an approximate plot of the *X*-band radar cross section of insects as a function of their mass. This particular curve is taken from E. F. Knott,[126] who based it on a much more complete presentation of insect cross sections given by J. R. Riley.[127] The reader is referred to Riley's paper for the complete list of insects that make up this plot. Riley states that the data indicate that "the radar cross section of an insect may be very approximately represented by that of a spherical water droplet of the same mass, and that this representation holds true over a mass range of 10,000:1." This is only approximate since the cross section of insects is aspect dependent. Insects have been observed to have echoes at broadside aspect of from 10 to 1000 times greater than when viewed end-on.[128] Insects generally travel with the wind. Because of the aspect dependence, when a radar looks into or away from the wind, the radar echoes from a swarm of insects will be smaller than the echo when looking broadside to the wind direction. Thus the configuration of a large area of insects on a PPI display will not be uniform with angle, but more like a figure-eight. (The orientation of the figure-eight can indicate the approximate direction of

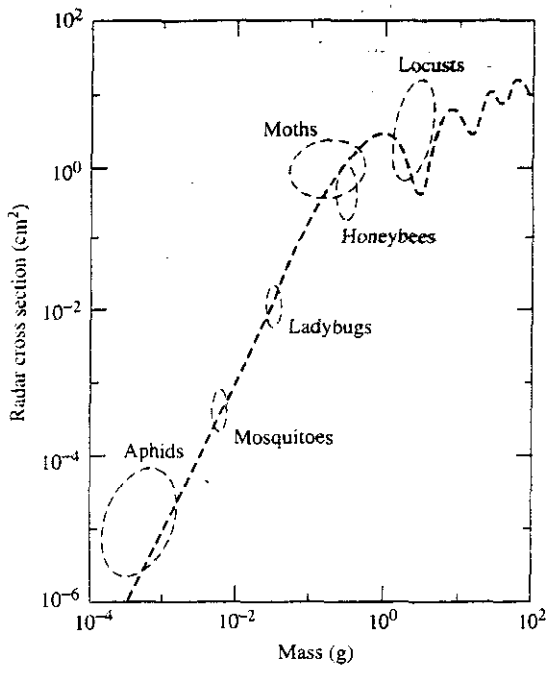


Figure 7.20 Sample of measured radar cross section of insects as a function of insect mass, as drawn by E. Knott[126] from data presented in a paper by J. Riley.[127] The dashed curve is the calculated cross section of spherical water drops, and is shown for comparison.
[Copyright, 1985 IEEE.]

the wind, with a 180° ambiguity.) The wing beat of insects is also observed to produce a distinctive modulation on the radar echo. It has been said,[128] "Male and female grasshoppers and locusts usually differ significantly in size and hence in wing-beat frequency, making it possible to distinguish sexes under favorable circumstances."

Radar echoes can be produced by insect concentrations that would scarcely be noticed visually. Insect echoes are more likely to be found at the lower altitudes, near dawn and twilight. The majority of insects do not fly at temperatures below 40°F (4.5°C) or above 90°F (32°C), so that large concentrations of insect echoes would not be expected outside this temperature range. As with clutter due to birds, sensitivity time control can reduce the adverse effects of clutter caused by insects.

Vaughn[119] summarizes his discussion of insects as follows:

> "Insect densities can easily exceed 10^{-5} to 10^{-4} m^{-3}, while 10^{-3} to 10^{-2} m^{-3} insects seem to be regularly encountered when a wind field convergence brings them together. The highest densities are very local, less than 100 km^2 probably being a typical area of concentration. Insects also trace many atmospheric boundary layer features, thus complicating data interpretation by radar meteorologists . . . . At a wavelength of less than 3–5 cm, a volume reflectivity due to insects of 10^{-8} cm^{-1} might reasonably be expected in a region of wind convergence; 10^{-11} to 10^{-10} cm^{-1} should be regularly expected elsewhere."

Table 7.6 is a summary of some relevant number densities compiled by Vaughn from a number of sources. His paper gives the references for the values shown in the table.

Table 7.6 Summary of number densities of insects; as presented by Vaughn[119] based on data from R. Rainey and C. G. Johnson. The altitude in the table is that at which the sample was taken. The cross section is at *X* band.

Sample	Altitude (m)	Insects (m^{-3})	σ (cm^2)	η (cm^{-1})
Medium sized butterflies (perids), heavy migration	surface	10^{-2}	10^{-2}–10^{0}	10^{-10}–10^{-8}
Major aphid migration	surface	10^{0}	10^{-3}–10^{-2}	10^{-9}–10^{-8}
Noctuid moths	surface	10^{-2}	10^{-1}–10^{0}	10^{-9}–10^{-8}
All insects in 1 h	surface	10^{1}	10^{-3}–10^{-1}	10^{-8}–10^{-6}
Stratified desert locust swarm	near surface	10^{0}	10^{-1}–10^{0}	10^{-7}–10^{-5}
Suction trap during passage of intertropical front, Africa	15	10^{-2}	10^{-3}–10^{0}	10^{-11}–10^{-8}
Aircraft flying across intertropical front	300	10^{-3}	10^{-3}–10^{0}	10^{-11}–10^{-8}
Mean over 50-km traverse. Mostly *Aiolopus simulatrix*	450	10^{-3}	10^{-2}–10^{0}	10^{-11}–10^{-9}
Spodoptera exempta, average over 12-h period	2	10^{-2}	10^{-2}–10^{0}	10^{-10}–10^{-8}

Clear-Air Turbulence Radar echoes from clear air (no birds, insects, clouds, dust, or precipitation) are caused by inhomogeneities in the refractive index (index of refraction) of the atmosphere that result from turbulence.[129,130] (The index of refraction is the square root of the dielectric constant.) The refractive index of the atmosphere is a function of the water vapor (humidity), temperature, and pressure. At microwave frequencies, inhomogeneities in the water vapor are usually more important than inhomogeneities in the temperature or pressure. Reflections from clear-air turbulence are small in terms of radar cross section and are detectable only by large radars or at close range.

The atmosphere can be considered turbulent everywhere, but its intensity varies widely both in space and time. It is only when turbulence is concentrated into regions of greater or lesser intensity than its surroundings (so that there is a contrast that can be detected) is it of interest as an electromagnetic scatterer. There are at least two types of turbulent atmospheric phenomena that result in clear-air echoes. One is the *convective cell*, or *plume*, that occurs in the lower part of the atmosphere. Atmospheric convective cells, also called *thermals*, are the means by which birds and gliders soar. The other turbulent phenomenon is the *atmospheric horizontal layer*, which can occur at any altitude and is the turbulent effect that aircraft encounter at the higher altitudes. To radar, the atmospheric layer is generally weaker than the convective cell.

The basic theory for scattering from homogeneous, isotropic turbulence was first given by Tatarskii.[131] The radar reflectivity (cross section per unit volume) of a homogeneous turbulent medium is given as

$$\eta = 0.38 C_n^2 \lambda^{-1/3} \qquad [7.41a]$$

where C_n^2 is the *structure constant* (a measure of the intensity of refractive-index fluctuations), and λ is the radar wavelength. At altitudes of several hundred meters, values of C_n^2 can be between 10^{-9} and 10^{-11} $m^{-2/3}$, which correspond to a volume reflectivity of about 0.8×10^{-9} to 0.8×10^{-11} m^{-1} at S band ($\lambda = 10$ cm). This is quite low, as can be seen by comparison with the volume reflectivity of rain in Fig. 7.18. The value of C_n^2 is variable and decreases with increasing height. At an altitude of 10 km, it can have values in the vicinity of 10^{-14} $m^{-2/3}$. An S-band radar with a 1° beamwidth and a 1-μs pulse width viewing a turbulent medium with $\eta = 10^{-9}$ (a very high value) sees a radar cross section at 10 km of about 3×10^{-3} m^2. Thus even with large values of C_n^2, echoes from clear-air turbulence are not likely to bother most radars.

Doppler weather radars, such as Nexrad[133] and the Terminal Doppler Weather Radar,[134] are used to detect wind shear, including that caused by the dangerous microburst, or downburst.* Wind shear caused by a downburst can be a hazard to aircraft during landing or take-off. It has sometimes been said that these radars detect clear-air wind shear, but the atmospheric wind-shear pattern is usually inferred from the radar echo of wind-driven rain rather than clear-air turbulence. When a downburst occurs without the presence of rain, wind-blown dust and insects provide the echoes that allow wind shear to be detected.

Other "Angel" Echoes Radar has been reported to detect the passage of an invisible sea breeze as it moves toward the shore.[103] Echoes have also been reported from the vicinity of forest fires, dump fires, and oil fires.[135,136] The reflectivity of smoke from such fires is too small to account for these echoes. Such echoes might be caused by numerous large particles and debris lifted into the air above the fires and/or by the generation of atmospheric turbulence due to the high heat from the fire.

Another example of clear-air echoes that are detectable by radar is gas seepage from underground oil and gas deposits.[137,138] Although these echoes are weak and the radar used for this purpose is small (a modified civil marine radar), the observations are performed at short range (500 to 1000 ft). This radar technique has been used commercially on many occasions for oil and gas exploration. It is speculated that the radar is detecting turbulent water vapor included with the gas seepage.

Ring echoes, or ring angels, are sometimes observed on PPI displays.[139] They start as a point and form a rapidly expanding ring. After one ring grows to a diameter of several miles, a second ring forms. Other rings can form in succession. They expand at velocities typically ranging from 20 to 30 kt and can attain diameters of 30 km or more. Over time, their appearance is similar to the outward moving ripples that occur when a pebble is dropped in a pond of water. These ring echoes have been associated with birds, such as starlings, flying away from their roosting areas at sunrise.

7.8 DETECTION OF TARGETS IN CLUTTER

Chapters 2 and 5 discussed the radar detection of targets when radar sensitivity is limited by receiver noise. When clutter is larger than receiver noise, different radar design methods

*Wind shear in radar meteorology is the local variation of wind velocity in range, altitude, or azimuth (or x, y, and z in a rectangular coordinate system). According to Theodore Fujita of the University of Chicago, a downburst is a strong downdraft that induces an outburst of damaging winds on or near the ground.

need to be employed for target detection. This section discusses the various techniques that might allow the detection of desired targets in the presence of undesired clutter echoes.

There is no one method that works equally well for the detection of all targets in various types of clutter. In brief, the major techniques available to the radar designer may be summarized as follows:

- The *doppler frequency shift* in MTI and pulse doppler radar is a powerful method for separating moving targets from stationary clutter.
- *High resolution in range and/or angle* reduces the amount of clutter with which the target must compete, thus increasing the signal-to-clutter ratio.
- Generally, for many types of clutter the *lower radar frequencies* produce smaller echo power. (An exception is an ionized media such as meteor trails and aurora).
- The use of *polarization discrimination* can increase the target-to-clutter ratio from some types of clutter, such as rain.
- *Clutter-echo decorrelation* by observing the target and clutter at different times or with different frequencies has some limited potential for allowing improved detectability.
- There are also techniques that help to *avoid saturation* of the receiver due to large clutter echoes. These can be important to good radar performance even though they do not increase the target-to-clutter ratio.

Doppler Frequency Shift (Target Radial Velocity) The doppler frequency shift of the echo signal is widely used for separating a moving target's echo signal from large, unwanted stationary clutter echoes. Doppler filtering allows echoes of moving targets to be separated from those of stationary clutter even though the clutter might be greater by many orders of magnitude. This is the basis for moving target indication radar (MTI), pulse doppler radar, and CW radar (Chap. 3). In spite of limitations, doppler filtering works well and can provide far greater suppression of the clutter echoes than any other technique.

Since radar echoes from land usually are much larger than those from the sea, the use of doppler filtering to detect aircraft is more demanding over land than over the sea. In the absence of ducted propagation or mountainous terrain, clutter usually does not extend more than a few tens of miles when the radar is at or near the surface. Airborne radars, however, can look down on extensive land or sea clutter and can experience surface clutter at relatively long ranges compared to what is observed from land or ship-based radars. When doppler filtering is employed to detect moving targets in rain, the rain itself is moving and the doppler processor must be designed differently from a doppler processor designed for stationary clutter. The MTD radar described in Sec. 3.6 is an example of how aircraft can be detected in the presence of moving rain clutter.

High Resolution The smaller the radar resolution cell, the less will be the competing clutter echo. The benefit of small resolution cells is indicated in the radar equations for detection of targets in surface clutter [Eq. (7.8)] and in rain [Eq. (7.37)]. Short pulses (or the equivalent pulse compression) and narrow-beamwidth antennas, therefore, can reduce

the amount of clutter with which the target echo signal must compete. High resolution usually is the preferred method for detection of ships.

With large resolution cells, the clutter statistics can be described by the Rayleigh probability density function (pdf). With small resolution cells, however, sea clutter is no longer Rayleigh and is likely to produce higher values (such as from sea spikes) than predicted by Rayleigh clutter. The result is that serious false alarms can occur when small targets are to be detected if the clutter is non-Rayleigh but the receiver is designed on the basis of Rayleigh statistics. Increasing the detection threshold can reduce the false alarms from non-Rayleigh clutter, but it results in the undesired lowering of the probability of detection. Other solutions are required.

A knowledge of clutter statistics is necessary for receiver design based on classical detection theory. It is also needed to design CFAR (Sec. 5.7) when an optimum detector is not practical. Unfortunately, the statistics of the sea clutter are variable, and it is not always possible to specify a single statistical description of the clutter when the statistics are not Rayleigh.

High resolution is usually a good method for enhancing the signal-to-clutter ratio, even when the statistics of the clutter are not Rayleigh. It is one of the few methods available for increasing the target-to-clutter ratio when the target is stationary (no doppler shift).

The clutter echo must be large compared to receiver noise in order for the clutter radar equations [Eqs. (7.8 and 7.37)] to apply. If this condition is not met, it is not necessarily bad since it means that radar detection will be noise limited rather than clutter limited—which is to be preferred. It is cautioned that the resolution cell size cannot be decreased without limit when considering detection of signals in clutter. With a small enough resolution cell size, the radar becomes noise limited and there is no need for a further decrease.

High-Resolution Clutter and Detector Design With sea clutter, non-Rayleigh clutter statistics might be expected to appear when the beamwidth is less than about one degree and the pulse width is less than about one μs (as indicated in the previous discussion of sea spikes in this chapter). Non-Rayleigh sea clutter is especially obvious when pulse widths are on the order of nanoseconds. Attempts have been made to model the statistics of high-resolution sea clutter, but without total success. The Rayleigh pdf generally underestimates the range of values obtained from real sea clutter. On the other hand, the log-normal pdf tends to overestimate the range of values. Several other analytical pdfs that lie between these two extremes have been suggested for modeling the statistics of sea clutter, some of which were mentioned in Sec. 7.5.

Any particular probability density function might be able to fit some particular set of data; but there does not seem to exist a single analytical form of probability density function that fits all the observed data and can be correlated with environmental parameters. For this reason it has not been practical to design a detector in a classical manner when sea clutter is dominant. (*Classical manner* means that the optimum detector is found based on the statistics and correlation properties of the background clutter and target signals.)

Billingsley[140] has said that on the basis of many measurements of various types of land clutter at low depression angles, "measured amplitude distributions almost never pass

rigorous statistical hypothesis tests for belonging to Weibull, log normal, or other theoretical distributions that have been tried." Thus the fitting of experimental data to arbitrary statistical models has not been successful as a method for describing clutter to aid in radar design.

When the clutter cannot be described by a specific pdf, *robust detection* and *distribution-free detection* are often considered. A robust detector is one that works satisfactorily when the clutter (or noise) statistics can be described as one of a family of pdfs or a number of different classes of pdf. It is not necessarily optimum for any of them. A distribution-free detector makes as few assumptions as possible regarding the statistics of the clutter or noise.[141] Its probability of false alarm is supposed to be constant under weak assumptions on the statistical character of the clutter. Distribution-free detectors have not been used much in practice because they are not easy to implement and the loss in detectability is larger than might be desired. The term *nonparametric detector* is sometimes used interchangeably with the term *distribution-free detector.*

The *median detector* and the *m-out-of-n detector* (Sec. 5.6) are examples of robust detectors that have been considered for detecting targets in high-resolution sea clutter. A median detector is one that bases its detection decision on the median value* of n received pulses. Trunk and George[142] showed that the *median detector* gave better performance in non-Rayleigh clutter than the conventional receiver (which can be called a *mean detector*). A median detector is robust in that the threshold values and detection probabilities do not depend on the detailed shape of the clutter pdf, but only on its median value. For instance, if one of the n received pulses happens to be very large, it can have a significant effect on the mean value; but it would have no greater effect on the median value than a pulse that is only slightly larger than the median. Trunk[143] has shown that the *trimmed-mean detector* is even better than the median detector for non-Rayleigh clutter. If there are n pulses available from which to make a detection decision, the trimmed-mean detector discards the smallest and the largest of the n pulses before finding the mean and making a detection decision. Although its performance may be attractive, the trimmed mean is more difficult to implement in practice since the n pulses must be rank-ordered to find the smallest and the largest.

According to Schleher,[144,145] the median detector is inferior to both the m-out-of-n detector and the logarithmic detector when clutter can be described by the Weibull pdf. The optimum value of m for an m-out-of-n detector will depend on the nature of the pdf. When the optimum m is used with non-Rayleigh clutter, Schleher states that the m-out-of-n detector is better than other detection criteria.

The logarithmic detector has an output voltage whose amplitude is proportional to the logarithm of the input envelope. A logarithmic characteristic provides a constant false-alarm rate (CFAR) when the clutter is described by the Rayleigh pdf (see the log-FTC discussion presented later in this section). The performance of the logarithmic detector in non-Rayleigh clutter is almost as good as the m-out-of-n detector and is probably easier to implement.[145] A variant sometimes useful for reducing the false-alarm rate in non-Rayleigh clutter is the *log-log detector.* In one implementation, this is a logarithmic receiver in which the slope of the logarithmic characteristic progressively declines by a

*The middle value of a statistical distribution, above and below which lie an equal number of values.

factor of 2 to 1 over the range from noise level to +80 dB.[146,147] In the log-log detector, the higher clutter values (tails of the distribution) are subjected to greater suppression.

Land Interclutter Visibility The patchiness of land clutter, when viewed with high-resolution radar at low grazing angles, usually is beneficial for target detection. Patches of large clutter are separated by areas with little or no clutter. Typical sizes of clutter patches can be several kilometers on a side. Billingsley[140] states that the macropatches of clutter might typically occupy about one half of the resolution cells of discernible clutter. However, for level terrain in which only isolated discrete clutter echoes appear, the percentage of cells with clutter might be as low as 25 percent. For high radar sites and/or steep terrain "in which relatively full illumination exists, it can approach 100%." Interclutter visibility is a result of masking caused by hills or higher levels of terrain which create shadow areas where clutter echoes are absent. Aircraft targets can be detected in regions with low or no clutter even though they might be undetectable when within patches of high clutter.[148] It has been said that at low grazing angles over land, interclutter visibility can be expected with beamwidths less than a degree and pulse widths less than one μs.[149]

Weather Reducing radar beamwidth and/or pulse width is generally considered good for decreasing the weather clutter with which the target must compete. There is, however, little information on the characteristics of rain with high-resolution radar.[150] Most forms of rain are not uniform when viewed with high resolution; but for purposes of radar design, it is usually assumed so. Reducing the radar resolution cell to improve signal-to-clutter ratio is an acceptable technique, but the designer should be aware that a small resolution cell size might cause the statistics of the clutter to become non-Rayleigh and increase the probability of false alarm.

Target Break-up with High-Resolution Waveforms In some cases, the range resolution cell might be smaller than the dimensions of the target, resulting in "break-up" of the target echo. This means that the cross section of each resolved area of the target will be less than the total cross section as seen by a lower resolution radar. A serious reduction of detectability need not occur because of target break-up. Reduction of clutter by high range-resolution is still a good tactic.

If echoes from the various parts of the target viewed with high-resolution radar are properly displayed to an operator, and if the operator knows the general shape of the target or even that it is a distributed target, there appears to be little, if any, loss in detectability. The operator seems to be able to provide a form of noncoherent integration. Even if the recognition ability of an operator is not utilized, detectability will likely be reduced only slightly with high resolution since many targets tend to consist of one or a few large scatterers and a large number of small scatterers.[151] The echo in the largest resolution cell and in surrounding cells might still be of significant value to produce adequate detection.

Hughes[152] has investigated detection strategies when the radar resolution is less than the target dimensions. He assumed, for sake of discussion, that a range cell was 10 ns (5 ft) and that a 100-ns window was examined. The distributed target was completely

within the 100-ns window. Clutter was homogeneous and described by a gaussian distribution. In one strategy, a threshold decision was made every resolution cell (10 ns), and any threshold crossing within the 100 ns window was declared a detection. This is similar to an m-out-of-n detector with $m = 1$. The other strategy, called the integrated detector, was to integrate 10 cells and make a detection decision every 100 ns. As reference case with which to judge these two procedures, Hughes took a 100-ns range cell with a detection decision made every 100 ns. The reference case was always significantly worse than either of the two strategies. Hughes concluded that the strategy that is better depends on the a priori knowledge of the target's nature that is available to the designer. If about two-thirds or more of the target echo energy comes from a single "flare" point, or individual scatterer, the m-out-of-n detector will be better. He stated that the correct strategy to use can be chosen if detailed knowledge of the target is known. Without such knowledge or when a variety of targets must be detected, the second strategy—integration before detection—is better.

Frequency Radar echoes from rain, sea (with horizontal polarization), and (to some extent) land decrease with decreasing frequency. Although this favors the lower frequencies, it is more difficult at the lower frequencies to achieve narrow beamwidths and high range-resolution. Thus some of the benefits of reduced clutter at the lower frequencies might be partially offset by the poorer resolution that results.

Rain The radar equation for detection of targets in rain, Eq. (7.37), shows that for a constant-gain antenna the range varies as the square of the wavelength. (For an antenna with a constant effective aperture, however, the range is directly proportional to wavelength.) There can be great benefit in using lower frequencies to reduce rain clutter. It is unlikely, for instance, that a radar at UHF or lower frequencies will be degraded by rain clutter. Unfortunately, as indicated above, in some applications it is not always possible to operate radars at frequencies low enough to make rain clutter negligible because of other undesirable constraints such as broad beamwidth, narrow bandwidth, a crowded frequency spectrum, external noise, and poor low altitude coverage.

Radars at L band or higher frequencies are likely to be affected by rain clutter. Some higher frequency radars might even be rendered useless in rain if no corrective measures are taken.

Sea The plot of σ^0 versus grazing angle that was shown in Fig. 7.13 for sea clutter indicates that at low grazing angles and horizontal polarization, the lower the frequency the lower will be the echo from the sea when everything else remains the same. There are some reservations, however, about how low in frequency one might wish to go. The lower the frequency, the more difficult it is to direct the radar energy at low angles. This is illustrated in Sec. 8.2, which shows that the elevation angle of the lowest interference lobe is $\lambda/4h_a$, where λ = wavelength and h_a = antenna height above the water. If the target of interest is a small boat, buoy, submarine periscope, swimmer, or low altitude (sea skimming) missile, the higher microwave frequencies might be preferred even though the sea clutter might be greater.

Another reason higher frequencies might be preferred in some overwater applications in spite of higher sea clutter is the better range and angle resolution that can be obtained. In addition to providing less clutter, the higher resolutions obtainable at the higher frequencies might be required for particular applications. An example is the civil marine radar, which is commercially available at both *X* and *S* bands. When radars at both frequencies are operated on a ship, it is the higher frequency radar (*X* band) that is generally depended upon, except in bad weather when rain clutter degrades the higher frequency much more than *S* band. The fact that sea clutter is less at the lower frequencies is but one of several factors to be considered in the selection of the best frequency for a particular radar application.

Land Although it was shown in Fig. 7.10 that land clutter at high grazing angles decreases with decreasing frequency, the frequency dependence is not usually a major consideration in radar design. At low grazing angles, Fig. 7.5 and the discussion of Sec. 7.3 show that for any particular type of terrain there might be significant variation of clutter with frequency, but the mean value of clutter taken over all types of terrain (excluding urban area) is essentially independent of frequency.

Polarization

Rain Circular polarization is often employed to reduce weather clutter in microwave air-surveillance radars. Discrimination of a target located in the midst of rain results because the echo from an aircraft (an asymmetrical scatterer) differs from that of a raindrop (a circularly symmetric scatterer) when using circular polarization. This difference in scattering can be used to enhance the echo from the target and suppress echoes from rain.

A circularly polarized wave incident on a spherical scatterer will be reflected as a circularly polarized wave of opposite sense of rotation (when viewed in the direction of propagation) and will be rejected by the same antenna that originally radiated it. On the other hand, when a circularly polarized wave is incident on an asymmetrical target such as an aircraft, it has been found experimentally that the reflected energy is more or less equally divided between the two senses of polarization rotation.[153] The two senses of circular polarization are (1) right-hand circular, when the electric field rotates clockwise when viewed in the direction of propagation and (2) left-hand circular, when the electric field rotates counterclockwise. Since about half of the reflected energy is of the same sense of circular polarization as that transmitted, it will be accepted by the antenna that originally radiated it. The circularly polarized echo from spherical raindrops is of the opposite sense of polarization (when viewed in the direction of its propagation) and will be rejected by the radar antenna; but an aircraft will have a significant amount of its reflected energy with a polarization that is accepted by the antenna. This then provides target-to-clutter enhancement.

Raindrops are seldom perfect spheres, especially when they are large. They deviate from a sphere when their radius is greater than about 1 mm, and have a shape more like that of an oblate spheroid with a flattened base (a hamburger bun).[154,155] This particular raindrop shape, which is quite different from that usually depicted by an illustrator or

cartoonist, is a result of aerodynamic forces as the drop falls.[156] The greater its deviation from a spherical shape, the less the echo signal will be rejected. Thus circular polarization is limited in the amount of rain cancellation it can provide and becomes less as the rain becomes heavier.

In addition to the nonspherical shape of raindrops, the ability to cancel rain echo using circular polarization is limited by practical considerations. It is difficult to obtain an antenna with pure polarization that does not accept energy from the orthogonal polarization. The cancellation of an orthogonally polarized signal by an exceptionally well designed, well maintained antenna might be limited to about 40 dB, a very good value.[157] Most antennas are not this good, and they can be even poorer because of the depolarizing effects of the nearby environment. Depolarization is also introduced when propagating through rain located between the radar and the target. The different reflection coefficients of horizontal and vertical polarization on reflection from the earth's surface also will cause a change in the signal's polarization when multipath reflection occurs from the surface. This factor has been said to limit the cancellation of rain clutter in certain cases to 13 dB over water[158] and 34 dB over desert.[159]

The various forms of depolarization that cause circular polarization to become elliptical suggest that better rain cancellation might be achieved if the optimum form of elliptical polarization were used rather than circular. With the optimum elliptical polarization, cancellation in some regions of heavy rain might be increased by as much as 12 dB.[160] However, the polarization that is optimum for one particular region of rain might not be optimum for some other region of rain, and might prove to be worse than the cancellation obtained with circular polarization. The optimum polarization will depend on the distance traveled and the nature of the rain. Thus the polarization on receive should be adaptive and made variable with range (time). Nathanson suggested that an improvement in cancellation of 6 to 9 dB can be obtained with adaptive polarization.[161] Adaptive polarization, however, has not been easy to implement successfully.

The radar cross section of aircraft is generally less with circular polarization than with linear polarization. As mentioned above, experimental measurements indicate that when an aircraft is illuminated with one sense of circular polarization, the echo power on a statistical basis is divided more or less equally between right-hand and left-hand circular polarization. Other experiments, however, have shown the circular polarized echo to be smaller than the linearly polarized echo by about 5 dB for aircraft and even greater reductions for missiles and satellites.[162]

An alternative to circular polarization is to transmit with a linear polarization (horizontal or vertical) and receive with the orthogonal linear polarization. Spherical scatterers will not produce an echo at the orthogonal linear polarization; but asymmetrical scatterers will have energy in the orthogonal component. It has been suggested that crossed linear polarization might give better rain rejection than conventional circular polarization.[163] Some measurements show, however, that the backscatter from aircraft with the orthogonal linear polarization can be as much as 7 dB below that reflected at the transmitted linear polarization.[164]

Although the use of circular polarization is often employed in surveillance radars during rainfall, at the higher microwave frequencies where the echo from rain can be a problem, cancellations of heavy rain might be 15 dB or less. In heavy rain that accompanies

thunderstorms cancellations of only 5 dB have been observed.[161] If detection of targets in rain is important, then circular polarization might not provide all that is needed. MTI designed for moving clutter will obtain much greater clutter rejection. Lowering the frequency by a factor of 2 or using a pulse-compression waveform with a 13-bit Barker code could likely provide in heavy rain as much cancellation as circular polarization.

Sea Over the sea, horizontal polarization at low and moderate sea states results in less sea clutter than vertical, as was indicated in Fig. 7.13. The majority of the radars that operate over the sea employ horizontal polarization. If a target, such as an oil slick or a swimmer's head, lies on the water, vertical polarization is preferred even though sea clutter may be larger than with horizontal polarization. Sea clutter is large with vertical polarization at low angles for the same reason the echo energy from targets on the water surface is large: these targets and the sea surface are illuminated with more energy than if horizontal polarization were used because of the behavior of the vertical polarization reflection coefficient (Sec. 8.2).

Land It was mentioned in Sec. 7.3 that there is no significant difference between horizontal and vertical polarizations from land clutter when viewed at low grazing angles. Although there are indications that horizontal polarization produces a lower clutter echo than vertical, the difference is not significant enough to affect the choice of polarization in most radar applications over land.

Time Decorrelation Unlike receiver noise, clutter echoes are generally correlated from pulse to pulse, and sometimes even from scan to scan. The techniques of *rapid antenna scan* for detection of small targets in the sea, and *time compression* for detection of moving targets in patchy land or sea clutter are examples of detection techniques that try to take advantage of the nature of correlated clutter echoes.

Antenna Scan Rate (Sea Clutter) As mentioned previously in this chapter, when clutter is stationary (not moving) there is no gain to be had by integrating pulses. The sea is always changing, but over a short time interval it might not change sufficiently for the radar echo to be decorrelated pulse to pulse. For a medium-resolution *X*-band radar observing the sea at low grazing angles, the time required for the sea to decorrelate is about 10 ms.[165] Any pulses received during this time will be correlated and no improvement in signal-to-clutter ratio will result by integrating pulses. The sea surface, however, will often change by the time of the next antenna scan so that sea clutter is generally decorrelated from scan to scan. To take advantage of the decorrelation of sea clutter from scan to scan, the antenna rotation rate can be increased, as was first demonstrated by J. Croney for civil marine radar.[166]

Consider, for example, a civil marine *X*-band radar with a 1.2° beamwidth, pulse repetition frequency of 6000 Hz, and a 20-rpm antenna rotation rate. There will be 60 pulses in the 10 ms it takes for the beam to scan past a target. No integration improvement is obtained from adding these 60 pulses together since they occur within a time that is less than the decorrelation time of *X*-band sea clutter. If the antenna rotation rate is speeded up to 600 rpm, two pulses will be received per scan. These will be correlated, but the two

echoes received on the next scan, one tenth of a second later, will be decorrelated from the two echoes from the previous scan. In 30 scans at 600 rpm, there will be 30 sets of two pulses each, and each set will be decorrelated from the others. Thus there can be an integration improvement corresponding to that obtained with 30 independent samples. Experimental measurements with a civil marine radar having a 1° beamwidth and 5000 Hz pulse repetition frequency show that increasing the antenna rotation rate from 20 to 420 rpm improves the target-to-clutter ratio from about 4 to 8 dB, depending on the sea state.[167] The smaller improvement corresponds to the higher sea state.

Radars that use a high antenna-rotation-rate to improve detection of small targets in sea clutter include the AN/APS-116,[168] and its successor the AN/APS-137. The antenna scan rate in these radars is 300 rpm (five scans per second). Scan-to-scan integration is performed with an m-out-of-n detector. With spiky clutter, analysis shows that the best ratio of m/n was 0.6, which is larger than the optimum value used in an m-out-of-n detector when only receiver noise is present.

The decorrelation time of sea clutter is approximated by the reciprocal of the width of the clutter doppler spectrum. Normally, the decorrelation time might be expected to be inversely proportional to frequency, if the decorrelation is due solely to doppler effects. However, experimental measurements show that when the decorrelation time is 10 ms at X band, it will be about 60 ms at L band instead of the 70 to 75 ms predicted on the basis of an inverse relationship of decorrelation time and frequency. Therefore, it is found that when clutter spectra are given in units of velocity rather than frequency, the velocity spectra are broader at the lower frequencies.[169]

Time Compression Even if a target seen with a high-resolution non-MTI radar can be resolved in the midst of patchy land clutter or spiky sea clutter, it might be difficult to recognize the target from the clutter on the basis of a single observation. If the target is in motion, however, an operator can eventually distinguish it from the stationary clutter. One method that has been proposed to accomplish this is time compression, in which the last N scans of a radar are stored and then repeatedly displayed speeded up to accentuate the target motion. (The number of scans N might be from 5 to 10.) An operator's attention is called to the moving targets by the flickering of the repeated speeded-up scans. Even with a fixed target, time compression can separate the short duration sea spikes from the more persistent target echoes. As a target at long range comes within the coverage of the radar, there will be a delay of N scans before time compression can be fully effective. In the time needed to accumulate the N scans, however, a trained and alert operator might be expected to detect the moving target even without time compression.

Frequency Agility If the RF frequency of a pulse of duration τ is changed by more than $1/\tau$, the echo from uniformly distributed Rayleigh type clutter will be decorrelated. This can be shown by calculating the change in frequency Δf that causes the difference in the phase between the clutter echo from the leading edge of the range-resolution cell and the clutter echo from the trailing edge of the range-resolution cell to be greater than 2π radians (as in Problem 7.15). Pulse-to-pulse frequency changes (called *frequency agility*) greater than $1/\tau$ Hz will therefore decorrelate the clutter and permit an increase in target-to-clutter ratio when the decorrelated pulses are integrated. This assumes the target's physical size is small compared to the size, $c\tau/2$, of the radar resolution cell. When the

clutter is non-Rayleigh, as when there are one or a few dominate scatterers, there might be little benefit in the use of frequency agility. Spiky sea clutter, for example, is highly non-Rayleigh and frequency agility is not expected to offer much improvement. If land clutter is dominated by only one or a few large clutter scatterers in a resolution cell, the benefit of frequency agility also is lessened. Weather clutter is probably more uniform than either sea or land clutter, and frequency agility might be more effective than with other forms of clutter.[170]

The effect on the target and the clutter by changing frequency from pulse to pulse is illustrated in the A-scope presentations of Fig. 7.21.[171] The left-hand trace is that of a single frequency at X band. The clutter is shown on the left side of each figure. Its amplitude decreases with increasing range. A single target is shown on the right side. Its appearance indicates that it is composed of multiple scatterers. The right-hand trace is with pulse-to-pulse frequency agility. Two changes are noted: The clutter is smoothed in appearance and the target echo is larger. The increase in target echo is due to some of the individual frequencies returning large echoes from the target, at least as compared to the particular single frequency used in the upper trace. If a different frequency were chosen in the upper trace, it might have been larger or it might have been smaller. But by using multiple frequencies, the composite of the echoes is not as likely to produce a small echo as might a single frequency. If the target on the right side of the A-scope presentation were located in the midst of the clutter in the left-hand trace, it might not be detected. However, the target signal in the right-hand trace would likely be detected. It seems from this example that the real benefit of frequency agility is in the greater target cross section rather than the suppression of clutter.

It should be cautioned that conventional doppler filtering, as in MTI or pulse doppler radar, is not possible with pulse-to-pulse frequency agility.

Although frequency agility is sometimes claimed to be an attractive method for obtaining increased target-to-clutter ratio, it has not had the full success that its proponents

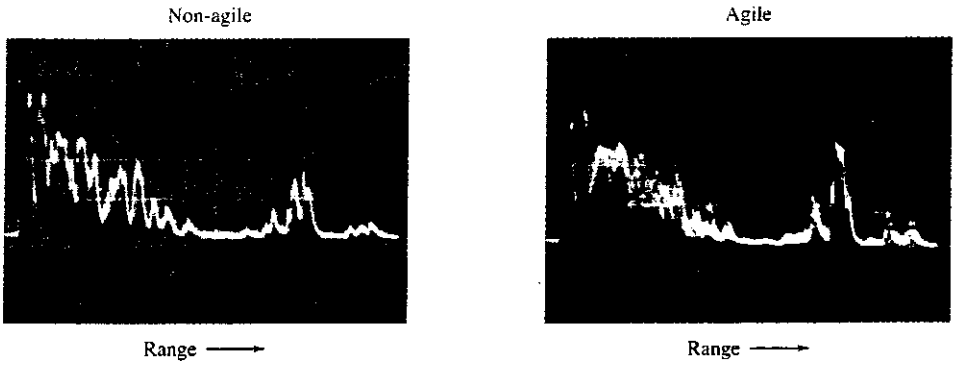


Figure 7.21 Effect of frequency agility on clutter and target. This A-scope presentation of clutter (left-hand side of each illustration) and a target (right-hand side of each illustration) was made with an X-band radar and a 0.2-μs pulse width. Left-hand trace is for a single frequency. Right-hand trace is for pulse-to-pulse frequency agility over a 100-MHz bandwidth.
(Courtesy of CPI Beverly Microwave Division, Beverly, MA.)

have desired since most clutter is non-Rayleigh. When a large frequency band is available, the bandwidth is usually better used to provide high range-resolution to decrease the clutter rather than to attempt to decorrelate the clutter by pulse-to-pulse frequency change.

Clutter Fence Reflections from nearby mountains and other large clutter might not be fully suppressed by conventional clutter reduction techniques. These echoes can be so large that they can enter the radar via the antenna sidelobes. At a fixed radar site it is possible to reduce the effects of such clutter by surrounding the radar with an electromagnetically opaque fence to prevent direct viewing of the clutter. The two-way attenuation provided by a typical clutter fence with a straight edge might be about 40 dB. Greater attenuation can be achieved by incorporating two continuous slots close to, and parallel with, the upper edge of the fence to cancel a portion of the energy diffracted by the fence. The increase in two-way attenuation by this method is 20 dB or more.[172]

A clutter fence can produce effects that are not always desirable. It can limit the accuracy of the elevation-angle measurement because of blockage and because of radar energy diffracted by the fence. Energy diffracted by the fence will interact with the direct energy and cause lobing of the radiation pattern in the angular region just above the fence. Radar energy backscattered by the fence towards the radar can sometimes damage the receiver front-end. In one design, the fence was tilted 15° away from the vertical to prevent damage to the receiver from occurring.[173]

Clutter fences are useful only when the targets are at higher elevation angles than the clutter. They have not been an attractive method to enhance target-to-clutter ratios for most applications.

Methods to Avoid Receiver Saturation by Clutter Echoes The finite dynamic range of radar receivers means that they can be saturated by large clutter signals, with the result that target echoes can be prevented from being detected even though they might be larger than clutter. Several methods have been used to minimize the effect of large clutter but they have no subclutter visibility in that they do not enhance the target-to-clutter ratio. They are useful, however, in preventing saturation or overload of the receiver, automatic processor, automatic tracker, or display.

Sensitivity Time Control (STC) The fourth-power relationship between the received echo-signal power and range means that clutter echoes at close ranges will be large and can saturate a radar receiver. A solution is to reduce the receiver gain at short ranges where the clutter echo signals are large, continually increase the gain as the pulse travels out in range, and finally operate at maximum gain (maximum sensitivity) at ranges beyond which clutter is expected. This is called *sensitivity time control* (STC). It has also been called *swept gain.* STC is used in air-traffic control radars to reduce nearby clutter from birds and insects as well as from land. The rate at which the gain is changed with time depends on the nature of the environment. It will be different, for example, when the radar is looking over water or desert than when the radar is looking over rugged or mountainous terrain. An air-surveillance radar might have more than one STC gain-versus-time characteristic that can be selected depending on the type of clutter the radar encounters. The

variation of gain with range might be as R^2 for rain, R^3 or R^4 for surface clutter, to R^7 over water at long range (as in a civil marine radar for targets below the lowest interference lobe[174]).

Even without the presence of clutter, STC can be used to compensate for the large change in the magnitude of the received target echo signal as a function of range. The change of target cross section with range has been said to overshadow other causes of echo variation.[175]

The use of STC causes the sensitivity of a 2D air-surveillance radar with a cosecant-squared antenna elevation radiation pattern (Secs. 2.11 and 9.11) to be reduced for targets at high angles and short range since the gain is lowered in these directions just as it is at low angles and long range. This can cause echoes from aircraft at short range and high altitude to be too weak to be detected. By modifying the cosecant-squared antenna pattern to direct more energy at the higher angles, it is possible to see these close-in aircraft even with STC applied. Multiple-elevation-beam (stacked beam) 3D radars can have a separate STC variation with range to match the clutter conditions found as a function of elevation angle.

STC usually cannot be used with pulse doppler and other radars that have prfs high enough to result in ambiguous range echoes. It can also cause degradation to pulse compression radars that employ very long uncompressed pulses, such as required with solid-state transmitters.

Raising the Antenna Beam Surface clutter can be decreased by raising the antenna beam in elevation and not illuminating the clutter. This technique can be applied only if the targets of interest (e.g., aircraft) are at a higher altitude than the clutter. Tilting an antenna beam upward reduces coverage at long range and low altitudes. A better technique is to use two beams in elevation, with one pointing higher than the other. Most civil air-traffic control radars use this method to reduce unwanted echoes from cars and trucks. In the ARSR-3 long range (nominally 200 nmi) air-traffic control radar, the gain of the upper beam is 16 dB less along the horizon than the gain of the lower beam along the horizon. The radar signal is transmitted on the lower beam and received only on the higher beam at short range so that surface clutter at short range is illuminated with less energy. After the pulse travels beyond the range of expected surface targets (typically 50 nmi in the ARSR-3), the receiver is shifted to the low beam to allow detection of aircraft at long range.

Log-FTC This is a receiver with a logarithmic input-output characteristic followed by a high-pass filter (sometimes called a *fast time-constant*, or FTC). When the input clutter or noise is described by the Rayleigh probability density function, the output clutter or noise is constant, independent of the input amplitude. Thus it acts to provide a constant false alarm rate, or CFAR.

The Rayleigh pdf can be written

$$p(v) = \frac{2v}{m_2} \exp\left(-\frac{v^2}{m_2}\right) \qquad v > 0 \qquad \text{[7.41b]}$$

where m_2 is the mean square value of v. The rms amplitude of the fluctuations about the mean (denoted here by δv_{in}) is proportional to the mean $\overline{v_{in}}$ or $\delta v_{in} = k\overline{v_{in}}$. (This is Problem 2.4.) A logarithmic receiver has the characteristic

$$v_{out} = a \log (bv_{in}) \qquad [7.42]$$

The slope of the logarithmic receiver characteristic at $\overline{v_{in}}$ is

$$\frac{\Delta v_{out}}{\Delta v_{in}} = \frac{a}{\overline{v_{in}}} \qquad [7.43]$$

If the input clutter fluctuations δv_{in} are small compared to the total range of the logarithmic characteristic, the output fluctuations δv_{out} are approximately

$$\delta v_{out} = \text{slope} \times \delta v_{in} = \frac{a}{\overline{v_{in}}} \delta v_{in} = ak \qquad [7.44]$$

Thus the output fluctuations are constant, independent of the input mean.

Although the output fluctuations about the mean are constant, the output mean is not. A high-pass filter removes the mean value of the output, leaving the fluctuation at a constant value on the display. The high-pass filter is equivalent to a differentiation.

The noise or clutter fluctuations that appear at the output of a logarithmic receiver are not symmetrical in amplitude since large amplitudes are suppressed more than normal. To make the output more like that of a linear receiver, the log-FTC may be followed by an amplifier with the inverse of the logarithmic characteristic (antilog).

A true logarithmic characteristic cannot be maintained down to zero input voltage since $v_{out} \to -\infty$ as $v_{in} \to 0$. At some point the receiver characteristic must deviate from logarithmic and go through the origin. Therefore, the practical receiver will have a linear characteristic at low signal levels and logarithmic at higher levels. This is called a *lin-log receiver.* The logarithmic receiver must be maintained at about 20 dB below the receiver rms noise level.[176]

Log-FTC prevents clutter from saturating the limited dynamic-range radar display and obscuring the presence of desired targets even though the targets may be of greater amplitude than the clutter. It is an example of a CFAR, but its effectiveness depends on the clutter having a Rayleigh pdf, which is not always the case.

The log-FTC was originally considered for the detection of targets in sea clutter;[176] however, it is probably more useful for operation in precipitation clutter. Precipitation is more likely to be described by a Rayleigh pdf than is sea or land clutter. When log-FTC when used for reducing the effect of rain, it has sometimes been called *weather fix.*

CFAR Circuitry for maintaining the false alarm rate constant (CFAR) has been described in Sec. 5.7 and by Taylor[177] in the *Radar Handbook.* As mentioned, CFAR does not provide an improvement in signal-to-clutter ratio, and it maintains the false-alarm constant at the expense of probability of detection. It also has the disadvantage of poorer range resolution and results in a loss in signal-to-noise ratio. It should be used only when necessary.

REFERENCES

1. Probert-Jones, J. R. "The Radar Equation in Meteorology." *Quart. J. Roy. Meteor. Soc.* 88 (1962), pp. 485–495.
2. Nathanson, F. E. *Radar Design Principles.* 2nd ed. New York: McGraw-Hill, 1991, p. 316.
3. Billingsley, J. B., and J. F. Larrabee: "Multifrequency Measurements of Radar Ground Clutter at 42 Sites," MIT Lincoln Laboratory, Lexington, MA, Technical Report 916, November 15, 1991, Vol. 1, Principal Results (ESD-TR-91-061); Vol. 2 Appendices A through D (ESD-TR-91-175); Vol. 3 Appendix E (ESD-TR-91-176). A condensation appears in Billingsley, J. B. "Radar Ground Clutter Measurements and Models, Part 1, Spatial Amplitude Statistics," paper No. 1 in *Target and Clutter Scattering and Their Effects on Military Radar Performance,* NATO Advisory Group for Aerospace Research and Development, AGARD Conf. Proceedings 501, September 1991. AD-A244 893.
4. Billingsley, J. B. "Ground Clutter Measurements for Surface-Sited Radar." MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 786, Revision 1, February 1, 1993.
5. Billingsley, J. B. "A Handbook of Multifrequency Land Clutter Coefficients for Surface Radar." MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 958. This will appear in the book *Low-Angle Radar Land Clutter* by J. B. Billingsley.
6. Moore, R. K., K. A. Soofi, and S. M. Purduski. "A Radar Clutter Model: Average Scattering Coefficients of Land, Snow and Ice." *IEEE Trans.* AES-16 (November 1980), pp. 783–799.
7. Schooley, A. H. "Some Limiting Cases of Radar Sea Clutter Noise." *Proc. IRE* 44 (August 1956), pp. 1043–1047.
8. Barton, D. K. *Modern Radar System Analysis.* Norwood, MA: Artech House, 1988.
9. Moore, R. K. "Ground Echo." In *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1990, chap. 12.
10. Ref. 2, Table 7.17.
11. Long, W. H., D. H. Mooney, and W. A. Skillman. "Pulse Doppler Radar." In *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1990, chap. 17, pp. 17.11–17.16.
12. Ulaby, F. T., R. K. Moore, and A. K. Fung. *Microwave Remote Sensing,* vol. II, Reading, MA: Addison-Wesley, 1982, Sec. 11-4.2.
13. Ulaby, F. T., W. H. Stiles, L. F. Dellwig, and B. C. Hansen. "Experiments on the Radar Backscatter of Snow." *IEEE Trans.* GE-15 (October 1977), pp. 185–189.
14. Ulaby, F. T., and W. H. Stiles. "Microwave Response of Snow." *Adv. Space Res.* (1981), pp. 131–149. (Also found in Refs. 15 and 17.)
15. Ulaby, F. T., R. K. Moore, and A. K. Fung. *Microwave Remote Sensing,* vol. III. Dedham, MA: Artech House, 1986, Sec. 21-7.

16. Waite, A. H., and S. J. Schmidt. "Gross Errors in Height Indication from Pulsed Radar Altimeters Operating over Thick Ice or Snow." *Proc. IRE* 50 (June 1962), pp. 1515–1520.

17. Ulaby, F. T., and M. C. Dobson. *Handbook of Radar Scattering Statistics for Terrain.* Norwood, MA: Artech House, 1989.

18. Borel, C. C., R. E. McIntosh, R. M. Narayanan, and C. T. Swift. "File of Normalized Radar Cross Sections (FINRACS)—A Computer Program for Research of the Scattering of Radar Signals by Natural Surfaces." *IEEE Trans.* GE-24 (November 1986), pp. 1020–1022.

19. Long, M. W., *Radar Reflectivity of Land and Sea,* 2nd ed. Norwood, MA: Artech House, 1983.

20. Bush, T. F., F. T. Ulaby, and W. H. Peake. "Variability in the Measurement of Radar Backscatter." *IEEE Trans.* AP-24 (November 1976), pp. 896–899.

21. Bowditch, *American Practical Navigator,* U. S. Hydrographic Office, H. O. Publication No. 9, 1966, Appendix R.

22. Pierson, W. J., G. Neumann, and R. W. James. *Observing and Forecasting Ocean Waves,* U. S. Naval Oceanographic Office, H. O. Publication No. 603, 1955.

23. Daley, J. "An Empirical Sea Clutter Model." Naval Research Laboratory, Washington, D.C., Memorandum Report 2668, October 1973.

24. Daley, J. C., J. T. Ransone, Jr., and W. T. Davis. "Radar Sea Return—JOSS II." Naval Research Laboratory, Washington, D.C., Rep. 7534, February 21, 1973.

25. Daley, J. C., J. T. Ransone, Jr., and J. A. Burkett. "Radar Sea Return—JOSS I." Naval Research Laboratory, Washington, D.C., Report 7268, May 11, 1971.

26. Daley, J. C., et al. "Upwind-Downwind-Crosswind Sea Clutter Measurements." Naval Research Laboratory, Washington, D.C., Report 6881, April 14, 1969.

27. Hansen, J. P., and V. F. Cavaleri. "High-Resolution Radar Sea Scatter, Experimental Observations and Discriminants." Naval Research Laboratory, Washington, D.C., Rep. 8557, March 5, 1982.

28. Lewis, B. L., and I. D. Olin. "Experimental Study and Theoretical Model of High-Resolution Radar Backscatter from the Sea." *Radio Science.* 15 (July–August 1980), pp. 815–828.

29. Macdonald, F. C. "Characteristics of Radar Sea Clutter, Pt. 1—Persistent Target-Like Echoes in Sea Clutter." Naval Research Laboratory, Washington, D.C., Report 4902, March 19, 1957.

30. Ewell, G. W., M. T. Tuley, and W. F. Horne. "Temporal and Spatial Behavior of High Resolution Sea Clutter 'Spikes.'" *Proc 1984 IEEE National Radar Conf.* pp. 100–104, IEEE Catalog no. 84CH1963-8.

31. U. S. Patent 3,971,997: "Sea Spike Suppression Technique," issued to B. L. Lewis and I. D. Olin, July 27, 1976.

32. Williams, P. D. L. "Limitations of Radar Techniques for the Detection of Small Surface Targets in Clutter." *The Radio and Electronic Engineer* 45 (August 1975), pp. 379–389.

33. Long, M. W. Ref. 19, Sec. 5.6.

34. Wetzel, L. B. "On Microwave Scattering by Breaking Waves." In *Wave Dynamics and Radio Probing of the Ocean Surface.* O. M. Phillips and K. Hasselmann (Ed.). New York: Plenum, 1986, pp. 273–284.

35. Longuet-Higgins, M. S., and J. S. Turner. "An 'Entraining Plume' Model of a Spilling Breaker." *J. Fluid Mech.* 63 (1974), pp. 1–20.

36. Croney, J., A. Woroncow, and B. R. Gladman. "Further Observations on the Detection of Small Targets in Clutter." *The Radio and Electronic Engineer* 45 (March 1975), pp. 105–115.

37. Croney, J. "Improved Radar Visibility of Small Targets in Sea Clutter." *The Radio and Electronic Engineer* 32 (September 1966), p. 135–148.

38. Wetzel, L. B. "Electromagnetic Scattering from the Sea at Low Grazing Angles." In *Surface Waves and Fluxes.* Vol. II Netherlands: Kluwer Academic, 1990, chap. 12, pp. 109–171.

39. Wetzel, L. B. "A Model for Sea Backscatter Intermittency at Extreme Grazing Angles," *Radio Science* 12 (September–October 1977), pp. 747–756.

40. Knott, E. F., J. F. Shaffer, and M. T. Tuley. *Radar Cross Section,* 2nd ed. Norwood, MA: Artech House, 1993, Secs. 5.9 and 6.3.

41. Katzin, M. "On the Mechanisms of Radar Sea Clutter." *Proc. IRE* 45 (January 1957), pp. 44–54.

42. Helmken, H. H., and M. J. Vanderhill. "Very Low Grazing Angle Radar Backscatter from the Ocean Surface." *Record of the IEEE 1990 International Radar Conf.* pp. 181–188, IEEE Catalog No. 90CH-2882-9

43. Dockery, G. D. "Method for Modelling Sea Surface Clutter in Complicated Propagation Environments." *IEE Proc.* 137, Pt. F (April 1990), pp. 73–79.

44. Nathanson, F. E. Ref. 2, Sec. 7.2

45. Kerr, D. E. (Ed.): *Propagation of Short Radio Waves.* MIT Radiation Laboratory Series. New York: McGraw-Hill, 1951, vol. 13.

46. Goldstein, H. "Frequency Dependence of the Properties of Sea Echo." *Phys. Rev.* 70 (Dec. 1 and 15, 1946), pp. 938–946.

47. Kinsman, B. *Wind Waves.* Upper Saddle River, NJ: Prentice Hall, 1965, Sec. 7.4.

48. Beckmann, P., and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces.* New York: Macmillan, 1963, Secs. 5.3 and 18.2.

49. Leung, H., and S. Haykin. "Is There a Radar Clutter Attractor?" *Appl. Phys. Lett.* 56 (February 1990), pp. 593–595.

50. Crombie, D. D. "Doppler Spectrum of Sea Echo at 13.56 Mc/s." *Nature* 175 (1955), pp. 681–682.

51. Maresca, J. W., and T. M. Georges. "Measuring RMS Wave Height and the Scalar Ocean Wave Spectrum with HF Skywave Radar." *J. Geophys. Res.* 85 (1980), pp. 2759–2771.

52. "Special Issue on High-Frequency Radar for Ocean and Ice Mapping and Ship Location." *IEEE J. Oceanic Engr.* OE-11, No. 2 (April 1986).

53. Wright, J. W. "A New Model for Sea Clutter." *IEEE Trans.* AP-16 (March 1968), 217–223.

54. Wetzel, L. B. "Sea Clutter." In *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGrall-Hill, 1990, Chap. 13, Sec. 13.4.

55. Wetzel, L. B. Ref. 38, Chap. 12.

56. Lyzenga, D. R., A. L. Moffett, and R. A. Shuchman. "The Contribution of Wedge Scattering to the Radar Cross Section of the Ocean Surface." *IEEE Trans.* GE-31 (October 1983), pp. 502–505.

57. Nathanson, F. E: Ref. 2, Chap. 7.

58. Horst, M. M., F. B. Dyer, and M. T. Tuley. "Radar Sea Clutter Model." *Inter. Conf. on Ant. and Prop.* IEE Pub. No. 169, Pts. 1 and 2, London, 1978. (Also available in Ref. 2, pp. 307–308.)

59. Morchin, W. *Radar Engineer's Sourcebook.* Norwood, MA: Artech House, 1993, Sec. 3.3.2.

60. Ulaby, F. T., R. K. Moore, and A. K. Fung. Ref. 15, Sec. 20–4.

61. Lewis, E. O., B. W. Currie, and S. Haykin. *Detection and Classifcation of Ice.* New York: John Wiley, 1987.

62. Haykin, et al. *Remote Sensing of Sea Ice and Icebergs.* New York: John Wiley, 1994.

63. Wylie, F. J. *The Use of Radar at Sea,* 5th ed. Annapolis, MD: Naval Institute Press, 1978.

64. Ringwalt, D. L., and F. C. Macdonald. "Terrain Clutter Measurements in the Far North." *Report of NRL Progress,* Naval Research Laboratory, Washington, D.C., pp. 9–14, December 1956.

65. Orlando, J. R., R. Mann, and S. Haykin. "Classification of Sea-Ice Images Using a Dual-Polarized Radar." *IEEE J. Oceanic Engr.* 15 (July 1990), pp. 228–237.

66. Curlander, J. C., and R. N. McDonough. *Synthetic Aperture Radar.* New York: John Wiley, 1991. Appendix C.

67. Orlando, J. R., and S. Haykin: "Real-Time Detection of Iceberg Shadows." *IEEE J. Oceanic Engr.* 15 (April 1990), pp. 112–118.

68. Pilon, R. O., and C. G. Purves. "Radar Imagery of Oil Slicks." *IEEE Trans.* AES-9 (September 1973), pp. 630–636.

69. Ulaby, F. T., and M. C. Dobson. Ref.17, Chap. 3.

70. Trunk, G. V., and S. F. George. "Detection of Targets in Non-Gaussian Sea Clutter." *IEEE Trans.* AES-6 (September 1970), pp. 620–628.

71. Schleher, D. C. "Radar Detection in Log-Normal Clutter." *Record of the IEEE 1975 International Radar Conf.*, pp. 262–267. Reprinted in *Automatic Detection and Radar Data Processing*, D. C. Schleher (Ed.). Dedham, MA: Artech House, 1980.

72. Sekine, M., and Y. Mao. *Weibull Radar Clutter.* London: Peter Peregrinus, 1990.

73. Boothe, R. R. "The Weibull Distribution Applied to the Ground Clutter Backscatter Coefficient." U.S. Army Missile Command Report No. RE-TR-69-15, June, 1969; reprinted in *Automatic Detection and Radar Data Processing*, D. C. Schleher (Ed.). Dedham, MA: Artech House, 1980, pp. 435–450.

74. Sekine, et al. "Weibull Distributed Ground Clutter." *IEEE Trans.* AES-17 (July 1981), pp. 596–598.

75. Fay, F. A., J. Clarke, and R. S. Peters. "Weibull Distribution Applied to Sea Clutter," *Radar 77, IEE Conf. Publ.* 155, 1977, pp. 101–104; reprinted in *Advances in Radar Techniques.* London: Peter Peregrinus, 1985, pp. 236–239.

76. Sekine, M., et al. "Weibull Distributed Sea Clutter." *IEE Proc.* 130, Pt. F (1983), p. 476.

77. Sekine, M., et al. "On Weibull Distributed Weather Clutter." *IEEE Trans.* AES-15 (November 1979), pp. 824–830.

78. Ogawa, H., et al. "Weibull-Distributed Radar Clutter Reflected from Sea Ice." *Trans. IEICE* E70 (1987), pp. 116–120.

79. Schleher, D.C. "Radar Detection in Weibull Clutter." *IEEE Trans.* AES-12 (November 1976), pp. 736–743.

80. Jakeman, E., and P. N. Pusey. "A Model for Non-Rayleigh Sea Echo." *IEEE Trans.* AES-24 (November 1976), pp. 806–814.

81. Ward, K. D., and S. Watts. "Radar Sea Clutter." *Microwave J.* 28 (June 1985), pp. 109–121.

82. Jao, J. K. "Amplitude Distribution of Composite Terrain Radar Clutter and the K-Distribution." *IEEE Trans.* AES-32 (October 1984), pp. 1049–1062.

83. Ward, K. D. "A Radar Sea Clutter Model and Its Application to Performance Assessment." *International Conf. Radar-82* October 18–20, 1982. IEE Conference Publication No. 216, pp. 204–207.

84. Ward, K. D., C. J. Baker, and S. Watts. "Martime Surveillance Radar, Part 1: Radar Scattering from the Ocean Surface." *IEE Proc.* 137, Pt. F (April 1990), pp. 51–62.

85. Baker, C. J. "K-Distributed Coherent Sea Clutter." *IEE Proc.* 138, Pt. F (April 1991), pp. 89–92.

86. Watts, S. and D. C., Wicks. "Empirical Models for Detection Prediction in K-Distribution Radar Sea Clutter." *Record of the IEEE 1990 Int. Radar Conf.*, pp. 189–194, IEEE Catalog No. 90CH-2882-9.

87. Watts, S. and K. D. Ward. "Spatial Correlation in K-Distributed Sea Clutter." *IEE Proc.* 134, Pt. F (October 1987), pp. 526–532.

88. Trunk, G. V., and S. F. George. "Detection of Targets in Non-Gaussian Sea Clutter." *IEEE Trans.* AES-6 (September 1970), pp. 620–628.

89. Hair, T., T. Lee, and C. J. Baker. "Statistical Properties of Multifrequency High-Range-Resolution Sea Reflections." *IEE Proc.* 138, Pt. F (April 1991), pp. 75–79.

90. Baker, C. J. "K-Distributed Coherent Sea Clutter." *IEE Proc.* 138, Pt. F (April 1991), pp. 89–92.

91. Pentini, F. A., A. Farina, and F. Zirilli. "Radar Detection of Targets Located in a Coherent K Distributed Clutter Background." *IEE Proc.* 139, Pt. F (June 1992), pp. 239–245.

92. Watts, S. "Radar Detection Prediction in K-Distributed Sea Clutter and Thermal Noise." *IEEE Trans.* AES-23 (January 1987), pp. 40–45.

93. Armstrong, B. C., and H. D. Grifiths. "CFAR Detection of Fluctuating Targets in Spatially Correlated K-Distributed Clutter." *IEE Proc.* 138, Pt. F (April 1991), pp. 139–152.

94. Trunk, G. V. "Radar Properties of Non-Rayleigh Sea Clutter." *IEEE Trans.* AES-8 (March 1972), pp. 196–204.

95. Tough, R. J. A., C. J. Baker, and J. M. Pink. "Radar Performance in a Maritime Environment: Single Hit Detection in the Presence of Multipath Fading and Non-Rayleigh Sea Clutter." *IEE Proc.* 137, Pt. F (February 1990), pp. 33–40.

96. Oliver, C. J. "Representation of Radar Sea Clutter." *IEE Proc.* 135, Pt. F (December 1988), pp. 497–500.

97. Sekine, et al. "Log-Weibull Distributed Sea Clutter." *IEE Proc.* 127, Pt. F (June 1980), pp. 225–228.

98. Shlyakhin, V. M. "Probability Models of Non-Rayleigh Fluctuations of Radar Signals (A Review)." *Soviet J. Communications Technology and Electronics.* 33 (January 1988), pp. 1–16.

99. Schleher, D. C. "Radar Detection in Log Normal Clutter." *Record of the IEEE 1975 International Radar Conf.* pp. 262–267, IEEE Publication 75 CHO 938-1 AES.

100. Farina, A., A. Russo, and F. A. Studer. "Coherent Radar Detection in Log-Normal Clutter." *IEE Proc.* 133, Pt. F (February 1986), pp. 39–54.

101. Farina, A., et al. "Theory of Radar Detection in Coherent Weibull Clutter." *IEE Proc.* 134, Pt. F (April 1987), pp. 174–190.

102. Sangston, K. J. "Coherent Detection of Radar Targets in K-Distributed, Correlated Clutter." Naval Research Laboratory, Washington, D.C., Report 9130, August, 1988.

103. Battan, L. J. *Radar Observation of the Atmosphere.* Chicago, IL: Univ. of Chicago, 1973.

104. Gunn, K. L. S., and T. W. R. East. "The Microwave Properties of Precipitation Particles." *Quart. J. Roy. Meteor. Soc.* 80 (October 1954), pp. 522–545.

105. Sauvageot, H. *Radar Meteorology.* Norwood, MA: Artech House, 1992.

106. Personal communication from Raymond Wexler.

107. Haddock, F. T. "Scattering and Attenuation of Microwave Radiation Through Rain." Naval Research Laboratory, Washington, D.C., (unpublished manuscript), 1948. (Mentioned in Gunn and East, Ref. 104)

108. Smith, P. L. Jr., K. R. Hardy, and K. M. Glover. "Applications of Radar to Meteorological Operations and Research." *Proc. IEEE.* 62 (June 1974), pp. 724–745.

109. Jameson, A. R. "A Comparison of Microwave Techniques for Measuring Rainfall." *J. Applied Meteorology* 30 (January 1991), pp. 32–54.

110. Nathanson, F. E. Ref. 2, Sec. 3.6.

111. Sauvageot, H. Ref. 105, Sec. 1.5.3.

112. Gunn, K. L. S., and J. S. Marshall. "The Distribution with Size of Aggregate Snowflakes." *J. Meteor.* 15 (1958), pp. 452–466.

113. Sekhon, R. S., and R. C. Srivastava. "Snow Size Spectra and Radar Reflectivity." *J. Atmos. Sci.* 27 (1970), pp. 299–307.

114. Puhakka, T. "On the Dependence of the Z-R Relationship on the Temperature in Snowfall." *Preprints 16th Radar Meteorology Conf.* Am. Meteor. Soc., April 22–24, 1975, Houston, TX, pp. 504–507.

115. Austin, P. M. "Measurements of the Distribution of Precipitation in New England Storms." *Proc. 10th Weather Radar Conf.*, Am. Meteor. Soc., pp. 247–254, 1963.

116. Saxton, J. A. "The Influence of Atmospheric Conditions on Radar Performance." *J. Inst. Navigation* (London) 11, pp. 290–303, 1958.

117. Nathanson, F. E., Ref. 2, Fig. 6.4.

118. Konrad, T. G., J. J. Hicks, and E. B. Dobson. "Radar Characteristics of Brids in Flight." *Science* 159 (Jan. 19, 1968), pp. 274–280.

119. Vaughn, C. R. "Birds and Insects as Radar Targets: A Review." *Proc. IEEE* 73 (February 1985), pp. 205–227.

120. Eastwood, E. *Radar Ornithology.* London: Methuen, 1967.

121. Williams, T. C. and J. M. Williams. "Open Ocean Bird Migration." *IEE Proc.* 137, Pt. F (April 1990), pp. 133–137.

122. Houghton, E. W., F. Blackwell, and T. A. Wilmot. "Bird Strike and the Radar Properties of Birds." *Int. Conf. on Radar—Present and Future,* Oct. 23–25, 1973, pp. 257–262, IEE Conference Publication no. 105.

123. Flock, W. L., and J. L. Green. "The Detection and Identification of Birds in Flight, Using Coherent and Noncoherent Radars." *Proc. IEEE* 62 (June 1974), pp. 745–753.

124. Blackwell, F., and E. W. Houghton. "Radar Tracking and Identification of Wild Duck During the Autumn Migration." *Proc. World Conf. on Bird Hazards to Aircraft, Canada* (1969), pp. 361–376.

125. Antonucci, J. "A Statitical Model of Radar Bird Clutter at the DEW Line." Rome Laboratory (EEAS), Air Force Systems Command, Hanscomb AFB, MA, Report no. RL-TR-91-85, May, 1991.

126. Knott, E. F. "*Radar Cross Section.*" In *Radar Handbook*, 2nd ed. M. Skolnik (Ed.) New York: McGraw-Hill, 1990, Chap. 11.

127. Riley, J. R. "Radar Cross Section of Insects." *Proc. IEEE* 73 (February 1985), pp. 228–232.

128. Schaefer, G. W. "Radar Observations of Insect Flight," *Insect Flight,* R. C. Rainey (Ed.) London: Blackwell Scientific Publications, 1976, Chap. 8.

129. Gossard, E. E., and R. G. Strauch. *Radar Observations of Clear Air and Clouds.* New York: Elsevier, 1983.

130. James, P. K. "A Review of Radar Observations of the Troposphere in Clear Air Conditions." *Radio Sci.* 15 (March–April 1980), pp. 151–175. (The entire March-April 1980 issue of *Radio Sci.* is devoted to "Radar Investigations of the Clear Air.")

131. Tatarskii, V. T. *Wave Propagation in a Turbulent Medium.* New York: McGraw-Hill, 1961.

132. Skolnik, M. "Atmospheric Turbulence and the Extension of the Radar Horizon." Naval Research Laboratory, Washington, D.C., Memorandum Rep. 2903, October 1974.

133. Heiss, W. H., D. L. McGrew, and D. Sirmans. "Nexrad: Next Generation Weather Radar (WSR-88D)." *Microwave J.* 33 (January 1990), pp. 79–98.

134. Michelson, M., Shrader, W. W., and J. G. Wieler. "Terminal Doppler Weather Radar." *Microwave J.* 33, (February 1990), pp. 139–148.

135. Plank, V. G. "A Meteorological Study of Radar Angels." U.S.A.F. Cambridge Research Center Geophys. Research Papers, no. 52, July, 1956, AFCRC-TR-56-211, AD 98752.

136. Rogers, R. R., and W. O. J. Brown. "Radar Observations of a Major Industrial Fire." *Bull. Am. Meteorological Soc.* 78 (May 1997), pp. 803–814.

137. Skolnik, M., D. Hemenway, and J. P. Hansen. "Radar Detection of Gas Seepage Associated with Oil and Gas Deposits." *IEEE Trans.* GRS-30 (May 1992), pp. 630–633.

138. Skolnik, M., and T. C. Bailey. "A Review of Radar for the Detection of Gas Seepage Associated with Underground Oil and Gas Deposits." *Proc. of the Conf. on Applications of Emerging Technologies: Unconventional Methods in Exploration for Petroleum and Natural Gas V,* Institute for the Study of Earth and Man, Southern Methodist Univ., Dallas, Texas, 1997, pp. 207–228.

139. Eastwood, E. Ref. 120, Chap. 9; also V. G. Plank, Ref. 135.

140. Billingsley, J. B. "Ground Clutter Measurements for Surface-Sited Radar," MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 786, Revision 1, February 1, 1993.

141. Caspers, J. W. "Automatic Detection Theory." In *Radar Handbook,* 1st ed. M. Skolnik (Ed.). New York: McGraw-Hill, 1970, chap. 15, Sec.15.6.

142. Trunk, G. V., and S. F. George. "Detection of Targets in Non-Gaussian Sea Clutter." *IEEE Trans.* AES-6 (September 1970), pp. 620–628.

143. Trunk, G. V. "Further Results on the Detection of Targets in Non-Gaussian Sea Clutter." *IEEE Trans.* AES-7 (May 1971), pp. 553–556.

144. Schleher, D. C. "Radar Detection in Weibull Clutter." *IEEE Trans.* AES-12 (November 1976), pp. 736–743. Correction in AES-13 (July 1977), p. 435.

145. Schleher, D. C. "Radar Detection in Log-Normal Clutter." *IEEE International Radar Conf.* Arlington, VA, April 21–23, 1975, pp. 262–267.

146. Croney, J., A. Woroncow, and B. R. Gladman, "Further Observations on the Detection of Small Targets in Sea Clutter." *The Radio and Electronic Engineer* 45 (March 1975), pp. 105–115.

147. Williams, P. D. L. "Observations on the Further Optimization of Radar Signal Processing for the Display and Detection of Targets in Sea Clutter." *Int. J. Remote Sensing* 5 (1984), pp. 489–496.

148. Tonkin, S. P., and R. A. McCulloch. "Gross Spatial Structure of Land Clutter." *IEE Proc.* 138, Pt. F. (April 1991), pp. 99–108.

149. Shrader, W. W., and V. Gregers-Hansen. "MTI Radar." *Radar Handbook,* M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 15, p. 15.13.

150. Gordon, W. B. "Analysis of Rain Clutter Data from a Frequency Agile Radar." *Radio Science* 17 (July-August 1982), pp. 801–816.

151. Queen, F. D., and J. J. Alter. "Results of a Feasibility Study for Determining the Yaw Angle of a Landing Aircraft." Naval Research Laboratory, Washington, D.C., Rep. 8480, May 27, 1981.

152. Hughes, P. K., II. "A High-Resolution Radar Detection Strategy." *IEEE Trans.* AES-19 (September 1983), pp. 663–667.

153. Gent, H., I. M. Hunter, and N. P. Robinson. "Polarization of Radar Echoes, Including Aircraft, Precipitation, and Terrain." *IEE Proc.* 110 (December 1963), pp. 2139–2148.

154. Sauvageot, H. *Radar Meteorology.* Norwood, MA: Artech House, 1992, Sec. 2.2.7.

155. Oguchi, T. "Electromagnetic Wave Propagation and Scattering in Rain and Other Hydrometeors." *Proc. IEEE* 71, (September 1983), pp. 1029–1078.

156. McDonald, J. E. "The Shape of Raindrops." *Scientific American* (February 1954).

157. Schneider, A. B., and P. D. L. Williams. "Circular Polarization in Radars." *The Radio and Electronic Engineer* 47, no. 1/2 (January/February 1976), pp. 11–29.

158. Kalafus, R. M. "Rain Cancellation Deterioration Due to Surface Reflections in Ground-Mapping Radars Using Circular Polarization." *IEEE Trans.* AP-23 (March 1975), pp. 269–271.

159. Beasley, E. W. "Effect of Surface Reflections on Rain Cancellation in Radars Using Circular Polarization." *Proc. IEEE* 54 (December 1966), pp. 2000–2001.

160. Hendry, A., and G. C. McCormick. "Deterioration of Circular-Polarization Clutter Cancellation in Anisotropic Precipitation Media." *Electronics Letters* 10 (May 16, 1974), pp. 165–166.

161. Nathanson, F. E. "Adaptive Circular Polarization." *IEEE 1975 International Radar Conf.*, April 21–23, 1975, pp. 221–225.

162. Nathanson, F. E. Ref. 2, Sec. 5.2.

163. Schneider, A. B., and P. D. L. Williams. Ref. 157.

164. Olin, I. D., and F. D. Queen. "Dynamic Measurement of Radar Cross Sections." *Proc. IEEE* 53 (August 1965), pp. 954–961.

165. Kerr, D. E. (Ed.). *Propagation of Short Radio Waves.* MIT Radiation Laboratory Series. New York: McGraw-Hill, 1951, vol. 13.

166. Croney, J. "Improved Radar Visibility of Small Targets in Sea Clutter. *The Radio and Electronic Engineer* 32 (September 1966), pp. 135–148.

167. Croney, J., and A. Woroncow. "Dependence of Sea Clutter Decorrelation Improvments Upon Wave Height." *IEE Int. Conf. On Advances in Marine Navigational Aids*, July 25–27, 1972, IEE (London) Conf. Publication no. 87, pp. 53–59.

168. Smith, J. M., and R. H. Logan. "AN/APS-116 Periscope Detecting Radar." *IEEE Trans.* AES-16 (January 1980), pp. 66–73.

169. Valenzuela, G. R., and M. B. Laing. "Study of Doppler Spectra of Radar Sea Echo." *J. Geophys. Res.* 75 (January 20, 1970), pp. 551–563.

170. Nathanson, F. E. Ref. 2, Sec. 6.6.

171. Fuller, J. B., and J. R. Martin. "Radar Subsystems," Varian Co. brochure, Beverly Division, Beverly MA, no date, Fig. 2. (Now known as CPI Beverly Microwave Division.)

172. Becker, J. E., and R. E. Millet. "A Double-Slot Radar Fence for Increased Clutter Suppression." *IEEE Trans.* AP-16 (January 1968), pp. 103–108.

173. Ruze, J., F. I. Sheftman, and D. A. Cahlander. "Radar Ground-Clutter Shields." *Proc. IEEE* 54 (September 1966), pp. 1171–1183.

174. Harrison, A. "Marine Radar Today—A Review." *The Radio and Electronic Engineer* 47 (April 1977), pp. 177–183.

175. Taylor, J. W., Jr. "Receivers." In *Radar Handbook,* M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 3, Sec. 3.6.

176. Croney, J. "Clutter on Radar Displays." *Wireless Engr.* 33 (April 1956), pp. 83–96.

177. Taylor, J. W., Jr. Ref. 175, Sec. 3.13.

PROBLEMS

7.1 Find the range at which the radar echo from low grazing angle surface clutter equals receiver noise for the following radar: peak power = 100 kW, azimuth beamwidth = 1°, elevation beamwidth = 20° [see Eq. (9.5c)], wavelength = 3 cm, pulse width = 0.1 μs, and receiver noise figure = 4 dB. The radar is located at a height of 1 km over a flat sea surface. You may assume the following variation of σ^0 with grazing angle Ψ:

$$\text{sigma zero} = -46 \text{ dB at } 1.0^\circ \text{ grazing angle}$$
$$= -42 \text{ dB at } 3.0^\circ$$
$$= -37 \text{ dB at } 10.0^\circ$$

7.2 The ARSR-3 is a long-range, ground-based, fan-beam, *L*-band radar that was used by the FAA for enroute air-traffic control. It had an azimuth beamwidth of 1.25° and a pulse width of 2 μs. (a) If the clutter cross section per unit area, σ^0, for surface land clutter is -20 dB, what is the MTI clutter attenuation required to detect a 2 m^2 target at a range of 30 nmi with an output signal-to-clutter ratio of 15 dB? (Clutter is considered to be much greater than receiver noise.) (b) What will be the radar cross section of rain clutter (in m^2) seen by the ARSR-3 radar at a range of 30 nmi when the rainfall rate is 4 mm/h? You may assume a flat earth and that the rain uniformly fills the radar resolution cell from the ground up to a height h of 3 km. Take the radar frequency to be 1.3 GHz. (The elevation angle coverage should not be needed here.)

7.3 Derive a radar equation for the detection of a target in surface clutter when the grazing angle is 90° (normal incidence). Assume the antenna employs a pencil beam. (This is not a usual radar detection situation because the clutter echo is large when the antenna beam is perpendicular to the surface.)

7.4 An S-band pencil-beam radar (3.2 GHz) with a 1.5° beamwidth and 3-μs pulse width can detect a 2 m^2 target at a range of 200 nmi in a clear atmosphere (no rain). If a single-pulse signal-to-clutter ratio of 10 dB is required, what would be the range of the radar when observing a 2-m^2 target in rain of 4 mm/h? (Assume that clutter is uniformly distributed and is much greater than receiver noise, attenuation in rain at this frequency can be neglected, the antenna beam is pointing at a low elevation angle but doesn't strike the ground, and rain completely fills the radar resolution cell.)

7.5 Sketch as a function of rainfall rate (from 1 to 40 mm/h) the variation of the pulse echo power from rain for a radar with the following characteristics (which are those of Nexrad):

peak power = 1 MW
frequency = 3.0 GHz
pulse width = 1.57 μs
antenna gain = 45.5 dB (pencil beam)
effective number of hits integrated = 80
range = 150 km

(Assume that the beam is completely filled with rain, the beam is low to the ground but does not intercept it, and the attenuation due to rain can be neglected.)

7.6 What will be the attenuation due to rain uniformly distributed throughout the radar coverage and what will be the radar cross section of rain at a range of 20 km for a radar pointing at a low elevation angle with a 2° by 4° beamwidth and a 2-μs pulse width at frequencies of 3, 10, and 35 GHz? The rain falls at a rate of 4 mm/h.

7.7 Briefly comment on how the radar parameters listed below affect radar performance when detection is limited by (1) surface clutter and by (2) receiver noise. The radar parameters

are: (a) pulse width, (b) antenna gain, (c) transmitter power, (d) number of pulses returned from the target, (e) system losses, and (f) the sensitivity of the maximum detection range to changes in the radar cross section. (It may help to make a table.)

7.8 Show that the radar cross section per unit area, σ^0, for a radar with a narrow pencil-beam antenna viewing a large perfectly reflecting flat surface at perpendicular incidence is approximately equal to G, where G = antenna gain. The flat surface is much larger than the extent of the footprint of the radar beam hitting the surface. In other words, derive Eq. (7.12). [The key to this problem is solving for the received power in two different ways, and then equating them to find σ^0. Start with the simple form of the radar equation for the received echo signal power P_r from a radar directed perpendicular to a flat surface. The cross section of the target is $\sigma^0 A_c$, where A_c is the area of that portion of the flat surface which is illuminated by the radar beam. Find an expression for A_c. The half-power beamwidth is θ_B, but you have to account for the change in illuminated area due to the two-way beamwidth. The second way to find the received power is recognize that the flat plate creates an image of the radar antenna a distance R behind the plate. The receive signal power is found by considering the geometry as a one-way communication link between the radar antenna and its image located a distance R behind the surface. Derive a one-way communication equation for the signal received at the image antenna when the radar antenna transmits to its image a distance $2R$ away. The two values of P_r, found from the two different models of propagation path, are the same and may be equated. You might need to use Eq. (9.5b) for the antenna gain.]

7.9 This problem concerns finding the radar cross section (in square meters) of various forms of distributed volumetric clutter at low altitudes as might be seen at a range of 10 km by an X-band radar with a 1° beamwidth and a 1-μs pulse width. The cross section per unit volume of the volumetric clutter is denoted by η. The clutter is assumed to occupy the entire radar resolution cell. (Be careful of units.) (a) Rain falling at 4 mm/h. (b) Migrating Shearwater birds with $\eta = 3 \times 10^{-9}$ cm^{-1}. Compare your answer with the radar cross section of a single Shearwater bird as given in Table 7.5. (c) Insects with $\eta = 10^{-10}$ cm^{-1}. (d) Clear-air turbulence with $C_n^2 = 10^{-10}$ m$^{-2/3}$.

7.10 List five options available to the radar designer for enhancing the detection of aircraft in rain. Briefly describe the operation of each and their chief limitation.

7.11 If an S-band radar is capable of detecting a 1 m^2 target at a range of 200 nmi, at what range will it be able to detect a single sparrow? (See Table 7.4.)

7.12 Assume an X-band radar (λ = 3.2 cm) has a range of 45 km in the absence of rain. If it is further assumed that the attenuation in rain is the only factor affecting range, what will be the radar range when rain is falling throughout the region at a rate of 4 mm/h? (You will probably have to solve for the range by trail and error. The ultimate in precision is not vital in the answer to this question.)

7.13 What radar parameters are under the control of the radar designer for increasing the range at which a target may be detected in rain? Which parameter do you think is most important for increasing the detectability of a target? Explain why.

7.14 The value of the cross section per unit volume of rain, η, varies as f^4, where f = radar frequency. Assume the rain completely fills the radar resolution cell. (a) What is the

variation of the echo from rain when the antenna gain G is independent of frequency? (b) What is the variation of the echo from rain when the antenna aperture area A_e is independent of frequency?

7.15 Show that the clutter echo seen by a pulse of width τ is decorrelated if the radar frequency is changed by an amount greater than $1/\tau$. [A decorrelated echo due to a change in frequency Δf means that the phase difference between the radar echo from the leading edge of the resolution cell and the trailing edge of the resolution cell changes by more than 2π radians when the frequency is changed an amount Δf. This occurs since phase is modulo 2π.) Start by obtaining an expression for the two-way phase difference ϕ in terms of the radar resolution cell ΔR and wavelength λ. Note that the resolution cell is $\Delta R = c\tau/2$. The change in phase $\Delta\phi$ has to be greater than 2π radians when the frequency is changed by Δf in order to decorrelate the echo. The answer is now staring at you.]

7.16 Explain why frequency agility is not compatible with MTI (doppler) radar processing for detection of moving targets in clutter. (You may take as a model the single delay-line canceler.)

7.17 Why is the Bragg-scatter model not suitable for describing sea clutter at the higher microwave frequencies?

7.18 Briefly indicate what you think might be the preferred radar method or methods for detecting the targets listed below when in the presence of clutter. Assume the radar is located on the surface of the earth.

a. Aircraft over normal land clutter.
b. Aircraft over normal sea clutter.
c. Aircraft in the presence of large numbers of birds and insects.
d. Aircraft in the presence of moving ground targets.
e. Aircraft in rain clutter but beyond the range of surface clutter.
f. Ships and large buoys with corner reflectors.
g. Small boats, periscopes, buoys, and swimmers.
h. Stationary targets in distributed clutter.

chapter

8

Propagation of Radar Waves

8.1 INTRODUCTION

Most of the discussion of the radar equation in Chap. 2 considered the radar energy to propagate in free space. In the real world, however, the earth's surface and atmosphere can have major effects on radar performance. In Sec. 2.12 we briefly mentioned a few of the propagation factors that can influence the range and coverage of a radar. Since propagation effects might extend the radar range significantly or reduce it drastically, it is important to account for the earth's environment when attempting to predict radar performance.

Free-space radar performance is modified by the following propagation effects:

- *Forward scattering* (reflection) of the radar energy from the surface of the earth, which enhances the radiated energy at some elevation angles and decreases it at others.
- *Refraction* (bending) of the radar energy by the earth's atmosphere, which can cause the radar energy to deviate from straight-line propagation.
- *Ducting* (trapping) of the radar energy, a form of severe refraction, which causes extended radar ranges (and, surprisingly, might not always be a good thing).
- *Diffraction* of radar waves by the earth's surface that causes energy to propagate beyond the normal radar horizon. It applies mainly at the lower frequencies that are seldom used for radar applications.

- *Attenuation* of radar waves by the clear atmosphere, which generally has little or no effect on microwave propagation.
- *External noise* that enters the radar receiver and increases the receiver noise level.
- *Backscatter from land, sea, and weather clutter, and attenuation in rain and other hydrometeors.* These are not discussed here since they were included in Chap. 7.

Most of the significant propagation effects at microwave radar frequencies occur within the line of sight of the radar. Diffraction effects when the radar is at a sufficiently low frequency, and ducting at almost any microwave frequency, can cause radar waves to bend around the surface of the earth and extend the radar range beyond the normal line-of-sight horizon.

Although the basic theory of radar propagation may be well understood, accurate quantitative predictions for a particular place and for a specific time in the future are not always easy to obtain because of the difficulty in acquiring the necessary information about the environment in which a radar operates. In some respects, predicting the effects of propagation is a little like forecasting the weather. The radar system designer is usually interested in a long-term statistical description of propagation effects so that the radar can be designed to fulfill its mission satisfactorily. Sometimes, however, the radar designer has to be content with only a qualitative knowledge of "average" propagation effects. The military tactical commander or the air-traffic controller is not as interested in the long-term statistical effects of propagation or average conditions, but is more interested in the current or short-term forecast of radar propagation conditions that might be encountered. For example, based on current measurements of the environmental factors that affect radar propagation, a military tactical planner might be able to determine a flight profile of an attacking aircraft that would minimize the range where it is first detected by a defender's radar.

8.2 FORWARD SCATTERING FROM A FLAT EARTH

To determine the type of effects the earth's surface has on radar propagation, we initially assume a plane, smooth, perfectly reflecting flat earth. The results obtained with this simplification are indicative of what is obtained with more realistic models. The geometry is shown in Fig. 8.1a. The radar antenna is located at a height h_a above the planar surface. Its antenna radiation pattern is assumed to be uniform in elevation angle. The target is at a height h_t and at a range R from the radar. The ground distance between the radar and the target is D (not shown in the figure). Energy radiated by the radar antenna arrives at the target via two separate paths. One is the direct path (AB) from radar to target; the other is the path (AMB) from the radar to the target that includes a forward-scatter reflection from the surface. The signal reflected by the target also arrives back at the radar by these same two paths. The magnitude of the resultant echo signal back at the radar antenna depends on the amplitudes and relative phases of the signals that propagate via the direct and surface-scattered paths. The modification of the field strength η (measured in volts/meter) caused by the presence of the earth's surface may be expressed by the ratio

$$\eta = \frac{\text{field strength at target in presence of earth's surface}}{\text{field strength at target if in free space}} \qquad [8.1]$$

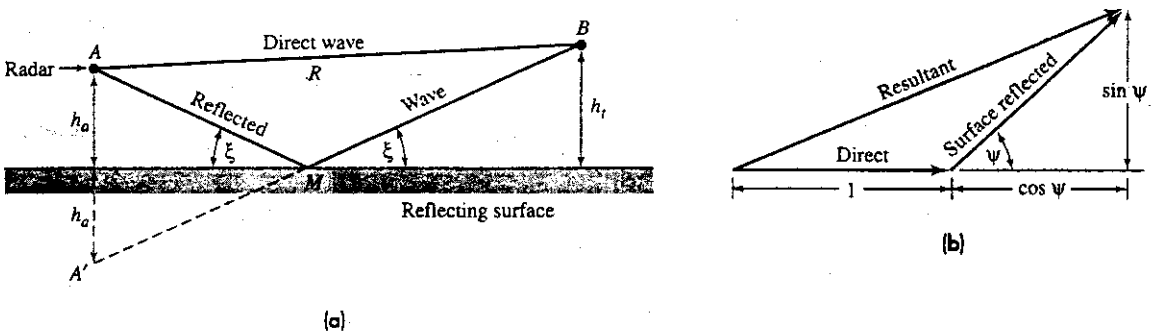


Figure 8.1 (a) Geometry illustrating radar propagation over a plane reflecting surface. (b) Vector addition of direct and surface-reflected signals, each of unity amplitude, with a phase difference of ψ.

It is assumed in this analysis that the path lengths of the direct and surface-reflected signals are almost (but not quite) equal so that the amplitudes of the two signals are essentially the same, except for any loss of signal suffered on reflection from the earth's surface. That is, if the two signals differ in amplitude from one another it is due to the surface reflection-coefficient being less than unity, and is not due to a significant difference in the $1/R^2$ factor. The slight range difference between the direct and surface-reflected paths, however, results in a difference in phase between the two which, when they combine at the target or at the radar, affects their sum. There is also a change in phase of the signal when it is reflected from the surface. This is represented by a reflection coefficient, which is a complex quantity $\Gamma = \rho e^{-\psi_r}$. The magnitude ρ describes the change in amplitude on reflection, the argument ψ_r describes the phase shift.

In this particular analysis we take the reflection coefficient to be $\Gamma = -1$. Thus the surface-reflected wave does not change its amplitude on reflection, but its phase is shifted by an amount $\psi_r = \pi$ radians. A reflection coefficient of $\Gamma = -1$ applies to a perfectly smooth, perfectly conducting surface if the radiation is horizontally polarized and the grazing angle is small.

The problem is easier to analyze if we replace the surface-reflected signal with the signal radiated from the image of the radar antenna that is below the surface, at A' in Fig. 8.1a. Instead of the surface-reflected path AMB, we consider the equivalent straight-line path $A'MB$. The path length AB is

$$AB = [D^2 + (h_t - h_a)^2]^{1/2} = D\left[1 + \frac{(h_t - h_a)^2}{D^2}\right]^{1/2}$$
$$\approx D\left[1 + \frac{1}{2} \cdot \frac{(h_t - h_a)^2}{D^2}\right] = D + \frac{(h_t - h_a)^2}{2D} \qquad [8.2]$$

In the above we assumed $|h_t - h_a| << D$. The surface-reflected path length AMB, or its equivalent, $A'MB$ is similarly

$$AMB = A'MB = [D^2 + (h_t + h_a)^2]^{1/2} \approx D + \frac{(h_t + h_a)^2}{2D} \qquad [8.3]$$

where we have assumed that $h_t + h_a << D$. Subtracting Eq. (8.2) from Eq. (8.3) results in the difference between the two paths $AMB - AB$, which is $\Delta = 2h_t h_a/D$. We make the assumption that the horizontal distance D can be replaced by the range R, so that the difference between the two paths is

$$\Delta \approx \frac{2h_a h_t}{R} \tag{8.4}$$

To recapitulate, the geometrical assumptions we have made are that $(h_t \pm h_a) << D \approx R$. The phase lag ψ_Δ associated with this difference is found by multiplying Eq. (8.4) by $2\pi/\lambda$. The total phase difference is then

$$\psi = \psi_\Delta + \psi_r = \frac{4\pi h_a h_t}{\lambda R} + \pi \tag{8.5}$$

At the target there are two signals, the direct and surface-reflected, which are of the same approximate amplitude with a phase difference between them of ψ, as given in Eq. (8.5). To obtain η, the vector addition of these two signals is divided by the signal amplitude that would have appeared if in free space. The value of η is found by applying the Pythagorean theorem to the sum of the two signal vectors, Fig. 8.1b, both of the same amplitude (normalized to unity) but with relative phase ψ. This gives

$$\eta = [(1 + \cos\psi)^2 + (\sin\psi)^2]^{1/2} = [2(1 + \cos\psi)]^{1/2} \tag{8.6a}$$

The value η^2 is the signal power density (W/m^2) at the target compared to the power density that would have been at the target if it were in free space, which becomes

$$\eta^2 = 2\left(1 - \cos\frac{4\pi h_a h_t}{\lambda R}\right) = 4\sin^2\left(\frac{2\pi h_a h_t}{\lambda R}\right) \tag{8.6b}$$

Because of reciprocity in propagation, the path from target to radar is the same as that from radar to target. The echo signal power density received at the radar, relative to what would have been received in free space, is the fourth power of η, or

$$\eta^4 = 16\sin^4\left(\frac{2\pi h_a h_t}{\lambda R}\right) \tag{8.7}$$

Lobing The radar equation describing the received echo power is multiplied by the factor η^4 as given by Eq. (8.7). Since the sine function varies from 0 to 1, the factor η^4 varies from 0 to 16. The effect of the earth's surface in this simplified example is to increase the received signal power at some elevation angles by as much as 16. At other elevation angles it can be zero. Because of the fourth-power relation between range and received echo-signal power, the radar range will vary from 0 to 2 times the range the radar would have if it were in free space. The result is that the radiation in elevation is broken up into lobes which increase the range at some elevation angles and decrease it at others, as shown in Fig. 8.2. This effect is sometimes called *lobing*.

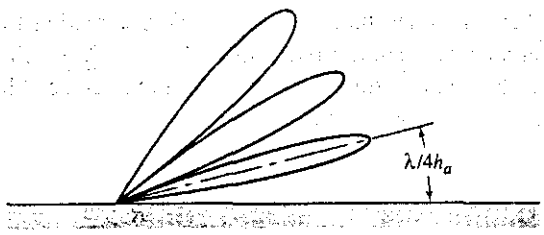


Figure 8.2 Vertical (elevation) lobe structure of the radar radiation caused by the presence of a planar reflecting surface.

The field strength in the presence of the earth's surface is a maximum when the argument of the sine term in Eq. (8.7) is equal to $\pi/2, 3\pi/2, \ldots, (2n + 1)\pi/2$, where $n = 0, 1, 2, \ldots$. The peaks of the lobes occur when

$$\frac{4h_a h_t}{\lambda R} = 2n + 1 \qquad \text{maxima} \tag{8.8}$$

and the nulls, or minima, occur when the sine term is zero, or when

$$\frac{2h_a h_t}{\lambda R} = n \qquad \text{minima} \tag{8.9}$$

From Eq. (8.8), the angle of the peak of the first (lowest) lobe ($n = 0$), is at

$$\theta_1 \approx h_t/R = \lambda/4h_a \tag{8.10}$$

Thus if it is desired to see targets at low angles, the wavelength must be small (high frequency) and/or the antenna height must be large.

To illustrate the effects of a flat, smooth, perfectly reflecting earth's surface on radar performance we include η^4 in the simple form of the radar equation [Eq. (1.6)], which then becomes

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4} \cdot 16 \sin^4 \left(\frac{2\pi h_a h_t}{\lambda R} \right) \tag{8.11a}$$

When the argument of the sine is small,

$$P_r \approx \frac{4\pi P_t G^2 \sigma (h_a h_t)^4}{\lambda^2 R^8} \tag{8.11b}$$

This represents the region below the peak of the first lobe. For targets at small angles (lower than the first lobe), the signal power is seen to vary as the inverse eighth power of the range, rather than as the inverse fourth power as occurs in free space.* The gain and wavelength appear in Eq. (8.11b) as the factor G/λ instead of $G\lambda$. The above applies for an antenna with a constant gain as a function of frequency. A different result is obtained if the effective aperture of the antenna is maintained constant with change in frequency.

*As shall be seen later, the variation of signal strength with range at low angles can be much more complicated than that given by Eq. (8.11b), especially under conditions of anomalous propagation.

Surface Reflection Coefficient In the above several simplifying assumptions were made. One was that the antenna elevation pattern was uniform. With an actual antenna the antenna gain as a function of elevation angle has to be taken into account since the gain in the direction of the target can be different from the gain in the direction of the surface-reflected ray. Another assumption was that the surface was smooth and perfectly reflecting. This is not always realistic since the reflection coefficient depends on the surface roughness, the dielectric properties of the surface, polarization of the radar energy, and the frequency. Figure 8.3a gives the magnitude of the reflection coefficient as a function of the grazing angle and frequency for vertical polarization over smooth sea water. Figure 8.3b is the phase of the reflection coefficient for vertical polarization and a smooth sea. The magnitude of the reflection coefficient for horizontal polarization is given in Fig. 8.3c. The phase of the reflection coefficient for horizontal polarization is approximately

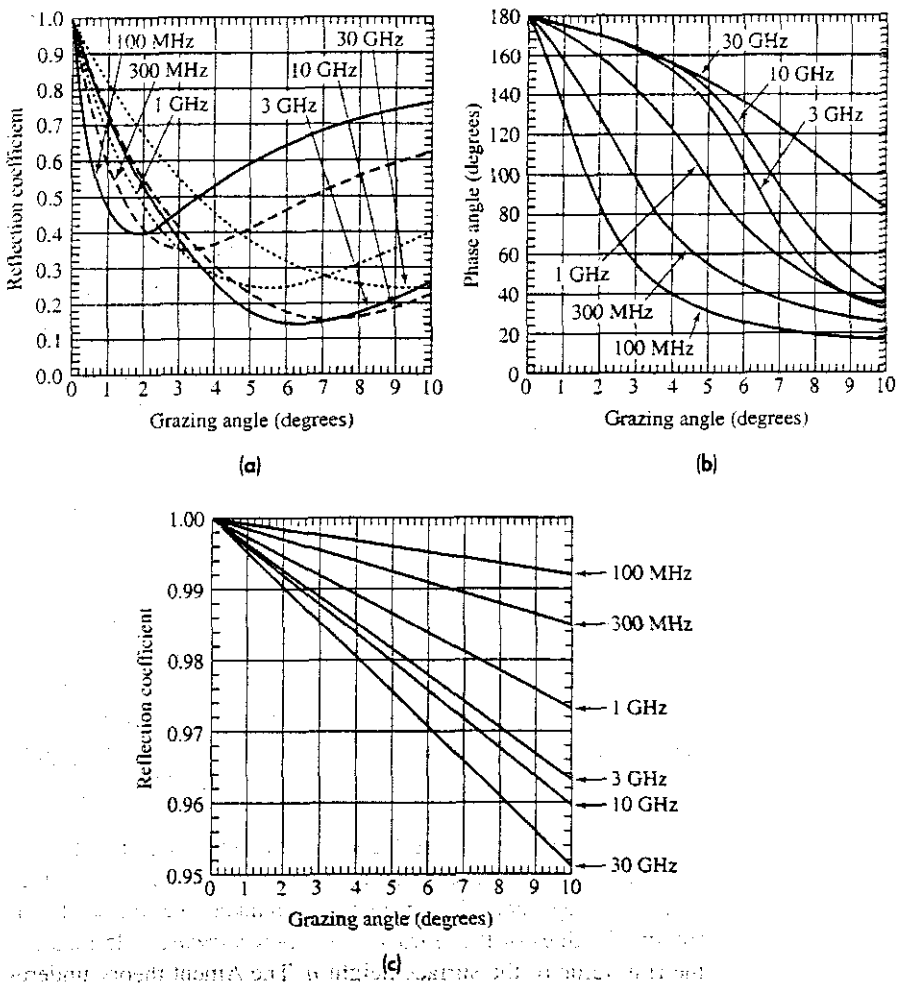


Figure 8.3
Reflection coefficient for a smooth sea as a function of grazing angle, frequency, and polarization. (a) Magnitude of the reflection coefficient for vertical polarization; (b) phase of the reflection coefficient for vertical polarization (phase of the reflected wave lags the phase of the incident wave); (c) magnitude of the reflection coefficient for horizontal polarization. The phase of the reflection coefficient for horizontal polarization is 180°, and is approximately independent of grazing angle and frequency. (From Lamont Blake, *Radar Handbook*.[20])

π radians, and doesn't vary much with frequency or grazing angle. The magnitude of the reflection coefficient is generally less for vertical polarization than for horizontal.

The minimum reflection coefficient for vertical polarization occurs at a grazing angle known as the *Brewster's angle.* When the reflection coefficient is less than unity and/or the phase of the surface reflection is not 180°, the nulls in the radiation pattern due to multipath will not be as deep and the peak value of the various lobes will decrease.

Surface roughness depends on the physical roughness *relative* to the radar wavelength. The lower the radar frequency (longer the wavelength), the smoother a surface will appear to the radar and the more likely the lobing pattern due to multipath will be highly pronounced. For example, the range in the direction of the lowest lobe of a VHF radar that is suitably sited over a sea surface might be increased almost by the theoretical factor of two indicated by the simple flat-earth model. At the higher microwave and millimeter wave frequencies, the less pronounced will be the effects of lobing.

The different values of reflection coefficients shown in Fig. 8.3 for vertical and horizontal polarization can result is different coverage patterns. The nulls with vertical polarization are not as deep and the maxima are not as great as with horizontal polarization. Vertical polarization might be specified when more uniform vertical coverage is desired and horizontal polarization might be preferred when greater range in the direction of the lobes is more important than more uniform coverage. Almost all air-surveillance radars, however, seem to employ horizontal polarization. The greater range at some elevation angles is a benefit that many radar manufacturers take advantage of when advertising the capabilities of their radars. The fact that there also are holes in the "long range" coverage with horizontal polarization is seldom deliberately mentioned.

Rough Surface Reflection Coefficient The theoretical curves of Fig. 8.3 assume a smooth reflecting surface. A smooth surface is sometimes defined by the *Rayleigh roughness criterion* which considers a surface to be smooth if $h \sin \psi < \lambda/8$, where ψ is the grazing angle and h is the difference between the extremes of the surface height. (Some take h to be approximately $4\sigma_h$, and some take it to be $3\sigma_h$, where σ_h is the standard deviation of the gaussian distribution of the surface heights.) A surface is smooth or rough to a radar signal depending on the grazing angle and the physical roughness relative to the radar wavelength.

The surface roughness can affect the reflection coefficient more than the electrical properties of the surface which enter into the reflection coefficients given by Fig. 8.3. Measurements by many workers have shown that the reflection coefficient for normal (nonsmooth) ground terrain is in the range 0.2 to 0.4 and is seldom greater than 0.5 at frequencies above 1500 MHz at low grazing angles.[1] An expression for the reflection coefficient ρ_r of a rough, perfectly conducting surface, such as the sea, was originally given by Ament[2] as

$$\rho_r = \rho_0 \exp\left[-2k^2\overline{h^2} \sin^2 \psi\right] \qquad [8.12]$$

where ρ_0 is the complex reflection coefficient for a smooth surface (Fig. 8.3), $k = 2\pi/\lambda$, λ = wavelength, $\overline{h^2}$ is the mean square surface height, ψ is the grazing angle, and h has a mean value of zero. Experimental data taken over the sea fit the expression of Eq. (8.12) for small values of the *surface roughness parameter* defined as $(\sigma_h \sin \psi)/\lambda$, where σ_h is the rms value of the surface height h. The Ament theory underestimates the experimental

data of Beard[3] when the roughness parameter is greater than 0.1. Miller, Brown, and Vegh[4,5] extended the theoretical analysis of Ament and obtained the following expression for the rough-surface reflection coefficient

$$\rho_r = \rho_0 \exp\left[-2k^2\overline{h^2}\sin^2\psi\right] \cdot I_0(k^2\overline{h^2}\sin^2\psi) \qquad [8.13]$$

where $I_0(z)$ is the modified Bessel function of zero order. Eq. (8.13) is Ament's equation multiplied by the I_0 factor. It fits experimental data for roughness parameters less than 0.3. Figure 8.4 plots Ament's expression along with its modification, Eq. (8.13), and a set of experimental data.

In the above the *specular*, or *coherent*, component of the scattered signal has been discussed. There is also a *diffuse*, or *incoherent*, component of reflection.[6] Its phase and amplitude are random, and scattering occurs over a wider range of angles than does the specular component. It has been reported to increase linearly with increasing surface roughness parameter, level off to a maximum, and then decrease as the inverse square root of the roughness parameter.

The surface-scattered energy that causes lobing of the antenna elevation pattern not only affects the coverage of a radar but it also can introduce serious errors in height-finding methods as well as degrade low-angle tracking (as was discussed in Sec. 4.5).

The effect of the surface-scattered wave on the coverage of the radar is indicated in the radar equation by the propagation factor F^4, where F is defined as

$$F = \frac{E_s}{E_0} \qquad [8.14]$$

where E_s is the field strength of the signal at the target (it includes the effects of the antenna pattern normalized to unity gain), and E_0 is the electric field strength that would occur in free space with loss-free isotropic antennas. It is similar to the parameter η defined by Eq. (8.1) except that F includes the effects of the antenna pattern on the elevation coverage. The propagation factor F was included in the numerator of the radar equation given in Chap. 2 as Eq. (2.61)

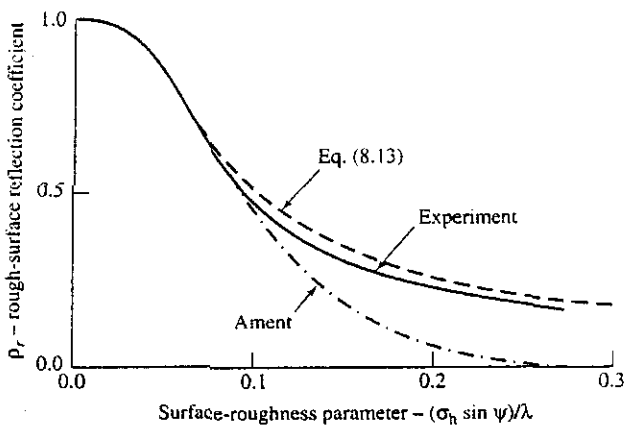


Figure 8.4 Surface reflection coefficient ρ_r as a function of the surface roughness parameter defined as $(\sigma_h \sin \psi)/\lambda$, where σ_h = rms value of the surface height h, ψ = grazing angle, and λ = radar wavelength. Top curve is the theoretical expression given by Eq. (8.13), bottom curve is the original expression given by Ament as in Eq. (8.12), and the middle curve is the experimental data of Beard. (After A. R. Miller and E. Vegh, Naval Research Laboratory Report 8898, July 31, 1985.)

8.3 SCATTERING FROM THE ROUND EARTH'S SURFACE

The use of a flat-earth model is quite suitable for understanding the general nature of the modifications that occur in the antenna coverage. The earth, of course, is not flat, and precise predictions of the effect of the surface on the antenna coverage must consider the round earth. This is especially true for coverage at low elevation angles near the horizon.

The reflection coefficient from a round earth is less than from a flat-earth surface because of the divergence, or widening, of the beam when scattered from a round surface. The so-called *divergence factor* describes the decrease in the scattered signal. The divergence of the beam, however, means that the reflected energy will be spread over a wider angular region than specular scattering from a flat surface. The grazing angle of specular reflection is easy to determine for a perfectly flat surface, but this same angle from a spherical surface is more difficult to compute. In the past it has been found using either approximations or numerical computations; but Miller and Vegh have provided a deterministic method for obtaining the grazing angle from a spherical surface.[7]

There exists in the literature the necessary information, graphs, and nomographs to compute the coverage of a radar when lobing occurs due to the presence of the earth's surface.[8,9] This can be tedious when done by hand, especially when there are many lobes generated in the coverage. Computer programs are available that considerably ease the burden of calculating and plotting the coverages, one of the first of which was by Lamont Blake.[10]

An example of a calculated coverage diagram for an *L*-band radar over sea water is shown in Fig. 8.5 for both horizontal and vertical polarization. At low angles the maximum range with vertical polarization in this example is decreased only slightly relative to the range with horizontal polarization. The effect of the Brewster angle on vertical polarization, however, can be seen in the reduced ranges at higher elevation angles. One of the major consequences of the lobed elevation pattern due to multipath is that tracking of an aircraft flying at a constant altitude will not be continuous. Echoes will be received when the target is in one of the lobes, but the target might not be detected when it is in a null between the lobes. For instance, a target flying at a constant height of 30,000 ft will first be seen by a radar, whose coverage pattern is given by Fig. 8.5a, at a range of 170 nmi. The target will be lost at 160 nmi, reappear at 146 nmi, be lost again at 136 nmi, and so forth. The coming and going of the radar echo can create problems with automatic detection and tracking systems. In such systems, allowance has to be made for the track to coast whenever the target is momentarily lost, rather than immediately drop the track as soon as it leaves the coverage of a lobe and have to initiate a new track when the target echo reappears.

As has been mentioned, the effect of the interference between the direct and surface-scattered waves is to cause the peak of the lowest lobe of the elevation coverage to be at an angle higher than zero degrees, as was seen by Eq. (8.10) for the example of a flat earth. The lowest lobe with a round earth likewise will be at some angle above the horizontal. The result will be a lack of coverage of low-altitude targets. Figure 8.6a is a sketch of the elevation coverage of a long-range enroute air-traffic control radar. The details of the lobing are not shown. (There are two beams, an upper and a lower. We need only be

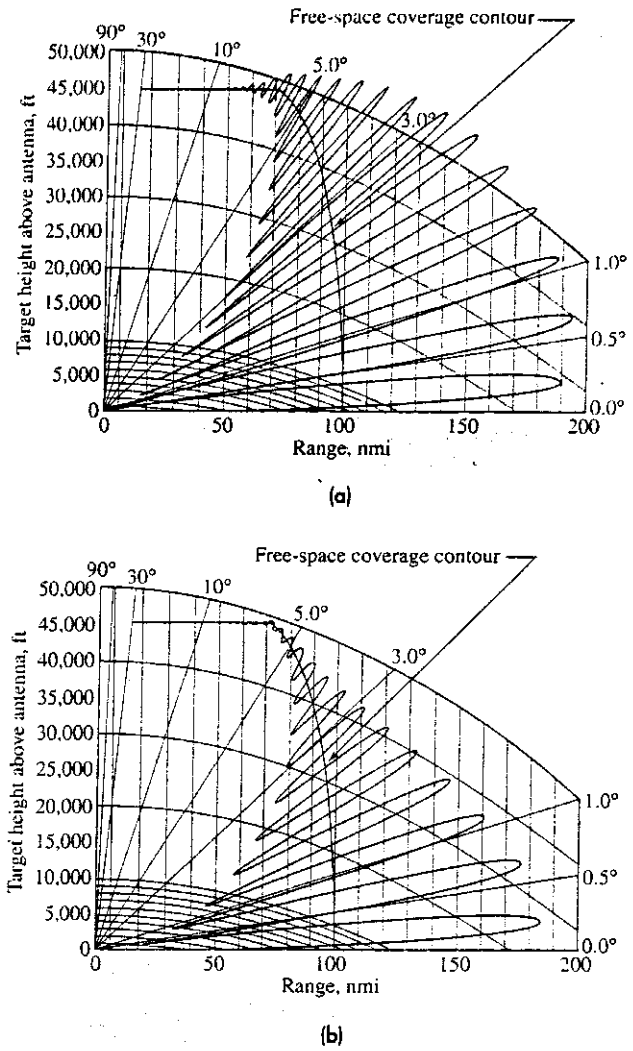
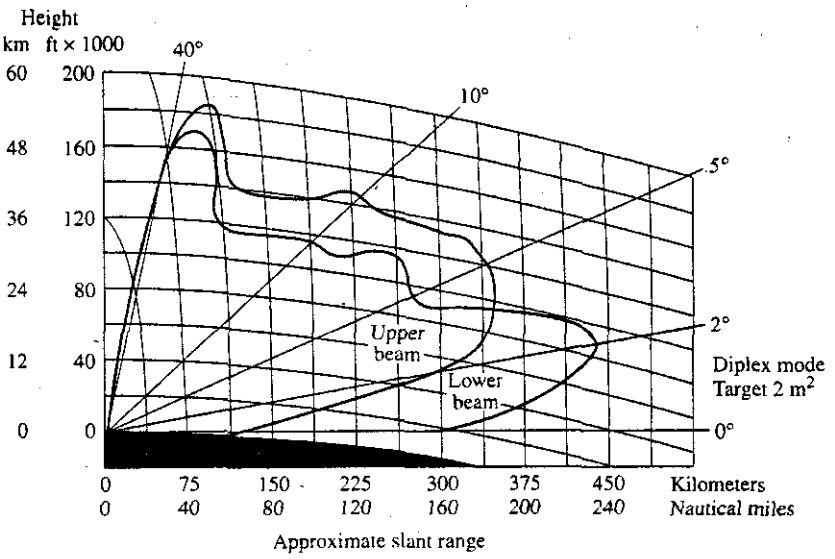


Figure 8.5 Example of a calculated vertical-plane coverage diagram for (a) horizontal polarization and (b) vertical polarization. Frequency = 1300 MHz, antenna height = 50 ft, antenna vertical beamwidth = 12° with the beam maximum pointing on the horizon, a sea surface with 4 ft wave height and free-space radar range of 100 nmi.

concerned with the lower.) The lower beam is tilted so that its half-power point, rather than its maximum, is on the horizon. This makes the lobes (which are not shown) less pronounced, but it also decreases the coverage at low altitudes. The radar is seen to have a maximum range of about 235 nmi. This is a good range for an air-surveillance radar, until it is noted that this range occurs at an altitude in the vicinity of 85,000 ft, which is much higher than commercial aircraft fly. The range for an aircraft flying at 30,000 ft, according to this coverage diagram, is about 185 nmi, and is 165 nmi when the altitude



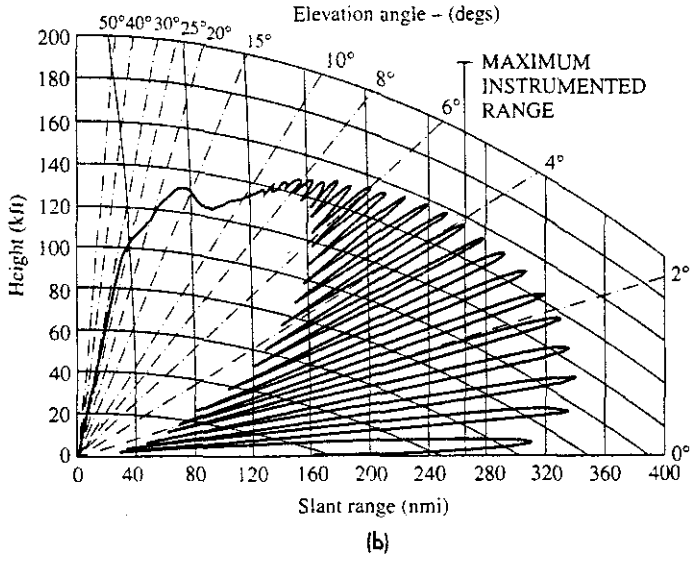


Figure 8.6 (a) Vertical coverage diagram of an *L*-band long-range en-route air-traffic control radar, illustrating the reduced range of the lower beam at very low angles due to the multipath from the earth's surface. (From the brochure *The ARSR-3 Story*, Westinghouse Defense and Electronic System Center, Baltimore, MD. Courtesy of Northrup Grumman Corporation.) (b) Calculated vertical coverage diagram of the Raytheon AN/SPS-49 radar. Long-range mode with antenna rotation rate of 6 rpm, 0.50 probability of detection, 10^{-6} probability of false alarm, Swerling Case 1 target with 1 m^2 radar cross section, sea state 3, and antenna height of 75 ft. This radar operates in the band from 850 to 942 MHz, with an antenna gain of 29 dB and an average transmitter power of 13 kW. (Courtesy of Vilhem Gregers-Hansen and the Raytheon Co.)

is 20,000 ft. Thus the quoted range of a radar might be much greater than the actual range at which an aircraft is first detected when flying at its more usual altitudes. If low-altitude performance at long range is important, the radar should be located on high ground so that its first lobe will be lowered in angle.

Another example of radar coverage, shown in Fig. 8.6b, is that of a long-range shipboard air-surveillance radar known as the AN/SPS-49.

Methods for Minimizing Lobing Effects In many instances, the added range produced by the multipath lobing is a desirable attribute in spite of the nulls which can cause loss of targets. In other cases, lobing might not be desired. This can be true for automatic tracking radars or when the target undertakes maneuvers that require the radar to obtain information at a high data rate. One method to reduce the effects of lobing is to tilt the antenna beam upward so that the radiated energy illuminating the surface is reduced. Thus it is customary in many air-surveillance radars to tilt the beam so that its lower half-power point lies on the horizon.

We have seen that the location of multipath nulls depends on the frequency and the antenna height. By changing the height of the antenna (switching the radar between antennas at different heights) or by changing the radar frequency, the nulls can be filled-in when the radar data from the two or more antennas at different heights are combined. The utility of *height diversity* is limited by the need for antennas at different heights and by the difficulty in obtaining more than two different heights. *Frequency diversity* has been demonstrated to effectively fill in the nulls and allow continuous tracking of the target.[11] This capability, however, requires a wide frequency band. A wide frequency range is highly desired in many applications since it has other advantages than just filling in the nulls.

A radar with *polarization diversity* utilizes two orthogonal polarizations, such as horizontal and vertical. It could, in principle, provide some filling in of the nulls, but it has seldom been used. The nature of the Brewster's angle seems to limit its utility.

A *fence* surrounding the radar antenna can prevent radiated energy from illuminating the ground and can thus reduce lobing. This has seldom been employed since a fence can be large and expensive, diffraction from the edge of the fence limits the amount of attenuation of ground clutter that can be achieved, and there can be a loss of detection of desired low altitude targets.

The land-based military air-surveillance radars used in World War II did not employ doppler processing for separating the unwanted clutter from the doppler-shifted echoes from moving targets. For this reason, when conditions allowed, the radars were sited where there might be natural shielding to attenuate the energy radiated in the direction of the surface. Modern military radars, however, have to see targets at low altitude and would not benefit from this technique.

Not much can be done to enhance the signals from targets that lie below the first lobe. This is why surface-based radars that might have large free-space ranges have significantly reduced detection ranges against targets low on the horizon.

The lobing effect was put to good use in World War II for obtaining the height of an aircraft by using VHF radars that were capable of measuring only azimuth angle and not elevation angle. The range of first detection seen by the lowest lobe was used as an indication of target height. This required good calibration of the radar's sensitivity, and

calibration flights with known aircraft to verify the relation between aircraft height and range of first detection. It also required the enemy to cooperate by not introducing new aircraft whose echo signals might be of quite different strength than those on which the radar was calibrated. In spite of its limitations, this method of height finding did what was required at the time.

8.4 ATMOSPHERIC REFRACTION—STANDARD PROPAGATION

Radar waves travel in straight lines in free space. Propagation in the earth's atmosphere, however, is not in free space. The atmosphere is not uniform; hence, it causes electromagnetic waves to be bent, or refracted. Normally the effect of bending caused by atmospheric refraction is favorable, in that it causes the radar horizon to be extended and increases the coverage of a radar beyond the geometrical horizon, Fig. 8.7a. On the other hand, the bending of the rays by the atmosphere can introduce an error in the measurement of the elevation angle, Fig. 8.7b.

Refractivity Refraction of radar waves in the atmosphere is due to the variation of the velocity of propagation with altitude. The *index of refraction* is a measure of the velocity of propagation and is defined as the velocity in free space divided by the velocity in the medium in question, the atmosphere in this case. (The index of refraction is the square root of the dielectric constant, a parameter that might be more familiar to the electrical engineer.) The difference in velocity of propagation in the atmosphere compared to that

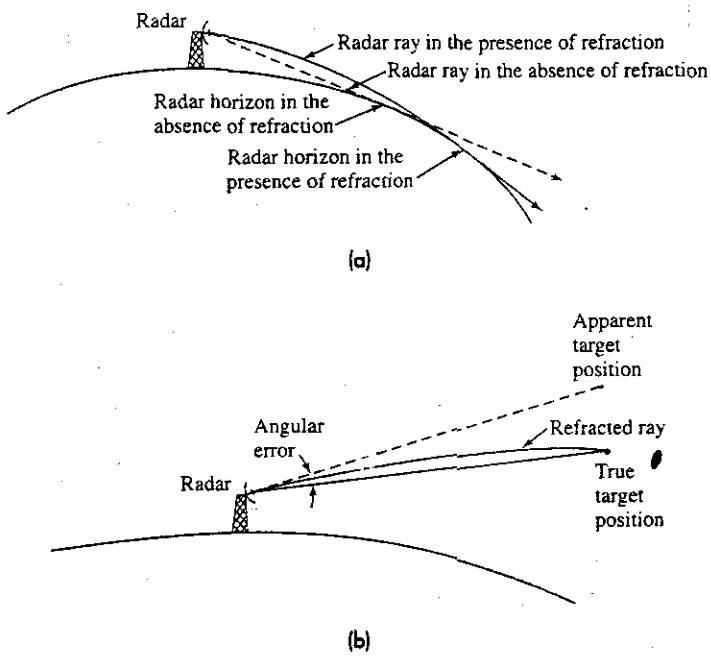


Figure 8.7 (a) Extension of the radar horizon due to refraction of radar rays by the atmosphere; (b) angular error caused by atmospheric refraction.

in free space is very small. According to the International Telecommunications Union, the average value of the surface index of refraction at mid-latitudes is 1.000315.[12] (Hitney,[13] on the other hand, gives 1.000350 as a typical value for the index of refraction at the earth's surface.) Rather than use the index of refraction, n, it is more convenient to use a modified parameter called the *refractivity*, N, which is defined as $N = (n - 1)10^6$. Thus an index of refraction $n = 1.000315$ corresponds to a refractivity $N = 315$.

At microwave frequencies, the refractivity N for air is given by the empirical relation[14,15]

$$N = (n - 1) \cdot 10^6 = \frac{77.6}{T}\left[p + \frac{4810e}{T}\right] \qquad [8.15]$$

where

p = barometric pressure, mbar (1 mm Hg = 1.3332 mbar)
e = partial pressure of water vapor, mbar
T = absolute temperature, K

Atmospheric refractivity depends on the pressure, temperature, and water vapor. Of these, water vapor is the most important at microwave frequencies. It strongly affects the speed of microwave propagation. Temperature variations are more significant than pressure variations. (It might be mentioned that although refractivity is generally not a function of frequency within the microwave region, atmospheric refraction at optical frequencies differs from that at microwave frequencies since it is more dependent on temperature than on water vapor.)

Since the barometric pressure p and the water-vapor content e decrease rapidly with height above the earth's surface, while the temperature decreases slowly, the refractivity normally decreases with increasing altitude. The decrease of N means that the velocity of propagation increases with altitude, causing the radar rays to bend downward. (Refraction of radar waves in the atmosphere is analogous to bending of light rays by an optical prism.) The result is an increase in the radar coverage, as was illustrated in Fig. 8.7a. The magnitude of the atmospheric refractivity at some particular altitude is not as important in determining the effect of refraction on propagation as is the small change of refractivity with height; that is, it is the *gradient* of refractivity that causes the rays to bend.

The major changes in atmospheric refractivity occur in the vertical dimension. There may be changes in the horizontal dimension as well; but these are generally small (especially over water) so that radar propagation can be considered independent of the azimuth direction, unless the radar ranges are very large. The path of radar waves in the atmosphere may be plotted using ray-tracing[16] techniques, provided the variation of refractivity with altitude is known.

Effective Earth's Radius A simple method to account for the effects of atmospheric refraction is to assume that the gradient of the index of refraction is constant with height, at least over the lower part of the atmosphere. This assumption allows the actual earth of radius a (a = 3440 nmi) and its nonuniform atmosphere to be replaced with an earth having a different radius (ka) and a uniform atmosphere in which radar waves propagate in straight

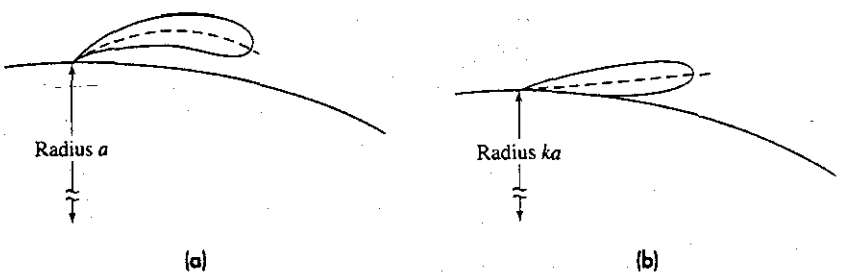


Figure 8.8 (a) Bending of the antenna beam due to refraction by the earth's atmosphere; (b) shape of the beam in the equivalent-earth representation with radius ka.

lines rather than along curved paths, Fig. 8.8. The factor k depends on the refractivity gradient at the surface. From Snell's law in spherical geometry, the value of k by which the earth's radius must be multiplied in order to plot the propagation paths as straight lines is

$$k = \frac{1}{1 + a(dn/dh)} \qquad [8.16]$$

where dn/dh is the rate of change of the refractive index with height. The vertical gradient of the refractive index is normally negative. The gradient of refractivity usually can vary from -79 to 0 N units per km of height.[12,13] The long-term average value of the gradient of N over the United States is approximately -39 N/km. When N is converted to n and substituted into the above equation, we get $k = 4/3$. The use of the $k = 4/3$ effective earth's radius to account for normal atmospheric refraction is convenient and widely used. It is only an approximation, however, and might not yield correct results when precise predictions are required. The term *standard refraction* is sometimes used to signify a value of $k = 4/3$ with the index of refraction decreasing uniformly with altitude with a gradient of $dn/dh = -39 \times 10^{-9}$/m.

The 4/3rd earth radius represents an average and should not be used where precision is important. The correct value of k depends on meteorological conditions and can be found by measurement. Bean[17,18] states that the average value of k measured at an altitude of 1 km varies from 1.25 to 1.45 over the continental United States during the month of February and from 1.25 to 1.90 during August. In general, higher values of k occur in the southern part of the country.

Distance to the Horizon The distance d to the horizon from a radar antenna at a height h may be shown from simple geometrical considerations to be

$$d = \sqrt{2kah} \qquad [8.17a]$$

where ka is the effective earth's radius and the height h above the surface is assumed to be small compared to the real earth's radius a. For a four-thirds earth, this relationship becomes

$$d \text{ (nautical miles)} = 1.23\sqrt{h(\text{ft})} \qquad [8.17b]$$

or,

$$d \text{ (km)} = 4.12\sqrt{h(\text{m})} \qquad [8.17c]$$

Equation (8.17) is often used as a measure of the line-of-sight coverage of a radar. This can lead to optimistic results since the propagation loss at the range d given by Eq. (8.17) can be quite high, as mentioned later in the discussion of diffraction in Sec. 8.6. The actual coverage of a radar can be less than given by the above simple geometric relation because of the large diffraction losses at the horizon. In spite of this, Eq. (8.17) has been widely used. It should be replaced, however, by diffraction calculations when it is important to know the maximum range at which a radar can detect low altitude targets.

Exponential Model of Refractivity A limitation in the use of the effective earth's radius model is that the gradient of refractive index dn/dh is not linear with altitude, but is better approximated by an exponential model, especially in the troposphere above an altitude of 1 km. The exponential decrease of refractivity with height has been given as

$$N = N_s \exp(-h/H_s) \qquad [8.18]$$

where

N_s = refractivity at the surface of the earth
h = height above sea level in km
H_s = scale height in km

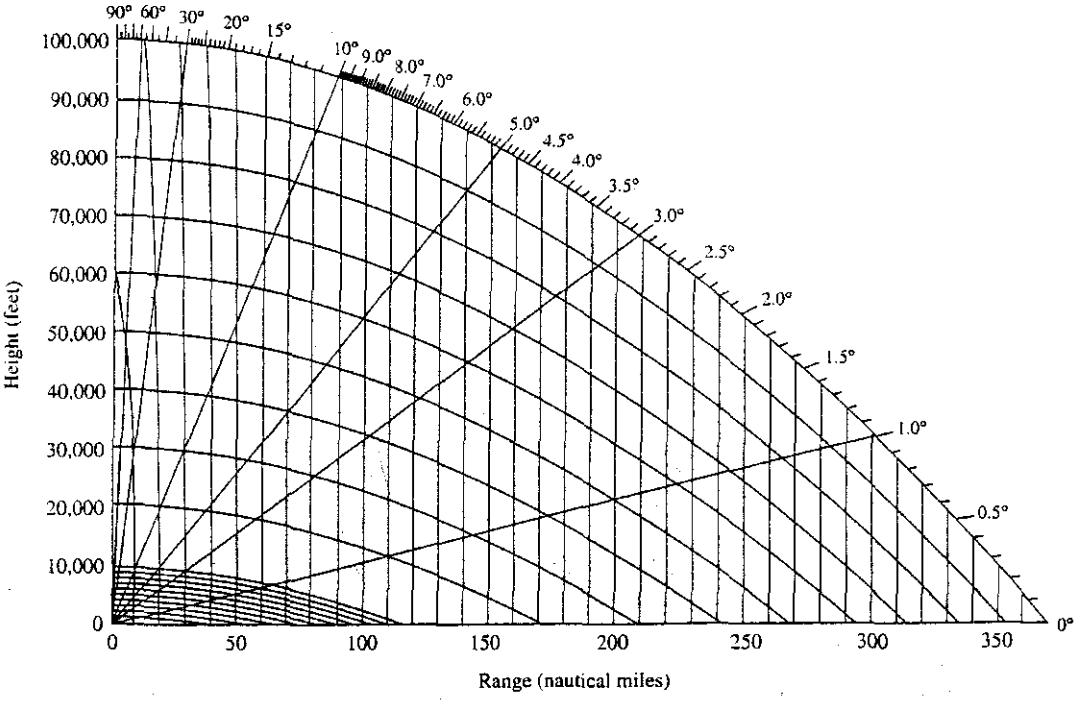


Figure 8.9 Radar range-angle-height diagram used to plot antenna coverage, based on the exponential model of refraction with $N_s = 313$.
(From Lemont Blake, *Radar Handbook*.[20])

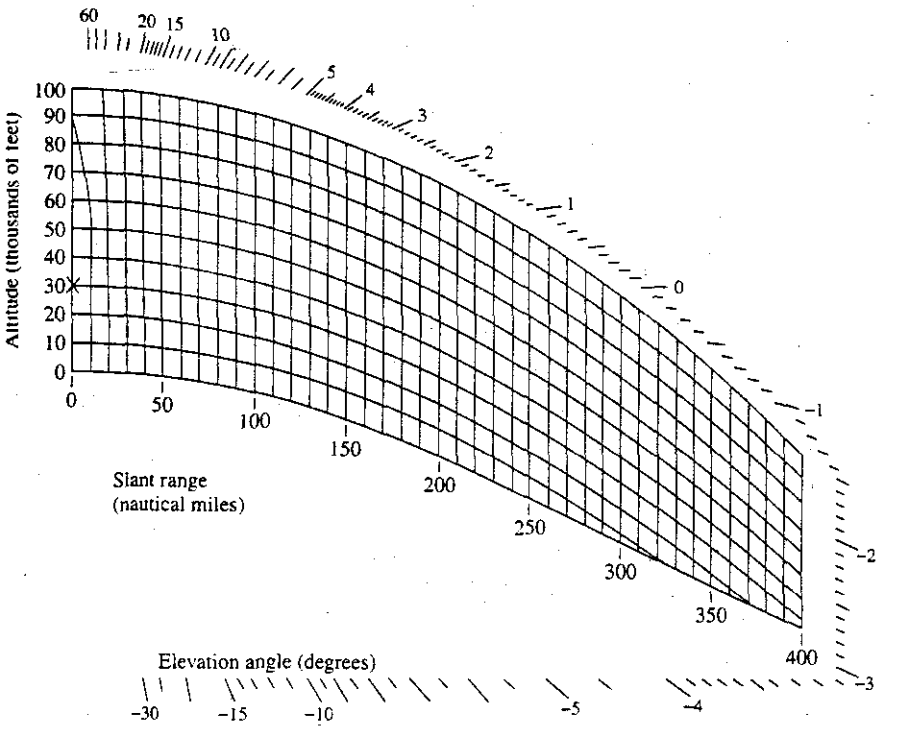


Figure 8.10 Radar range-height-angle diagram when the antenna is at an altitude of 30,000 ft and a U.S. Standard Atmosphere at 45° N latitutde in spring or fall. Due to W. G. Tank.[21]

At mid-latitudes, the average value of N_s has been said to be 315 and the average value of H_s as 7.35 km.[19] Other values have been proposed, however.

An example of a range-height-angle chart (due to Lamont Blake) used to plot antenna coverage patterns based on the exponential refractivity model of Eq. (8.18) is shown in Fig. 8.9.[20] The surface refractivity is $N_s = 313$, and the scale height $H_s = 6.95$ km. In charts such as this, it is customary to plot height in feet and range in nautical miles.

Range-height-angle charts are often based on an antenna at ground level. When the radar antenna is elevated, as when on a tall mountain or when performing as an Airborne Early Warning (AEW) radar, these charts have to be modified. Figure 8.10 is an example due to W. G. Tank[21] for a radar at an altitude of 30,000 ft and a U.S. Standard Atmosphere. This can be thought of as a nomogram for determining the angle of arrival by first selecting the radar altitude (30 kft in this example), and then laying a straight edge from the radar location to the target point. The angle of arrival is read on the angle scales around the edges of the figure. A slightly different approach for plotting the range-height-angle charts when the antenna is not at ground level was suggested by Bauer,[22] based on the CRPL exponential atmosphere.

Standard Atmosphere A *U.S. Standard Atmosphere* is a hypothetical vertical distribution of atmospheric temperature, pressure, and density which by international agreement and for historical reasons, is roughly representative of year-round mid-latitude (45°N) conditions.[23,24] At microwave frequencies, a model of the moisture as a function of altitude

must be added, which results in a refractivity $N = 316 \exp(-Z/26.5)$, for $Z \leq 25$, where Z = altitude in thousands of feet.

Radar Measurement Errors Due to Refraction As was illustrated in Fig. 8.7b, refraction causes the radar rays to bend and results in an apparent target height that is different from the true height. An example of the angular error as a function of height and elevation angle as calculated by Shannon[25] is shown in Fig. 8.11. He assumed $N_s = 313$ and $H_s = 7$ km. At an altitude of 40 kft and an elevation angle of 3° the angle error is 2.67 mrad (17.45 mrad equals 1°).

When precision measurements are required, corrections should be made to the radar data to obtain accurate elevation angle, target height and range.[26] Surface observations often are sufficient for ascertaining the effects of refraction, but there can be cases when the variation of refractivity with height is not simple (as in the case of ducting discussed in the next section). In these cases, the variation of refractivity with height has to be measured and ray-tracing methods used to determine the measurement errors.

Even without ducting or other nonlinear refractivity profiles, errors can be significant. A comparison of heights obtained with the 4/3rd earth-radius model and the exponential model can be found in Brown[27] (and repeated in Murrow[28]). For example, at a range of 100 nmi with $N_s = 315$, the exponential model predicts a height 200 ft greater than the 4/3rd earth-radius model, when the elevation angle is 0.5°. It results in a 500 ft greater height compared to the 4/3rd earth model when the elevation angle is 2.0°. At the same range of 100 nmi and an elevation angle of 2.0°, the exponential model gives a

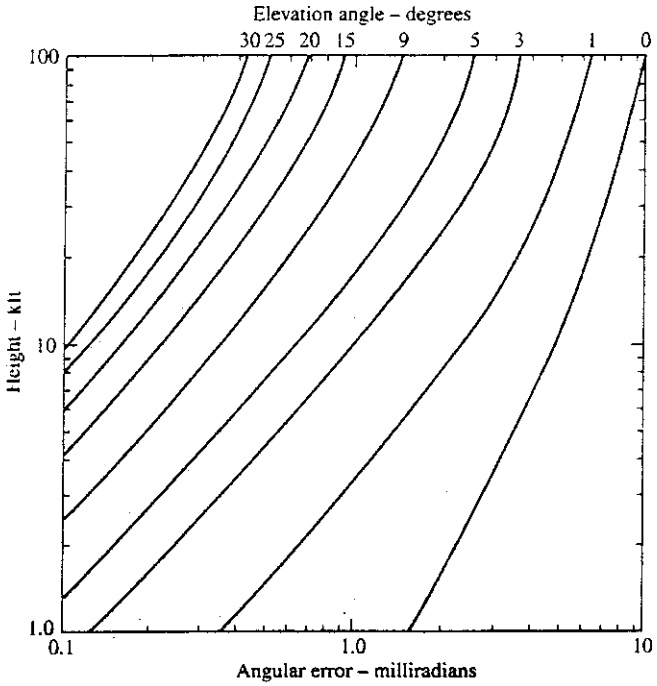


Figure 8.11 Calculated angle error (abscissa) due to atmospheric refraction for a standard atmosphere as a function of elevation angle and target height (ordinate.) (After Shannon.[25])

height error of 2.32 kft when $N_s = 370$ and an error of 1.38 kft when $N_s = 280$. The need for highly accurate knowledge of atmospheric refractivity limits the accuracy with which target height can be determined. This is the reason radar has not been used in the past when precision height measurements are needed.

In measurements to determine how well one can predict the effect of refraction, the accuracy of an AN/FPS-16 tracking radar was checked against photo theodolites at the U. S. Army Electronic Proving Ground, Fort Huachuca, Arizona.[29] The AN/FPS-16 is a highly precise tracker with rms mechanical and electronic measurement errors of ±0.1 mrad for azimuth and elevation angles and ±15 ft for range. At 40 nmi, the three-dimensional rms precision of this radar is ±33 ft. Location measurements were made on aircraft at a range of 40 nmi and elevation angles from 0.7 to 2.8°. It was found that at an elevation angle of 2.5°, the rms elevation error was equal to the inherent ±0.1 mrad precision of the radar. Above 2.5°, the predicted elevation errors were smaller than the inherent precision. Below 2.5°, the angle errors due to refraction were larger than the inherent radar precision.

There is also an additional time delay in propagating within the atmosphere that leads to an error in range. Shannon calculated this for the same conditions as for the angle error in Fig. 8.11. At 40 kft and an elevation angle of 3°, the range error is 97.5 ft. Shannon's plot can be found reproduced in Nathanson.[30]

A correction for the range error due to refraction, suggested by G. Robertshaw,[31] is given by the expression

$$R_c \cong 0.42 + 0.0577\, R_t \left(\frac{N_s}{h}\right)^{1/2} \qquad [8.19]$$

where

R_c = range correction, m
R_t = radar range, km
N_s = surface refractivity in N-units
h = radar altitude in kft

This expression was derived from ray-trace computations performed for radar altitudes of 15 kft, 35 kft, and 65 kft, based on a total of three months of measured atmospheric data over a period of three years at two sites in Germany, one in Saudi Arabia, and one in South Korea. It was said to have "the advantage of great simplicity with a tolerable loss of accuracy" and that "it will provide adequate range correction for airborne radar."

Measurement of Refractivity There are two basic methods for determining the refractivity profile of the atmosphere: one is the radiosonde, or equivalent, and the other is the refractometer.

Radiosonde The atmospheric refractivity profile may be obtained indirectly by the use of the empirical relationship given by Eq. (8.15) which equates refractivity with the properties of the atmosphere. The three measurements of water vapor pressure (humidity), atmospheric pressure, and temperature can be obtained with conventional

weather-observation instruments launched into the atmosphere on a balloon. Such an instrument is called a *radiosonde.* Radiosonde measurements are routinely obtained daily throughout the world. The accuracy of radiosonde weather measurements, however, is generally not as good as might be desired for radar propagation predictions. Furthermore, the temperature, pressure, and humidity are sampled sequentially rather than simultaneously, and the sampling interval for each measurement is approximately 100 m in altitude. This may be satisfactory for weather observations, but it can sometimes cause the radiosonde to miss the sharp changes in refractivity that are characteristic of the strong, but shallow atmospheric layers that give rise to nonstandard propagation conditions.

Helicopter Probes[32] In addition to having less than desired resolution in altitude, the conventional balloon-borne radiosonde makes one vertical pass and obtains only one vertical profile. The wind can cause it to move over a considerable horizontal distance during its flight and its measurements might not be representative of a true vertical profile. Furthermore, the measurements produced by conventional radiosondes launched from ships at sea might be contaminated by the microclimate created by the ship. A better method to probe the lower atmosphere where nonstandard propagation is prevalent is to mount instruments on a helicopter which can simultaneously provide the water vapor, pressure, and temperature as a function of range as well as altitude. A helicopter equipped by the Johns Hopkins University Applied Physics Laboratory (APL) employed more precise meteorological sensors than are normally found with weather radiosondes to obtain in almost real time the atmospheric refractivity. All three meteorological measurements were made at a data rate of 0.5 s. The altitude could be determined with a resolution of 0.3 m from 0 to 1000 m. These measurements were fed to a computer that calculated and displayed a continuously updated plot of modified refractivity versus altitude. The helicopter traveled with an airspeed greater than 30 m/s. Soundings were made with the helicopter ascending or descending at a rate of 1.5 to 3 m/s.

Small Rocket Probes[32] The meteorological measurements needed to determine refractivity profiles can be obtained with a simple, low-cost expendable rocket. It can be used when it is not practical or convenient to use a helicopter. Such a rocketsonde developed and employed by APL carried an instrument package weighing 453 g to an altitude of 150 to 800 m, where it deployed the instruments using a 1 m diameter parachute. Measurements can be made with a vertical resolution of 2 m, which are then telemetered back.

Refractometer An alternative to using meteorological measurements to indirectly derive the atmospheric refractivity is to use the more accurate and more responsive *microwave refractometer.* This instrument directly measures the index of refraction (or its square, the dielectric constant) by comparing the resonant frequencies of two identical microwave cavities. The resonant frequency of a cavity depends on its dimensions and its contents. One cavity is vented so as to sample the atmosphere; the other is hermetically sealed and acts as the reference. The cavities are fed by the same microwave source that is swept in frequency. The difference between the resonant frequencies of the sampled and reference cavities is a measure of the index of refraction. The microwave refractometer has much greater accuracy than the indirect measurement based on Eq. (8.15). It can

measure changes in refractivity of less than 0.1 N unit; and when used to determine variations about an undetermined mean, it can have a time constant that allows detection rates of up to 100 Hz.[33] Although the refractometer may provide excellent refractivity measurements, it might not be suitable for all applications since it requires an aircraft or helicopter. It is usually too costly to be expendable, as is a radiosonde or a rocketsonde.

8.5 NONSTANDARD PROPAGATION

The previous section described the effect on radar propagation of standard, or normal, refractive conditions. Refractive effects, however, can be much more complex than described by the standard exponential model, and can cause significant changes in radar propagation. Such conditions are known as *anomalous*, or *nonstandard, propagation.* As a rough generalization, when nonstandard propagation conditions occur, the maximum ranges of a surface radar for detecting low-altitude or surface targets might be extended from two to five times what would be expected with a uniform atmosphere.[34]

Normal refraction occurs when the refractive gradient with height, dN/dh, is between 0 and -79 N units per km of height. (Note that we have said previously that the long-term mean gradient over the continental United States is about -39 N/km.) When the gradient equals -157 N/km, the effective earth's radius as given by Eq. (8.16) becomes infinite. Rays that are initially horizontal will then follow the curvature of the earth. Under such conditions, the radar range is significantly increased and detection beyond the radar horizon can result. Refractive gradients between -79 and -157 N/km result in what is called *superrefraction.* When the gradients exceed -157 N/km, the curvature of the propagating ray exceeds the curvature of the earth and ducts can form that trap the radar energy. The trapped energy within the duct can propagate to ranges well beyond the normal horizon.

If the refractive gradient were to increase with height, instead of the more usual decrease, the propagating rays would curve upward and the radar range would decrease as compared to normal conditions. This is called *subrefraction.* Its occurrence is rare; but when it does occur, its effect can be serious. It has been suspected of causing ship accidents when using marine radar. The term nonstandard propagation, or anomalous propagation, applies to any of the above propagation conditions other than normal. Table 8.1 summarizes these refractive conditions.

There are three general classes of ducts that will be briefly described. These are evaporation ducts, which occur at the surface of the sea; surface-based ducts; and elevated ducts. The later two can occur over land as well as water. To propagate energy within the duct, the angle the ray makes with the duct should be small, usually less than about one-half degree. Therefore, only those rays launched nearly parallel to the duct are trapped.

Evaporation Ducts In a maritime environment, standard refractive conditions seldom appear, so that nonstandard conditions of some form are often present.[35] The most common type of anomalous propagation over the ocean and other large bodies of water is the evaporation duct. It is found at the surface, and is a relatively common occurrence. The air in

Table 8.1 Summary of Refractive Propagation Conditions

Refractive Condition	Gradient: N units per km
Subrefraction	Positive gradient
No refraction (uniform atmosphere)	0
Standard refraction (4/3rd earth radius)	−39
Normal refraction	0 to −79
Superrefraction	−79 to −157
Trapping, or ducting	−157 to $-\infty$

contact with the sea surface is usually saturated with water vapor so that its relative humidity is almost 100 percent. The air several meters above the sea surface is not usually saturated so there will be a decrease in humidity from the surface value to the ambient value determined by the general meteorological conditions well above the surface. The rapid decrease of water vapor causes a rapid decrease of refractivity that results in the formation of a low-lying duct that traps the radar energy so that it propagates close to the sea surface. Ducting can cause the radar ranges for targets at or near the sea surface to be considerably greater than the free-space range.

Duct Height The "height" used to characterize an evaporation duct is not the height below which an antenna must be located in order to obtain extended propagation. It is more a measure of the strength of the duct. Evaporation duct heights, which typically might have values from 6 to 30 m, vary with the geographic location, season, time of day, and wind speed. Of the factors that can affect the strength of a duct, the wind speed seems to be of special importance. In one set of experimental observations conducted in the Atlantic trade wind region off the east coast of Antigua with X- and S-band radars, it was said that the wind was the only meteorological factor that was correlated with the rate of attenuation in the duct.[36] Wind speeds from 8 to 15 kt produced a moderately strong duct of low height. Winds from 20 to 30 kt produced a greater duct height, but in this case the duct was weaker according to the meteorological predictions. The attenuation in the duct, however, decreased with increasing wind. Passing squalls and rain showers did not affect the duct or decrease the propagated signal strength. The duct heights varied from 20 to 50 ft and were found to exist all the time during these experiments.

Evaporation duct heights cannot be readily determined by standard radiosondes or refractometers. They can be inferred, however, from theoretical models based on meteorological measurements. One such model due to H. Jeske[37,38] and modified by Paulus[39] utilizes the sea-surface temperature, air temperature, relative humidity, and wind speed—the last three being obtained at some reference altitude (often taken to be 6 m).[38,40] Table 8.2 gives the calculated average duct height for various areas of the world, based on this formulation using meteorological measurements from a 15-year subset of data from the National Climatic Data Center.[41] The histograms for evaporation-duct heights for three of

the areas in Table 8.2, plus the worldwide average are shown in Fig. 8.12. (These were originally given in Ref. 41 in 2-m increments of height, but here a smooth curve was drawn.)

The predicted evaporation-duct height using the Jeske-Paulus model does not always agree with actual observations of radar propagation. This might be due, in part, to the difficulty in making the relevant meteorological observations at sea with sufficient accuracy. Also, the theory and the accompanying simplifying assumptions on which the predictions are based might not be applicable under all conditions, especially when large duct heights are predicted. The nature of an evaporation duct can be more complex and varying than are accounted for in a simple model.

An improved method of determining the evaporation duct height has been proposed by Babin, Young, and Carton, which they call *Model A*.[42] They take advantage of the availability of high-speed desktop computers, not available when the Jeske-Paulus model was first formulated, to eliminate some of the assumptions that were originally made, incorporate more atmospheric boundary-layer physics, and decrease the use of empirical relationships. This model not only provides a more accurate duct height, but it also gives the standard deviation of duct heights based on the accuracy of the sensors used to obtain the atmospheric data.

Frequency Dependence The thicker the duct, the lower the frequency that can be propagated; but there is a limit. The lower frequency limit for propagation within an evaporation duct is said to be about 3 GHz.[13] An approximate model for propagation in an evaporation duct is that of a waveguide with the sea as the bottom wall and a leaky top wall. Thus there is a low-frequency cutoff for propagation in this type of leaky waveguide.

According to Hitney,[13] a rough guide to the lowest frequency that can be trapped by a duct of a given height is as follows: 3 GHz requires a duct height of at least 25 m; 7 GHz, 14 m; 10 GHz, 10 m; and 18 GHz, 6 m. As the frequency increases, however, the propagation loss increases due to attenuation caused by the high concentration of water

Table 8.2 Calculated average height of the evaporation duct[41]

Area	Average Duct Height, m
North Atlantic	5.3
Canadian Atlantic	5.8
East Atlantic	7.4
North Pacific	7.8
Mediterranean	11.8
West Atlantic	14.1
Persian Gulf	14.7
Indian Ocean	15.9
Tropics	15.9
World Average	13.1

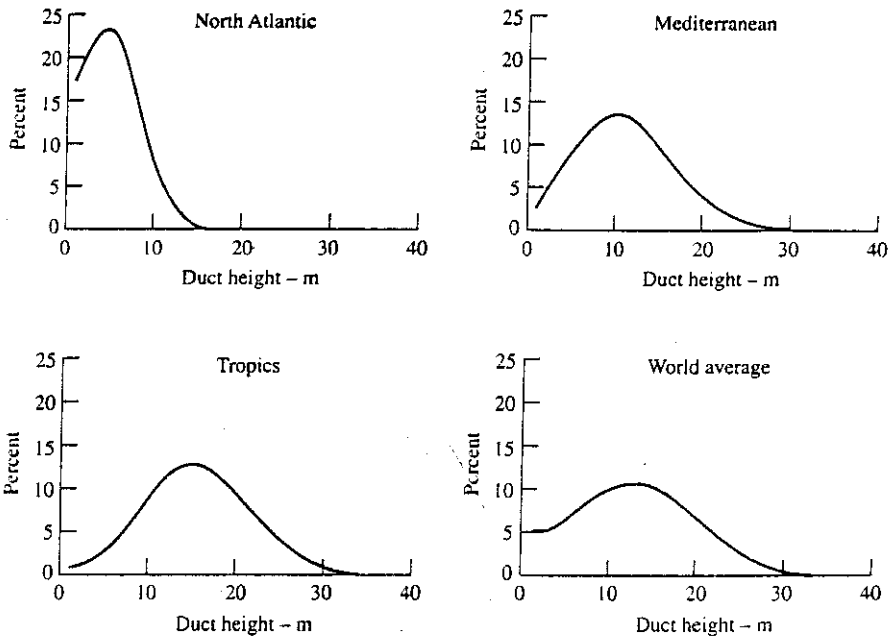


Figure 8.12
Examples of the statistics of evaporation-duct height for three areas of the world and the worldwide average, as given by Hitney et al.[41] These were calculated based on historical average meteorological measurements.

vapor within the duct as well as the attenuation due to a rough sea surface. For this reason Hitney suggests that the optimum frequency for duct propagation is around 18 GHz. Significant ducting has been observed, however, at frequencies as high as 94 GHz. Experiments at 94 GHz conducted over a 40.6 km over-the-horizon path along the Southern California coast found that the median loss with ducted propagation was 60 dB less than the loss that would have been obtained if there were no ducting (as with a standard atmosphere).[43]

Multiple-Mode Propagation If the duct height is large enough, more than one mode can be propagated. (A *mode* is a configuration of electric and magnetic field distribution, similar to the modes of propagation in a conventional waveguide.) Multiple propagation modes have two consequences: (1) the signal strength will not vary uniformly throughout the duct and (2) there can be more than one antenna height suitable for low-loss propagation. In normal propagation within a uniform or a standard atmosphere, the propagation loss decreases with increasing antenna height. The higher the antenna the better. On the other hand, with a single mode of propagation in an evaporation duct, the loss will increase with height within the duct. Propagation is better with a low-sited antenna. With a duct height that supports multiple modes of propagation, there can be more than one choice of antenna height that provides low-loss propagation. In one particular *X*-band experiment, the minimum attenuation occurred with an antenna height of 2 m.[36] Increasing the antenna height increased the loss until a maximum loss was found at about 10 m. Further increases of height decreased the attenuation again until a secondary minimum was

obtained at 20 m height, after which the attenuation increased (at least to a height of 30 m). On a ship, because of the pounding of the waves, it might not be practical to site a radar antenna 2 m over the sea if it were desired to obtain the benefits of ducted propagation. Instead, in this particular case, the antenna could be located at 20 m height, with a slightly greater loss being the price to be paid for the more convenient antenna location. These values apply to a particular location at a particular time. On most ships one probably would not want to have a variable antenna height to maximize propagation in the duct. What this example illustrates is that a fixed high-sited antenna might be a suitable compromise, rather than try to place an antenna in the duct at a low height where it can be subjected to destruction by the force of the waves.

Theoretical models of ducted propagation confirm the general nature of the above experimental observations.[44] Each mode of propagation of a guided wave has a particular configuration of electric field strength. When multiple modes are present, the fields of each mode can interfere constructively or destructively with the others. Consequently, the field strength along the duct or across the duct might not be uniform, which can lead to variations in the radar echo signal strength. When multiple modes of propagation are present in the duct, theory predicts considerable variation of the signal strength, called *fades*, because of the interference among the several modes. Fades can be of the order of 20 dB. In one example, it was shown that fading was strongly dependent on the height of the radar and the target, and for centimeter wavelengths the fading occurred at intervals of from two to three miles. Because of the strong dependence with height, the negative effect of holes in the coverage might be reduced by having more than one antenna, each at a different height, to provide height diversity. Frequency diversity might also reduce the effects of the holes in the coverage, if the frequencies are widely separated.

Ducting Within or Near the Horizon Most of the above has been concerned with propagation in a duct beyond the normal horizon. Within the horizon, the refractive effects of the duct can lead to a modification of the normal lobing pattern caused by the interference of the direct and surface-reflected rays discussed in Sec. 8.2.[45] The relative phase between the direct and surface-reflected rays can be different in the presence of the duct, and focusing[46] can change the relative amplitudes of the two components. Focusing by the atmosphere might even cause the amplitude of the surface-reflected ray to sometimes exceed that of the direct ray. The effect of the duct on line-of-sight propagation is to reduce the angle of the lowest lobe, bringing it closer to the horizon.

Although the evaporation duct provides a significant increase in signal strength for ranges well beyond the horizon, compared to what would be received if there were no ducting, it results in reduced signal strength at or near the horizon. This is illustrated in Fig. 8.13 which shows the propagation loss as a function of range for an *X*-band radar (where propagation loss is defined here as the ratio of transmitted to received power assuming that the antenna pattern is normalized to unity gain).[47] The presence of a duct increases the signal at the longer range compared to no duct, but it decreases the signal at the first lobe (about 10 dB for the 16 m duct) and it shifts the first null to longer range than when no duct is present. This confirms what was said in the previous paragraph about the evaporation duct lowering the lowest lobe.

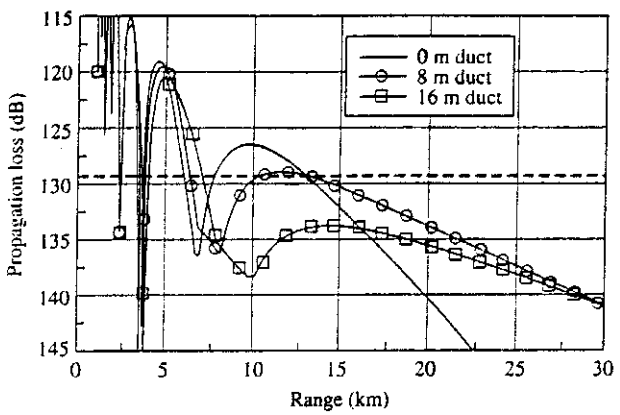


Figure 8.13 Propagation loss between radar and target as a function of range for an X-band radar located 23.5 m above the sea with a point target 4.9 m above the sea. The radar has a free-space range of 7.5 km. The dashed horizontal line represents the detection threshold for this radar. The "0 m duct" corresponds to a standard atmosphere. (From Anderson,[47] Copyright 1995 IEEE.)

Surface-Based Ducts[48,49] A surface-based duct is one whose base is at the earth's surface. There are three types of such ducts depending on the relationship of the trapping layer to the earth's surface. One example is the evaporation duct discussed in the above. The evaporation duct, however, is usually considered separately (as was done here) because of its unique characteristics and because it is a nearly permanent worldwide feature.[38] The second is a surface duct created from an elevated trapping layer, and the third is a surface duct created from a surface-based trapping layer. In this subsection, only the latter two will be considered, since the first has been discussed.

Surface-based ducts are formed when the upper air is exceptionally warm and dry compared with the air at the surface. There are several meteorological conditions that may lead to their formation. Over land, a surface-based duct can be caused by the radiation of heat from the earth on clear nights, especially in the summer when the ground is moist. The earth loses heat and its surface temperature falls, but there is little or no change in the temperature of the upper atmosphere.[50] This leads to a temperature inversion at the ground and a sharp decrease in moisture with height. Thus over land, ducting is most noticeable at night and usually disappears during the warmest part of the day.

Another cause of surface-based ducts is the movement (advection) of warm dry air, from land, over cooler bodies of water. Examples of such advections are the Santa Ana wind of Southern California, the sirocco of southern Mediterranean, and the shamal of the Persian Gulf.[38] Warm dry air that is blown out over the cooler sea is cooled at its lowest layers to produce a temperature inversion. At the same time, moisture from the sea is added to produce a moisture gradient, and a surface-based duct can be formed. This type of ducting tends to be on the leeward side of land masses, it occurs either during the day or night (but more likely to occur in the late afternoon or evening when the warm afternoon air drifts out over the sea), it might extend out over the ocean for several hundreds of kilometers, and it can last for long periods of time (several days).

The height of surface-based ducts generally does not exceed a few hundred meters. Propagation within a surface-based duct is relatively insensitive to frequency, and long-range propagation can occur at frequencies exceeding 100 MHz. Ducted propagation has

even been reported for frequencies down to 20 MHz.[51] Most of the reports of unusual long-range radar detection are due to this form of duct. A classic example is the often cited detection in the Indian Ocean during World War II by a 200 MHz radar located at Bombay, India. This radar, located 225 ft above sea level, frequently detected echoes from points in Arabia at ranges from 1000 to 1500 miles during the hot season. It was quite common to plot ships out to ranges of 200 miles, and cases were reported to 700 miles. By contrast, during the monsoon season when propagation conditions were more normal, the radar was able to plot ships out to a range of only about 20 miles.

An example of the effects of surface-based ducts is shown in Fig. 8.14. This is a plot of the two-way propagation factor (ordinate) seen in the vertical plane at a range of 50 nmi. The abscissa is the height above sea level. The radar frequency is 900 MHz and antenna height is 85 ft. The dashed curve represents the calculated propagation factor for a standard atmosphere. (According to Eq. (8.10) the peak of the lowest lobe due to multipath is predicted to be at 0.18°, and the peak of the second lobe is at 0.55°.) The solid curve was calculated for a nonstandard atmosphere whose refractivity profile was measured in the Persian Gulf at local noon on a day in August. The duct height was of the order of 700 to 800 ft. The signal strength below the duct height is considerably increased compared to the no-ducting standard-atmosphere situation. The increase in energy at the lower altitudes is accompanied by a decrease in energy in the lowest lobe below about 0.5°. The maximum of the two-way radar signal in the lowest lobe is seen to decrease by about 10 dB with ducting present.[52]

Surface-based ducts may also be formed by the diverging downdraft of air under a thunderstorm.[50] The relatively cool air which spreads out from the base of a thunderstorm results in a temperature inversion in the lowest few thousand feet. The moisture gradient is also appropriate for the formation of a duct. Although duct formation by a thunderstorm might not be as frequent as other ducting mechanisms, it may be used as a means of detecting the presence of a storm. An operator carefully observing a radar display can detect the storm by the sudden increase in the number and range of ground echoes. The conditions appropriate for the formation of a thunderstorm duct might have a duration of from one-half hour to a few hours.

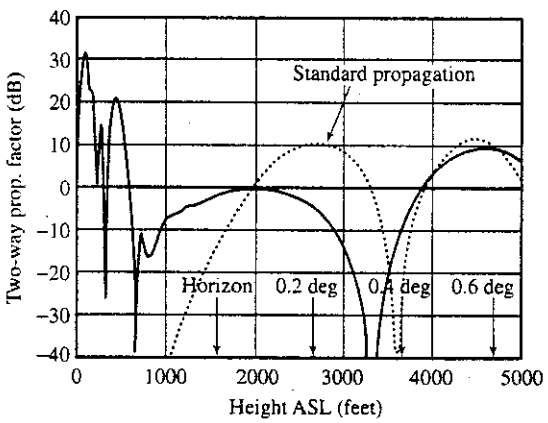


Figure 8.14 The solid curve is the two-way propagation factor as a function of height above sea level (ASL) for a surface-based duct taken in the Middle East at a range of 50 nmi measured at local noon on a day in August. The dashed curve represents the propagation factor calculated for a standard atmosphere.
(Courtesy of John Walters and Vilhelm Gregers-Hansen, NRL Radar Division.)

Elevated Ducts[38] The base of an elevated duct lies above the surface of the earth. Elevated ducts can occur in the tradewind regions between the mid-ocean high-pressure cells and the equator. Semipermanent high-pressure areas are centered at approximately 30° north and 30° south latitudes over the ocean. The region between these high-pressure areas and the equator are called the tradewinds. Two such tradewind areas that have been studied lie between Brazil and the Ascension Islands[53] and between Southern California and Hawaii.[54] Within the high-pressure regions, there is a slow-sinking of high-altitude air (called large-scale *subsidence*) which meets low-level maritime air (marine boundary layer) flowing toward the equator. The general sinking of air from high altitudes causes compression which results in adiabatic heating and a decrease in moisture content. This leads to warmer, drier air lying above cooler, moist air to produce a temperature inversion (an increase in temperature that causes a decrease in refractivity with height). The result is a strong duct along the interface of the temperature inversion along the top of the marine boundary layer.

In some cases, a stratus cloud layer will form at the base of the temperature inversion, and the duct altitude can be identified by the height of the cloud tops that are suppressed by the temperature inversion.[55] When the temperature inversion occurs below the altitude at which clouds are formed, a haze layer in the air below the temperature inversion can be observed.

The altitude of the tradewind duct varies from hundreds of meters at the eastern part of the tropical oceans to thousands of meters at the western end. Thus, the height rises gradually in going from east to west. Over the coast of Southern California, elevated ducts were found to occur an average of 40 percent of the time and have an average maximum elevation of about 600 m. On the western side of the Pacific Ocean along the coast of Japan, elevated ducts are said to occur 10 percent of the time and have an average maximum height of 1500 m.[38] The thickness of elevated ducts can range from near zero to several hundred meters. Frequencies as low as 100 MHz can propagate in the thicker elevated ducts. Since the elevated duct is due to meteorological effects, there can be seasonal as well as diurnal variations. It has been said, however, that elevated ducts give rise to strong persistent anomalous propagation throughout most of the year over at least one third of the oceans.

Ray-optics theory indicates that both the radar antenna and the target must be within the duct to obtain the benefits of the low propagation loss provided by the duct. In practice, it has been observed that this is not always necessary since enhanced propagation can occur with the radar and/or the target outside the duct (as defined by its classical duct thickness). This is likely due to the oversimplification of the duct model when defined by a smooth surface. Both the upper and lower boundaries of a duct can be irregular, allowing energy to "leak" or scatter into or out of the duct. It has been reported[56] that the presence of a very strong evaporation duct (height less than 100 ft) along with an elevated duct (at 2000 ft) can result in a significant increase in signal strength of a 3 GHz signal beyond the horizon at heights (3000 ft) which are an order of magnitude or more greater than the evaporation duct height.

The capability of elevated ducts to propagate to long range in the tradewind region is illustrated by a particular flight of an instrumented Naval Research Laboratory aircraft from San Diego, California, to Oahu, Hawaii. These were communications, rather than

radar, experiments. A 220-MHz signal was transmitted from an antenna located at an elevation of about 800 ft near San Diego. The transmitting antenna was within the duct. The signal was received by an aircraft throughout the entire path and was even detected after the aircraft was on the ground in Hawaii.[54] This long range, however, occurred only once during the fourteen runs conducted during these experiments. In all these runs, significant increases in range were achieved compared to what would be expected from tropospheric scatter propagation.

Meteorological conditions necessary for an elevated duct are similar to those for a surface-based duct. Under the proper conditions, one can turn into the other.[38]

Although enhanced propagation can occur when the target and the radar are properly located with respect to the duct, it is also possible to obtain reduced or no coverage above or below the duct, compared to that expected with a standard atmosphere. This lack of coverage due to the duct is called a *radar hole.*

Subrefraction The gradient of refractivity may, at times, be such as to bend electromagnetic rays upward rather than downward, causing a decrease in range when compared with standard refractive conditions.[57] This is called *subrefraction,* or *substandard propagation.* It occurs when the index of refraction increases with altitude, instead of decreasing as is the more usual situation. Subrefraction can occur when warm, moist air flows over a cool ocean surface or a over a cooler air mass just above the ocean surface.

An interesting example of subrefraction on radar performance has been reported by Brookner et al.[58] for an S-band marine radar located near the entrance to the Delaware Bay between the States of Delaware and New Jersey. The radar was operated by the Pilots Association of the Bay & River Delaware. Subrefraction was said to occur throughout the year in the Delaware Bay region. The radar typically had a range of 20 nmi for ships of interest, but when subrefraction occurred the range was reduced in half to 10 nmi. A one-way loss was observed that was greater than 20 dB relative to free-space propagation. Reduced conditions could last for several hours before returning to normal. Subrefraction results in an effective earth's radius less than one, and can be one half of the actual earth's radius. Some subrefraction conditions can produce a skip zone. For example, an approaching ship might be first detected at 20 nmi range, be undetected starting at about 12 nmi, and not seen again until it is within 6 to 8 nmi of the radar. It has been said that operating the radar antenna at a higher height above the surface can reduce the effects of subrefraction, but frequency diversity has little effect.

In some cases fog can lead to subrefraction. When fog forms, part of the water in the air changes from the gaseous to the liquid state, but the total amount of water remains unchanged. Water in liquid form contributes far less to the refractive index than water in the gaseous (vapor) form. The formation of fog near the surface results in a reduction of water vapor and a corresponding lowering of refractivity in the region of fog. The result is an upward bending of the radar rays, and a shortening of the radar range. Because other factors can enter, the presence of fog is neither a necessary nor a sufficient condition for the occurrence of substandard propagation.

It was mentioned in both references 57 and 58 that under some subrefraction conditions (and in the absence of fog), a ship that is not seen by radar might be seen visually.

This results because water vapor makes a significant contribution to the atmospheric index of refraction at microwave frequencies, but has little effect at optical frequencies.

Nonstandard Propagation over Land Much of the discussion about nonstandard propagation in this section has been based on overwater paths. Similar effects can occur over land; but there has been less written about overland effects than the effects that occur over water.

Over land, nonstandard propagation can be caused by radiation of heat from the earth on clear nights, especially in summer when the ground is moist. When the earth loses heat, its surface temperature falls, but there is little or no change in the temperature of the upper atmosphere. This leads to conditions favorable to superrefraction; that is, a temperature inversion at the ground and a sharp decrease of moisture with height. Thus over land masses, the phenomena of superrefraction and ducting are most noticeable at night and disappear during the warmest part of the day. Ducting observed over the Arizona desert during the winter when the atmosphere was clear and dry (low moisture content) has been attributed to this type of temperature inversion caused by nocturnal cooling of the ground.[59] In these experiments, the duct increased in height and intensity as the night progressed. It was also stated that "The cyclic variation in the meteorological conditions occurs with the same general character night after night and, in fact, year after year" and that "The field-strength variations reflect this consistency."

Terrain generally consists of high and low regions. Diffraction (forward scattering) from the high regions can have a significant effect on radar propagation and sometimes can be the dominant propagation mechanism.[13] In other cases, ducting may be dominant, but the profile of the terrain along the propagation path might reduce the strength of the duct.

The computation of radar propagation over irregular land surfaces is much more difficult than computations of propagation over the sea. The variation of refractivity with range must be taken into account when radar waves propagate over irregular terrain. The irregular features, including trees and other structures, are not easy to model realistically. The mathematical techniques must include diffraction as well as refraction effects. One approach has been to map the irregular terrain boundary into a rectangular domain where a numerical solution can be generated by applying well-established numerical methods.[60] The parabolic equation method, mentioned later in this section, is one of the computation techniques that have been applied to predict propagation over land since it can account for the variation of refractivity with range.

Effect of Ducting on Surface Clutter Measurements Most clutter measurements do not take account of the effects of nonstandard propagation. Ducting effects, when they exist, are usually inherent in the clutter data since it is difficult to separate them. Ducting can be, therefore, a major source of inaccuracy when trying to make quantitative measurements of the radar cross section per unit area (σ^0) of surface clutter.

Another reason why measurements of clutter are suspect when ducted propagation exists is that the precise grazing angle of the radar ray is not known. Even if the grazing angle were known, it is not easy to accurately determine the attenuation caused by ducted propagation (which is needed to determine the radar cross section of clutter).

Unfortunately, most of the available clutter data does not account for nonstandard propagation conditions. If the refractivity conditions of the atmosphere were accurately known (as with extensive measurements), it is possible in principle to extract the value of grazing angle* and the clutter cross section during ducted propagation.[61,62] This, however, is not easy to do in practice.

In the curves of sea clutter data shown in Fig. 7.13, the value of sigma zero (clutter cross section per unit area) for surface-based *X*-band radars did not drop off sharply at low gazing angles, as might usually be expected due to multipath from the surface. On the other hand, it was found that the *X*-band clutter echo did drop off with decreasing grazing angle, as expected when the radar was in an aircraft. It is suggested that this behavior might be due to the surface-based radars experiencing ducted propagation while the airborne radars flying above the evaporation duct did not.

It has also been observed in some cases with large radars looking at low angles over the ocean under ducted conditions, that echoes at long range from atmospheric clear-air turbulence can be much larger than the echoes from sea clutter.[63]

Occurrence of Ducting Ducting is essentially a fine-weather phenomenon (with the exception of thunderstorm ducts). Since tropical climates, other than at the equator, are noted for their fine weather, it is not surprising to find the most intense ducting occurring in such regions.[64] In temperate climates, ducting is more common in summer than in winter. It does not occur when the atmosphere is well mixed, a condition generally accompanying poor weather. When it is cold, stormy, rainy, or cloudy, the lower atmosphere is well stirred up and propagation is likely to be normal. Both rough terrain and high winds tend to increase atmospheric mixing, reducing the occurrence of ducting. Although windy weather which causes the atmosphere to be well mixed can inhibit the formation of ducts, experiments in the Atlantic trade wind region indicated that the wind was the most important meteorological factor required for the appearance of an evaporation duct.[36] Thus, evaporation ducts might be weak or even not exist if there were no wind at all.

Consequences of Ducted Propagation on Radar Performance Ducts can provide extended ranges against surface targets or low-flying aircraft considerably beyond the ranges that would be expected from a radar within a standard atmosphere. As previously mentioned, the radar antenna and the target must be in, or near to, the duct to experience extended range. Although ducts can significantly increase the range of a radar, the consequences of ducted propagation are not necessarily good. In fact, in most cases the negatives tend to outweigh the positives.

The ability to see at long ranges because of ducted propagation cannot be readily predicted in advance, and the increased ranges cannot be relied upon since they are not always present. One would certainly not want to depend on ducted propagation to extend the range of a radar when the proper ducting conditions for long range might not be available when needed. Furthermore, increased ranges in some directions are balanced by decreased ranges in other directions. There can be significant "radar holes" which prevent

*The grazing angle for sea clutter that can be found in this manner is the angle to the undisturbed sea surface. In practice, the angle the radar ray makes with the sea will depend on the slope of the large water waves on which ride the shorter waves as well as other surface disturbances.

a radar from seeing targets in some direction that would normally be detected if surface-based or elevated ducts were not present. The loss of detectability caused by radar holes can affect airborne radars as well as ground-based and shipborne radars. An aircraft or missile flying just above a duct might not be detected until it is at short range. A radar in an aircraft, such as for airborne air-surveillance, might not be able to detect targets that are below the duct even if they were well within the range of the radar. This can be avoided by proper control of the altitude of the aircraft carrying the radar, but it requires knowledge in real time of the local refractive conditions that can affect radar propagation.

Ducted propagation can make possible the detection of unwanted clutter echoes at long ranges that might otherwise not be detected in a normal atmosphere. This can place a severe burden on MTI radars designed on the assumption that clutter will not appear beyond a certain range. Also, multiple-time-around clutter echoes that arrive from beyond the maximum unambiguous range might not be eliminated if the doppler processing employs pulse-to-pulse staggered repetition periods.

In areas of the world where surfaced-based ducts can significantly extend the radar horizon for a large portion of the time, the greatly increased clutter echoes that are obtained can seriously degrade the performance of a radar not designed to cope with it. Such radars that experience the detrimental effects of ducting should be designed with a large dynamic range so as to avoid receiver saturation by large clutter echoes, the radar should have additional MTI or pulse doppler improvement factor to eliminate the larger than normal clutter, and the radar waveforms and processing should be designed to cancel multiple-time-around clutter echoes that originate from long ranges. The last mentioned is accomplished by using a constant prf (instead of pulse-to-pulse prfs) with processing that uses the required number of "fill pulses."[65] Fill pulses are those given zero weight in the digital MTI processor (that is, they are discarded) so as to eliminate pulse repetition intervals that do not have multiple-time-around clutter.

Again it should be mentioned that ducted propagation theoretically requires that the radar and target be within or close to the duct. It is seldom that a conventional surface-based radar, for example, will experience serious ducting effects if its beam is pointed to an elevation angle greater than about 0.5°.

Modified Refractivity In this section the refractivity, denoted by N, has been used to describe the property of the atmosphere to bend radar waves. It is sometimes convenient, however, to employ a modified refractivity, defined as

$$M = N + (h/a) \times 10^6 \qquad [8.20]$$

where h = height above the earth's surface, and a = earth's radius (in the same units as h). The modified refractivity M takes account of the curvature of the earth. It is useful in identifying ducting, since the trapping of radar waves occurs for all negative gradients of M. The modified refractivity, rather than N, is commonly employed by propagation engineers for determining the effects of refraction.

Prediction of Refractive Effects The theory of ducted propagation is not as complete as one might like. Available theoretical models sometimes fail to adequately describe what is taking place in nature. Nevertheless, there have been developed usable prediction

methods based on local meteorological measurements that provide estimates of how ducted propagation might affect the radar coverage and how the radar platform might be positioned to minimize the effects of nonstandard propagation. In the following are briefly described the several methods mentioned by Hitney[66] for determining the effects of refraction by a nonstandard atmosphere.

Ray Tracing This uses geometrical optics to determine the paths taken by radar waves as they propagate through the atmosphere.[67] In ray tracing the modified refractivity profile is assumed to vary only with height. The height is divided into small increments Δh, and Snell's law is applied with the small-angle approximation to determine the bending of a ray as it leaves a region of refractivity N at a height h and passes into a region of refractivity $N + \Delta N$ at height $h + \Delta h$. The refractivity at each increment of Δh is assumed to be linear. Since refractivity does not depend on frequency (within the normal range of radar frequencies), a single ray-tracing diagram can also be considered to be independent of frequency. Ray tracing is relatively simple compared to other models; but it does not provide the magnitude of the field strength. It also requires that the refractive index not change significantly in a distance comparable to a wavelength and the spacing between neighboring rays must be small in order to produce correct results when the rays diverge, converge, or cross. Also, diffraction effects are not taken into account.

Waveguide Model The trapping layer can be considered as a waveguide with multiple modes of propagation.[68] Those readers familiar with microwave propagation in metallic waveguides know that the dimensions of the waveguide usually are chosen so that only one dominant "mode" can propagate. In a rectangular guide this occurs when the broad dimension is slightly greater than a half-wavelength. If the dimensions are much larger than this, more than one mode of propagation can take place. Each mode has a different field configuration along the guide than the others. The various modes interfere or reinforce with one another as the energy travels along the guide, which is why a single mode of propagation usually is preferred. The theory and practice of propagation in waveguides was developed primarily in the 1930s and 1940s. The theory of guided wave propagation in other than metallic waveguides, however, is much older. It was first devised to explain how long-wavelength radio waves can propagate around the earth along the structure formed by the earth's surface and the ionosphere. Such a situation results in the energy propagating in more than one mode. A similar type of propagation occurs when sound waves travel in water when one surface of the guide is the bottom of the sea and the other is the sea surface. In a metallic waveguide, the walls present sharp boundaries that reflect the propagating waves incident upon them. Waves can also be reflected from a stratified medium in which the refractivity varies continuously and which has no sharp boundary. This is what happens in ducted propagation.

The theory of guided propagation in a layered medium has been applied to radar propagation in atmospheric ducts. This is a physical optics approach that takes into account propagation loss and diffraction. The theory predicts a cutoff frequency, below which no propagation can take place in the guide medium. The waveguide model can be employed when the vertical refractivity profile is independent of range (a vertically stratified, horizontally homogeneous atmosphere). Thus it is not suited when the atmospheric refraction

varies with range. It can be used for ducted propagation beyond the horizon. The mode theory applies best when only a few modes are present, and when both the radar and the target are well within the duct. It tends to predict greater loss than indicated by actual measurements when the radar, target, or both, are not within the duct. One reason for this is that the duct is usually *leaky*. The upper boundary of the atmospheric duct is not necessarily smooth, but can be ragged so that the simple model of a plane surface is not realistic. Another limitation is that mode theory when applied over land does not usually include the effects of scattering from the nonsmooth terrain which is characteristic of most of the earth's surface.

The effects of multiple modes on the propagation within a duct has been mentioned earlier in this section, in the subsection "Multiple-Mode Propagation." It was said there that fades due to multiple modes can be of the order of 20 dB and can occur at intervals of 2 to 3 mi in range.

Parabolic Equation Model Unlike ray tracing or the waveguide model, the parabolic equation method can handle refractive index changes that are inhomogeneous in both the horizontal and vertical directions. Thus it is useful whenever the refractivity profile varies with range; such as at land/ocean interfaces and in propagation over irregular terrain. This approach solves the Helmholtz wave equation with a parabolic equation approximation, such as a numerical method called the Fourier split-step algorithm.[69] The *split-step parabolic equation* provides an efficient method for modeling atmospheres where the vertical refractivity profile changes along the propagation path. It works well within the horizon and over the horizon as well as near the horizon, so that a single model can be used to make computations in all regions of interest. It has considerable advantage over prior methods, but it might require large computer resources of memory and run time. It has been noted[66] that rough surface effects are difficult to handle rigorously with this model. In addition to being applied over the ocean, the parabolic equation method has been applied to propagation over terrain.[70–72]

Hybrid Method The purpose of a hybrid method is to provide the benefits of the split-step parabolic equation (PE) method without the extensive computations. One example is the Radio Physical Optics (RPO) model that uses a combination of ray optics and split-step PE methods.[73] These two are complementary in that split-step PE works well for small angles and the ray optics works well at the higher angles not covered by the PE method. Hitney states that this hybrid method can be 100 times faster than the pure split-step PE method for stressful cases.

Computer-Based Propagation Assessment Methods There exist several computer software programs for determining the effects of propagation based on knowledge of the environmental factors that influence atmospheric refractivity. These have been used mainly by the military to determine the actual coverage of their radars under nonstandard propagation conditions. Much of the pioneering work in this area was performed by the SPAWAR Systems Center, San Diego (formerly called NRaD).

Figure 8.15 illustrates in a simple manner one example why the military tactical planner would be interested in knowing the radar propagation conditions to be expected.[74] The

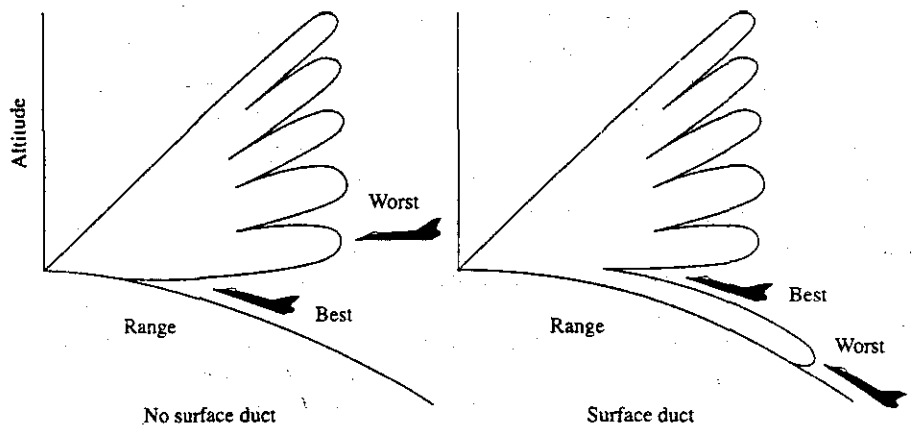


Figure 8.15 Sketch illustrating the effect of atmospheric ducting on the flight altitude of an attacking aircraft.
(From J. H. Richter.[74])

left-hand side of the figure shows the radar coverage that might be obtained under standard propagation conditions with no surface duct present. An attack aircraft could avoid detection at long range by flying at low altitude. If, on the other hand, a surface duct were present, as depicted on the right-hand side of the figure, a low-flying attacker would be detected at much longer range because of the ducted propagation. A better tactic under such circumstances would be to fly at an altitude just above the duct.

IREPS One of the first computer-based propagation prediction programs for operational use was *IREPS*, which stands for *Integrated Refraction Effects Prediction System.* It was developed by the SPAWAR Systems Center as an operational assessment tool for the U.S. Navy and was installed on major ships of the Fleet.[75,76] It was used with a PC, and had an interactive display. IREPS provided the following products:

1. *Propagation conditions summary.* The plot of refractivity as a function of height was shown graphically for the location, date, and time of day, along with a plain-language narrative assessment of what effects might be expected over surface-to-surface, surface-to-air, and air-to-air paths. The presence and vertical extent of any ducts were indicated, and wind speed and evaporation duct height were listed numerically. A correction for the measurement error of the target elevation angle due to bending by refraction was also given.
2. *Display of radar coverage.* These were vertical coverage diagrams (range versus height) for specific radars and other electromagnetic systems.
3. *Display of one-way path loss with range.*
4. *AEW aircraft stationing aid.* This showed the distortion of normal propagation caused by refractive effects at a given range for various combinations of radar and target altitudes. It provided the optimum location for the radar, so as not to lose targets

because of radar holes and allowed the aircraft to minimize detection by a hostile intercept receiver.

5. *Surface-search radar-range table.* This provided predictions of the detection range for an operator-selected surface-search radar against an operator-selected table of surface targets.
6. *Electronic support measures (ESM) intercept-range table.* A display was given of the maximum intercept range for an operator-selected ESM system (intercept receiver) against various operator-selected radars or other emitters.

In addition, IREPS could display the 50 percent probability of detection range of various size targets for an operator-selected forward-looking infrared (FLIR) system operating at various altitudes.

There were several methods by which IREPS obtained information about refractive conditions, depending on what was available. Refractivity could be obtained from radiosonde measurements (balloon borne instruments to determine upper atmosphere pressure, temperature, and water vapor pressure); aircraft-borne microwave refractometers; or inputs from surface meteorological measurements made on board the ship on which IREPS was located. When none of these were available, IREPS utilized a stored library of historic refractivity and climatology statistics as a function of the latitude, longitude, season, and time of day. When historic data was used, the output was a prediction of propagation performance in probabilistic terms. Also stored were the necessary system parameters for the various radars, communications, electronic warfare, and other electromagnetic systems whose predicted propagation performance was required. IREPS could make propagation predictions for frequencies from 100 MHz to 20 GHz.

Hitney[66] indicated that one of the most important uses of IREPS was for selecting the best flight profile for an attack aircraft attempting to penetrate the coverage of a hostile radar. Under normal (nonducting) propagation conditions, an attack aircraft flies most of the distance to its target at low altitude in order to be below the coverage of the hostile radar. If, on the other hand, a surface-based duct was predicted by IREPS, the defending radar detection range against low-altitude targets might be greater than when ducting was absent. Under such circumstances, it is usually better for the attacker to fly at an altitude slightly above the duct. Hitney has said that "the use of IREPS coverage diagrams in strike [attack] warfare flight profile selection has been verified operationally to be effective 85% of the time."

TESS, or Tactical Electronic Support System[66] This is designed for naval tactical decision making and is similar to IREPS in that it uses the same basic propagation assessment models and displays, but with better environmental information and some improved propagation models. For example, it employs the hybrid RPO model mentioned previously so as to take account of range-dependent refractive effects. It uses real-time satellite data, and has the ability to overlay this data with other meteorology analyses and forecasts. TESS is designed for surface ships that have officers trained in the environmental sciences, including aircraft and helicopter carriers and amphibious assault ships.

EREPS, or Engineer's Refractive Effects Prediction System[38] This is also derived from IREPS, but is designed for the use of engineers instead of naval tactical decision makers.

For example, the engineer is more interested in the long-term performance of a radar, usually in statistical terms, rather than in single-event performance prediction that IREPS was designed to provide as a tactical decision aid. EREPS is more flexible than IREPS in that it increases the user's ability to edit the various parameters and determine how changes in radar parameters affect performance. It also allows the use of a high-fidelity range-dependent propagation model such as the RPO program, including the use of the binary files of propagation loss versus range as generated by the RPO program.

AREPS or Advanced Refractive Effects Prediction Program (AREPS) This software program is an advanced version of IREPS that computes and displays radar probability of detection, electronic support measures (ESM) vulnerability, UHF/VHF communications capability, and simultaneous radar detection and ESM (intercept) vulnerability.[77] It is Windows based and is available from the SPAWAR Systems Center as a CD or from the Internet.

Other versions of computer-based propagation prediction methods have been reported by Ferranti Computer Systems, Ltd. of the United Kingdom[78] and by the Ukraine.[79] The original Ferranti system was called IMP, or Identification of Microwave Propagation. As described, it appears to have less capability than IREPS. The Ukrainian system is said to incorporate scattering from atmospheric turbulence.

8.6 DIFFRACTION

In the previous section we discussed how radar waves can propagate beyond the geometrical horizon of the earth by means of atmospheric refraction. Another mechanism that permits electromagnetic waves to extend beyond the geometrical horizon is *diffraction.* Radio waves are diffracted around the curved earth in a manner similar to the way light is diffracted by a straight edge (a topic usually covered in college physics courses). The ability of electromagnetic waves to propagate beyond the horizon by diffraction depends upon the frequency (the lower the better). At microwave radar frequencies there is very little energy diffracted by the earth's surface so that microwave radar coverage cannot be significantly extended beyond the line of sight by this propagation mechanism. Diffraction is important, however, for understanding HF surface-wave radar and for predicting the signal strength at or near the radar horizon at any radar frequency.

Frequency Dependence of Diffraction Figure 8.16 is a theoretical plot of the electric field strength (relative to free space) incident on a target as a function of the distance from the radar transmitting antenna. Both the transmitting antenna and the target are at a height of 100 m, and there is no refraction to extend the horizon. The geometrical horizon at 71.4 km is also the horizon distance for optical frequencies ($\lambda \rightarrow 0$). It represents the approximate boundary between propagation and no propagation for optics and infrared. As the frequency decreases (wavelength increases), energy propagates farther into the region beyond the geometrical horizon. As the wavelength increases, however, it is noted that there is a reduction of energy at the geometrical horizon as well as just within the

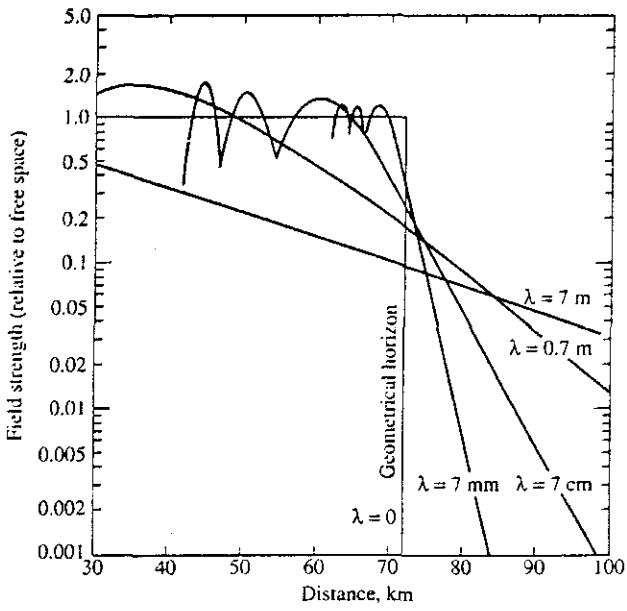


Figure 8.16 Theoretical field strength (relative to free-space field strength) as a function of the distance from the transmitting antenna. Vertical polarization, $h_a = h_t = 100$ m, $k = 1$, ground conductivity = 10^{-2} mho/m, dielectric constant = 4. (After Burrows and Attwood,[8] courtesy Academic Press, Inc.)

horizon. In the absence of refraction effects, microwave radar seldom has the ability to detect low-altitude targets beyond the geometrical horizon, or even at the horizon, unless it has sufficient power to overcome the loss caused by diffraction.

If low-altitude radar coverage is desired beyond the geometrical horizon in the diffraction region, the frequency should be as low as possible and there should be excess power to compensate for the diffraction loss. As an example, the loss in signal strength in the diffraction region at a frequency of 500 MHz is roughly 1 dB/mi at low altitudes. (It is even greater at higher frequencies.) Therefore, to penetrate 10 mi beyond the horizon within the diffraction region, the radar power at 500 MHz must be increased by 20 dB over that required for free-space propagation. Even if lower frequencies were available for radar to take advantage of the lower diffraction loss, the range resolution is poorer (because of narrower bandwidths), beamwidths are wider, the spectrum is crowded, and external noise increases with decreasing frequency.

If, on the other hand, the low-altitude coverage is to be optimized within the horizon in the interference region, Fig. 8.16 shows that the radar frequency should be as high as practical (consistent with other constraints on the choice of radar frequency).

Coverage at the Horizon The maximum "line-of-sight" distance d between a radar at a height h_a and a target at height h_t is given by the expression

$$d = \sqrt{2kah_a} + \sqrt{2kah_t} \qquad [8.21]$$

where k is the effective earth radius, and a is the earth actual radius. In this equation, the line of sight between the radar and the target is assumed to be tangent to the earth's surface. This equation is often used to describe the maximum range at which a target at an

altitude h_t can be seen above the radar horizon. The diffraction loss at the horizon, however, might be from 10 to 30 dB below that if in free space.[80] Thus caution should be used when employing the above equation to describe the low-altitude coverage of a radar when propagation is near the horizon.

Surface-Wave HF Radar The diffraction loss at HF frequencies (3 to 30 MHz) is much lower than that at microwave frequencies. For this reason, the potential of extended coverage at low altitudes beyond the horizon has been examined many times in the past using HF radar and the surface wave, or ground wave, mode of propagation where the diffraction loss is low. It has been said[81] that for every nautical mile increase in radar range beyond 75 nmi over the sea in the diffraction region requires an increase of radar energy of approximately 0.3, 0.5, and 0.6 dB per nautical mile at frequencies of 5, 10, and 15 MHz, respectively. The loss depends on the surface conductivity. Over land the loss is much higher than over the sea; which is why HF surface-wave radars are seldom considered for operation over land.

The lower the frequency the lower will be the propagation loss. The lower the frequency, however, the greater will be the external noise. (External noise can be many orders of magnitude greater than receiver noise). External noise is due to either atmospheric, cosmic, or anthropogenic noise. The radar cross section of most aircraft at HF is usually much larger than at microwave frequencies; but at a sufficiently low frequency (depending on the size of the target), the cross section falls in the Rayleigh region where it varies as the fourth power of the radar frequency. It then decreases rapidly with decreasing frequency. For example, the decrease in cross section of a fighter aircraft signifying the Rayleigh region might begin somewhere around 15 to 20 MHz. For a large bomber aircraft, the Rayleigh region might begin at a frequency of 3 to 6 MHz.

There will be an optimum frequency for a surface-wave radar. Below the optimum frequency there is an increase in radar power because of increased external noise and lower radar cross section. Above this frequency, increased power is also needed because of the increase in diffraction loss with increasing frequency. The optimum frequency will depend on the type of target and the variability of external noise with time of day and season. Because of the variability of external noise and the need to detect small as well as large targets, the choice of HF radar frequency is often a compromise. In one analysis of the optimum frequency for a particular HF radar, the region from 5 to 10 MHz seemed to be the best place to operate.[81]

Vertical polarization is used in HF surface-wave radar since its energy extends down to the surface, which is what is desired to detect targets at or near the surface. The energy radiated by a horizontally polarized antenna, on the other hand, decreases as the surface is approached. When surface-wave radars are considered for shore-based operation looking out over the sea, the antennas can be of large extent in the horizontal dimension. They might be from 300 to 1000 m (more or less). A 500-m antenna at 10 MHz has a beamwidth of about 4°. The antenna is an array which is either electronically steered in azimuth angle or which forms a number of multiple fixed receiving beams covering the area of interest. With multiple receiving antenna beams, the transmitting antenna is much smaller in size than the receiving array antenna since it utilizes one broad beam to cover the region viewed by the multiple, narrow receiving beams. The advantage of this antenna

arrangement is that it is easier to make a large receiving antenna than a large transmitting antenna. The multiple beams of the receiving antenna allow the simultaneous processing of the radar echo signals from each beam, which allows a faster data rate. Because of expense, HF radar antennas seldom have significant directivity in the elevation plane other than that obtained from a monopole radiator in front of a backscreen. These radars need doppler processing in order to detect desired moving targets in the midst of the large echo from the land or sea.

Because of the exponential diffraction loss, land-based HF surface-wave radars that look over the sea might have average powers of from several tens of kW to well over 100 kW. The pulse repetition frequencies are usually low so that ground echoes received via skywave propagation from long range do not interfere with target echoes from the near-range. The ranges of such radars might be from 50 to 150 nmi depending on the type of target. (Most of the numbers presented here are not meant to be interpreted as rigid bounds.)

HF surface-wave radars for shipboard operation have to be much smaller in size than land-based systems and be satisfied with lower power transmitters.[82] Their ranges would be correspondingly reduced. The main naval application for such radars might be to detect low-altitude antiship missiles at greater ranges than can be detected by microwave radar.

In spite of some attractive attributes of HF surface-wave radar there are many difficulties in their application (which mainly have been for military purposes). They have poor resolution in range and angle, no indication of target elevation angle (or height), they require large antennas, and on board ship, they can cause mutual interference with HF communications. If the chief interest is in detecting targets at over-the-horizon ranges, the radar cannot immediately differentiate between a target that is at the same range as the over-the-horizon target but at a higher altitude and within line of sight. Military radars for air defense require some form of target identification, usually a cooperative IFF (identification friend or foe) system that operates at microwave frequencies. Thus there is no convenient means for target recognition when using HF radar. Also, the powerful HF signals radiated by a military radar can be detected by a hostile intercept receiver at very long distances. Thus there are reasons why the HF surface-wave radar has not seen operational application even though it has the capability of seeing over the horizon.

In general, the long-range HF over-the-horizon radar that employs skywave propagation has greater range and greater coverage than a surface-wave radar, and is not that much larger than the largest surface-wave systems.[83] Much smaller HF surface-wave radars have also been operated at within-the-horizon ranges for remote sensing of sea state,[84] surface currents,[85,86] and icebergs.[87,88]

8.7 ATTENUATION BY ATMOSPHERIC GASES

Water vapor and oxygen in the clear atmosphere can attenuate radar energy when the radar frequency is at or in the vicinity of one of the resonant frequencies of these molecules. Figure 8.17 shows the attenuation due to both water vapor and oxygen as a function of frequency. There is a resonance peak of water vapor at a frequency of 22.2 GHz, and another in the millimeter wave region at 184 GHz.[89,90] The attenuation due to water

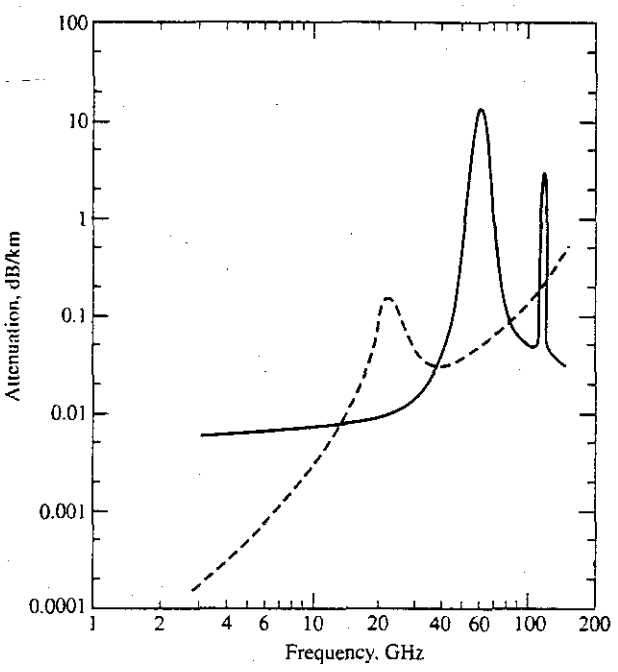


Figure 8.17 Attenuation of electromagnetic energy by atmospheric gases in an atmosphere at 76 cm pressure. Dashed curve is absorption due to water vapor in an atmosphere containing 1 percent water vapor molecules (7.5 g water/m^3). The solid curve is the absorption due to oxygen. (From Burrows and Attwood[8] and Straiton and Tolbert.[89])

vapor will depend on the amount of moisture in the atmosphere, which can vary with time and place. Although the attenuation is only about 0.2 dB/km at a frequency of 22 GHz, absorption can be sufficient to deteriorate the effectiveness of radars that operate at the original K-band frequency of 24 GHz. When radars were first developed at K band during World War II, it was not realized that there was a nearby absorption band. To avoid this problem, the original K band was split into a lower band, K_u, and an upper band, K_a, Table 1.1. Radars are almost never found any more at the original K band. The oxygen molecule has resonances at 60 GHz and 118 GHz. The 16 dB/km attenuation at 60 GHz makes this region unusable except for very short range radars and radars that operate in space outside the atmosphere.

Atmospheric attenuation generally has negligible effect on radar performance at the normal microwave frequencies. It begins to be increasingly important at frequencies above 10 GHz. The large attenuations experienced at millimeter wavelengths is one of the chief reasons why long-range radars are seldom found above 40 GHz.

The effect of attenuation, when it is large enough to be a problem, is accounted for in the radar range equation by inserting into the numerator the multiplicative factor $\exp[-2\alpha R]$, where α is the one-way attenuation coefficient measured in units of distance^{-1}; and R is the range to the target. Instead of α it is more usual to express the one-way attenuation, especially when plotted in graphs, as decibels per unit distance—which is what is done here. This is equivalent to the quantity 4.34α, where the constant 4.34 accounts for the conversion from the natural logarithm to the base 10 logarithm.

If the attenuation per unit distance α is not constant, $\exp[-2\alpha R]$ should be replaced by $\exp[-2\int\alpha(R)dR]$, with the integration taken from 0 to the target range R.

Atmospheric attenuation decreases with increasing altitude (there are fewer molecules to absorb the radar energy). When the antenna beam is pointed at some elevation angle, the variability of attenuation with altitude must be taken into account when determining the total attenuation along the propagation path. With a ground-based radar the attenuation is greatest when the antenna points along the horizon, and is least when it points to the zenith. For example, at the water vapor absorption line located at 22.2 GHz, when the energy is directed at the horizon (elevation angle of 0°), the total attenuation in propagating completely through the earth's troposphere and back again is 80 dB, a formidable number. When the energy propagates at the zenith (elevation angle of 90°), the total two-way attenuation through the entire troposphere is only 1.3°. If the elevation angle were greater than 10°, the total attenuation is less than 7 dB, so that when the radar is looking at the higher elevation angles, attenuation might not be important. Figure 8.18 gives

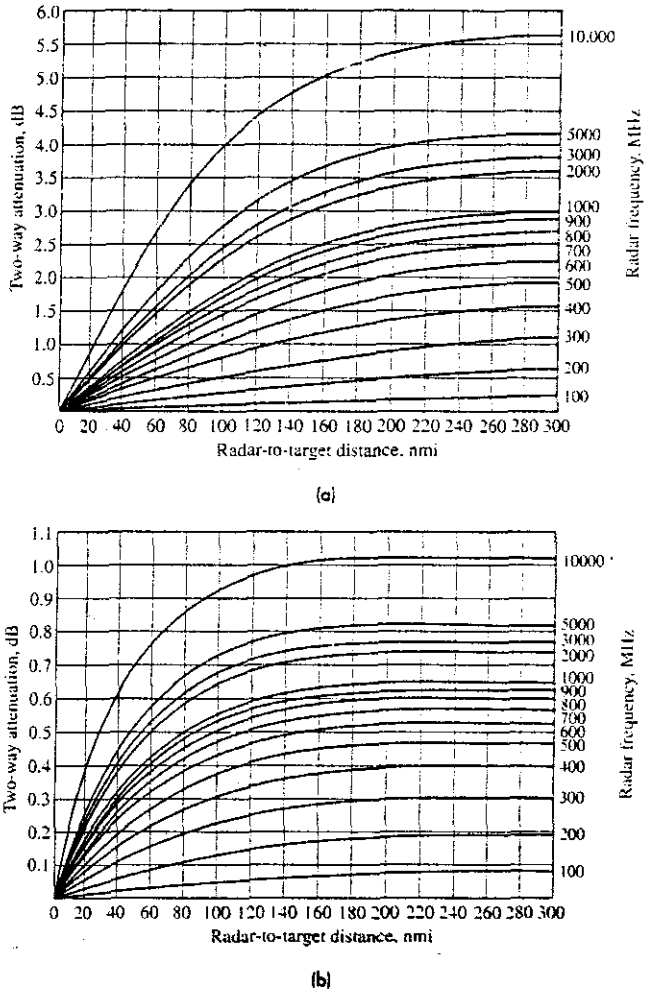


Figure 8.18 Two-way atmospheric attenuation as a function of range and frequency for (a) 0° elevation angle and (b) 5° elevation angle. (From L. V. Blake.[91])

examples of the two-way attenuation in the atmosphere as a function of range and frequency for elevation angles of 0 and 5°.[91] It might be noted that even if there were no loss at 0° elevation, the null at the horizon due to multipath prevents significant propagation of radar energy at or near this angle (except under ducting conditions or at low frequencies with vertical polarization.)

8.8 ENVIRONMENTAL, OR EXTERNAL, NOISE

The inherent internal noise of the radar receiver is what usually limits the detectability of targets by microwave radars (in the absence of clutter echoes). At frequencies at either end of the microwave spectrum, however, the limitation on sensitivity is usually the external noise that appears at the antenna terminals from some outside source. The reradiation noise due to atmospheric absorption usually determines the receiver sensitivity at the upper end of the microwave spectrum and at millimeter waves. At VHF and lower frequencies the receiver sensitivity is usually set by cosmic noise, the noise due to the combined effects of lightning strokes throughout the world, and anthropogenic noise. The minimum noise occurs at the middle of the microwave region, in the vicinity of *S* band. Generally, external noise is not a factor in radar performance unless the radar frequencies are outside the range of the usual microwave frequencies. Harmful external noise at the antenna terminals due to deliberate hostile jamming, however, can cause serious degradation to an unprepared military radar system, but this is not the subject of this chapter.

Atmospheric Absorption Noise It is known from the theory of blackbody radiation that any body that absorbs energy reradiates the same amount of energy that it absorbs, else it would increase in temperature. As mentioned in Sec. 8.7, water vapor and oxygen absorb (attenuate) radar energy. This absorbed energy must then reradiate as thermal noise. If L is the loss of radar energy when propagating through the atmosphere, and T_a is the ambient temperature of the absorbing atmosphere, the effective noise temperature of the reradiated energy is $T_e = T_a(L - 1)$. (The effective temperature is defined in Sec. 11.2.) Atmospheric absorption noise, just like atmospheric attenuation, is of potential concern only at the higher radar frequencies. Figure 8.19, shown later, is a composite plot of the several sources of electromagnetic noise as a function of frequency. Atmospheric absorption noise is the dominant effect at the right side of the figure. The maximum value of atmospheric absorption noise occurs for an elevation angle of 0°, the minimum values for an elevation angle of 90° (looking straight up).

Cosmic Noise There is a continuous background of noiselike electromagnetic radiation from extraterrestrial sources in our own galaxy (the Milky Way), extragalatic sources, and radio stars. Cosmic noise generally decreases with increasing frequency and can usually be ignored at frequencies above UHF. The magnitude of cosmic noise depends upon the portion of the celestial sphere to which the antenna points. It is a maximum when looking toward the center of the Milky Way, and a minimum when observing along the pole about which the Milky Way revolves. The maximum and minimum *brightness temperature* due to cosmic noise is shown in the left-hand portion of Fig. 8.19. The brightness

temperature can affect the system noise temperature and the sensitivity of the radar receiver (especially at the lower frequencies). In the absence of any radio stars, the background cosmic noise left over from the "big bang" at the start of the universe is the minimum noise level that might be expected. Its value is 2.7 K, which is too small to bother any radar receiver.

The sun is a relatively strong emitter of noise if the radar antenna beam directly views the solar disk. It might also be detectable with radars having poor sidelobes and very sensitive receivers. The minimum level of solar noise is due to blackbody radiation at a temperature of 6000 K. Solar storms (sunspots and flares), however, can increase the solar-noise level several orders of magnitude over that of the quiet or undisturbed sun. Radar stars are too weak to be a serious source of interference. Both radio stars (in conjunction with sensitive receivers) and the sun have been used as sources to calibrate the beam-pointing (boresight) of large antennas.[92–94]

Atmospheric Noise (Lightning) A single lightning stroke radiates considerable RF noise power, especially at the lower frequencies. At any one moment there are an average of 1800 thunderstorms in progress in different parts of the world. From these storms about 100 lightning strokes take place every second somewhere in the world.[95] The combined effect of all the lightning strokes is to give rise to a noise spectrum that is especially large at broadcast and shortwave frequencies. Noise radiated by lightning strokes throughout the world is called *atmospheric noise* (not to be confused with the noise produced by atmospheric absorption, as mentioned previously). The spectrum of atmospheric noise falls off rapidly with increasing frequency and is usually of little consequence above 50 MHz. It is seldom, therefore, an important consideration in radar design; except possibly for radars in the lower portion of the VHF region.

Anthropogenic Noise In preparing the previous (second) edition of this text, the publisher reminded me on several occasions to avoid the use of sexist terminology. I believe I was successful in doing so, except for one term: that of *man-made noise*. My reason for wanting to continue to use it was not that men usually make such noise (which they do), but I could not find a suitable substitute. Terms such as "human-made noise," "people-made noise," or "population-made noise" might be nonsexist, but they just didn't sound right. The publisher took pity on my inability to find a suitable substitute and relented; so on p. 463 of the 2d edition, the term "man-made noise" was allowed to appear. After the book was published, I came upon an excellent substitute, which is "anthropogenic noise." *Anthropogenic* is an adjective that means *relating to, or resulting from the influence of humans on nature*. It is a proper replacement for the no longer acceptable term *man-made noise*.

Electromagnetic emissions that appear as noise or interference to other electromagnetic services, originate from many possible sources: higher harmonics and other incidental radiation from transmitters, automobile ignition, electric razors, power tools, automatic garage door openers, fluorescent lights, industrial processing equipment, and power transmission lines.[96,97] Anthropogenic noise is more prevalent in urban and industrial areas than in rural areas. It decreases with increasing frequency and is seldom a factor in the design of microwave radars. It can be of concern, however, for VHF and lower-UHF

systems. Because of its variability in time and space, it is difficult to be precise about the quantitative nature of this form of noise. In the UHF radar frequency band, one source indicates[98] that the noise temperature of anthropogenic noise from a business region (industrial park, large shopping center, busy street, or highway) might be about 500 K, and for a residential region (at least two dwelling units per acre and no nearby highways) it might be about 200 K. At VHF and UHF, anthropogenic noise varies almost inversely as the cube of the frequency.

Interference and noise can also be experienced from other users of the electromagnetic spectrum, such as other radars and communications of all varieties. These sources of interference are generally different from anthropogenic noise, and is considered a problem in electromagnetic compatability (EMC).

Composite Graph Figure 8.19 is a composite graph of the several forms of environmental noise that might affect radar.[99] Only the minimum and maximum resultants from the various component factors are shown. Atmospheric noise and cosmic noise dominate at the lower frequencies, and atmospheric absorption noise dominates at the higher frequencies. The minimum noise levels occur from about 1 to 5 GHz (L to C band). Anthropogenic noise is not included on this graph since it is so variable and generally is of little consequence to radars that operate within the usual microwave radar bands. It can be quite important, however, for radars that operate at VHF or lower frequencies.

Earth Thermal-Noise The temperature of the earth is nominally 290 K; hence, it will radiate thermal noise. If an antenna beam illuminates the ground, it will receive a portion

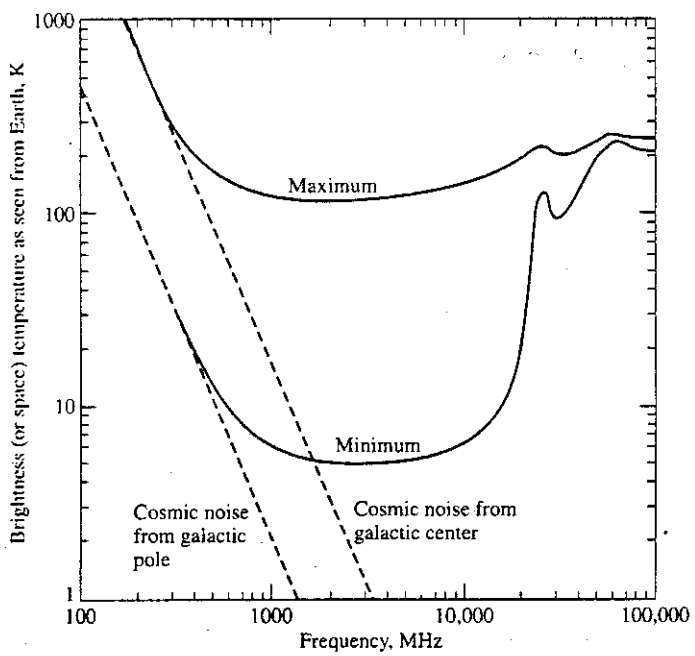


Figure 8.19 Maximum and minimum brightness temperatures of the sky as seen by an ideal single-polarization antenna on earth. (After Green and Lebenbaum.[99])

of the thermal noise radiated by the earth. The actual noise temperature seen by the receiver depends on whether the entire main beam of the antenna views the ground or if only a portion of the beam does so. The radiated noise depends on the emissivity of the ground as well as its temperature. Thus the brightness temperature seen by the antenna will be less than the actual temperature when the emissivity of the ground is less than unity, as will happen if the antenna views a water surface. The effect of the thermal noise radiation from the ground will affect only those systems with very sensitive receivers, generally those with noise figures of a fraction of a decibel. Radar receivers are seldom that sensitive; hence, the thermal noise radiated by the ground usually does not bother conventional radars.

8.9 OTHER PROPAGATION EFFECTS

Radar Siting Minimizing of environmental noise effects sometimes can be an important consideration in determining where a radar should be located. The siting of a radar also depends on the masking (obscuring or screening of targets) due to terrain and the backscatter from terrain features. It is especially important to know how the terrain affects the performance of ground-based radars when there is either severe masking, severe clutter echoes, or when low altitude targets have to be detected with high reliability.

The air-surveillance radars employed during World War II had no doppler processing, so that an aircraft target located in the same radar resolution cell as a clutter echo was likely to be undetected because of the large clutter echoes. As mentioned previously, when siting military air-surveillance radars at that time, it was desired that a site be chosen such that the surrounding terrain would shield clutter echoes from the radar. In effect, the terrain was used to mask the low-lying clutter echo and prevent it from entering the radar receiver so that aircraft at high altitudes could be seen without the presence of clutter. That was satisfactory in World War II, since heavy bomber aircraft generally flew at high altitudes. Modern aircraft and missiles, however, fly at low altitudes to deliberately avoid a radar's coverage. The military radar designer can respond by siting the antenna as high as practical and employing doppler processing to separate moving targets from fixed clutter echoes. Thus the siting of ground-based military air-defense systems has to take account of the local terrain so as to minimize the regions where the detection ranges of low-altitude targets are significantly reduced because of terrain masking.

The effect of terrain is an important part of the siting of civil air-traffic control radars. Computer software has been developed for providing detailed information about the terrain and its effect on civil air-traffic control radars so that the performance of a potential radar site can be determined before the radar is installed.[100] This program, called the Radar Support System (RSS), provides information that allows the radar designer to optimize the performance of a site-specific sensor by selecting the optimum radar height, optimum beam-tilt-angle, sensitivity time control (STC) characteristics, and to determine the likelihood of false alarms due to echoes from highway traffic or large clutter echoes. The input to RSS includes a digital model of the terrain obtained from databases available from the United States Geological Survey (USGS) or the Defense Mapping Agency. The height resolution is one meter and the steps in latitude and longitude are three seconds. There is

also a USGS database of land use and land cover with the same latitude/longitude quantization. It identifies the land as one of ten possibilities; such as urban, agricultural, forest, wet land, barren land, and so forth. Also included is a cultural database that provides three-dimensional models of the significant buildings near the radar site (such as an airport), and includes an estimate of the construction material of each structure. The RSS determines plots of the line-of-sight visibility from the radar, areas which are screened from the radar, the radar cross section of all visible terrain cells, and the probability of detection for specified targets.

Similar computer models and terrain databases have been applied to a military overland situation when the radar is located off shore.[101] Based on the Defense Mapping Agency's Digital Terrain Elevation Data and propagation programs like those mentioned in the previous section for standard and nonstandard atmospheric refraction conditions, it determines terrain and target visibility, masking, clutter echo strength, and plots the results on a maplike presentation to show how radar coverage is affected by the environment.

Atmospheric Lens-Effect Loss Weil[102,103] has shown there is another effect of atmospheric refraction on radar propagation, in addition to what was discussed in Secs. 8.4 and 8.5. The variation of standard refractivity with altitude causes the atmosphere to act as a negative lens that decreases the radiated energy density incident on a target. This loss is independent of radar frequency. Unlike atmospheric attenuation, the lens-effect loss continues to increase at ranges beyond the sensible atmosphere, but it approaches a limiting value asymptotically. For a CRPL exponential atmosphere with surface refractivity equal to 313, the loss is less than 1 dB at a range of 200 nmi and 0° elevation angle, and less than 0.18 dB at an elevation angle of 5°. The limiting values at very long range (10,000 nmi) are 2.9 dB at 0° elevation and 0.27 dB at 5° elevation. The lens-effect loss is an additional loss that is additive to the atmospheric attenuation. It is usually small enough to be neglected, except at low elevation angles and long range.

Ionospheric Propagation at Microwave Frequencies The ionosphere is a partially ionized region of the upper atmosphere that extends from about 50 km to 2000 km in altitude.[104] It is generated by high-energy particles that travel from the sun to ionize the atoms of the thin upper atmosphere. The refraction, or bending, of electromagnetic radiation by the ionosphere allows long-range propagation by shortwaves (the HF region) that is well know to radio amateurs. It is also the basis for HF over-the-horizon radar which allows the detection of aircraft, ships, and ballistic missiles at ranges extending out to 2000 nmi and beyond. The ionosphere is usually considered transparent to microwave radiation, but this is not fully correct. It can adversely affect in several ways the propagation of microwave radiation that travels through it.

Faraday Rotation of Polarization An electromagnetic wave experiences a rotation of its plane of polarization when traveling in an ionized medium (the ionosphere) and a magnetic field (that of the earth). This is known as *Faraday rotation.* If a ground-based radar is to detect satellites and other space objects, and if the radar employs linear polarization, the polarization of the echo signal will be different from that transmitted. A loss in signal can result. If, for example, the polarization were rotated by 90° because of

Faraday rotation, the received signal would be zero since the polarization is orthogonal to that of the transmitting antenna (which is assumed to be the same antenna as that used for receiving). The amount of polarization rotation varies inversely as f^2, where f = radar frequency. The rotation is determined by the total electron content of the ionosphere seen over the radar propagation path, which depends on the radar location, time of day, time of year, and the sunspot cycle. The effect is greater when the radar beam is pointing north or south and minimum when pointing east or west (directions are relative to magnetic north).

Radars for the detection of extraterrestrial targets that operate at UHF or lower frequencies can encounter loss due to the large polarization rotations of their echo signals. One solution for radars that are subject to Faraday rotation is to transmit on a single linear polarization (for example, vertical) and receive on two orthogonal linear polarizations (horizontal and vertical) in order to avoid loss in signal because of Faraday rotation. The echo signals in each polarization receive-channel are processed separately and then combined. This technique has been used in UHF radars for the detection of space objects such as BMEWS (ballistic missile early warning system), the AN/FPS-85 space surveillance radar, and the Pave Paws missile warning radar. At one time it was thought that radars at L band did not experience sufficient polarization rotation to require receiving with dual polarizations. Faraday rotation, however, can sometimes be significant enough at L band to require compensation.[105] The maximum one-way polarization rotation at a frequency of 1 GHz, for example, has been said to be 108° for a radar in the United States that views a target at an elevation angle of 30°.[106] Faraday rotation can also affect spaceborne radars viewing ground targets.

Other Ionospheric Effects The ionosphere will introduce a time delay that is inversely proportional to f^2. At 1 GHz, the maximum delay is said to be 0.25 μs.[105] There can also be refraction of the beam, loss by absorption, and frequency dispersion that introduces distortion into wideband signals. These effects are generally small at microwave frequencies for most radar applications that propagate through the ionosphere. There is one important example, however, where compensation had to be made in a microwave radar to avoid degradation of performance because of dispersion when propagating through the ionosphere. This occurred for the Cobra Dane high resolution L-band radar (1175 to 1375 MHz).[107]

The Cobra Dane radar, located on the island of Shemya at the southern tip of the Aleutian island chain in Alaska, was designed to gather intelligence information on Soviet ballistic missile systems. Its high range-resolution waveform employed a 1000-μs linear FM pulse with a 200-MHz bandwidth. Stretch pulse compression was incorporated to achieve a range resolution of about one meter. The time delay in propagating through the ionosphere was different at the low-frequency end of the 200-MHz bandwidth compared with the time delay at the high end of the bandwidth. This difference in propagation time was sufficient to introduce phase distortion and broaden the compressed pulse width unless compensated. The necessary corrections were obtained by predistorting the transmitted pulse by the inverse of the distortions introduced by the ionosphere. In addition to the ionospheric corrections, there had to be a correction applied to account for the fact that the doppler-frequency shift also was not constant over the 200-MHz bandwidth of the high-resolution waveform.

REFERENCES

1. Bachynski, M. P. "Microwave Propagation Over Rough Surfaces." *RCA Review,* 20, no. 2 (June 1959), pp. 308–335.
2. Ament, W. S. "Toward a Theory of Reflection by a Rough Surface." *Proc. IRE,* vol. 41 (January 1953), pp. 142–146.
3. Beard, C. I. "Coherent and Incoherent Scattering of Microwaves from the Ocean." *IRE Trans.* AP-9 (September 1961), pp. 470–483.
4. Miller, A. R., R. M. Brown, and E. Vegh. "New Derivation for the Rough-Surface Reflection Coefficient and the Distribution of Sea-Wave Elevations." *IEE Proc.,* 131, Pt. H (1984), pp. 114–116.
5. Miller, A. R., and E. Vegh. "Family of Curves for the Rough Surface Reflection Coefficient." *IEE Proc.* 133, Pt. H (December 1986), pp. 483–489.
6. Barton, D. K. "Low-Angle Radar Tracking." *Proc. IEEE* 62 (June 1974), pp. 687–704.
7. Miller, A. R., and E. Vegh. "Exact Result for the Grazing Angle of Specular Reflection from a Sphere." Naval Research Laboratory, Washington, D.C., Memorandum Rep. 6867, August 9, 1991.
8. Burrows, C. R., and S. S. Attwood. *Radio Wave Propagation.* New York: Academic, 1949.
9. Shibuya, S. *A Basic Atlas of Radio-Wave Propagation.* New York: John Wiley, 1987.
10. Blake, L. "Machine Plotting of Radio/Radar Vertical-Plane Coverage Diagrams." Naval Research Laboratory, Washington, D.C. Rep. 7098, June 25, 1970 (AD 709897).
11. Skolnik, M. "Improvements for Air-Surveillance Radar." *Proc. 1999 IEEE Radar Conf.* pp. 18–21, IEEE Catalog no. 99CH36249.
12. ITU-R Recommendation P.453–5, "The Radio Refractive Index: Its Formula and Refractivity Data," 1995.
13. Hitney, H. V. "Refractive Effects from VHF to EHF, Part A: Propagation Mechanisms." Paper no. 4A in *Propagation Modeling and Decision Aids for Communications, Radar, and Navigation Systems.* AGARD-LS-196, NATO, September 1994.
14. Bean, B. R., and E. J. Dutton. "Radio Meteorology," National Bureau of Standards Monograph 92, March 1, 1966.
15. Shibuya, S. Ref. 9, p. 28.
16. Bean and Dutton. Ref. 14, Secs. 3.1 and 3.2.
17. Bean, B. R. "The Geographical and Height Distribution of the Gradient of Refractive Index." *Proc. IRE,* 41 (April 1953), pp. 549–550.
18. Bean, B. R. et al. "A World Atlas of Atmospheric Radio Refractivity," U. S. Dept. of Commerce, ESSA Monograph 1, 1966.

19. Hall, M. P. M., L. W. Barclay, and M. T. Hewitt. *Propagation of Radiowaves.* London: Institution of Electrical Engineers, 1996, Sec. 6.2.

20. Blake, L. V. "Prediction of Radar Range." *Radar Handbook,* 2nd ed. M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 2, Fig. 2.18.

21. Tank, W. G. "Atmospheric Effects." *Airborne Early Warning.* W. C. Morchin (Ed.). Boston, MA: Artech House, 1990, Chap. 3. See also Morchin, W. *Radar Engineer's Sourcebook.* Boston, MA: Artech House, 1993, Sec. 15.4.3.

22. Bauer, K. W. "Range-Height-Angle Charts with Lookdown Capability." *Microwave J.* 24 (October 1981), pp. 89–92.

23. Jursa, A. S. (Ed.). *Handbook of Geophysics and the Space Environment.* U.S. Air Force Geophysics Laboratory, 1985. (Available from National Technical Information Service, Springfield, VA, 22161.)

24. Valley, S. L., (Ed). *Handbook of Geophysics and Space Environments.* New York: McGraw-Hill, 1965.

25. Shannon, H. H. "Recent Refraction Data Corrects Radar Errors." *Electronics* 35, no. 49 (Dec. 7, 1962), pp. 52–56.

26. Blake, L. V. "Ray Height Computation for a Continuous Nonlinear Atmospheric Refractive-Index Profile." *Radio Science* 3 (January 1968), pp. 85–92.

27. Brown, B. P. "Radar Height Finding." *Radar Handbook,* 1st ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1970, Chap. 22, Sec. 22.3.

28. Murrow, D. J. "Height Finding and 3D Radar." in *Radar Handbook,* 2nd ed. M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 20, Sec. 20.2.

29. Barnett, K. M., and S. H. Brown. "Accuracy of Calculated Radar Refraction Errors." *IEEE Trans.* AP-13 (November 1965), p. 986.

30. Nathanson, F. E. *Radar Design Principles,* 2nd ed. New York: McGraw-Hill, 1991, Fig. 6.26.

31. Robertshaw, G. "How Accurate is Range Correction?" *Microwaves & RF* 25 (March 1986), pp. 129–132.

32. Rowland, J. R., and S. M. Babin. "Fine-Scale Measurements of Microwave Refractivity Profiles with Helicopter and Low-Cost Rocket Probes." *Johns Hopkins APL Tech. Dig* 8, no. 4 (1987), pp. 413–417.

33. Bean, B. R., and E. J. Dutton. Ref. 13, Sec. 2.3.

34. Anderson, L. J., and E. E. Gossard. "Prediction of Oceanic Duct Propagation from Climatological Data." *IRE Trans.* AP-3 (October 1955), pp. 163–167.

35. Anderson, K. D. "Radar Measurements at 16.5 GHz in the Oceanic Evaporation Duct." *IEEE Trans.*, AP-37 (January 1989), pp. 100–106.

36. Katzin, M., R. W. Bauchman, and W. Binnian. "3- and 9-Centimeter Propagation in Low Ocean Ducts." *Proc. IRE* 35 (September 1947), pp. 891–905.

37. Jeske, H. "State and Limits of Prediction Methods of Radar Wave Propagation Conditions Over the Sea." In *Modern Topics in Microwave Propagation and Air-Sea Interaction*, A. Zancla (Ed.). D. Reidel Publishing, 1973.

38. Patterson, W. L., et al. "Engineer's Refractive Effects Prediction System (EREPS)." Version 3.0, Naval Command, Control, and Ocean Surveillance Center, RDT&E Division, San Diego, CA, Tech. Document 2648, May, 1994.

39. Paulus, R. A. "Practical Applications of an Evaporation Duct Model." *Radio Science* 20 (July–August 1985), pp. 887–896.

40. Jeske, H. "The State of Radar-Range Prediction Over Sea." *AGARD Conf. Proceedings* 70(2) (1971), pp. 50.1–50.10.

41. Hitney, H. V., A. E. Barrios, and G. E. Lindem. "Engineer's Refractive Effects Prediction System (EREPS) User's Manual." Naval Ocean Systems Center, San Diego, CA, July 1988, (AD 203443)

42. Babin, S. M., G. S. Young, and J. A. Carton. "A New Model for the Oceanic Evaporation Duct." *J. Applied Meteorology* 36 (March 1997), pp. 193–204.

43. Anderson, K. D. "94-GHz Propagation in the Evaporation Duct." *IEEE Trans.* AP-38 (May 1990), pp. 746–753.

44. Joseph, R. I., and G. D. Smith. "Propagation in an Evaporation Duct: Results in Some Simple Analytic Models." *Radio Science* 7 (April 1972), pp. 433–441.

45. Früchtenicht, H. W. "Notes on Duct Influences on Line-of-Sight Propagation." *IEEE Trans.* AP-22 (March 1974), pp. 295–302.

46. Giger, A. J. *Low-Angle Microwave Propagation*. Norwood, MA: Artech House, 1991, Sec. 2.5.

47. Anderson, K. D. "Radar Detection of Low-Altitude Targets in a Maritime Environment." *IEEE Trans.* AP-43 (June 1995), pp. 609–613.

48. Hitney, H. V. Ref. 13, Sec. 6.2.

49. Patterson, W. L., et al. Ref. 38, pp. 12–14.

50. Battan, L. J. *Radar Observation of the Atmosphere*. Chicago: University of Chicago Press, 1973.

51. Pappert, R. A., and C. L. Goodhart. "A Numerical Study of Tropospheric Ducting at HF." *Radio Science* 14 (September–October 1979), pp. 803–813.

52. Appreciation is expressed to John Walters of the U. S. Naval Research Laboratory for supplying this information, and to Vilhelm Gregers-Hansen who originally created the figure.

53. Ringwalt, D. L., and F. C. McDonald. "Elevated Duct Propagation in the Tradewinds." *IRE Trans.* AP-9 (July 1961), pp. 377–383.

54. Guinard, N. W., J. Ransone, D. Randall, C. Purves, and P. Watkins. "Propagation Through an Elevated Duct: Tradewinds III." *IEEE Trans.* AP-12 (July 1964), pp. 479–490.

55. Purves, C. G. "Geophysical Aspects of Atmospheric Refraction." Naval Research Laboratory, Washington, D.C. Rep. 7725, June 7, 1974.

56. Hitney, H. V., R. A. Pappert, C. P. Hattan, and C. L. Goodhart. "Evaporation Duct Influences on Beyond-the-Horizon High Altitude Signals." *Radio Science* 13 (July–August 1978), pp. 669–675.

57. Kerr, D. E. (Ed.). *Propagation of Short Radio Waves.* MIT Radiation Laboratory Series, vol. 13 New York: McGraw-Hill, 1951.

58. Brookner, E., E. Ferraro, and G. D. Ouderkirk. "Radar Performance During Propagation Fades in the Mid-Atlantic Region." *IEEE Trans.* AP-46 (July 1998), pp. 1056–1064.

59. Day, J. P., and L. G. Trolese. "Propagation of Short Radio Waves Over Desert Terrain." *Proc. IRE* 38 (February 1950), pp. 165–175.

60. Donohue, D. J., and J. R. Kuttler. "Modeling Radar Propagation Over Terrain." *Johns Hopkins APL Tech. Dig.* 18, no. 2 (1997), pp. 279–287.

61. Dockery, G. D. "Method of Modeling Sea Surface Clutter in Complicated Propagation Environments." *IEE Proc.* 137, no. 2, Pt. F (April 1990), pp. 73–79.

62. Reilly, J. P., and G. D. Dockery. "Influence of Evaporation Ducts on Radar Sea Return." *IEE Proc.* 137, no. 2, Pt. F (April 1990), pp. 80–88.

63. Heimken, H. F. "Low-Grazing-Angle Radar Backscatter from the Ocean Surface." *IEE Proc.* 137, no. 2, Pt. 7 (April 1990), pp. 113–117.

64. Booker, H. G. "Elements of Radar Meteorology: How Weather and Climate Cause Unorthodox Radar Vision Beyond the Geometrical Horizon." *J. IEE* 93, Pt. IIIA (1946), pp. 69–78.

65. Nathanson, F. E., Ref. 30, Sec. 9.5.

66. Hitney, H. V. "Refractive Effects from VHF to EHF, Part B: Propagation Models." Paper no. 4B in *Propagation Modeling and Decision Aids for Communications, Radar, and Navigation Systems.* AGARD-LS-196, NATO, September 1994.

67. Kerr, D. E. Ref. 57.

68. Budden, K. G. *The Wave-Guide Mode Theory of Wave Propagation.* Englewood Cliffs, NJ: Prentice-Hall, 1961.

69. Dockery, G. D. "Modeling Electromagnetic Wave Propagation in the Troposphere Using the Parabolic Equation." *IEEE Trans.* AP-36 (October 1988), pp. 1464–1470.

70. McArthur, R. J. "Propagation Modeling Over Irregular Terrain Using the Split-Step Parabolic Equation Method." *International Conf. Radar 92, IEE Conf. Publication 365,* London, 1992, pp. 54–57.

71. Barrios, A. E. "Parabolic Equation Modeling in Horizontally Inhomogeneous Environments." *IEEE Trans.* AP-40 (July 1992), pp. 791–797.

72. Barrios, A. E. "A Terrain Parabolic Equation Model for Propagation in the Troposphere." *IEEE Trans.* AP-42 (January 1994), pp. 90–98.

73. Hitney. H. V. "Hybrid Ray Optics and Parabolic Equations Methods for Radar Propagation Modeling." *Int. Conf. Radar 92, IEE Conf. Publication 365*, London (1992), pp. 58–61.

74. Richter, J. H. "Electromagnetic Wave Propagation Assessment." *AGARD Highlights* 92/1 (March 1992), pp. 6–14.

75. Hitney, H. V., and J. H. Richter. "Integrated Refractive Effects Prediction System (IREPS)." *Naval Engineers J.* 2 (April 1976), pp. 257–262.

76. Patterson, W.L., et al. *IREPS 3.0 User's Manual.* Naval Ocean Systems Center (NOSC), San Diego, CA, Tech. Document 1151, September 1987. See also Revision PC-2.0, Tech. Document 1874, August 1990.

77. Patterson, W. L. *Advanced Refractive Effects Prediction Sysem (AREPS), Version 1.0 User's Manual.* SPAWAR Systems Center, San Diego, Technical Document 3028, April 1998.

78. Gelsenheyner, S. "Computerized Data Help Predict Anomalous Propagation." *Microwave System News* 11 (April 1982), pp. 45–46.

79. Belobrova, M. V., et al. "Software Suite for Diagnosing USW Propagation Over the Sea." *Radiophysics and Quantum Electronics* 33 (June 1991), pp. 961–965.

80. Skolnik, M. I. "Radar Horizon and Propagation Loss." *Proc. IRE* 45 (May 1957), pp. 697–698.

81. Millman, G. H., and G. R. Nelson. "Surface Wave HF Radar for Over-the-Horizon Detection." *Record of the IEEE 1980 International Radar Conf.* pp. 106–112, IEEE Publication 80CH1493–6 AES, New York.

82. Powers, R. L., L. M. Lewandoski, and R. J. Dinger. "High Frequency Surface Wave Radar—HFSWR." *Sea Technology* 37 (November 1996), pp. 25–32.

83. Headrick, J.M., and M. I. Skolnik. "Over-the-Horizon Radar in the HF Band." *Proc. IEEE* 62 (June 1974), pp. 664–673.

84. Lipa, B. J., and D. E. Barrick. "Extraction of Sea State from HF Radar Sea Echo: Mathematical Theory and Modeling." *Radio Science* 21 (January–February 1986), pp. 81–100.

85. Lipa, B. J., D. E. Barrick, J. Isaacson, and P. M. Lilleboe. "CODAR Wave Measurements From a North Sea Semisubmersible." *IEEE J. Oceanic Engineering* 15 (April 1990), pp. 119–125.

86. Graber, H. C., B. K. Haus, R. D. Chapman, and L. N. Shay. "HF Radar Comparisons with Moored Estimates of Current Speed and Direction: Expected Differences and Implications." *J. Geophys. Res.* 102, no. C8 (August 15, 1997), pp. 18,749–18,766.

87. Walsh, J., B. J. Dawe, and S. K. Srivastava. "Remote Sensing of Iceburgs by Ground-Wave Doppler Radar." *IEEE J. Oceanic Engineering* OE-11 (April 1986), 276–284.

88. Srivastava, S. K., and J. Walsh. "Over-the-Horizon Radar." Chapter 7 in *Remote Sensing of Sea Ice and Ice Bergs*, S. Haykin, E. O. Lewis, R. K. Raney, and J. R. Rossiter (Eds.). New York: John Wiley, 1994.

89. Straiton, A. W., and W. Tolbert. "Anomalies in the Absorption of Radio Waves by Atmospheric Gases." *Proc. IRE* 48 (May 1960), pp. 898–903.

90. Straiton, A. W. "The Absorption and Reradiation of Radio Waves by Oxygen and Water Vapor in the Atmosphere." *IEEE Trans.* AP-23 (July 1975), pp. 595–597.

91. Blake, L. V. "Prediction of Radar Range." In *Radar Handbook*, 1st ed. M. Skolnik (Ed.). New York: McGraw-Hill, 1970, Chap. 2, Sec. 2.7.

92. Baars, J. W. M. "The Measurement of Large Antennas with Cosmic Radio Sources." *IEEE Trans.* AP-21 (July 1973), pp. 461–474.

93. Graf, W., R. N. Bracewell, J. H. Deueter, and J. S. Rutherford. "The Sun as a Test Source for Boresight Calibration of Microwave Antennas." *IEEE Trans.* AP-19 (September 1971), pp. 606–612.

94. Evans, G. E. *Antenna Measurement Techniques.* Boston, MA: Artech House, 1990, Secs. 2.6 and 3.3.3.

95. Schonland, B. F. J. *The Flight of Thunderbolts.* 2nd ed. Oxford, London: Clarendon Press, 1964.

96. Skomal, E. N. "Man-Made Noise in the M/W Frequency Range." *Microwave J.* 18 (January 1975), pp. 44–47.

97. Skomal, E. N. *Man-Made Radio Noise.* New York: Van Nostrand, 1978.

98. Ralston, J., J. Heagy, and R. Sullivan. "Environmental/Noise Effects on VHF/UHF UWB SAR." Institute for Defense Analyses, Alexandria, VA, IDA Paper, P-3385, September 1998.

99. Greene, J. C., and M. T. Lebenbaum. Letter in *Microwave J.* 2 (October 1959), pp. 13–14.

100. Pieramico, A. F., D. A. Rugger, and L. R. Moyer. "The Radar Support System (RSS): A Tool for Siting Radars and Predicting their Performance." *ATC Systems* 2, no. 1 (January/February 1996), pp. 32–40.

101. Lin, C. C., and J. P. Reilly. "A Site-Specific Model of Radar Terrain Backscatter and Shadowing." *Johns Hopkins APL Tech. Digest* 18, no. 3 (July–September 1997), pp. 432–447.

102. Weil, T. A. "Atmospheric Lens Effect; Another Loss for the Radar Range Equation." *IEEE Trans.* AES-9 (January 1973), pp. 51–54.

103. Shrader, W. W., and T. A. Weil. "Lens-Effect Loss for Distributed Targets." *IEEE Trans.* AES-23 (July 1987), pp. 594–595.

104. Goodman, J. M., and J. Aarons. "Ionospheric Effects on Modern Electronic Systems." *Proc. IEEE* 78 (March 1990), pp. 512–528.

105. Brookner, E., Hall, W. M., and R. H. Westlake. "Faraday Loss for L-band Radar and Communcations Systems." *IEEE Trans.* AES-21 (July 1985), pp. 459–469.

106. Flock, W. L. "Propagation Effects on Satellite Systems at Frequencies Below 10 GHz." *NASA Reference Publication 1108* (December 1983), Sec. 2.4.

107. Filer, E., and J. Hartt. "Cobra Dane Wideband Pulse Compression System." *IEEE EASCON '76*, pp. 61-A to 61-M.

PROBLEMS

8.1 In this problem, you may assume a flat earth. (a) What are the elevation angles (in degrees) of the two lowest elevation-pattern multipath interference lobes for an L-band (1300 MHz) radar antenna located at a height 50 ft above a perfectly conducting flat surface? (b) What is the height (in meters) of the peak of the first (lowest) lobe above the earth's surface at a range of 3 nmi? (c) Repeat (a) and (b) for an X-band (9375 MHz) radar antenna. (d) What can you conclude from the above about the detection of low-altitude targets and radar frequency? (e) When might the X-band ship navigation-radar of part (c) have trouble detecting navigation buoys because of multipath lobing, especially if the ship is sailing in calm waters? (You may assume that the echo from the buoy is due to a corner reflector mounted at the top of the buoy 6 m above mean sea level.)

8.2 Under what conditions might the received echo-signal power from a point target located over a flat conducting surface (such as a smooth sea) vary inversely as the eighth power of the range?

8.3 Figure 8.1 illustrates the two paths between the radar and a point target (the direct path and the surface-reflected path). Assume a radar transmits a single short-pulse with pulse duration much less than the time difference between the signals transiting these two paths; i. e., the pulse width is small compared to $2\Delta/c$, where Δ was given by Eq. (8.4). (You may think of the pulse as a delta function that propagates in space.) (a) Sketch the nature of the echo signal received back at the radar after reflection by the point target. (There will be more than one echo returned to the radar.) (b) Derive an expression for the time separation between the pulses? (c) How can this type of short-pulse transmission be used to measure the height of a target?

8.4 The lobes in the elevation pattern due to multipath (such as in Fig. 8.6b) cause loss of target signal when the target is within the null regions of the pattern. What might the radar system designer do to avoid the loss of signal due to the multipath nulls?

8.5 Many air-surveillance radars operate with the antenna beam pointed slightly upward so that the lower half-power point of the elevation pattern is directed along the horizon rather than have the maximum antenna gain along the horizon. Discuss the pros and cons of having the antenna half-power point at the horizon instead of the maximum antenna gain at the horizon.

8.6 What radar characteristics are important for detecting targets at low altitudes?

8.7 The caption of Fig. 8.6b states that this is a plot of the elevation pattern of a 900-MHz radar at an antenna height of 75 ft. Using Eq. (8.8) verify that this is correct based on the antenna pattern shown in Fig. 8.6b. Assume a flat earth and an effective earth's radius of 4/3. (Use of the lowest lobe, however, will probably not give as correct an answer as will the next higher lobe.)

8.8 (a) Show that the distance d to the horizon from a radar at a height h above a spherical earth of effective radius ka is $d = \sqrt{2kah}$. (b) If $k = 4/3$ and the actual radius of the earth $a = 3440$ nmi, what is the distance in nautical miles to the horizon for a radar at a height of 10,000 ft? (c) How much would the distance to the horizon in (b) be increased if the

atmospheric refraction were such that $k = 1.8$ instead of 1.33? (d) If a radar had a free-space range exactly equal to the distance to the horizon, d, why might it not be able to see a target located at the horizon?

8.9 The factor k that describes the modification to the earth's radius to account for atmospheric refraction was given by Eq. (8.16) as

$$k = \frac{1}{1 + a(dn/dh)}$$

where a = radius of the earth and dn/dh is the rate of change of the index of refraction with height. (a) What value of dn/dh results in $k = \infty$? (b) What does it mean when $k = \infty$?

8.10 In Sec. 8.4 of the text it is reported[25] that the range error due to atmospheric refraction is 97.5 ft for a target at 40,000 ft and an elevation angle of 3°. Compare this to the answer you get for the range error when using Eq. (8.19), based on different experimental data. (You might need to use Fig. 8.9. Take the surface refractivity to be 313.)

8.11 (a) Why are radars seldom operated at or near a frequency of 22 GHz or near 60 GHz? (b) What is the two-way attenuation of a radar signal (in dB) in the clear atmosphere at a frequency of 5 GHz when propagating 200 nmi (and back) at an elevation of 0°? (c) What is the two-way attenuation when the elevation angle is increased to 5°? (d) Why are aircraft targets not likely to be detected at long range at zero degrees elevation angle?

8.12 A shipboard military radar for detecting low-altitude missile targets over water might be based on either (1) microwave ducted propagation or (2) HF surface-wave propagation to extend the detection range beyond the horizon. Compare these two radar methods with respect to their effectiveness in performing this task. Include in your comparison the effect of their relative size, reliability for detection under all conditions, accuracy, and anything else you think is appropriate.

8.13 Equation 8.11b assumes that the antenna gain remains constant with frequency (which means the beamwidths remain constant), so the received echo signal power P_r varies as λ^{-2}, where λ = wavelength. How would the echo signal power P_r vary with wavelength if the antenna aperture A_e remained constant with frequency?

chapter

9

The Radar Antenna

9.1 FUNCTIONS OF THE RADAR ANTENNA

The radar antenna is a distinctive and important part of any radar. It serves the following functions:

- Acts as the transducer between propagation in space and guided-wave propagation in the transmission lines.
- Concentrates the radiated energy in the direction of the target (as measured by the antenna gain).
- Collects the echo energy scattered back to the radar from a target (as measured by the antenna effective aperture).
- Measures the angle of arrival of the received echo signal so as to provide the location of a target in azimuth, elevation, or both.
- Acts as a spatial filter to separate (resolve) targets in the angle (spatial) domain, and rejects undesired signals from directions other than the main beam.
- Provides the desired volumetric coverage of the radar.
- Usually establishes the time between radar observations of a target (revisit time).

In addition, the antenna is that part of a radar system that is most often portrayed when a picture of a radar is shown. (More can be learned about the nature of a radar from a picture of its antenna than from pictures of its equipment racks.)

With radar antennas, big is beautiful (within the limits of mechanical and electrical tolerances and the constraints imposed by the physical space available on the vehicle that carries the antenna). The larger the antenna, the better the radar performance, the smaller can be the transmitter, and the less can be the total amount of prime power needed for the radar system. The transmitting antenna gain and the receiving effective aperture are proportional to one another [as given by Eq. (1.8) or Eq. (9.9)] so that a large transmitting gain implies a large effective aperture, and vice versa. As was mentioned in Chap. 1, in radar a common antenna generally has been used for both transmission and reception.

Almost all radar antennas are directive and have some means for steering the beam in angle. Directive antennas mean narrow beams, which result in accurate angular measurements and allow closely spaced targets to be resolved. An important advantage of microwave frequencies for radar is that directive antennas with narrow beamwidths can be achieved with apertures of relatively small physical size.

In this chapter, the radar antenna will be considered as either a transmitting or a receiving antenna, depending on which is more convenient for explaining a particular antenna property. Results obtained for one may be readily applied to the other because of the reciprocity theorem of antenna theory.[1]

Antenna designers have a variety of directive antenna types from which to choose including the reflector antenna in its various forms, phased arrays, endfire antennas, and lenses. They all have seen application in radar at one time or other. These antennas differ in how the radiated beam is formed and the method by which the beam is steered in angle. Steering the antenna beam can be done mechanically (by physically positioning the antenna) or electronically (by using phase shifters with a fixed phased array). The relatively simple *parabolic reflector*, similar to the automobile headlight or the searchlight, in one form or other has been a popular microwave antenna for conventional radars. As will be discussed later in this chapter, a parabolic reflector can be a paraboloid of revolution, a section of a paraboloid, a parabolic cylinder, Cassegrain configuration, parabolic torus, or a mirror scan (also called polarization-twist Cassegrain). There have also been applications of spherical reflectors, but only for special limited purposes.

The mechanically rotating array antenna was the basis for most of the lower frequency air-surveillance radars that saw service early in World War II. They were eventually replaced by parabolic reflector antennas when radar frequencies increased to the microwave region during and just after World War II. In the 1970s the mechanically scanned planar array antenna reappeared, but at microwave frequencies with slotted waveguide radiators or printed-circuit antennas rather than dipoles. The mechanically scanned planar array is found in almost all 3D radar antennas, low sidelobe antennas, and in airborne radars where the antenna is fitted behind a radome in the nose of the aircraft. (A planar aperture allows a larger antenna to be used inside the radome than is practical with a parabolic reflector.) An example of a very low sidelobe rotating planar array used in the AWACS radar is shown later in Fig. 9.49.

Starting in the mid-1960s the electronically steered phased array antenna began to be employed for some of the more demanding military radar applications. It is the most interesting and the most versatile of the various antennas, but it is also more costly and more complex.

9.2 ANTENNA PARAMETERS

Several of the important antenna parameters were introduced in the discussion of the radar equation in Chapters 1 and 2. Here we review and expand on those we have introduced previously and add some other parameters not yet discussed.

Directive Gain *Gain* is a measure of the ability of an antenna to concentrate the transmitted energy in a particular direction. There are two different, but related, definitions of antenna gain. One is the *directive gain*, sometimes called *directivity*. The other is the *power gain*, and is often simply called *gain*. Both gain definitions need to be understood by the radar systems engineer since both are used. The directive gain is descriptive of the nature of the antenna radiation pattern, and is usually the definition of gain that interests the antenna engineer. The power gain is related to the directive gain, but it takes account of loss in the antenna itself. The power gain is more appropriate for use in the radar range equation and is therefore of more interest for the radar engineer.

We will denote the directive gain by G_D. (In other literature it is sometimes denoted by D.) The directive gain of a transmitting antenna may be defined as

$$G_D = \frac{\text{maximum radiation intensity}}{\text{average radiation intensity}} \qquad [9.1]$$

where the radiation intensity is the *power per unit solid angle* radiated in the direction (θ,ϕ), and is denoted $P(\theta,\phi)$. Its units are watts per steradian. A plot of the radiation intensity as a function of the angular coordinates is called a *radiation-intensity pattern*. The power density, or *power per unit area*, when plotted as a function of angle is called the *power pattern*. The power pattern and the radiation-intensity pattern are identical when each is plotted on a relative basis; that is, when the maximum is normalized to a value of unity. When plotted on a relative basis, they are called the *antenna radiation pattern*, or simply *radiation pattern*.

Since the average radiation intensity over the entire solid angle of 4π steradians is equal to the total power radiated by the antenna divided by 4π, the directive gain of Eq. (9.1) can be written as

$$G_D = \frac{4\pi(\text{maximum power radiated per unit solid angle})}{\text{total power radiated by the antenna}} \qquad [9.2]$$

This equation indicates the procedure by which the directive gain may be found from the antenna radiation pattern. The maximum power radiated per unit solid angle is obtained by inspection, and the total power radiated is found by integrating the volume under the radiation pattern. From Eq. (9.2) the directive gain can be expressed as

$$G_D = \frac{4\pi P(\theta,\phi)_{\max}}{\int\int P(\theta,\phi)\, d\theta d\phi} = \frac{4\pi}{B} \qquad [9.3]$$

where B is called the *beam area* and is defined by

$$B = \frac{\int\int P(\theta,\phi)\, d\theta d\theta}{P(\theta,\phi)_{\max}} \qquad [9.4]$$

The beam area B is the solid angle through which all the radiated power would pass if the power per unit solid angle over the entire beam area were equal to $P(\theta,\phi)_{max}$. It defines, in effect, an equivalent antenna pattern. If θ_B and ϕ_B are the half-power (radian) beamwidths in the two orthogonal planes, the beam area B is approximately equal to $\theta_B\phi_B$. Substituting into Eq. (9.3) gives

$$G_D \approx \frac{4\pi}{\theta_B\phi_B} \qquad [9.5a]$$

This is only an approximation and should be used with caution. Another approximation, an improvement on the above, is

$$G_D \approx \frac{\pi^2}{\theta_B\phi_B} \qquad [9.5b]$$

This expression assumes a gaussian beamshape with θ_B and ϕ_B defined as the half-power beamwidths.[2] (This form of the directive gain has been popular with radar meteorologists.) Beamwidths in Eqs. (9.5a and b) are in radians. If the beamwidths are in degrees in Eq. (9.5a), the 4π in the numerator is replaced by 41,253. Equation (9.5a), however, is overly optimistic in that it provides too high a value of directive gain. It applies for a rectangular beam with no sidelobes.

Warren Stutzman,[3] however, recommended that for practical antennas the following "is an excellent approximation for general use:"

$$G \approx \frac{26{,}000}{\theta_B\phi_B} \qquad [9.5c]$$

where the half-power beamwidths are in degrees. He states that "gain [G] is used here instead of directivity, not because of losses . . . [but] to indicate that the formula is appropriate to real antenna hardware, where gain is the parameter used in performance descriptions." This equation states that a one-degree pencil-beam antenna has a gain of 44 dB. It should not be a substitute for more exact analysis, calculation, or measurement; but it is far better than Eqs. (9.5a and b) when nothing further is known about the antenna other than its beamwidths in the principal planes.

Power Gain The power gain, which we denote by G, is similar to the directive gain except that it takes account of dissipative losses in the antenna. (It does not include loss arising from mismatch of impedances or loss due to polarization mismatch.) It can be defined similarly to the definition of directive gain, Eq. (9.2), if the denominator is the net power accepted by the antenna from the connected transmitter, or

$$G = \frac{4\pi(\text{maximum power radiated per unit solid angle})}{\text{net power accepted by the antenna}} \qquad [9.6a]$$

An equivalent definition is

$$G = \frac{\text{maximum radiation intensity from subject antenna}}{\text{radiation intensity from a lossless isotropic radiator with the same power input}} \qquad [9.6b]$$

Whenever there is a choice, the power gain should be used in the radar equation since it includes the dissipative losses introduced by the antenna. The directive gain, which is always greater than the power gain, is more closely related to the antenna beamwidth. The difference between the two antenna gains is usually small for reflector antennas. The power gain and the directive gain are related by the radiation efficiency ρ_r as follows

$$G = \rho_r G_D \tag{9.7}$$

The radiation efficiency is also the ratio of the total power radiated by the antenna to the net power accepted by the antenna at its terminals. The distinction between the two definitions of gain often can be ignored in practice, especially when the dissipative loss in the antenna is small.*

The definitions of power gain and directive gain described in the above were given in terms of a transmitting antenna. Because of reciprocity the pattern of a receiving antenna is the same as the pattern of a transmitting antenna, so the receiving antenna can be described by a gain just as can the transmitting antenna. This is why one can talk of a receiving gain even though gain was defined in terms of a transmitting antenna. The effective aperture of a receiving antenna, on the other hand, has no similar attribute in a transmitting antenna.

It should be kept in mind that the accuracy with which the gain of a radar antenna can be measured is usually about ± 0.5 dB.[4] Thus one should not specify or quote antenna gains to an accuracy much better than this unless there is a reason to be more accurate.

Antenna Radiation Pattern In the above, antenna gain meant the maximum value. It is also common to speak of gain as a function of angle. Quite often the ordinate of a radiation pattern is given as the gain as a function of angle, normalized to unity. It is then known as *relative gain*. Unfortunately the term *gain* is used to denote both the maximum value and the gain as a function of angle. Uncertainty as to which usage is meant can usually be resolved from the context.

An example of an antenna radiation pattern for a paraboloidal reflector antenna is shown in Fig. 9.1.[5] This particular pattern might not be representative of a well-designed modern high-gain antenna, but it does illustrate the various features that a simple reflector-antenna radiation pattern might have. The *main beam* is shown at zero degrees. The remainder of the pattern outside the main beam is the *sidelobe* region. As the angle increases from the direction of maximum gain, there is an irregularity in this particular radiation pattern at about 22 dB below the peak. This is called a *vestigial lobe* or "shoulder" on the side of the main beam. It does not appear in all radiation patterns and is not desired since it is indicative of phase errors in the aperture illumination. Normally, when errors in the aperture illumination are small, the first sidelobe appears near where this vestigial lobe is indicated rather than where the first sidelobe is indicated in the figure.

The near-in sidelobes generally decrease in magnitude as the angle increases. The decrease is determined by the shape of the aperture illumination (as described in the next

*In some types of phased array antennas, the losses in the phase shifters and the power dividing networks can be quite high so that the difference between the power gain and the directive gain can be significant. In such cases, the directive gain is what is usually quoted and the losses are accounted for elsewhere.

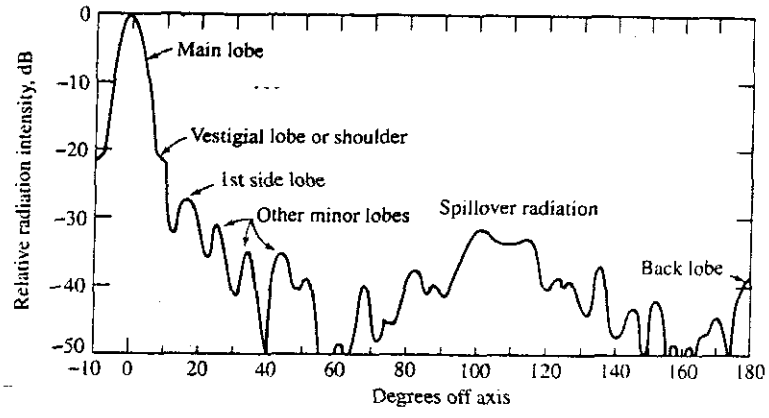


Figure 9.1 Radiation pattern for a particular paraboloid reflector antenna illustrating the main beam and sidelobe radiation. (After Cutler et al.,[5] Proc. IRE.)

section). Eventually, the sidelobes due to the aperture illumination are masked by sidelobes due to the random errors in the aperture [Sec. (9.12)]. With a conventional reflector antenna, there usually will be spillover radiation from that part of the feed radiation pattern that is not intercepted by the reflector (in the example of Fig. 9.1, this appears from about 100 to 115°). This radiation pattern also has a pronounced lobe in the backward direction (180°) due to diffraction around the edges of the reflector as well as direct leakage through the mesh reflector surface (if the surface is not solid).

The radiation pattern shown in Fig. 9.1 is plotted as a function of one angular coordinate, but the actual pattern is a plot of the radiation intensity $P(\theta,\phi)$ as a function of two angles. The two angle coordinates commonly employed with a ground-based radar antenna are azimuth and elevation, but other appropriate angle coordinates also can be used.

A complete three-dimensional plot of the radiation pattern can be complicated to display and interpret, and is not always necessary. For example, an antenna with a symmetrical pencil-beam pattern can be represented by a single plot in one angle coordinate because of its circular symmetry. The radiation intensity pattern for rectangular or rectangular-like apertures can often be written as the product of the radiation-intensity patterns in the two coordinate planes; for instance,

$$P(\theta,\phi) = P(\theta,0)\ P(0,\phi) \qquad [9.8]$$

Thus when the pattern can be expressed in this manner, the complete radiation pattern in two coordinates can be determined from the two single-coordinate patterns in the θ and in the ϕ planes.

Effective Aperture The effective aperture of a receiving antenna is a measure of the effective area presented to the incident wave by the antenna. As was given previously as Eq. (1.8), the transmitting gain G and receiving effective area A_e of a lossless antenna are related by

$$G = \frac{4\pi A_e}{\lambda^2} = \frac{4\pi \rho_a A}{\lambda^2} \qquad [9.9]$$

where λ = wavelength, ρ_a = antenna aperture efficiency, A = physical area of the antenna, and $A_e = \rho_a A$. The aperture efficiency depends on the nature of the current illumination across the antenna aperture. With a uniform illumination, $\rho_a = 1$. The advantage of high efficiency obtained with a uniform illumination is tempered by the radiation pattern having a relatively high peak-sidelobe level. An aperture illumination that is maximum at the center of the aperture and tapers off in amplitude towards the edges has lower sidelobes but less efficiency than the uniform illumination.

Sidelobe Radiation Sidelobe radiation is radiation from an antenna that is not radiated by the main beam. It is possible in theory to have an antenna radiation pattern with only a single main beam and no sidelobes, but not only is this impractical to achieve, it is not desirable since the main beam would be unusually wide (as mentioned later in Sec. 9.13). In general, the lower the sidelobes, the lower will be the antenna gain and the aperture efficiency, and the greater will be the width of the main beam.

The sidelobe level of an antenna may be described by the value of the peak sidelobe, the rms value of all the sidelobe radiation (usually of the dB values), or some other suitable measure. The peak sidelobe is a good measure of sidelobe behavior for purposes of the radar system engineer. For a line-source aperture with a uniform illumination, the peak sidelobe (which is the first sidelobe) is 13.2 dB down from the maximum value of the main beam, or a value of -13.2 dB. This is usually too high for most radar applications even though its aperture efficiency is unity, that is, 100 percent.

Low antenna sidelobes are desired in a radar so as to avoid detecting large targets when they are illuminated by the antenna sidelobes. Any echoes received from the sidelobes will not be indicated by their true angle. (The angle assigned to an echo from a sidelobe will be the angle at which the main beam points at the time of detection rather than the angle of the sidelobe which illuminates the target.) Low sidelobes are also useful for minimizing the effect of strong jamming and interference and to reduce the large clutter echoes that can enter the antenna sidelobes of a high-prf pulse doppler radar (as was discussed in Sec. 3.9).

The highest sidelobe of an antenna is usually the first sidelobe adjacent to the main beam. A typical parabolic reflector antenna fed from a waveguide horn might have a peak sidelobe of -23 to -28 dB. Peak sidelobes of -40 to -50 dB are possible with specially designed array antennas. Peak sidelobes less than -50 dB are sometimes called *ultralow sidelobes.*

Aperture Efficiency The aperture efficiency is based on the maximum radiation intensity, which usually occurs at the center of the main beam. It is not like the radiation efficiency that is a measure of the energy dissipated as the signal travels through the antenna. A radiation efficiency less than unity means that energy is lost. On the other hand, an aperture efficiency less than unity means that the radiated energy is redistributed in angle rather than be dissipated. For example, consider a line-source antenna aperture-illumination proportional to $\cos^2(\pi z/2)$, where z is the distance from the center of the aperture, $-D/2 \leq z \leq +D/2$, and D is the aperture dimension. In this example the amplitude of the aperture illumination is one-half cycle of the square of the cosine over the aperture. Its radiation pattern has a first sidelobe of -32 dB compared to the -13.2 dB of a uniform

illumination. The gain is reduced by 0.67 and its beamwidth is increased by 1.63 compared to the radiation pattern from a uniformly illuminated antenna. With a scanning antenna, as might be used in a surveillance radar, the reduction in gain is compensated in part by the increased number of pulses received because of the wider beamwidth.

There are other antenna properties that might be more important to the radar systems engineer than the aperture efficiency. Thus the aperture efficiency might be reduced in order to obtain other benefits. These might include low-sidelobe levels, a radar antenna beam which maximizes the radiated energy within a specified angular region,[6,7] shaped beams such as the cosecant-squared pattern, and monopulse antenna patterns optimized for good angle-tracking accuracy. The aperture efficiency might be important to the antenna designer, but to the radar systems engineer it is often something to be traded in order to achieve some more important antenna characteristic.

Polarization The polarization of an electromagnetic wave is defined by the orientation of the electric field. Most radar antennas are *linearly polarized*, with the orientation of the electric field being either horizontal or vertical. Air-surveillance radars generally employ horizontal polarization. Most tracking radars are vertically polarized. *Circular polarization* occurs when the electric field rotates at a rate equal to the RF frequency. It is sometimes used to enhance the detectability of aircraft targets in the midst of rain (Sec. 7.8). There is also *elliptical polarization*, where the electric field also rotates at the RF frequency; but unlike circular polarization, the amplitude of the elliptically polarized electric field varies during the rotation period. Circular and linear polarizations are special cases of elliptical polarization. Although some radar applications seem to prefer a particular polarization (based sometimes on tradition), in many applications there is often not a strong requirement for one polarization over the other. Even the use of circular polarization for rain is not absolutely necessary since orthogonal linear polarizations can be used instead.

9.3 ANTENNA RADIATION PATTERN AND APERTURE ILLUMINATION

The electric-field intensity $E(\phi)$ (units of volts per meter) produced by the electromagnetic radiation emitted from a line-source antenna is a function of the amplitude and phase of the distribution of current across the aperture. The angle ϕ, shown in Fig. 9.2, is with respect to the normal to the center of the antenna aperture. $E(\phi)$ may be found by adding vectorially the individual contributions radiated from the various current elements that constitute the line-source antenna aperture. The mathematical summation at a point in space of all the contributions radiated by the current elements contained within the aperture gives the field intensity in terms of an integral that is difficult to evaluate in the general case.[8] It reduces, however, to a conventional inverse Fourier transform when the distances from the antenna are large enough for the radiation to be considered a plane wave. This occurs in the so-called *far field* of the antenna, when the range $R > D^2/\lambda$, where D is the size of the aperture and λ is the radar wavelength, with D and λ being in the same units. (Although antenna engineers call this region the far field, optical physicists call it the *Fraunhofer region*.) In radar, the target is almost always in the far field.

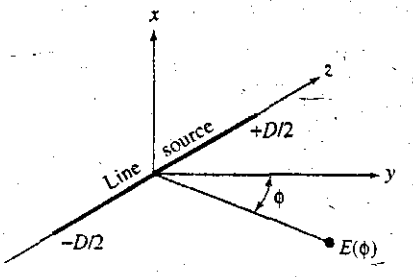


Figure 9.2 Coordinate system for a line source lying along the z axis. Field-intensity pattern $E(\phi)$ lies in the yz plane.

Electric-Field Intensity and the Fourier Transform In Fig. 9.2, the width of the aperture in the z coordinate is D and the angle in the y,z plane as measured from the y axis is ϕ. The aperture is a line source, or linear antenna, in that its dimension in the z-coordinate is much larger than its dimension in the x-coordinate, and the latter (x-coordinate dimension) is small compared to a quarter wavelength. We are interested in the electric-field intensity in the y,z plane. Assuming $D >> \lambda$ and $R > D^2/\lambda$, the variation of the electric field intensity with angle ϕ in the far field is proportional to

$$E(\phi) = \int_{-D/2}^{+D/2} A(z) \exp\left(j2\pi \frac{z}{\lambda} \sin \phi\right) dz \qquad [9.10]$$

where the *aperture illumination* $A(z)$ is the current at a distance z from the center of the radiating line-source antenna. The aperture illumination, also called the *current distribution*, can be a complex quantity including both an amplitude $|A(z)|$ and a phase $\psi(z)$, so that $A(z) = |A(z)| \exp [j\psi(z)]$. The phase of the aperture illumination becomes important if the beam is to be steered in a direction other than broadside or if the antenna is focused (which is rare in radar applications). Here we assume $\psi(z) = 0$.

The electric-field intensity given by Eq. (9.10) applies to a one-dimensional radiating antenna lying along the z axis. If the aperture were two-dimensional and situated in the x,z plane, the aperture illumination, $A(z)$, would be the integral of $A(x,z)$ over the variable x. The electric field in the far field is a function only of the angle ϕ and does not depend on the range R except for the usual $1/R$ factor. The expression of Eq. (9.10) can be extended to two dimensions by considering the aperture illumination to be a function of x as well as z. The plot of the magnitude of the electric-field intensity $|E(\theta,\phi)|$ is called the *electric-field intensity pattern* of the antenna. The plot of the square of the field intensity $|E(\theta,\phi)|^2$ normalized to unity is the *power pattern* or the *radiation pattern.*

The integral describing the electric-field intensity of a radiating source in the far field is an inverse Fourier transform, which most electrical engineers are familiar with since the Fourier transform relates the frequency spectrum and waveform of a temporal signal. The Fourier transform of a time waveform $s(t)$ is the frequency spectrum

$$S(f) = \int_{-\infty}^{+\infty} s(t) \exp(-j2\pi ft)\, dt \qquad [9.11]$$

and the inverse Fourier transform of the spectrum $S(f)$ is the time waveform

$$s(t) = \int_{-\infty}^{+\infty} S(f) \exp(j2\pi ft)\, df \qquad [9.12]$$

This happens to be of the same form as the field-intensity expression of Eq. (9.10). Since the aperture illumination is zero beyond $z = \pm D/2$, the limits of Eq. (9.10) can extend from $-\infty$ to $+\infty$ to make it consistent with the inverse Fourier transform of Eq. (9.12). Thus the mathematical model that relates the time waveform and its spectrum is analogous to the mathematical model that relates the radiated field-intensity and the aperture illumination. The (spatial) antenna pattern $E(\phi)$ is related mathematically to the (temporal) waveform $s(t)$, and the aperture illumination $A(z)$ is related to the spectrum $S(f)$. What is known from signal theory about the role of the spectrum $S(f)$ in determining the nature of the signal $s(t)$ is applicable to how the aperture illumination $A(z)$ affects the radiation in space $E(\phi)$. The converse is also true.

In the above, the antenna was viewed as transmitting. As was stated earlier, the property of antenna reciprocity means that the variation of the radiated field on transmit as a function of angle will be similar to the variation of the received signal as a function of angle when the same antenna is used for both transmit and receive.

In the remainder of this section, the antenna field intensity will be examined for various analytical aperture illuminations. The phase distribution is assumed zero or constant so that only the effects of the amplitude variation across the aperture need be considered. The aperture over which the integral of Eq. (9.10) is taken is defined as the projection of the antenna surface on a plane perpendicular to broadside. In this formulation of antenna radiation based on the inverse Fourier transform of Eq. (9.10) it does not matter whether the illumination is produced by a reflector antenna, a lens, or an array so long as the illumination is that in the plane of the aperture.

One-Dimensional Aperture Illumination Consider in Fig. 9.2 a uniform (constant) aperture illumination extending from $-D/2$ to $+D/2$, and zero outside. This represents the illumination across a line source or the projected illumination in one of the principal planes of a uniformly illuminated rectangular aperture. If the constant value of the amplitude of the aperture illumination is A_0, the variation of the electric-field intensity as a function of angle ϕ is computed from Eq. (9.10) as

$$E(\phi) = A_0 \int_{-D/2}^{+D/2} \exp\left(j2\pi \frac{z}{\lambda} \sin\phi\right) dz = \frac{A_0 D \sin[\pi(D/\lambda)\sin\phi]}{\pi(D/\lambda)\sin\phi} \qquad [9.13]$$

Normalizing to make $E(0) = 1$ results in

$$E(\phi) = \frac{\sin[\pi(D/\lambda)\sin\phi]}{\pi(D/\lambda)\sin\phi} \qquad [9.14]$$

This is of the form $(\sin x)/x$. The square of the above is the antenna radiation pattern or power pattern. It is shown by the solid curve in Fig. 9.3. The first sidelobe adjacent to the main beam is 13.2 dB below the peak value of the main beam. The angular distance between the two nulls defining the main beam is $2\lambda/D$ radians, and the beamwidth as

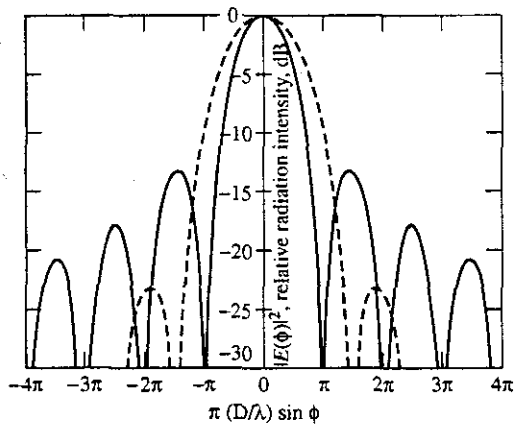


Figure 9.3 Solid curve is the radiation pattern produced by a uniform aperture illumination of a line source of dimension D. The dashed curve is the radiation pattern of an aperture illumination proportional to the cosine function. Both curves are normalized to unity maximum gain.

measured between the half-power points is $0.88\lambda/D$ radians, or $51\lambda/D$ degrees. (Quite often the half-power beamwidth for this uniform illumination is approximated by λ/D radians.) The field-intensity pattern of Eq. (9.14) is positive over the entire main lobe, but changes sign in passing through the first zero, returning to a positive value in passing through the second zero, and so on. The odd-numbered sidelobes are therefore out of phase with the main beam, and the even-numbered ones are in phase.

The normalized field-intensity pattern for an aperture illumination $A(z)$ proportional to one-half cycle of the cosine function, given by $\cos(\pi z/D)$, with $|z| \leq D/2$, is from Eq. (9.10)

$$E(\phi) = \frac{\pi}{4}\left[\frac{\sin(\psi + \pi/2)}{\psi + \pi/2} + \frac{\sin(\psi - \pi/2)}{\psi - \pi/2}\right] = \frac{\pi^2 \cos\psi}{\pi^2 - 4\psi^2} \qquad [9.15]$$

where $\psi = \pi(D/\lambda)\sin\phi$. The square of the above is shown as the dashed curve in Fig. 9.3. In this figure, the peak gains of both patterns (for the uniform and the cosine illuminations) are normalized to unity. In reality, however, the maximum gain of the pattern from the cosine function is 0.9 dB less than that of the maximum gain of the uniform illumination. Notice that the peak sidelobe of the pattern from the cosine function is much lower than the peak sidelobe from the uniform illumination. Its beamwidth, however, is increased and its maximum gain decreased. The greater the taper of the aperture illumination as it approaches the edges of the antenna aperture, the lower will be the sidelobe level, but at the cost of a wider beamwidth and a lower maximum gain.

Table 9.1 lists some on the characteristics of the radiation patterns produced by various one-dimensional (line-source) antenna aperture-illuminations.[8] These aperture illuminations are expressed in analytic form so that the solution of the inverse Fourier transform of Eq. (9.10) can be conveniently determined. They are not necessarily what might be employed by the antenna designer, but they do illustrate how variations in the form of the aperture illumination affect the antenna pattern. More complicated distributions, which cannot be found from available tables of Fourier transforms, may be determined by computer computation. Using the Schwartz inequality, Silver[8] showed that the uniform

Table 9.1 Radiation-pattern characteristics produced by various aperture distributions

λ = wavelength; D = aperture wideth

Type of distribution, $\lvert z\rvert < 1$	Relative gain	Half-power beamwidth, deg	Intensity of first sidelobe, dB below maximum intensity
Uniform; $A(z) = 1$	1	$51\lambda/D$	13.2
Cosine; $A(z) = \cos^n(\pi z/2)$:			
$n = 0$	1	$51\lambda/D$	13.2
$n = 1$	0.810	$69\lambda/D$	23
$n = 2$	0.667	$83\lambda/D$	32
$n = 3$	0.575	$95\lambda/D$	40
$n = 4$	0.515	$111\lambda/D$	48
Parabolic; $A(z) = 1 - (1 - \Delta)z^2$:			
$\Delta = 1.0$	1	$51\lambda/D$	13.2
$\Delta = 0.8$	0.994	$53\lambda/D$	15.8
$\Delta = 0.5$	0.970	$56\lambda/D$	17.1
$\Delta = 0$	0.833	$66\lambda/D$	20.6
Triangular; $A(z) = 1 - \lvert z\rvert$	0.75	$73\lambda/D$	26.4
Circular; $A(z) = \sqrt{1 - z^2}$	0.865	$58.5\lambda/D$	17.6
Cosine-squared plus pedestal:			
$0.33 + 0.66\cos^2(\pi z/2)$	0.88	$63\lambda/D$	25.7
$0.08 + 0.92\cos^2(\pi z/2)$, Hamming	0.74	$76.5\lambda/D$	42.8

aperture illumination produces the maximum gain. When either the $\cos^n$ (cosine raised to the nth power) or the parabolic distributions shown in this table are examined, it is seen that, as has mentioned before, the more tapered the illumination the lower is the peak sidelobe, the wider the beamwidth, and the lower the maximum gain. Note that relative gain in Table 9.1 is the same as the aperture efficiency ρ_a defined previously by Eq. (9.9).

The cosine-squared on a pedestal listed one line from the bottom of the table is a representative illumination for conventional antennas. This is close to the so-called 25-dB Taylor illumination discussed later in Sec. 9.11. The Hamming illumination produces the lowest peak sidelobe for a cosine-squared on a pedestal illumination (in this case the peak lobe is not the one closest to the main beam). The reduction of the spatial sidelobes of an antenna by a tapered aperture illumination is similar to the windowing employed in digital processing to reduce filter sidelobes and to the filter weighting in pulse compression to reduce the time sidelobes of the compressed pulse (Sec. 6.5).

Having the proper aperture illumination is an important requirement for achieving suitable antenna radiation patterns. There is more to consider, as will be discussed in Sec. 9.13, which is on the subject of very low sidelobes.

Two-Dimensional Aperture Illumination To extend the above discussion of radiation from a line source to an aperture with two dimensions, the angles θ,ϕ are defined by the coordinate system shown in Fig. 9.4. The antenna is in the x,y plane. This is the coordinate

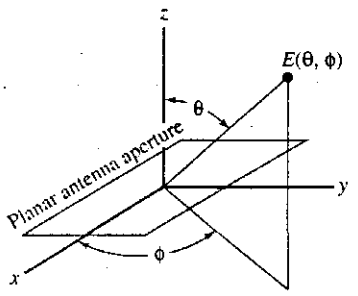


Figure 9.4 Coordinate system for a planar antenna lying in the *xy* plane.

system usually used by antenna theorists, and differs from the azimuth-elevation coordinates preferred by radar engineers. With the coordinate system of Fig. 9.4, the two-dimensional field-intensity pattern from an aperture is given by[8]

$$E(\theta,\phi) = \iint A(x,y) \exp[j(2\pi/\lambda) \sin\theta\,(x\cos\phi + y\sin\phi)]\,dxdy \qquad [9.16]$$

This integral is not easy to solve analytically, so that numerical techniques are sometimes used. Equation (9.16) is easier to use when the aperture illumination is separable; that is, when $A(x,y) = A_1(x)\,A_2(y)$, where $A_1(x)$ is the projection of the aperture illumination along the x axis and $A_2(y)$ is the projection along the y axis. Silver[9] showed that when aperture illuminations are separable, the field-intensity patterns are also separable. Thus the two-dimensional pattern of Eq. (9.16) can be written as the product of the one-dimensional patterns in the two principal planes, as was indicated earlier by Eq. (9.8). One principal plane occurs when $\phi = 0$, and the other when $\phi = 90°$ in the coordinate system of Fig. 9.4. When the patterns are separable, the pattern in the xz plane is the same as would be produced by a linear antenna with aperture illumination $A_1(x)$, and the pattern in the yz plane is the same as that produced by the linear aperture illumination $A_2(y)$.

Circular Aperture[10–12] Instead of the rectangular coordinates used in Eq. (9.16), polar coordinates are used to describe the aperture illumination $A(r,\theta)$ of a circular aperture. The radial distance from the center of the circular aperture is r, and θ is the angle measured in the plane of the aperture with respect to a reference. When the aperture illumination depends only on the radial distance and is independent of the angle θ, the field intensity is proportional to

$$E(\theta) = 2\pi\int_0^{r_0} A(r)\,J_0[2\pi(r/\lambda)\sin\phi]\,rdr \qquad [9.17]$$

where $A(r)$ is the aperture illumination as a function of the radial distance, r_0 is the radius of the circular aperture, and ϕ is the angle with respect to the normal to the circular aperture. If the aperture illumination is uniform [$A(r) = 1$], this reduces to

$$E(\theta) = 2\pi r_0^2\,J_1(\xi)/\xi \qquad [9.18]$$

where $\xi = 2\pi(r_0/\lambda)\sin\phi$ and $J_1(\xi)$ is the first-order Bessel function. A normalized plot of the square of this equation is shown in Fig. 9.5. The first sidelobe is 17.5 dB below

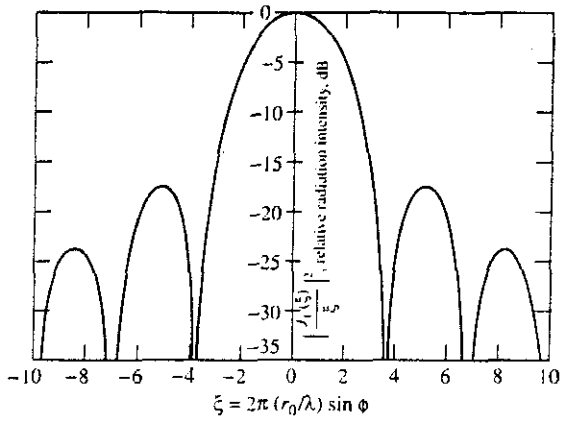


Figure 9.5 Radiation pattern for a uniformly illuminated circular aperture of radius r_0.

the main-beam maximum, and the beamwidth in degrees is $58.5\lambda/D$. Note that this is the same as from a one-dimensional (line source) antenna with a circular aperture-illumination (as was listed in Table 9.1) since the projection of the uniform circular-aperture-illumination on its diameter is a circular one-dimensional illumination.

Tapering of the amplitude illumination in the radial dimension of a circular aperture reduces the peak sidelobe, but at the expense of broader beamwidth and less antenna gain. Consider the family of circularly symmetrical aperture illuminations[8] given by

$$A(r) = [1 - (r/r_0)^2]^p \qquad \textbf{[9.19a]}$$

where $p = 0, 1, 2, \ldots$. This aperture illumination depends only on r and not on θ. The field-intensity pattern is[12]

$$E(\theta) = \pi r_0^2 \, 2^p p! \, \frac{J_{p+1}(\xi)}{\xi^{p+1}} \qquad \textbf{[9.19b]}$$

where ξ is defined as it was for Eq. (9.18). When $p = 0$, the illumination is uniform and the radiation pattern reduces to that given by Eq. (9.18). For $p = 1$, the gain is 0.75 of the gain of a uniformly illuminated aperture, the half-power beamwidth is broadened to $72.6\lambda/D$, and the first sidelobe is 24.6 dB below the maximum. The sidelobe level is 30.6 dB down for $p = 2$, and the gain is 0.56 relative to that of the uniform illumination.

Aperture Blocking[13] An obstacle in front of an antenna can alter the effective aperture illumination and distort the radiation pattern. This is called *aperture blocking* or *shadowing*. Examples are the feed and its supports in a reflector antenna; masts on board a ship; and nearby buildings, trees, and other obstructions to a land-based radar. The subreflector, as well as the feed, of a Cassegrain antenna (Sec. 9.4) also blocks the aperture illumination. Aperture blocking lowers the antenna gain, raises the sidelobes, and fills in the nulls. It would not be unusual for a low-sidelobe antenna with −30 to −40 dB peak sidelobe level to be increased to a sidelobe level of from −15 to −20 dB when its beam is obstructed.

The effect of aperture blocking can be approximated by subtracting from the antenna pattern of the undisturbed aperture the antenna pattern produced by the shadow of an obstacle. This procedure is possible because of the linearity of the Fourier transform that relates the aperture illumination and the radiation pattern. An example of the effect of aperture blocking caused by the feed in a paraboloid-reflector antenna is illustrated in Fig. 9.6.[14] This relatively simple method for determining the effect of blocking a portion of the radiated energy is only an approximation.

The reduction of the antenna gain η due to the blockage of a circular obstacle of radius r_b placed in front of a circular aperture of radius r_0 whose aperture illumination is given by Eq. (9.19a) is

$$\eta = \left[1 - \delta^2\{[(1 - \delta^2)^p]p + 1\}\right]^2 \qquad [9.20]$$

where $\delta = r_b/r_0$ and p is defined by Eq. (9.19a). This equation, due to Sciambi,[15] would apply to a circular feed at the focus of a parabolic reflector or to the subaperture of a Cassegrain antenna. (Sciambi included in his paper the aperture illumination of Eq. (9.19a) on a pedestal, but it was omitted here for simplicity.) When $\delta = r_b/r_0$ is small, then $\eta \approx [1 - (p + 1)\delta^2]^2$. Based on this approximation, the new sidelobe level due to aperture blocking can be written as

$$sl_b = \left(\frac{\sqrt{sl} + (p + 1)\delta^2}{1 - (p + 1)\delta^2}\right)^2 \qquad [9.21]$$

where sl is the original sidelobe level relative to the main-beam peak when there is no aperture blockage. The sidelobes sl_b and sl are power ratios less than one. In obtaining the above, the maximum value of the obstacle pattern based on Eq. (9.19b) was taken as $\pi r_b^2/2$ (with $p = 0$ and replacing r_0 with r_b).

As an example, consider a parabolic aperture illumination as in Eq. (9.19a) with $p = 1$ and $\delta = r_b/r_0 = 0.1$ (one-percent of the antenna area is blocked). The reduction in gain due to blockage from Eq. (9.20) is 0.96 (about 0.2 dB), and the peak sidelobe of the antenna from Eq. (9.21) is increased from −24.6 dB to −21.9 dB. When $\delta = 0.2$ (4 percent blockage), the peak sidelobe is increased to −16.4 dB. Thus with antennas having

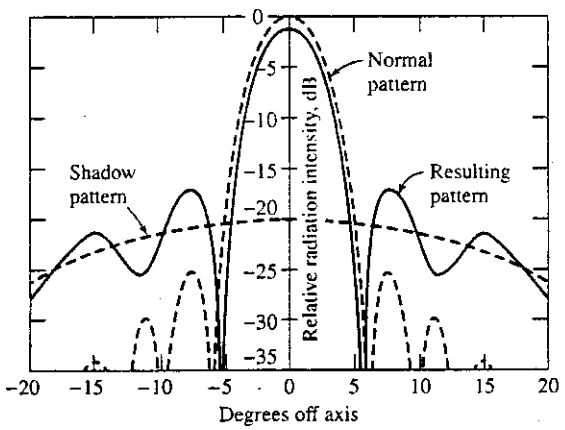


Figure 9.6 Effect of aperture blocking caused by an obstacle (such as a feed) in a parabolic-reflector antenna. (From C. Cutler[5] Proc. IRE.)

conventional sidelobes of −23 to −28 dB, the aperture blockage should not be more than about 1 percent in order to maintain decent sidelobes. With low peak sidelobes of −40 dB, the aperture blockage should be less than 0.05 percent, and with ultralow sidelobe levels (Sec. 9.13), no aperture blockage at all can be tolerated.

9.4 REFLECTOR ANTENNAS

The parabola, sketched in Fig. 9.7, works well as a reflector of electromagnetic energy and has been the basis for many radar antennas. The parabolic surface is illuminated by a source of radiated energy called the *feed*, which is placed at the focus of the parabola. The parabola converts the spherical wave radiated from the feed to a plane wave because (1) any ray radiated from the focus that intersects the parabolic surface is reflected in a direction parallel to the axis of the parabola, and (2) the distance traveled by any ray from the focus to the parabola and by reflection to a plane perpendicular to the parabola's axis is the same for all rays no matter what angle they emanate from the focus. (This description is only an approximation based on geometric optics. In practice a plane wave does not emerge after reflection until the wave travels a sufficient distance to be in the far field of the antenna, but this need not be of concern at present.)

Paraboloid There are several ways in which the parabola is used for antennas. Rotating the parabolic curve shown in Fig. 9.7 about its axis produces a surface which is a parabola of revolution called a *circular parabola;* or, more usually, a *paraboloid.* When properly illuminated by a source at the focus, the paraboloid generates a nearly symmetrical pencil-beam antenna pattern. This has been a popular antenna for tracking radars. (The paraboloid reflector is sometimes called a *dish.*)

Section of a Paraboloid Instead of a circular shape, consider the reflector antenna to have an elliptical shape (as though an elliptical section were cut from the symmetrical paraboloid). This produces an asymmetrical beam shape known as a *fan beam.* It is often used

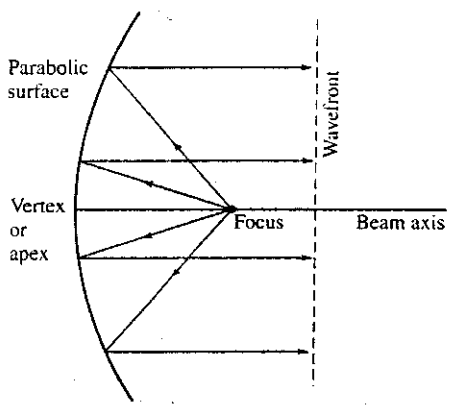


Figure 9.7 Contour of a parabolic-reflector antenna.

for two-dimensional (range and azimuth angle) air-surveillance radars, as shown in the example of Fig. 9.8. Sometimes this type of asymmetrical antenna has a different curvature in the horizontal and vertical planes so as to shape the beams differently in the two planes. This might be the case for an antenna used with an air-surveillance radar where the azimuth beamwidth is required to be narrow, and the vertical beam is shaped to provide a broader coverage, such as the cosecant-squared pattern discussed in Sec. 9.11.

Feeds for Paraboloids The "ideal" feed for a paraboloid reflector would be a source at the focus with a radiation pattern that (1) had no phase variation with angle, (2) produced on the reflector surface the desired aperture amplitude illumination, and (3) had a directivity that allowed all of the feed radiation to be intercepted by the aperture without spillover. The ideal may be approximated but never fully accomplished. The radiation pattern produced by the feed is called the *primary pattern* and that radiated by the aperture is called the *secondary pattern.*

A simple half-wave dipole with a parasitic reflector to direct most of its energy towards the antenna aperture can be used as the feed for a paraboloid, Fig. 9.9a. A dipole, however, is of limited utility as a reflector feed since it is difficult to shape the primary pattern and it is limited in power handling, especially at the higher microwave frequencies. An open-ended waveguide is usually preferred over a dipole for microwave-radar reflector feeds. A circular paraboloid, for example, might be fed by a circular, open-ended waveguide operating in the TE_{11} mode. A rectangular guide operating in the dominant TE_{10} mode, however, does not result in a perfectly symmetrical secondary pattern since its dimensions in the E and H planes are different.

When more directivity is required from the feed than is available from an open-ended waveguide, some form of waveguide horn can be used. A horn can be made to provide the asymmetrical feed illuminations (the primary pattern) for a fan beam generated by a

Figure 9.8 Signaal LW08 2D (Jupiter) L-band fan-beam air and surface surveillance radar with an elliptical shape fan-beam antenna fed by a horn (on the left). The reflector surface is a mesh, the horizontal antenna mounted on the top is for IFF (identification friend or foe). At the middle right can be seen a fin which is added to the back of the antenna to counterbalance the wind forces on the reflector in the position of the worst yawing moment.
(Courtesy Hollandse Signaalapparten B. V., The Netherlands.)

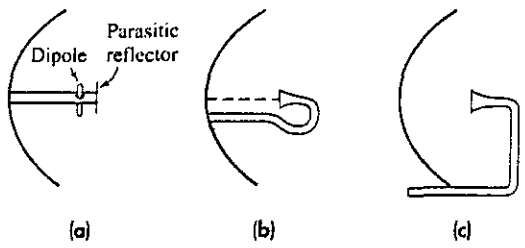


Figure 9.9 Examples of the placement of the feeds in parabolic reflectors. (a) Rear feed using half-wave dipole with parasitic reflector; (b) rear feed using horn; (c) front feed using horn.

section of a paraboloid. Feeds for reflector antennas can come in many varieties in addition to the simple horn and open-ended waveguide.[16]

As an approximate rule of thumb, the intensity of the radiation from the feed toward the edge of the reflector should be about 10 dB down from the maximum radiation. The aperture illumination at the edges of the reflector surface will be even less than this because of the longer path length from the feed to the edge compared to the path length from the feed to the center of the reflector. When the primary feed pattern is 10 dB down toward the edges, the first sidelobe in the secondary pattern usually is in the vicinity of 22 to 25 dB.

The *f/D ratio* of a reflector antenna is the focal length f divided by the aperture diameter D. Most practical reflector antennas have f/D ratios ranging from 0.3 to 0.5. A small ratio means a deep reflector that is difficult to illuminate properly. A large f/D ratio results in a shallow reflector. The shallow reflector is easier to support and to position mechanically, but the feed must be supported farther from the reflector. The farther the feed is from the reflector, the narrower must be the primarily pattern (to avoid spillover loss) and the larger must be the feed. A large f/D is preferred for tracking radars and when the beam must be offset in angle by displacing the feed from the focus.

Calculations of the antenna efficiency based only on the aperture illumination established by the primarily pattern from a feed as well as the spillover indicate theoretical efficiencies of about 80 percent compared to an ideal uniformly illuminated aperture without spillover. In practice, however, phase variations across the aperture, poor polarization characteristics, and antenna mismatch reduce the overall antenna efficiency to the order of 55 to 65 percent for ordinary paraboloidal-reflector antennas.

Feed Support The dipole and the waveguide horn (or open-ended waveguide) can be arranged to feed the paraboloid from the rear as shown in Fig. 9.9a and b. Other types of rear-feed systems have also been used. Figure 9.9c illustrates what is called a front feed using a horn radiator at the focus. It is well suited for supporting horn feeds, but the supports obstruct the aperture.[17] These obstructions due to the feed and its supports reduce the antenna gain, increase the sidelobes, and cause some of the radiated energy to be cross polarized. Analytical expressions and design curves for determining the adverse effects of aperture blockage have been proposed.[18] There is also an impedance mismatch at the feed due to some of the energy reflected by the antenna surface re-entering the feed and its transmission line. Both aperture blockage and mismatch due to reflection can be eliminated by the offset feed.

Offset-Fed Reflector[19,20] As seen in Fig. 9.10, the feed in this arrangement is placed at the focus of the parabola, but the horn is tipped (upwards in the figure) with respect to the parabola's axis. The lower half of the parabolic surface is removed, leaving that portion shown by the solid curve in the figure. The feed is therefore outside the path of the energy reflected from the antenna surface. There is no pattern deterioration due to aperture blocking nor is there any significant amount of radiation intercepted by the feed to produce an impedance mismatch (high VSWR).

Although the offset feed eliminates aperture blockage and mismatch of rear and front feeds, it introduces problems of its own. Its f/D ratio (focal length divided by diameter) is greater than that of conventional paraboloids so that the feeds are larger. Furthermore, this type of antenna is generally more difficult to support mechanically. Because of the increased asymmetry of this geometry, when illuminated by a conventional linearly polarized feed, cross-polarized radiation lobes are produced which can reduce radar system performance by indicating false targets. It has been said[19] that when circular polarization is used, the offset-fed reflector does not depolarize the radiated field, but the beam will be squinted relative to the electrical boresight of the antenna. With the increased importance of operating satellite communications with dual orthogonal polarizations, there have been improvements made in the cross-polarization properties of offset-fed reflector antennas.[21,22] Cross-polarized sidelobes of a single-reflector offset-fed antenna can be made comparable to the co-polarized sidelobes, and can be much lower if a dual-reflector antenna is used.

Cassegrain Antenna This is a dual-reflector antenna, Fig. 9.11, with the feed at or near the vertex of the parabola rather than at its focus. The larger (primary) reflector has a parabolic contour and the (secondary) subreflector has a hyperbolic contour. One of the two foci of the hyperbola is the real focal point of the system. The feed is located at this point, which can be at the vertex of the parabola or, more usually, in front of it. The other focus is a virtual focal point and is located at the focus of the primary parabolic surface. Parallel rays coming from a target are reflected by the parabola as a convergent beam and are re-reflected by the hyperbolic subreflector so as to converge at the position of the feed. There exists a family of hyperbolic surfaces that can serve as the subreflector. The larger the subreflector, the nearer it will be to the primary reflector and the shorter will be the axial dimension of the antenna assembly. A large subreflector, however, results in large aperture blocking, which may not be desirable. A small subreflector reduces aperture blocking, but it has to be supported at a greater distance from the primary reflector.

The chief advantage of the Cassegrain configuration is that the feed at or near the apex of the parabola does away with the need for long transmission lines out to a feed at

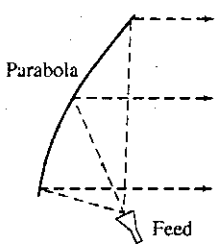


Figure 9.10 Parabolic reflector with offset feed.

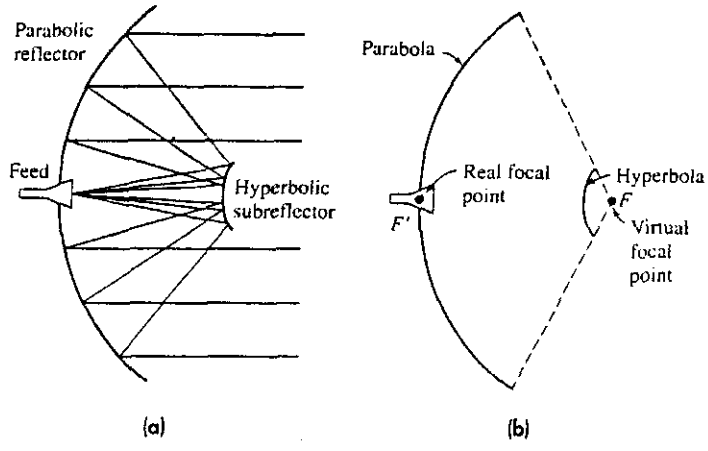


Figure 9.11 (a) Cassegrain antenna showing the hyperbolic subreflector, the feed at the vertex of the main parabolic reflector, and the paths of the rays from the feed; (b) geometry of the Cassegrain antenna.

the normal focus of the parabola. Furthermore, it allows greater flexibility in the size of the feed system. It is popular for monopulse tracking radars since the microwave hardware for generating the sum and difference patterns can be located behind the reflector without increasing aperture blocking. It has also been a good structure for experimental systems that use radar and other electromagnetic systems for different purposes, such as in the MIT Lincoln Laboratory Haystack Hill microwave research system. In that system there were separate RF systems for radar, radiometer, and space communications which operated at various frequencies. Each was constructed in replaceable modules, 8 by 8 by 12 ft in size, which were mounted directly behind the primary reflector.[23]

The antenna noise temperature (Sec. 11.2) of a Cassegrain configuration is usually smaller than that of a conventional front-focus antenna since there are no lossy transmission lines between the receiver and the feed. Also, the sidelobes caused by the spillover of the feed radiation from the subreflector illuminate the cold sky rather than the warm earth. Low antenna noise temperature is important for antennas used for radio astronomy or space communications, but it is generally not an issue in radar since extremely low-noise receivers are not always desirable, especially for military applications.

Aperture Blocking in the Cassegrain Antenna The hyperbolic subreflector of the Cassegrain antenna causes aperture blocking. Aperture blocking can be reduced by decreasing the size of the subreflector. This requires that the feed be made more directive or moved closer to the subreflector in order to minimize the spillover from the subreflector. A more directive feed means a larger feed that partially shadows the primary reflector and contributes to blockage. Thus blockage includes the obstacle presented by the feed as well as the subreflector. Minimum total aperture blocking occurs when the area of the subreflector and the projected area of the feed are equal.[24]

Polarization-Twist Reflector[24] The technique diagrammed in Fig. 9.12 can reduce aperture blocking if the application permits the antenna to operate with only a single polarization. The subreflector consists of a horizontal grating of wires, called a *transreflector.*

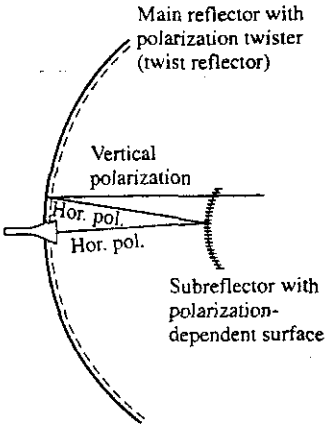


Figure 9.12 Polarization-twist Cassegrain antenna. Aperture blocking by the subreflector is reduced with this design.

It will pass vertically polarized radiation with negligible attenuation but will reflect horizontal polarization radiated by the feed. At the primary reflector the horizontally polarized radiation reflected by the subreflector is rotated 90° by the *twist reflector.* The twist reflector consists of wires oriented 45° to the incident polarization and placed one-quarter wavelength from the reflector's surface. Half the energy incident on the wires oriented at 45° passes through the grating; the other half is reflected. The one-quarter wavelength spacing of the wire grid from the reflector surface results in half-wavelength total travel of the component reflected from the surface. When combined with the component reflected from the wire grid, the resultant polarization is rotated 90°, and is therefore vertically polarized. This vertically polarized component is perpendicular to the horizontal-wire grid of the subreflector and passes through with negligible attenuation. The twist reflector as described above is narrowband, but it can be made to have very wide bandwidths.[25,26]

Gregorian Antenna The Gregorian antenna uses a dual-reflector similar to the Cassegrain except that the subreflector is an ellipsoid with one of its foci at the focus of the primary paraboloidal reflector. The ellipsoid lies beyond the focus of the paraboloid, instead of closer to it as does the subreflector of the Cassegrain. Also in the Gregorian configuration, the concave side of the secondary reflector ellipsoid faces the primary reflector, which differs from the Cassegrain in which the secondary reflector has its convex side facing the feed. The Gregorian has not seen as much application to radar as has the Cassegrain. There are other multireflector antennas; but they also have not had significant radar application.

Parabolic Cylinder Another method for obtaining an asymmetrical antenna pattern is to use a *parabolic cylinder,* shown in Fig. 9.13. This antenna surface is generated by moving the parabolic contour parallel to itself. A line source, such as a linear array, located at the focus of the cylinder is used to illuminate the parabolic-cylinder reflector (the focus is a line rather than a point). The beamshape and beamwidth in the plane containing

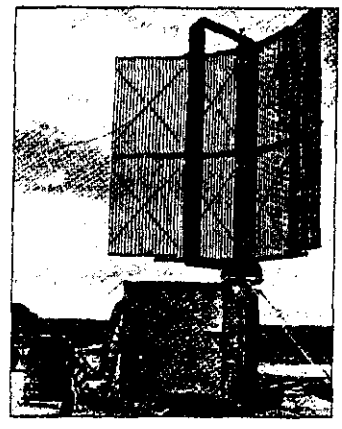


Figure 9.13 Example of a vertically oriented parabolic cylinder reflector antenna with a linear-array feed. This is the antenna used for the U.S. Marine Corps AN/TPS-63 air-surveillance radar.
(Courtesy Northrup Grumman Corp.)

the linear feed are determined by the illumination of the line-source feed, while the beamwidth in the perpendicular plane is determined by the illumination across the parabolic profile. The reflector is usually made slightly longer than the linear feed to avoid spillover and diffraction effects.

An advantage of the parabolic cylinder is that the large number of individual radiators on its linear-array (line-source) feed provides more control of the aperture illumination than does a single point-source feeding a paraboloid. The aperture illuminations required for low-sidelobe radiation patterns are more readily achieved with a parabolic cylinder than a paraboloid or a section of a paraboloid because of the control that can be applied at each of the radiating elements of the linear-array feed. The line feed, however, shapes the radiated beam in one plane only. Shaping of the beam in the orthogonal plane is determined by the reflector. Precise elevation-beam shaping is the purpose of the parabolic cylinder shown in Fig. 9.13, where the cylindrical antenna is oriented in the vertical so that the elevation radiation pattern can be shaped to minimize the radiation that strikes the ground.

The parabolic cylinder can generate an asymmetrical fan beam with a much larger ratio of the two orthogonal beamwidths than can a section of a paraboloid. Aspect ratios greater than 8:1 are practical with a parabolic cylinder but are difficult to achieve with a section of a paraboloid. Also, there is usually less depolarization on reflection from a parabolic cylinder than from a paraboloid.

9.5 ELECTRONICALLY STEERED PHASED ARRAY ANTENNAS

Background A phased array is a directive antenna made up of a number of individual antennas, or radiating elements. Its radiation pattern is determined by the amplitude and phase of the current at each of its elements. The phased array antenna has the advantage of being able to have its beam electronically steered in angle by changing the phase of the current at each element. The beam of a large fixed phased-array antenna therefore can

be rapidly steered from one direction to another without the need for mechanically positioning a large and heavy antenna. A typical phased array radar for microwave radar might have several thousand individual radiating elements using, for example, ferrite or diode phase shifters that allow the beam to be switched from one direction to another in several microseconds, or less.

Electronically steerable phased arrays are of interest because they can provide:

- Agile, rapid beam-steering.
- Potential for large peak and large average power. Each element can have its own transmitter. The power-aperture product can be large, especially at the lower frequencies.
- Multiple-target tracking. This can be accomplished either by generating multiple, simultaneous, independent beams or by rapidly switching a single beam to view more than one target in sequence.
- A convenient means to employ solid-state transmitters.
- Convenient shape for flush mounting or for blast hardening.
- Control of the aperture illumination because of the many antenna elements available.
- A lower radar cross section, if properly designed.
- Operation with more than one function (a multifunction radar), especially if all functions are best performed at the same frequency.

The chief disadvantages of a phased array radar are that it is complex and can be of high cost. Although an advantage of a phased array is that it can perform multiple functions in a sequential (time-shared) manner, its ability to employ multiple functions requires serious compromises for some applications.

A *linear array* consists of antenna elements arranged in a straight line in one dimension. It was mentioned in the last section that a linear array can be used as the feed for a parabolic cylinder antenna. A *planar array* is a two-dimensional configuration of antenna elements arranged to lie in a plane. In both the linear and planar arrays, the element spacings usually are uniform (equal spacing). The planar array may be thought of as a linear array of linear arrays. Most phased arrays of interest for radar are planar, but in this section we will start with the linear array as the model since it is simpler to analyze. A *broadside array* is one in which the direction of maximum radiation is perpendicular to, or almost perpendicular to, the plane (or line) of the antenna. An *endfire array* has its maximum radiation parallel to the array or at a small angle to the plane of the array.

Radiation Patterns of Phased Arrays Consider, as in Fig. 9.14, a receiving linear array made up of N elements equally spaced a distance d apart. The elements are assumed to be isotropic radiators in that they have uniform response for signals from all directions. Although isotropic radiators are not realizable in practice, they are a convenient concept in array theory. The outputs received from all N elements are summed via lines of equal length to produce a sum output voltage E_a. Element 1 will be taken as the reference with zero phase. From simple geometry, the difference in path length between adjacent elements for signals arriving at an angle θ with respect to the normal to the antenna, is

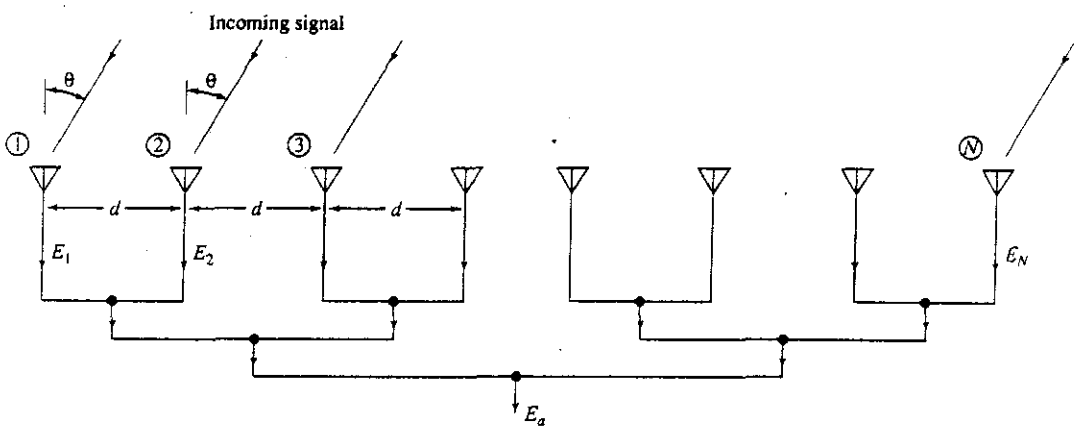


Figure 9.14 *N*-element receiving, parallel-feed, linear array, with equal lengths of transmission lines between each antenna element and the antenna output (at the bottom on the figure).

$d \sin \theta$. This gives a phase difference between adjacent elements of $\phi = 2\pi(d/\lambda) \sin \theta$, where λ = wavelength of the received signal. It is assumed that there is no further amplitude or phase weighting of the received signals. For convenience, we take the amplitude of the received signal at each element to be unity. The sum of all the voltages from the individual elements, when the phase difference between adjacent elements is ϕ, can be written

$$E_a = \sin \omega t + \sin(\omega t + \phi) + \sin(\omega t + 2\phi) + \cdots + \sin[\omega t + (N-1)\phi] \quad [9.22]$$

where ω is the angular frequency of the signal. The sum can be written[27]

$$E_a = \sin\left[\omega t + (N-1)\frac{\phi}{2}\right] \frac{\sin(N\phi/2)}{\sin(\phi/2)} \quad [9.23]$$

The first factor is a sinewave of frequency ω with a phase shift $(N-1)\phi/2$. (If the phase reference were taken at the center of the array instead at the left-hand side, this phase shift would be zero. In any event this factor is not as important as the second factor.) The second factor is an amplitude of the form $(\sin NX)/(\sin X)$. The magnitude of Eq. (9.23) represents the *field-intensity pattern*, or

$$|E_a(\theta)| = \left| \frac{\sin[N\pi(d/\lambda) \sin \theta]}{\sin[\pi(d/\lambda) \sin \theta]} \right| \quad [9.24]$$

The field-intensity pattern has zeros when the numerator is zero. This occurs when $N\pi(d/\lambda) \sin \theta = 0, \pm\pi, \pm 2\pi, \ldots, \pm n\pi$, where n = integer. The denominator, on the other hand, is zero whenever $\pi(d/\lambda) \sin \theta = 0, \pm\pi, \pm 2\pi, \ldots, \pm n\pi$. When the denominator is zero, it is seen that the numerator is also zero, and the value of $|E_a(\theta)| = 0/0$ is indeterminate. By applying L'Hopital's rule (differentiating numerator and denominator separately) it is found that $|E_a(\theta)|$ is a maximum and is equal to N when $\sin \theta = \pm n\lambda/d$. The maximum at $\theta = 0$ defines the main beam of the field-intensity pattern. The other maxima are

called *grating lobes* and are of the same magnitude as the main beam. They are generally undesirable in that they can cause ambiguities by being mistaken for the response of a target in the main beam. Grating lobes can be avoided if the spacing d between elements is equal to or less than λ. (There is still a grating lobe at $\theta = \pm 90°$ when $d = \lambda$, but practical radiating elements are not isotropic and can have negligible radiation $\pm 90°$.)

Equation (9.24) indicates that $E_a(\theta) = E_a(\pi - \theta)$; which means that an array of isotropic elements has a similar pattern in the rear of the antenna as in the front. The same is true for an array of dipole antennas. To avoid ambiguities between echoes from the front and the rear, the backward radiation can be eliminated by placing a reflecting screen behind the array so that only radiation over the forward half of the array antenna ($-90° \leq \theta \leq +90°$) need by considered. The field-intensity pattern with a back screen will be different from that of Eq. (9.24).

The normalized radiation pattern of an array of isotropic elements, which is sometimes called the *array factor*, is

$$G_a(\theta) = \frac{|E_a|^2}{N^2} = \frac{\sin^2 [N\pi(d/\lambda) \sin \theta]}{N^2 \sin^2 [\pi(d/\lambda) \sin \theta]} \qquad [9.25]$$

If Nd = D, the antenna dimension, and if the sine in the denominator can be replaced by its argument (implying that the angle θ is small), the pattern of the uniformly illuminated array is similar to the pattern of a uniformly illuminated line-source antenna, as was given by Eq. (9.14). The half-power beamwidth of this uniformly illuminated array of N elements when d = $\lambda/2$ is approximately

$$\theta_B = \frac{102}{N} \qquad [9.26]$$

When N is sufficiently large, the first (and largest) sidelobe is 13.2 dB below the main-beam maximum value.

When the radiating elements are not isotropic, the antenna radiation pattern of Eq. (9.24) has to be modified by the radiation pattern $G_e(\theta)$ of an individual directive element, so that

$$G(\theta) = G_e(\theta) \frac{\sin^2 [N\pi(d/\lambda) \sin \theta]}{N^2 \sin^2 [\pi(d/\lambda) \sin \theta]} = G_e(\theta)\, G_a(\theta) \qquad [9.27]$$

This is the product of the *element factor* $G_e(\theta)$ times the *array factor* $G_a(\theta)$, the latter being the pattern of an array composed of isotropic elements. Grating lobes caused by element spacings greater than half-wavelength may be eliminated by using directive elements whose pattern is zero or small in directions of undesired grating lobes. For example, if the element spacing $d = 2\lambda$, grating lobes occur at $\pm 30°$ and $\pm 90°$, in addition to the main beam at $\theta = 0°$. If the individual radiating elements, for example, have a radiation pattern whose null width (defining its main beam) is less than 60°, the grating lobes produced by the array factor will be suppressed. When this occurs, the antenna beam cannot be steered beyond the coverage of the individual elements that make up the array.

Equation (9.27) assumes that the radiation pattern of each element is the same. This is not true in practice, however. The radiation from an element in an array will be affected

by the mutual coupling among elements and the coupling due to the outward-traveling wave. An element in the center of the array sees a different electromagnetic environment from an element at the edge of the array. The radiation patterns of the elements will not be the same and will depend on the mutual coupling. Thus the pattern of an individual element depends on where it is located within an array. In order to obtain a more accurate representation of the radiation pattern of an array antenna, the pattern of each element within the array should be measured (or otherwise determined) in the presence of all others. Because the element pattern is not the same for each element, the radiation pattern of Eq. (9.27) is only an approximation, but one which has been widely employed.

Two-Dimensional Radiation Pattern In a two-dimensional, rectangular planar array whose aperture illumination can be separated into two orthogonal planes such as the horizontal and the vertical planes, the radiation pattern may then be written as the product of the radiation patterns in these two planes (sometimes called *principal planes* of the antenna). If the radiation patterns in the two principal planes are $G_1(\theta_a)$ and $G_2(\theta_e)$, the two-dimensional antenna pattern in this case is

$$G(\theta_a, \theta_e) = G_1(\theta_a)\, G_2(\theta_e) \qquad [9.28]$$

The angles θ_a and θ_e are not necessarily the elevation and azimuth angles normally associated with radar antennas. The normalized radiation pattern of a uniformly illuminated rectangular array of isotropic elements with spacing d is

$$G(\theta_a, \theta_e) = \frac{\sin^2 [N\pi(d/\lambda) \sin \theta_a]}{N^2 \sin^2 [\pi(d/\lambda) \sin \theta_a]} \frac{\sin^2 [M\pi(d/\lambda) \sin \theta_e]}{M^2 \sin^2 [\pi(d/\lambda) \sin \theta_e]} \qquad [9.29]$$

where N = number of (vertical) columns of the array that give rise to the (azimuth) angle θ_a and M = number of (horizontal) rows that generate the (elevation) angle θ_e. The above assumes the spacing between elements in the two directions is the same; but if they are not, the required modification is simple. Since array elements are not isotropic, the two-dimensional element factor should multiply this equation to obtain the antenna pattern.

Beam Steering and Array Feed Networks The beam of a linear array can be steered in angle by changing the relative time delays between the elements. Consider, as in Fig. 9.15a, two elements of a many-element array spaced a distance d apart. The signal from a direction θ_0, relative to the normal to the two elements, arrives at element 2 before it arrives at element 1. If the signal is delayed at element 2 for a time $\Delta T = (d/c) \sin \theta$, it will be in time coincidence (congruent) with the signal at element 1. If they are added together, it is as though the "main beam" of this simple two-element array was pointed in the direction θ_0. Beam steering occurs by changing the time delay. Inserting variable true-time-delays at each element of a many-element phased array, however, can be quite complicated and is generally unattractive with available technology. Instead, it is much simpler to employ a (modulo 2π) phase shift equal to $\phi = 2\pi f_0\, \Delta T = 2\pi(d/\lambda) \sin \theta_0$, where f_0 = frequency. The signals are then in phase rather than coincident in time. This is illustrated by Fig. 9.15b.

In a linear array, the phase shift that needs to be inserted at each of the elements in order to have all the signals with the same phase is $m\phi$, where m, an integer from 0 to

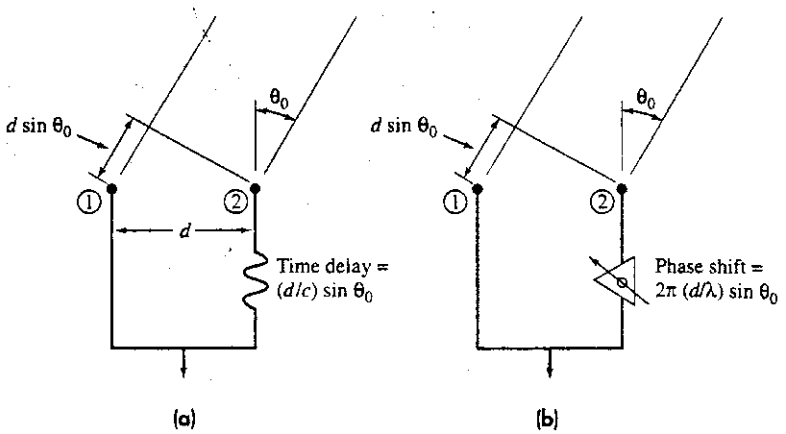


Figure 9.15 Two array elements spaced a distance d apart with a received signal arriving at an angle θ_0 measured with respect to the broadside direction. (a) Beam steering based on true time-delay; (b) beam steering using a phase shifter that is variable over the range from 0 to 2π radians.

$N - 1$, is the number of the element relative to the reference element. This means that the phase difference between elements is ϕ. The normalized radiation pattern of a linear array of isotropic elements is

$$G(\theta) = \frac{\sin^2 [N\pi(d/\lambda)(\sin\theta - \sin\theta_0]}{N^2 \sin^2 [\pi(d/\lambda)(\sin\theta - \sin\theta_0]} \qquad [9.30]$$

The maximum of this pattern occurs when $\sin\theta = \sin\theta_0$; hence, θ_0 is the direction at which the main beam points. As before, the element pattern should multiply this equation to get the antenna radiation pattern. Thus the beam can be steered in an array by changing the phase shift at each element.

Feeding an Array Variable phase shifters may be used at each element of a linear array to steer the beam as illustrated by the simple four-element array of Fig. 16a. This is called a *parallel-fed array.* The difference in phase between elements is $\phi = 2\pi(d/\lambda) \sin\theta_0$. When a series of power splitters, such as hybrid junctions, are used to create a tree-like structure as in the figure, it is sometimes called a *corporate feed,* since it vaguely resembles (when turned upside down) the organization chart of a corporation. Equal lengths of line between the elements and the transmitter/receiver are desired, but that is not always possible. The phase at each element should be the same (other than that introduced by the phase shifter). If the power is equally divided among all the elements and if the loss at each element is L_{ps}, then the entire loss in the parallel feed network is also given by L_{ps}.

A *series-fed* linear array is shown in Fig. 9.16b. Each phase shifter has the same phase, which means that only one steering command (the phase ϕ) need be generated, as compared to the $N - 1$ phase commands needed for the parallel-fed array. This is an advantage since it simplifies the computer that has to generate the phase commands. A serious disadvantage, however, of the series-fed array is its high loss. If the loss of each phase shifter is L_{ps}, then the loss through the array feed network is $(N - 1)L_{ps}$. Since it is not unusual for the loss of a phase shifter to be a significant fraction of a dB and since

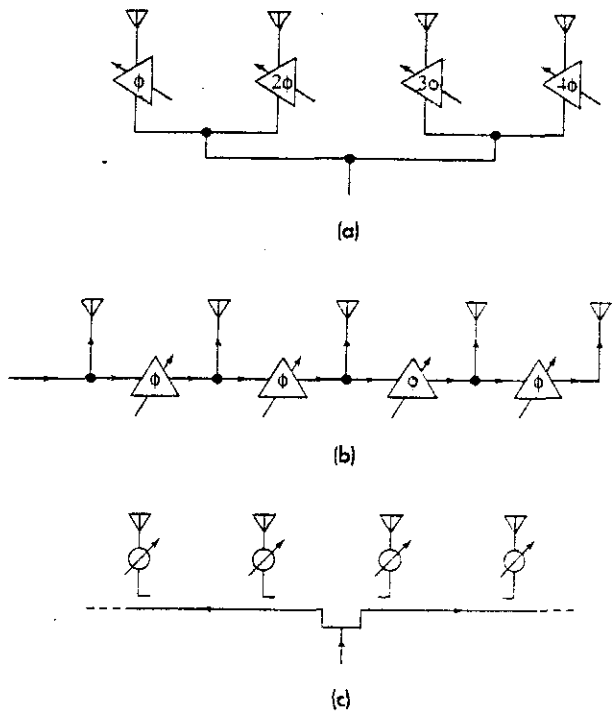


Figure 9.16 Steering of a linear array with variable phase shifters: (a) parallel-fed; (b) series-fed from one end; (c) series-fed from the center.

there might be many tens of elements in a linear array, the loss with a series feed is generally unacceptable.

There are two ways in which a series-fed array can be acceptable. One is when frequency scanning (Sec. 9.7) is used with low-loss waveguides connecting the elements. (The frequency scan array does not employ lossy phase shifters.) The other is when power amplifiers and low-noise receivers are placed between the phase shifter and the radiating element. There is still loss, but it is at low power level on transmit since it occurs before the power amplifier. On receive, the loss due to the phase shifter occurs after signal amplification so that it doesn't seriously affect the receiver overall noise figure.

The series-fed array, even when it is configured to have an acceptable low loss, is sensitive to changes in frequency. It has the properties of a frequency-scan array in that the direction of its beam will change with a change in frequency. Compensation for the shift of the beam position due to a change in frequency can be made in the computer by having it indicate the true beam-pointing direction for any given frequency. Inadvertent beam-steering with frequency can be avoided if the array is fed from the center as indicated in Fig. 9.16c. Although the beam is not shifted in angle, its beamshape will change with frequency.

The phase shifters in a two-dimensional parallel-fed planar array of $M \times N$ elements require $M + N - 2$ separate control signals. A two-dimensional series-fed array, however, requires but two control signals.

Grating Lobes Using an argument similar to the nonscanning array described previously, grating lobes will appear at an angle (or angles) θ_g whenever the denominator of Eq. (9.30) is zero, which means that

$$\pi \frac{d}{\lambda}(\sin \theta_g - \sin \theta_0) = \pm n\pi \qquad [9.31]$$

or

$$|\sin \theta_g - \sin \theta_0| = n\frac{\lambda}{d}$$

From this equation it is found that the element spacing d should be no greater than half wavelength in order to avoid grating lobes. With $d = \lambda/2$, a grating lobe will only appear at $\theta_g = -90°$ when the main beam is steered to $\theta_0 = +90°$. Practical phased arrays, however, cannot scan $\pm 90°$. If the scan were limited to $\pm 60°$, Eq. (9.31) states that the element spacing should not be greater than 0.54λ.

Change of Beamwidth with Steering Angle As the beam of a phased array scans in angle θ_0 from broadside, its beamwidth increases as $1/(\cos \theta_0)$. This may be shown by assuming the sine in the denominator of Eq. (9.30) can be replaced by its argument, so that the radiation pattern is of the form $(\sin^2 u)/u^2$, where $u = N\pi(d/\lambda)(\sin \theta - \sin \theta_0)$. The $(\sin^2 u)/u^2$ antenna pattern is reduced to half its maximum value when $u = \pm 0.443\pi$. Denote by θ_+ the angle corresponding to the half-power point when $\theta > \theta_0$, and denote by θ_- the angle corresponding to the half-power point when $\theta < \theta_0$; that is, θ_+ corresponds to $u = +0.443\pi$ and θ_- to $u = -0.443\pi$. The $\sin \theta - \sin \theta_0$ term in the expression for u can be written[28]

$$\sin \theta - \sin \theta_0 = \sin (\theta - \theta_0)\cdot\cos \theta_0 - [1 - \cos (\theta - \theta_0)] \sin \theta_0 \qquad [9.32]$$

The second term on the right-hand side of this equation can be neglected when θ_0 is small (beam is near broadside), so that $\sin \theta - \sin \theta_0 \approx \sin (\theta - \theta_0) \cos \theta_0$. With this approximation, the two angles corresponding to the half-power (3 dB) point of the antenna pattern are

$$\theta_+ - \theta_0 = \sin^{-1} \frac{0.443\lambda}{Nd \cos \theta_0} \approx \frac{0.443\lambda}{Nd \cos \theta_0}$$

$$\theta_- - \theta_0 = \sin^{-1} \frac{-0.443\lambda}{Nd \cos \theta_0} \approx \frac{-0.443\lambda}{Nd \cos \theta_0}$$

The half-power beamwidth is

$$\theta_B = \theta_+ - \theta_- \approx \frac{0.886\lambda}{Nd \cos \theta_0} \qquad [9.33]$$

Thus when the beam is scanned an angle θ_0 from broadside, the beamwidth in the plane of scan increases as $(\cos \theta_0)^{-1}$. This expression, however, is not valid when θ_0 is large, and the array performance can be much worse. In addition to the approximation made in

this derivation not being valid at large angles, mutual coupling effects can increase as the beam is scanned from broadside. At a scan angle of 60° from broadside, the beamwidth of a practical phased array antenna increases by more than the factor of 2 predicted from Eq. (9.33) and the sidelobe levels increase more than expected from simple theory.

Equation (9.33) applies for a uniform line-source distribution, which seldom is used in radar. With a cosine-on-a-pedestal aperture illumination of the form $a_0 + 2a_1 \cos(2\pi n/N)$ for a linear array of N elements with spacing d, the beamwidth is approximately[29]

$$\theta_B \approx \frac{0.886\lambda}{Nd \cos \theta_0}[1 + 0.636(2a_1/a_0)^2] \qquad [9.34]$$

where a_0 and a_1 are constants, and the parameter n in the aperture illumination represents the position of the element. Since the illumination is assumed to be symmetrical about the center element, n takes on values of $0, \pm 1, \pm 2, \ldots, \pm(N-1)/2$. The antenna aperture illuminations cover the span from uniform illumination to a tapered illumination that drops to zero at the ends of the array. (The effect of the array is assumed to extend a distance $d/2$ beyond each end element.) Although the above applies to a linear array, similar results are obtained for a planar aperture; that is, the beamwidth varies approximately inversely as $\cos \theta_0$.

A consequence of the beamwidth increasing with scan angle is that the antenna gain also decreases with scan angle as $\cos \theta_0$.

9.6 PHASE SHIFTERS

The shift in phase of a signal of wavelength λ transiting a line of length l at a velocity v is

$$\phi = 2\pi l/\lambda = 2\pi f l/v = 2\pi f l\sqrt{\mu\epsilon} \qquad [9.35]$$

where the frequency $f = v/\lambda$, μ = permeability and ϵ = permittivity. Usually the velocity of propagation v of electromagnetic waves is taken to be the velocity of light c; but with phase shifters it can be different. Here we have assumed for simplicity that the velocity of propagation corresponds to that in a TEM transmission line such as a coaxial cable, so that $v = 1/\sqrt{\mu\epsilon}$. The velocity of propagation of TE and TM waves propagating in waveguides is a bit more complicated than the above, but it is still proportional to $1/\sqrt{\mu\epsilon}$. Based on the far right-hand side of this equation, the various methods for obtaining a change in the phase shift may be summarized as follows:

- *Frequency f.* This is a relatively simple method for electronically scanning a beam. It was the first practical method for electronic beam steering and was at one time widely employed for many phased array radars. In spite of its simplicity, it is no longer popular since it restricts the use of bandwidth for other than beam-steering purposes and it is only practical for electronically steering the beam in one angular coordinate. Frequency scanning has been superseded by the development of other methods for phase shifting that do not have its limitations.

- *Line length l.* This may be accomplished by electronically switching in or out various lengths of transmission line to achieve the desired phase shift. Diodes are often used as the switches.
- *Permeability* μ. Ferrite, or ferrimagnetic, materials exhibit a change in permeability, and therefore a change in phase, when the applied magnetic field is changed. They have been popular for use at the higher microwave frequencies.
- *Permittivity* ϵ. The permittivity, or dielectric constant, of ferroelectric materials changes with a change in applied voltage. A change in the current of an electrical discharge also results in a change in the electron density which produces a change in permittivity.
- *Velocity v.* Changes in μ and ϵ cause the velocity of propagation to change; but a change in velocity can be had directly by changing the broad dimension of a rectangular waveguide, the so-called "*a*" dimension. By varying the "*a*" dimension of a rectangular waveguide, the proper phase change can be applied across an entire row of radiators of a linear array antenna to scan a beam in one angular coordinate. This form of rapid one-dimensional scanning was used for many years in *X*-band landing radars. It was called a *delta-a scanner* or an *eagle scanner.* The beam could be mechanically scanned over an angle of about 60° at a rate of 10 times per second.

All of the above have been employed or seriously considered as phase shifting devices for phased arrays. There are many other devices that can be used to obtain a phase shift for phased array radars, as has been mentioned in previous editions of this text; but the most popular are those that use ferrites or diodes.

Early electronic phase shifters were analog. Their phase shift could be continuously adjusted. Later they were replaced by digital phase shifting in which the values of phase took on discrete values, generally in binary steps. For example an *N*-bit phase shifter covers 360° of phase change in 2^N steps. Four-bit phase shifters with phase increments of 22.5° are commonly used, but digital phase shifters can have much finer quantization if needed. Although the analog phase shifter permits continuous variation of the phase shift, the relationship between its control current (or voltage) and phase is usually not linear, so that setting an analog phase shifter to a precise value of phase might not be as easily accomplished as obtaining similar or better accuracy with a digital device. Digital phase shifters have come to be the preferred method. Phase shifters have also been known as *phasors.*

Phase shifters for most phased array radar applications should be:

- Able to change phase rapidly (a few microseconds)
- Capable of handling high peak and high average power
- Require control signals that operate with little drive power (generally, one wouldn't want to use more power to drive the phase shifters than the total power that is radiated by the antenna)
- Low loss (a fraction of a dB if it is not used in an active aperture radar)
- Insensitive to changes in temperature
- Of small size (to fit within an element spacing of about a half-wavelength)

- Low weight (especially for airborne or mobile radars)
- Low cost (since the cost of a phase shifter is multiplied by the total number of phase shifters in the system).

There have been many types of phase shifters examined for radar application, and they possess these properties in varying degrees. No one type of phase shifter is sufficiently universal to meet the requirements of all applications.

Diode Phase Shifters[30–34] The semiconductor diode works well as a switching device for radar phase shifters. They are capable of relatively high power and low loss, and they can be switched rapidly from one state to another (low impedance to high impedance, or vice versa). They are not significantly affected by normal changes in temperature; they can be switched with low control power; and they are compact in size. They lend themselves well to microwave integrated circuitry and are capable of being used over the entire range of frequencies of interest to radar, except their loss increases and their power handling decreases at the higher microwave frequencies.

There have been three methods by which diodes have been used: (1) digitally switched lines, (2) hybrid coupled, and (3) loaded-line. Each will be briefly discussed.

Digitally Switched Lines A digital phase shifter can be obtained using a cascade of switched lines of length $\lambda/2$, $\lambda/4$, $\lambda/8$, and so forth. An N-bit phase shifter has N line lengths. Figure 9.17, for example, is a four-bit cascade of digitally switched phase shifters capable of switching in or out lengths of line equal to $\lambda/16$, $\lambda/8$, $\lambda/4$, and $\lambda/2$ to obtain a quantization level of $\lambda/16$, which corresponds to a phase increment of $360/16 = 22.5°$. Each phase bit consists of two lengths of line that provide the differential phase shift, and two single-pole, double-throw switches made up of four diodes. In this diagram, when the upper two switches are open, the lower two are closed, and vice versa. In the "zero" phase state, the phase shift is not zero, but is some residual amount ϕ_0, so that the two states are ϕ_0 and $\phi_0 + \Delta\phi_0$. The difference $\Delta\phi_0$ is the desired phase increment. The residual phase of the zero state has to be calibrated out in the radar system.

Hybrid Coupled The hybrid-coupled phase bit, as shown in Fig. 9.18, uses a 3-dB hybrid junction with balanced reflecting terminations connected to the coupled arms. Two switches (diodes) control the phase change. The 3-dB hybrid junction has the property that a signal at port 1 is divided equally in power between ports 2 and 3, and no signal power appears at port 4. The diodes act to either pass or reflect the incident signals.

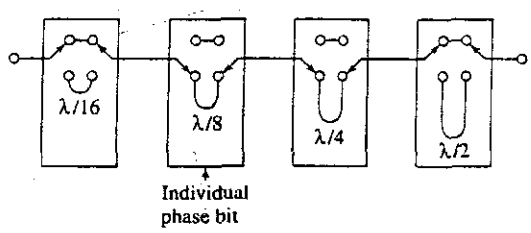


Figure 9.17 Digital phase shifter with four-bit diode-switched line lengths with $\lambda/16$ quantization. Particular arrangement shown gives 135° of phase shift (3/8 wavelength).

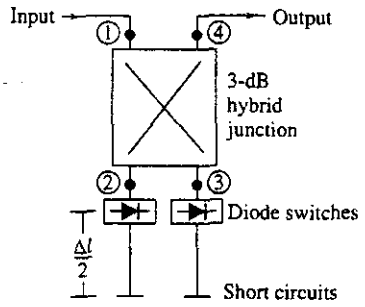


Figure 9.18 Hybrid-coupled phase bit.

depending on the bias applied to the diode. When the diodes allow the signals to pass, they are reflected by short circuits located farther down the transmission lines. The reflected signals combine at port 4, but no reflected signal appears at port 1. If the diode impedances are such as to reflect rather than pass the signals, the total path length traveled is less. The difference Δl is the two-way path length with the diode switches open and closed, and is chosen to correspond to the desired increment of digitized phase shift. An N-bit phase shifter can be obtained by cascading N such hybrid junctions and diode switches, with different lengths of lines for each bit.

Loaded Line This is a little different from the two diode phase shifters mentioned above. As shown in Fig. 9.19, it consists of a transmission line periodically loaded with spaced, switched impedances, or susceptances. Diodes are used to switch between the two states of susceptance. The spacing between diodes is one-quarter wavelength at the operating frequency. Adjacent quarter-wave-spaced loading-susceptances are equal and can take either of two values. If the magnitude of the susceptance is small compared to the characteristic impedance of the line, the quarter-wave spacing will result in cancellation of the reflections from any pair of symmetrical susceptances so that there will be matched transmission for either of the two susceptance conditions. Each pair of diodes spaced a quarter-wave apart produces an increment of the total phase required. Shunt capacitive elements increase the electrical length of the line and shunt inductive elements decrease its length. The number of pairs of shunt susceptances determines the total transmission phase shift. To obtain high power-handling capability, many such sections with small phase increments can be used so there are a large number of diodes available to share the power.

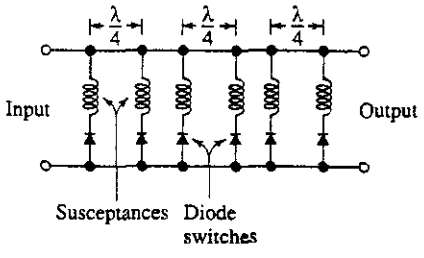


Figure 9.19 Periodically loaded-line phase shifter.

The advantage of the loaded line is its ability to handle larger power than other diode-based phase shifters. If the largest practical phase shift per diode pair is $\lambda/16$ (or 22.5°), 32 diodes would be needed to shift the phase 360°.

Comparison of Diode Phase Shifters . The hybrid-coupled phase shifter generally has less loss than the other two, uses the least number of diodes, and can be made to operate over a wide band. The switched-lines phase shifter uses more diodes than the other types and has an undesirable phase-frequency response that can be corrected at the expense of a higher insertion loss. It is often used in solid-state TR modules where the phase shifting is done at low power ahead of the power amplifier on transmit and following the receiver front end on reception. For a four-bit phase shifter with a total phase change of 360°, the loaded line requires 32 diodes, the switched line 16, and the hybrid-coupled shifter needs only 8 diodes. The theoretical peak power capability of the switched-line device is twice that of the hybrid-coupled circuit since voltage doubling is produced by the reflection of the hybrid junction. The switched-line phase shifter has the greatest insertion loss, but its loss does not vary with the amount of phase shift as it does in the other two types of circuits.

Diode phase shifters have been built in practically all types of transmission-line media, including waveguide, coax, and stripline. Microstrip is useful for medium power devices because of its ease of manufacture and circuit reproducibility, as well as its reduced size, weight, and cost of production. Diode chips can be mounted directly on the substrate without the parasitic reactances of the diode package.

A multiple-bit diode phase shifter need not be constructed with just one type of phase shifting device. The loaded-line is often preferred for small phase increments because of its compact size. It is not as suitable for large phase increments because it is difficult to match in both states when large. For example, a four-bit phase shifter might use a loaded-line configuration for the 22.5 and 45° bits, and the hybrid-coupled reflection circuit for the 90 and 180° bits to obtain the minimum insertion loss with suitable bandwidth and power-handling capability.

PIN Diodes The PIN diode has been a popular choice for use in diode phase shifters since it can handle higher power than other diodes; it can be designed to have relatively constant parameters in either or its two states; and it can have switching times from a few microseconds for high-voltage diodes to tens of nanoseconds with low-voltage operation. (Typically, switching times of the order of one or a few microseconds are quite suitable for most radar applications.) As sketched in Fig. 9.20, the PIN diode consists of a thin slice of high-resistivity intrinsic semiconductor material sandwiched between heavily doped low-resistivity P^+ and N^+ regions. The intrinsic region acts as a slightly lossy dielectric at microwave frequencies, and the heavily doped regions are good conductors. When d-c biased in the reverse (nonconducting) state, it resembles a low-loss capacitor since it is essentially an insulator situated between two conductors. Its parallel-plate capacitance is determined by the dielectric of the intrinsic region and is independent of the reverse-bias voltage. The series resistance is determined by the resistivity and geometry of the metallic-like P and N regions. In the forward-bias (conducting) state, when appreciable current is passed, the injection of holes and electrons from the P and N regions,

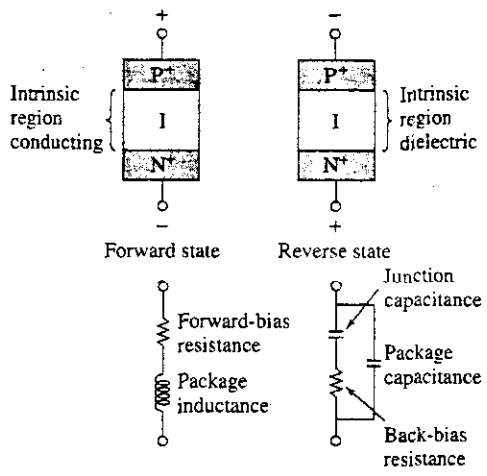


Figure 9.20 PIN diodes and simplified equivalent circuit for forward and reverse states.

respectively, creates an electron-hole plasma in what was formerly the dielectric region. Thus the slightly lossy dielectric is changed to a fairly good conductor with the application of forward bias. The capacitive component of the circuit disappears and the equivalent circuit becomes a small resistance that decreases with increasing forward current. The resistance can vary from thousands of ohms at zero bias to a fraction of an ohm with tens of milliamperes bias current. With forward bias, the diode resembles a resistance of low value.

Varactor Phase Shifters The varactor, or variable capacitance semiconductor, also can be used as the switch in a diode phase shifter. Its capacitance is varied by a change in voltage under reverse bias. It is capable of very rapid switching, of the order of a nanosecond, but its average-power handling is limited to a few tens of milliwatts as compared to about 100 W for a PIN diode.[35] The varactor peak-power rating is about 100 times less than that of the PIN. Instead of being used as a switch as is the PIN diode, the varactor can be employed as an analog (continuously variable) voltage-tuned phase shifter. This property has been used as an added module in a 6-bit digital diode phase shifter to provide a continuously variable phase change of from 0 to 11°.[36]

Monolithic Microwave Integrated Circuit (MMIC) Phase Shifters[37,38] The diode phase shifters that have been discussed thus far have generally been implemented as hybrid microwave integrated circuits (MIC) in that the passive components are deposited on the surface of a low-loss dielectric substrate and the active semiconductor devices are either bonded or soldered to the passive circuit. Phase shifters can also be constructed using monolithic microwave integrated circuit (MMIC) technology in which the entire circuit of passive elements, active devices, and interconnections are incorporated into a single semiconductor substrate. A MMIC phase shifter can be of much smaller size and weight than similar devices in hybrid MIC. They provide better reliability, they are highly reproducible because of the absence of wire bonds, and they can be produced economically

with high-volume production. Because of their small size they can be integrated on a single chip with other circuit functions such as power amplification, low-noise receiver front-end, and switching to form a compact T/R (transmit/receive) module for use in active array antennas. These advantages are accompanied, however, by the loss of flexibility in circuit tuning and troubleshooting that is available in hybrid MIC. This loss of flexibility to adjust (or tweak) the circuits means that more attention has to be paid to the use of computer-aided design to insure that the device, once manufactured, will do its job.

MMIC devices have usually utilized gallium arsenide (GaAs) metal semiconductor field-effect transistors (MESFET) as the switches for digital phase shifting. Although silicon technology has been extensively employed at the lower and middle microwave frequency regions, GaAs is preferred at the higher frequencies where MMIC techniques are employed. MESFETs are capable of rapid switching speeds of less than a fraction of a nanosecond, and they operate with relatively low d-c bias power.

The high-pass/low-pass phase-shifter configuration, Fig. 9.21, has been used with MESFETs as the switching elements for MMIC phase shifters. Its small size results from lumped elements being used, rather than distributed elements. A change in phase is obtained by switching between a high-pass filter and a low-pass filter. The insertion of a high-pass filter produces a phase advance and the insertion of a low-pass filter produces a phase delay.

Phase shifters using monolithic microwave integrated circuits at the higher microwave frequencies have been reported[38] to have wide bandwidth and constant phase shift over the band, insertion losses of 5 to 10 dB, and a maximum dimension of from a few millimeters to about a centimeter. The high loss has only a small effect since it occurs at a low power level before the power amplifier on transmit or after the low-noise amplifier on receive.

Ferrite Phase Shifters A ferrite is a ceramic-like metal-oxide insulator material that possesses magnetic properties while maintaining good dielectric properties.[39–41] Its dielectric constant is in the range from 10 to 20. In contrast to ferromagnetic materials such as iron, ferrites are insulators rather than conductors and have a high resistivity which

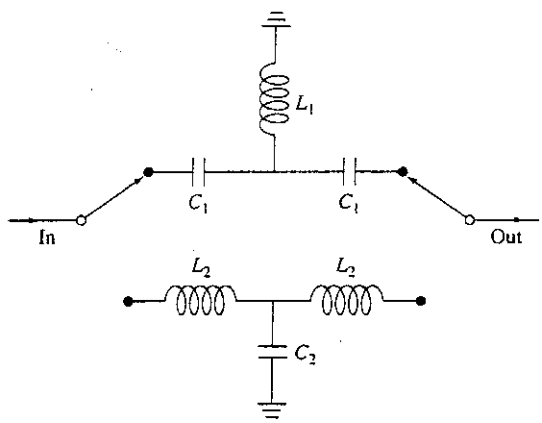


Figure 9.21 High-pass/low-pass phase shifter using T-type networks, shown switched to the high-pass filter.[38]

allows electromagnetic waves to propagate with low loss through the material. The term *ferrimagnetism* was introduced to describe the novel magnetic properties of these materials now known as ferrites.

Ferrite phase shifters are two-port devices that may be either analog or digital with either reciprocal or nonreciprocal characteristics. They are generally used at the higher microwave frequencies since their loss decreases with increasing frequency. Ferrites are generally preferred over diode phase shifters for radars above *S* band (except when the phase shifters are used before the power amplifier on transmit and after the low-noise amplifier on receive). At *S* band, either ferrites or diodes might be used. Below *S* band, the diode phase shifter usually is preferred.

The physics of propagation of electromagnetic energy in ferrite materials is not easy to describe, and will not be attempted here. The basic operation occurs by the interaction of electromagnetic waves with the spinning electrons of the ferrite material to produce a change in the microwave permeability of the ferrite, and therefore a change in phase. The magnetic permeability of a ferrite is anisotropic in that it must be represented by a complex tensor rather than a scalar. For this reason, the value of permeability and the resulting phase shift in a ferrite can depend on the direction of propagation. Some types of ferrite phase shifters, therefore, are *nonreciprocal* in that their phase change depends on the direction of propagation. This is different from the semiconductor phase shifters discussed earlier in this section, which were reciprocal devices. Nonreciprocal phase shifters have to be set differently for receiving than for transmitting.

There have been many different types of ferrite phase shifters developed, but those of most interest for radar include the latching, flux drive, and dual-mode phase shifters. One of the first successful ferrite phase shifters, the Reggia-Spencer shifter, will be described so as to illustrate some of the properties of ferrites, their limitations, and how these limitations were overcome in later types of ferrite devices. In spite of its shortcomings, the Reggia-Spencer device was used in an operational phased array radar at very high power (at the time, in the 1960s, when there was no better device available).

Reggia-Spencer Phase Shifter This device consisted of a rod or bar of ferrimagnetic material suspended at the center of a section of rectangular waveguide. A solenoid was wound around the waveguide to provide a longitudinal magnetic field. A change in phase was obtained by changing the current flowing through the solenoid coil. It was a reciprocal, analog phase shifter that had a high *figure of merit* (defined as the *change of phase per dB of loss*) and was more compact than previous experimental ferrite phase shifters. It had two serious limitations, however. First, the location of the ferrite rod at the center of the waveguide meant it was out of contact with the metal waveguide walls. Thus it was difficult to conduct the dissipated heat away. Second, the time required to switch from one phase state to another was relatively long; hundreds of microseconds rather than the microsecond or two that is characteristic of diode phase shifters. Furthermore, this phase shifter was sensitive to changes in temperature so it usually had to be operated in a temperature-controlled environment. There were also hysteresis effects that had to be accommodated when the phase had to be changed.

The lack of a convenient thermal path to dissipate heat was overcome in one design[42] by having the axially located garnet bar directly cooled by a low-loss liquid dielectric that

was allowed to flow along the surface of the garnet material. (A garnet is a ferrite with a different crystal structure than other ferrites.) The flow was confined by completely encapsulating the garnet in a teflon jacket so that the cooling liquid was in direct contact with the garnet bar. A *C*-band Reggia-Spencer phase shifter with this method of cooling operated over an 8 percent bandwidth at a peak power of 100 kW, average power of 600 W, insertion loss of 0.9 dB, and a VSWR of 1.25. The device was 2.4 by 2.1 by 8.2 inches and weighed 1.5 lb. It required, however, 125 μs to switch its phase, and at a switching rate of 300 Hz it used 16 W of switching power.

The long switching times for the Reggia-Spencer phase shifter were due to (1) the large inductance of the solenoid that provided the magnetic field and (2) the "shorted turn" effect caused by the metallic waveguide around which the solenoid was wrapped generating eddy currents in the metallic waveguide wall. There were things that could be done to reduce the switching time, but the Reggia-Spencer switching times were always much longer than those of other phase shifters.

Latching Ferrite Phase Shifter[40,43] A latching ferrite phase shifter overcomes many of the limitations of the Reggia-Spencer device by taking advantage of the hysteresis loop of a magnetic material so as to latch, or lock, its permeability to one of the two remanent magnetization points on the ferrite material's *B-H* curve. It does not need a continuous holding current to maintain the phase shift; hence, its drive power might be an order of magnitude less than that of the Reggia-Spencer shifter. It is not as temperature sensitive, it has a much faster switching speed, and there is less of a problem caused by hysteresis in the ferrite. It also lends itself to implementation as a digital phase shifter. Figure 9.22a illustrates one bit of a latching ferrite phase shifter mounted in a waveguide. The ferrite is in the form of a rectangular toroid. The contact of the toroid with the walls of the waveguide allows the generated heat to be dissipated. The toroid, however, results in this device being nonreciprocal.

Figure 9.23 is a hysteresis loop, or *B-H* curve, for a magnetic material such as a ferrite. It is plot of the magnetization, or magnetic induction (units of flux density, or webers/m^2) as a function of the applied magnetic field (ampere-turns/m) for a toroidal-shaped section of ferrite. The applied magnetic field is proportional to the current in the drive wire, which can be considered a solenoid of one turn. When a sufficiently large pulse of current is passed through the drive wire threading the center of the toroid, the magnetization is driven to saturation. When the current is then reduced to zero, there exists a remanent magnetization B_r. Similarly, when a large current pulse of opposite polarity is passed through the drive wire, the ferrite becomes saturated with the opposite polarity, and when the current is reduced to zero the remanent magnetization of opposite sign is obtained. Thus a toroidal ferrite may take on two values of magnetization, $\pm B_r$, obtained by pulsing the drive wire with either a positive or a negative current pulse. The difference in the two states of remanent magnetization produces the differential phase shift. Only a short-duration current pulse is needed to set the phase of a latching phase shifter.

The amount of differential phase shift depends on the ferrite material and the length of the toroid. A digital latching phase shifter may be obtained by placing in cascade a number of separate toroids of the proper lengths. The lengths of each toroid are selected to provide a differential phase shift of 180°, 90°, 45°, 22.5°, and so on, depending on the

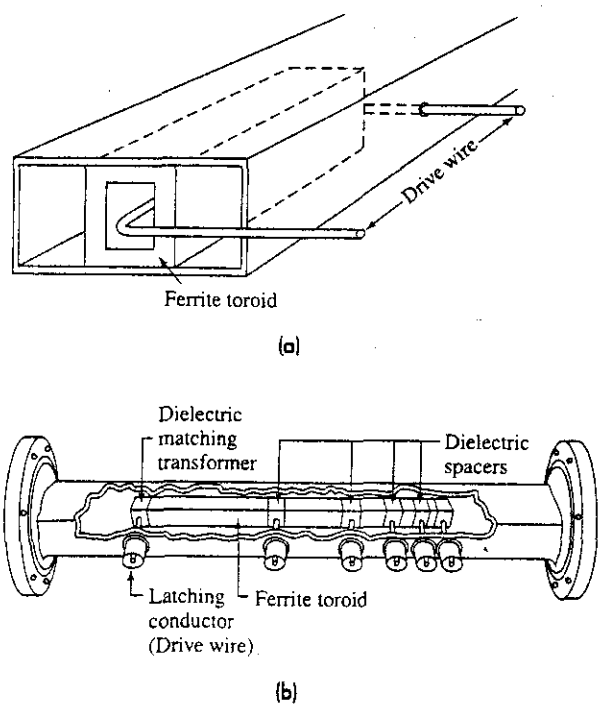


Figure 9.22 (a) Single bit of a latching ferrite phase shifter mounted in waveguide, showing the drive wire through the center of the toroid that establishes the magnetic field to latch the phase shift; (b) sketch of a five-bit latching ferrite phase-shifter. (From Wicker and Jones,[44] Courtesy IEEE.)

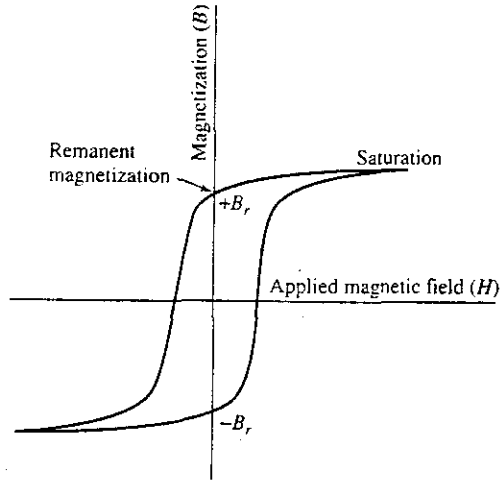


Figure 9.23 Hysteresis loop, or *B-H* curve, of a ferrite toroid.

number of bits required. The sketch in Fig. 9.22b illustrates a five bit latching ferrite phase shifter.[44] A separate drive wire is used for each bit. Impedance matching is provided at the input and output toroids. Filling the center slot of the toroids with a high dielectric-constant material produces a higher value of phase change per dB of loss (higher figure of merit) and lower switching power, but the lower will be the peak power that the device can handle before breakdown. The individual toroids usually are separated by thin dielectric spacers to avoid magnetic interaction. The drive wire is oriented for minimum RF coupling. It is also important that there be the proper mechanical contact between the toroid and the waveguide wall since an air gap can cause the generation of higher-order modes that result in relatively narrow-frequency-band, insertion-loss spikes. These unwanted air gaps have to be eliminated with care since excessive mechanical pressure on the material can cause magnetostriction that changes the magnetic properties of the material, especially if garnets are used. This type of latching phase shifter has also been called a *twin-slab toroidal phase shifter* since the major action is due to the two vertical branches of the toroids. The horizontal branches of the toroid do not contribute to the phase shift, but they are needed is to complete the magnetic circuit.

Introducing the applied magnetic field from within the waveguide via the single-turn drive wire eliminates the shorted-turn effect and avoids the long switching times that were characteristic of the original Reggia-Spencer phase shifter. Switching times of the order of microseconds become practical. Hysteresis was a nuisance to be tolerated in a Reggia-Spencer phase shifter, but the latching ferrite phase shifter takes advantage of the hysteresis loop to produce two discrete values of phase shift without the need for continuous holding power.

Different phase shifts must be used for transmit and receive with a nonreciprocal phase shifter. The nonreciprocal latching phase shifter, therefore, must be reset just after transmission is completed in order to receive the echo signals. The switching speeds of the latching phase shifter, which are of the order of microseconds, permit the rapid switching required. The phase shift for reception is obtained by simply reversing the polarity of the drive pulses that were used to set the phase for transmission. This reverses the direction of magnetization of the ferrite toroid, which is equivalent to reversing the direction of propagation. Although nonreciprocal phase shifters can be employed in many radar applications, they cannot be used in space-fed reflectarrays (Sec. 9.9) since the electromagnetic energy rapidly changes direction during both transmission and reception in such an antenna. Their use is also not practical in high-duty-cycle pulse doppler radars or in very short-range radars.

A nonreciprocal digital latching five-bit ferrite phase shifter was used in the S-band 3D radar known as the RAT 31/S built by Alenia of Rome, Italy.[45] It had the following characteristics: peak power = 7 kW, average power = 70 W, insertion loss <0.9 dB, VSWR < 1.3, bandwidth = 3 percent, switching time ≤ 2.5 μs, temperature tracking of the insertion phase = 0.6°/°C, and temperature range from 0 to 60°C. The rms value of the standard deviation of the insertion loss was 0.03 dB. The deviation of the insertion phase from its anticipated value was compensated in the path between phase shifter and antenna element. The rms value of the deviation of the phase of the smallest bit (nominally 12.25°, but actually an average value of 12.40°) was 1.13°; and that of the largest bit (nominally 180°, but actually an average of 199.8°) was 3.89°.

Twin-Toroid Latching Phase Shifter[46,47] The latching phase shifter described above has been improved by the use of the twin toroid, the geometry of which is sketched in Fig. 9.24a. The two toroids are separated by dielectric which concentrates the RF energy in the center of the waveguide. The active ferrite regions (in which the nonreciprocal interaction with the RF field occurs) are the two vertical ferrite arms that are in contact with the dielectric in the center. The differential phase shift of the twin-toroid ferrite can be made independent of frequency, and it is capable of wide bandwidths. The twin-toroid phase shifter is said[48] to be easier to construct than the single-toroid device. Hord[46] gives the following characteristics for an *X*-band twin-toroid phase shifter: loss = 0.4 dB, switching time = 3 μs, switching energy = 100 μJ, and size = 0.27 by 0.18 by 2.3 inches. Bandwidth can be 10 percent or greater.

A variant of the twin-toroid employs what is called a grooved waveguide, as is illustrated by the cross section view of Fig. 9.24b.[49] Note that there are different gaps between the waveguide ridges. It has been said that this geometry increases the differential phase shift by 20 percent, decreases the insertion loss for 360° differential phase shift by about 10 to 30 percent (thus providing a better figure of merit), and allows better thermal conductivity and an increase in average-power capability.

Flux Drive[50] The toroid ferrite phase shifter can be operated in an analog fashion to obtain digital phase-shift increments by varying the current of the drive pulse to provide different values of remanent magnetization. This is called *flux drive*. It has the further advantage of having reduced temperature sensitivity. A single long section of ferrite toroid is used that is capable of providing the total differential phase shift of 360°. The required digital phase increment is obtained by operating on a minor hysteresis loop, as indicated in Fig. 9.25. If $B_r(1)$, for example, were the remanent magnetization needed to produce a phase change of 180° (relative to the remanent magnetization $-B_r$), the amplitude and width of the driving pulse would be selected so as to rise to the point (1) on the hysteresis curve.

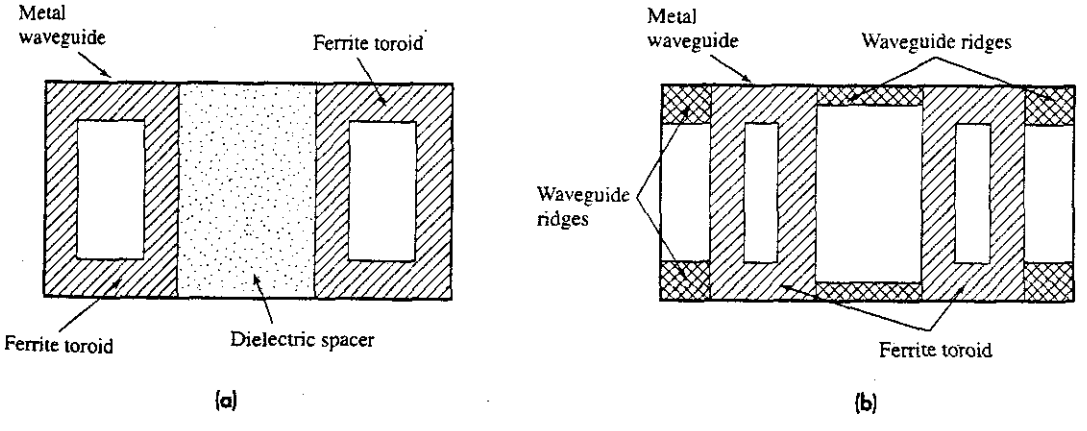


Figure 9.24 (a) Cross section of a twin-toroid ferrite phase shifter. (b) Cross section of a twin-toroid ferrite phase shifter with a grooved waveguide.

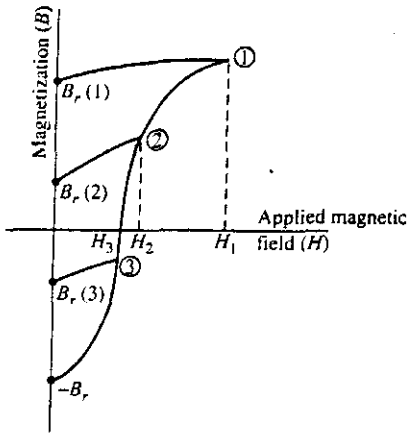


Figure 9.25 Hysteresis loop showing the operation of flux drive, where a single ferrite toroid is excited by discrete current pulses to produce digital phase-shift increments from what is basically an analog device.

When the pulse current decays to zero, the magnetization falls back to the remanent value $B_r(1)$ along the indicated curve. The difference in phase between $B_r(1)$ and $-B_r$ determines the differential phase increment. With a different value of current pulse, a different value of remanent magnetization and a different phase shift are obtained. In this manner, the ferrite toroid is basically an analog device that can provide any phase increment. It acts as a digital phase shifter if the drive currents are digital. The length of the toroid can be made 15 to 20 percent greater than the normal value to allow for some shrinkage of the total available increment of magnetization due to temperature changes. When the drive output impedance is small, the effect of the temperature-caused variations in magnetization will be small.

Dual-Mode Ferrite Phase Shifters[46,51] This is a latching phase shifter that is reciprocal, but without the limitations of the reciprocal Reggia-Spencer phase shifter. It is a variant of the Faraday rotation phase shifter[52] and the mechanical Fox phase shifter[53] (Faraday rotation is the rotation of the polarization, or electric field, when the wave propagates in a ferrite material in the presence of a magnetic field.) An outline of the configuration of a dual-mode ferrite phase shifter is sketched in Fig. 9.26. In the center portion is the ferrite bar that supports the propagation of circularly polarized waves. The bar is metallized to form a ferrite-filled waveguide that is accessible for dissipating the heat generated by the loss in the ferrite. A solenoid (not shown in the figure) is wound around the ferrite rod so as to apply an axial magnetic field that rotates the circular polarization to provide a phase change. A linearly polarized signal that enters the rectangular waveguide at the left-hand port is converted to circular polarization by a nonreciprocal circular polarizer (which is a ferrite quarter-wave plate, or quadrapole-field ferrite polarizer[54]). The applied axial magnetic field rotates the circular polarized wave in the ferrite bar, an action that imparts the desired phase shift. After propagating through the ferrite, the phase-shifted circular polarized wave is converted back to linear polarization by a second nonreciprocal polarizer. In a similar manner, a wave incident from the right is converted to circular polarization of the opposite sense by the nonreciprocal quarter-wave plate, and a

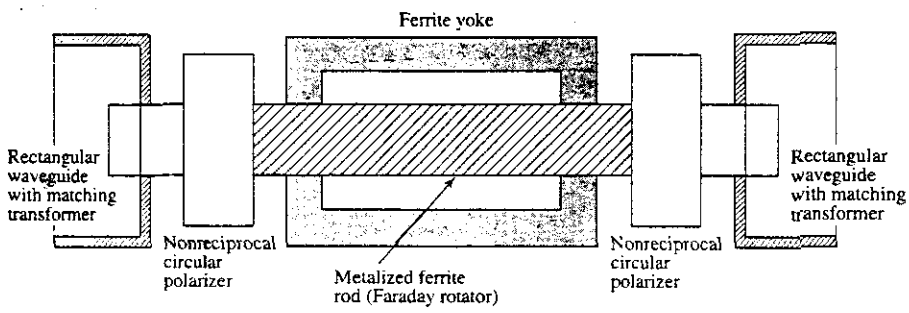


Figure 9.26 Outline of the configuration of a dual-mode ferrite phase shifter. The metalized ferrite rod in the center is based on the Faraday rotator.

phase shift occurs. Since both the sense of polarization and direction of propagation are reversed, the phase shift for a signal traveling from right to left is the same as that of a signal traveling from left to right. The magnetic circuit is completed externally by a temperature-stable ferrite yoke to permit latching of the magnetic field. Flux drive can be used to control the value of the remanent magnetization.

The dual-mode ferrite phase shifter is lightweight and capable of high average power. It also has a good figure of merit. Switching times are from 10 to 100 μs, which is longer than what is achieved with nonreciprocal ferrite phase shifters. Its longer switching times are due to the shorted-turn effect of the thin metallic film covering the ferrite rod. Hord[46] states that an X-band dual-mode phase shifter can have an insertion loss = 0.6 dB, switching speed = 100 μs, switching energy = 400 μJ, length = 1.6 in., and diameter = 0.48 in. Whicker and Young[55] indicate that dual-mode phase shifters are capable of 10 percent bandwidth, 1 kW peak and 100 W average power, and with latching (switching) speeds of 20 to 40 μs. Dual-mode phase shifters were used in the AN/TPN-19 X-band landing radar, where the phase shifter weight was 3.7 oz and had a phase error of 15°.[56]

Polarization-Insensitive Phase Shifters It has been said[57] that the dual-mode phase shifter can be made to be insensitive to polarization; that is, have the same phase shift for differently polarized electromagnetic waves so that they can be used in phased array antennas that employ more than one polarization. Polarization insensitive phase shifters are of interest when the radar must use dual orthogonal polarizations to avoid the large loss of signal caused by Faraday rotation of the plane of polarization when VHF or UHF radar waves propagate through the ionosphere, when circular polarization is used for detecting targets in the rain, or in any other situation where a choice of more than one polarization is desired.[58]

Rotary-Field Phase Shifter[59] This is similar to the dual-mode phase shifter mentioned above in that it also acts as a Faraday rotator to impart a phase shift. It is a reciprocal

device, but is nonlatching. It is very accurate, being capable of phase errors of one degree or less, which is considerably better than many other types of phase shifters. Such accuracy is required for low sidelobe array antennas. Linear polarization in the input rectangular waveguide is converted to circular. The circular polarized wave propagates in the ferrite rod which completely fills a circular waveguide. A phase shift is obtained in the ferrite by a constant-magnitude magnetic bias that is rotated in space by the application of currents to two orthogonal windings on a ferromagnetic yoke fitted over the ferrite. This is accomplished with a pair of coils wound on a motor-like stator. The quadrapole field generated by the two coils can be smoothly rotated to any desired angle. The accuracy of the differential phase shift is determined by the ratio of the control currents in the two coils. The ferrite rod then acts as a half-wave plate whose orientation determines the amount of phase shift, which is similar to the function of the mechanical rotation of the half-wave plate (or 180° differential phase shift section) used in the original Fox phase shifter.[52] A rotation of the half-wave plate by an angle θ results in a 2θ-radian change in the time phase of the signal. After propagating through the ferrite rod and experiencing a phase change, the circular polarization is converted back to linear.

Boyd[60] states that at X band, a rotary-field phase shifter might have 0.5 dB loss, 10 percent bandwidth, rms phase error less than one degree, and switching time of 50 μs. The control power to the stator windings is less than 0.5 W. These devices are capable of moderate to high power, are less temperature sensitive than other ferrite phase shifters, have a phase shift that varies little with frequency over a wide band, are highly accurate, and their low weight makes them suitable for airborne application.

Other Phase Shifters There have been many other phase shifters developed in the past, including other types of ferrite devices, electromechanical shifters, traveling wave tubes used as phase shifters, plasma devices, and ferroelectric phase shifters in which the dielectric constant of a ferroelectric material is a function of the applied electric field. The ferroelectric phase shifter has been said[61] to have high power capability, low drive power, voltage control of phase, and low production costs; but it has been difficult to obtain suitable ferroelectric materials to satisfy the important requirements of a phase shifter.

9.7 FREQUENCY-SCAN ARRAYS[62-64]

Because of its relative simplicity, the frequency-scan array was at one time the most popular form of phased array and was widely used. Its beam was steered by simply changing the radar frequency. It was especially popular for scanning a beam in one angular coordinate, such as with 3D air-surveillance radars.* A frequency-scan array has, however, significant limitations. The use of frequency for beam steering prevents the frequency domain from being used for other important purposes in radar, such as high range-resolution, electronic counter-countermeasures, and pulse-to-pulse frequency agility.

*A 3D air-surveillance radar is one which mechanically rotates in azimuth and scans one or more pencil beams in elevation or has multiple fixed beams in elevation for the purpose of measuring elevation angle. Other radars, of course, also can obtain three-dimensional data, but they are not usually "3D radars" in the sense of this definition.

Beam Steering by Change of Frequency The frequency-scanned array is almost always series fed as depicted in Fig. 9.27. Although series-fed arrays using phase shifters have high loss (as was mentioned in Sec. 9.5), it is not the case here since only waveguide connects the elements. The loss in propagating through waveguide transmission line is low.

We next derive the relationship between the radar frequency and the beam steering angle. The difference in phase between two adjacent elements in the series-fed array of Fig. 9.27 is

$$\phi = 2\pi fl/v = 2\pi l/\lambda \qquad [9.36]$$

where f = frequency of the electromagnetic signal, l = length of line connecting adjacent elements (generally l is much greater than the distance between elements), v = velocity of propagation, and λ = radar wavelength. [Eq. (9.36) is basically the same as Eq. (9.35)]. For convenience in this simplified analysis, the velocity of propagation is taken to be c, the velocity of light. This applies for coaxial lines and other transmission lines which propagate a TEM mode. Waveguides, which are more often used than TEM lines for the end-fed transmission line, can have a velocity of propagation that varies with frequency (i.e., it is dispersive).

As described in Sec. 9.5, if the beam is to point in a direction θ_0, the phase difference ϕ between elements spaced a distance d apart must be equal to $2\pi(d/\lambda)\sin\theta_0$. In a frequency-scan array, it is advantageous for practical reasons to add an integral number m of 2π radians of relative phase change. Since phase is modulo 2π and m is an integer, $2\pi m = 0$; so that the inclusion of $2\pi m$ has no effect on the phase difference between elements. The addition of the $2\pi m$ radians phase is achieved with the length of line l that connects adjacent array elements. The reason for adding the fixed $2\pi m$ phase shift is that it allows a given scan angle to be obtained with a much smaller frequency change than if a line of length $d = \lambda/2$ were used. This will become evident from Eq. (9.38). Equating the phase difference $\phi + 2\pi m$ between adjacent elements that is required to scan the beam to an angle θ_0, to the phase shift [Eq. (9.36)] introduced by a transmission line of length l, results in

$$2\pi(d/\lambda)\sin\theta_0 + 2\pi m = 2\pi l/\lambda \qquad [9.37a]$$

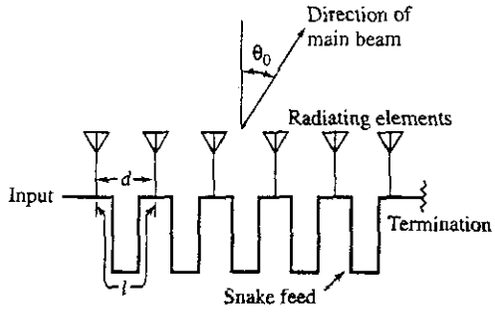


Figure 9.27 Series-fed, frequency-scan linear array.

or

$$\sin \theta_0 = -\frac{m\lambda}{d} + \frac{l}{d} \qquad [9.37b]$$

When $\theta_0 = 0$, the beam points to broadside and the above equation yields $m = l/\lambda_0$, where λ_0 is the wavelength that points the beam to broadside. If the frequency corresponding to beam pointing at broadside is denoted f_0, the direction of beam pointing can be written

$$\sin \theta_0 = \frac{l}{d}\left(1 - \frac{\lambda}{\lambda_0}\right) = \frac{l}{d}\left(1 - \frac{f_0}{f}\right) \qquad [9.38]$$

From the above, the wavelength excursion $\Delta\lambda$ required to scan the beam over an angular region $\pm\theta_s$ is

$$\Delta\lambda = 2\,\lambda_0(d/l)\sin\theta_s \qquad [9.39]$$

This equation shows that the greater the ratio l/d, the smaller will be the wavelength excursion $\Delta\lambda$ required to cover a given angular region $\pm\theta_s$. The ratio l/d is usually called the *wrap-up factor.* (The beam position is symmetrical with wavelength, but it is not symmetrical as a function of frequency.) To scan the beam $\pm 45°$ from the broadside direction requires a fractional wavelength change of 0.28 when the wrap-up factor is 5, and a 0.07 change when the wrap-up factor is 20 (fractional wavelength $= \Delta\lambda/\lambda_0$.)

Equations (9.38) and (9.39) apply for a TEM transmission line where the velocity of propagation is equal to the velocity of light. It is more usual, however, for waveguides to be used as transmission lines in this type of radar. Since the velocity of propagation in a waveguide depends on the frequency, a different and more complicated expression for the beam pointing angle results when waveguides are used instead of TEM lines. The velocity vs. frequency characteristic of waveguides can be used to good advantage to scan an angular region with less frequency change than indicated for a TEM line with the same wrap-up factor.

Grating lobes can occur in a frequency scan array, just as in other array antennas, when the electrical spacing between elements is too large. Equation (9.31) therefore applies. If we assume that a grating lobe can be tolerated at $\theta_g = -90°$ when the main beam is steered to the maximum scan angle $\theta_0 = +\theta_m$, then the following relationship applies

$$|1 + \sin\theta_m| < \lambda/d \qquad [9.40]$$

The onset of a grating lobe can limit the maximum angle the beam can be scanned.

Bandwidth Limitation Equation (9.39) illustrates the need for large frequency tunability of the radar transmitter in order to employ frequency scanning, especially if l/d is small. A large bandwidth might cause a potential problem of interference with other electromagnetic systems. Interference among frequency-scan radars operating in the same band, however, might not be that serious a problem since such a radar dwells at any one frequency for only a short time. Of more significance, however, is the reduction in signal bandwidth that can be used with a frequency scan antenna as the l/d ratio increases. If a wideband signal is used, distortion of the main beam will result.

With a series-feed, such as was indicated by Fig. 9.27, the signal travels a total distance $(N - 1)l$ from one end of the array to the other, where N is the number of elements in the linear array and l is the length of the transmission line between adjacent elements. For example, consider an S-band frequency-scan radar ($\lambda_0 = 10$ cm) with a linear array antenna of 101 elements spaced one-half wavelength apart. The wrap-up factor is assumed to be 10. The width of the antenna is taken to be 5 m, so that the total length of the feed line is 10×5 m = 50 m. An impulse incident at the input of a TEM feed line would require 0.167 μs to travel down the 50 m feed line and reach the other end (assuming propagation at the velocity of light). This build-up time, or time to fill the array, has a similar effect on the radar's signal bandwidth as does the transient response time of the more familiar signal filter. Thus the time t_D for the signal to travel from one end of the antenna to the other will limit the bandwidth to $1/t_D$, or 6.7 MHz in this example. If the wrap-up factor were 20 instead of 10, the limitation on the bandwidth would be 3.3 MHz. Thus the greater the wrap-up factor, the less frequency excursion that need be used to provide a given angular coverage, but the more narrowband will be the radar. Another way to see the effect of too wide a bandwidth is to note that if the signal has a wide frequency spectrum, the beam will be smeared in angle. Frequency scanning, therefore, is generally not compatible with high-resolution radar that might require large bandwidths.

Various Forms of Frequency-Scanned Radars There have been several methods by which frequency scanning has been employed, mostly as 3D air-surveillance radars that scan in one angular coordinate (elevation). Each of the following has been used in radar systems.

Single Scanning-Pencil-Beam The original 3D frequency-scan radars used a single pencil-beam antenna to scan in elevation as they mechanically rotated 360° in azimuth, Fig. 9.28a. The antenna beam dwells at a particular frequency at one elevation beam position before moving on to the next position and a different frequency. If the pencil beam is narrow (order of 1 or 2°), and if long range and large elevation-angle coverage are required, the revisit times with a single scanning beam will likely be excessively long (perhaps more than a minute). For example, if a pencil-beam antenna had azimuth and elevation beamwidths of 1.5°, elevation coverage 30°, range coverage to 200 nmi, and 10 pulses integrated from each target, then the time to cover 360° in azimuth is 66.7 s. This is far too long. Thus a single narrow-beamwidth scanning-beam radar operating as described is not suitable as a long-range radar because of its long revisit time. This applies to any type of electronically steered single-beam phased array, not just frequency scan.

The scan time can be reduced by increasing the pulse repetition frequency (decreasing the time between pulses) as a function of elevation angle. A low prf would be used at low elevation angles to obtain the desired maximum unambiguous range. As the elevation angle increases, the maximum range decreases since aircraft do not fly above a certain altitude. The prf can be increased correspondingly at the higher elevation angles to decrease the total time it takes for the antenna to cover 360°.

The number of pulses per dwell can be reduced to decrease the total scan time. Fewer pulses mean a larger transmitter peak power and perhaps longer pulse widths to make up for the smaller energy available from the fewer number of echo pulses received from a target. The limit occurs when there is only one pulse per dwell, which has serious

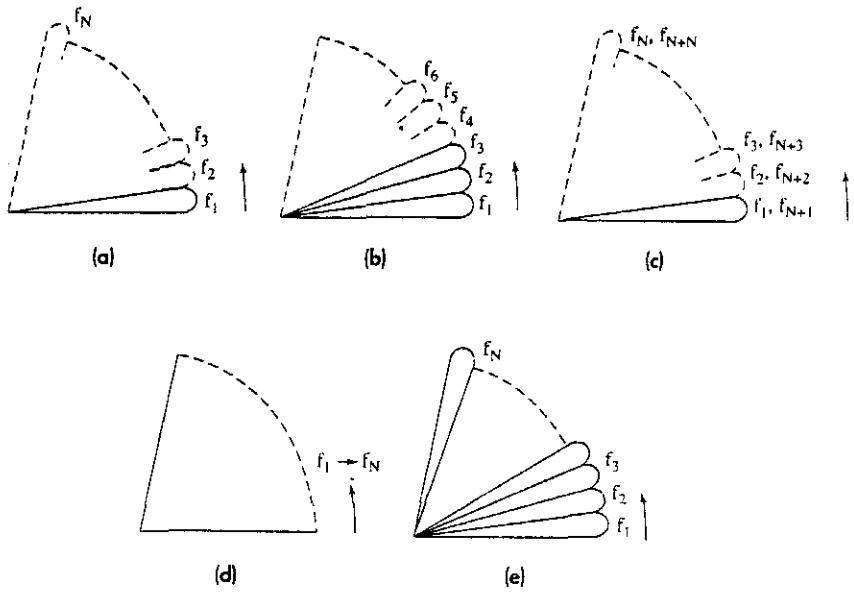


Figure 9.28 Several beam configurations for frequency-scan phased arrays that scan in a single angle coordinate, usually elevation. (a) Single-beam scanning; (b) multiple-beam scanning; (c) multiple-frequency frequency-scanning (radiating on more than one frequency at each beam position); (d) within-pulse scanning, transmit; and (e) within-pulse scanning, receive.

consequences for target detection. A single pulse per beam position can result in a large loss of two-way antenna gain if the target is near the half-power position of the antenna pattern when the pulse is transmitted. Another problem when using only one or a few pulses for target detection is that good doppler processing as required to eliminate clutter in an MTI radar cannot be achieved. Large MTI improvement factors (Sec. 3.7) require a large number of pulses to be processed (a longer time on target). Thus 3D long-range radars that employ one or a few pulses per beam position generally have poor, or no, clutter rejection.

Multiple-Beam Scanning One method to reduce the time for the antenna to cover its surveillance volume is to simultaneously, or almost simultaneously, transmit more than one beam (at more than one frequency), Fig. 9.28b. In the example of this figure, three pulses might be radiated at frequencies f_1, f_2, and f_3, respectively, so as to cover three contiguous elevation beam positions. The next set of three pulses is transmitted at frequencies f_4, f_5, and f_6. In this example there are three pulses radiated nearly simultaneously, but the number of frequencies radiated as one burst typically might vary from 3 to 9. The time required to scan 360° in azimuth is reduced in proportion to the number of simultaneous beams (frequencies) radiated.

A disadvantage of this approach is that there must be a separate receiver and signal processor for each of the n beams radiated simultaneously. In a military radar that must use sidelobe cancelers to reduce the effects of jamming, a separate set of sidelobe cancelers is required for each beam. When there are a large number of simultaneous beams, the cost of the radar can increase significantly.

Even though multiple contiguous beams are formed, one might not be able to perform an accurate elevation-angle measurement by comparing the amplitudes received in

adjacent beams. The reason is that amplitude-comparison angle measurement requires that the cross section of the target seen in the two adjacent beams to be the same. In a frequency-scan antenna, adjacent beams are at different frequencies. If the radar cross section of a target is a sensitive function of frequency, there can be a change in amplitude of the received echo signal due to changes in radar cross section that can result in an error when measuring angle by comparison of the amplitudes in adjacent beams.

Multiple-Frequency Frequency-Scan (or Multiple Mode) In the frequency-scan radars described so far, each elevation beam corresponds to a fixed frequency. This is not good for a military radar since measurement of the frequency by a hostile intercept receiver can provide the elevation angle of the radar beam. Effective hostile jamming can be achieved in some types of frequency scan radars by concentrating the jammer's power over a narrow range of frequencies (angles) rather than force the jammer to cover the entire frequency range of the radar transmitter. Furthermore, the rigid relationship between frequency and elevation angle does not allow a target to be observed at two or more frequencies, something that is desired in order to decorrelate the target cross section for improved detection performance.

It is possible, however, for a frequency-scan array to radiate more than one frequency at the same elevation angle, Fig. 9.28c, assuming that the antenna and the rest of the radar are sufficiently broad band.[64] This can be seen from an examination of Eq. (9.37b). The factor m in this equation is an integer. As the frequency is increased, a beam will scan from the endfire direction (in the direction of the input) through broadside and to the endfire direction in the other direction (pointing to the termination of the array series feed). When the frequency is increased further, another beam will eventually form, corresponding to a higher value of m. The range of frequencies corresponding to one value of m has sometimes been called a *scan band.*

Only one beam at a time will be radiated if the grating lobe relation of Eq. (9.40) is satisfied. It can be shown that if an array radiates at a particular angle corresponding to a value m_1 in Eq. (9.37b) when the frequency is f_1, then for some other value of m, say m_2, a beam will be radiated at the same angle when the frequency is $f_2 = (m_2/m_1)f_1$. As an example, consider an array with spacing $d = 0.6\lambda_0$ and $l/d = 15$, which corresponds to $m_1 = l/\lambda_0 = 9$. From Eq. (9.37b), the array will scan over a region $\pm 30°$ as the frequency changes from $0.968f_0$ to $1.035f_0$, where f_0 is the frequency corresponding to the broadside position of the beam ($\theta_0 = 0$). As the frequency is increased further, the factor $m_2 = 10$ applies and the same angular region is scanned as the frequency varies from $1.075f_0$ to $1.149f_0$. For $m_3 = 11$, the corresponding frequency range is $1.183f_0$ to $1.264f_0$. The beams corresponding to different values of m are related to grating lobes discussed previously.

In addition to allowing better target detection and better electronic counter-countermeasures, the ability to operate similar frequency-scan radars in different parts of the frequency band can help in reducing mutual interference among nearby radars.

Within-Pulse Scanning In the frequency-scan systems discussed above, the antenna beam dwells at each angular resolution cell (beamwidth) for one or more pulse-repetition intervals before moving to the next resolution cell. Another method that has been used for

frequency scanning arrays is to radiate a single frequency modulated pulse that covers a frequency range wide enough to scan the beam over the entire elevation coverage, as in Fig. 9.28d. During each pulse, the antenna beam rapidly scans through all elevation angles. This is sometimes called *within-pulse scanning*. The transmitted waveform is similar to that of a linear FM pulse-compression radar, but it serves a different purpose. The frequency of an echo signal reflected from a target will be a function of its elevation angle. The receiver employs a bank of filters, each tuned to a different carrier frequency which in turn depends on the target's elevation angle, Fig. 9.28e. The number of filters depends on the antenna beamwidth and the total angular coverage. The bandwidth Δf_B of each filter is determined by the frequency change required to scan the antenna one beamwidth, which is then

$$\Delta f_B = \frac{df}{d\theta_0}\,\theta_B \approx \frac{df}{d\theta_0}\,\frac{\lambda}{D} \qquad [9.41]$$

where θ_B = beamwidth = λ/D, and D = aperture dimension. Rearranging Eq. (9.38), differentiating and substituting into the above gives

$$\Delta f_B = \frac{f}{f_0}\,\frac{\cos\theta_0}{(l/d)(D/c)} = \frac{f\cos\theta_0}{f_0 t_D} \qquad [9.42]$$

where t_D is the time for the signal to propagate through the transmission line of length (l/d) D. Thus in the vicinity of broadside ($\theta_0 = 0$), the received signal has a bandwidth $\Delta f_B \approx 1/t_D$, over which the signal is linearly frequency modulated. This bandwidth can be used for achieving a modest amount of pulse compression on receive. Normally, pulse compression cannot be used with frequency scan; but with within-pulse frequency scan, pulse compression processing can reduce the pulse to a width no smaller than the time t_D it takes for the signal to travel from one end of the array to the other. Another way to look at this is that the finite beamwidth of the scanning antenna causes the target to be illuminated with a changing frequency, which can be compressed on receive.

Back-to-Back 2D/3D Antennas A 3D frequency-scan radar has the advantage of being able to obtain a measurement of the elevation angle (target height). A 2D radar does not obtain elevation angle, but since it has a longer time on target and more received pulses than does a 3D radar it has better MTI processing. In a military radar, a 2D radar can take advantage of the wide frequency range of a 3D frequency-scan transmitter to operate with greater flexibility in the choice of frequency so as to reduce the effectiveness of electronic jamming. The advantages of both 2D and 3D radars can be had by employing a 2D (fan beam) antenna back-to-back with a 3D (scanning pencil beam) antenna on the same rotating antenna mount. The agile-frequency transmitter can be switched between the two antennas to obtain the type of performance desired. A separate transmitter can be used for each antenna so as to achieve simultaneous, rather than time-shared, operation of the two.

Phase-Frequency Planar Array Frequency scanning has been employed in the past for obtaining beam steering in one angle coordinate, with phase shifters to steer in the orthogonal angular coordinate. This is called a *phase-frequency* array in contrast to a

phase-phase array which uses phase shifters to steer in both angle coordinates. In an $N \times N$ planar array a phase-frequency array needs only N phase shifters in addition to the fixed frequency-scan transmission lines, while a phase-phase array needs N^2 phase shifters. The phase-frequency array may be considered as a number of frequency-scan arrays placed side by side. The phase-frequency array was used in the early days of phased arrays when phase shifters might be said to have been primitive. As phase shifter technology improved, the all-phase-shifter array became more popular.

Frequency-Scanned Reflector Antennas[65,66] A very different form of frequency-scanned antenna, compared to those described above, is one that employs a frequency-sensitive grating as the reflector surface. Those familiar with optics might recall the *diffraction grating* which has had an important history as an optical device. The optical diffraction grating is a planar or curved surface obtained by ruling many closely spaced parallel grooves on an optical surface (such as a polished metal mirror). It has the property that a beam incident on the grating will reflect at an angle that depends on the frequency. A similar property can be obtained at microwave frequencies with a periodic array of thin conducting elements etched on a dielectric substrate placed over a surface. When a beam is incident on such a surface, the angle at which it is reflected will depend on the frequency of the incident beam, Fig. 9.29. The dielectric substrate over which the periodic array is etched can be designed to convert most of the incident energy to the diffracted direction governed by its frequency rather than have it reflect in the specular direction given by Snell's law. The shape of the reflecting surface is not parabolic as is common for reflector antennas, but is determined by the properties of the grating and the need to suppress the direct wave whose reflection would be given by Snell's law. As seen in Fig. 9.29, the feed is offset from the normal to the reflector surface in order to achieve the frequency scanning properties. In one experimental demonstration, a 10 percent change in frequency resulted in a beam scan of 10°. The power reflected in the direction of the angle given by Snell's law (equal to the incident angle) was suppressed approximately 20 dB below the power in the frequency-scanned beam.

Frequency Scan in Two Coordinates In principle, an array can be made to frequency scan a beam in two angular coordinates (a TV raster type of scan) by employing an array of slightly dispersive arrays fed from a single highly dispersive array. It has been said that a 90 by 20° sector can be scanned using a 30 percent frequency change.[67] Two-coordinate frequency scan is almost never used, however, since it requires a very large tunable frequency range of operation and it results in a very narrow signal bandwidth.

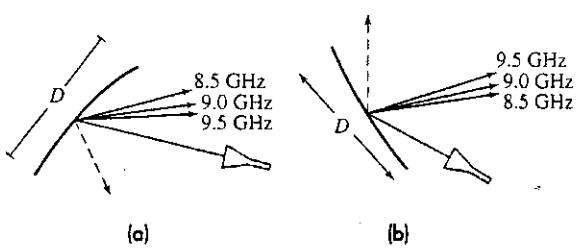


Figure 9.29 Geometries of a frequency-scanned reflector antenna employing a diffraction grating. In (a) the feed is positioned between the specular reflected beam and the diffracted beam; in (b) the specular and diffracted beams are reflected on the same side of the feed.
[From Johansson et al.,[65] Copyright 1989 IEEE.]

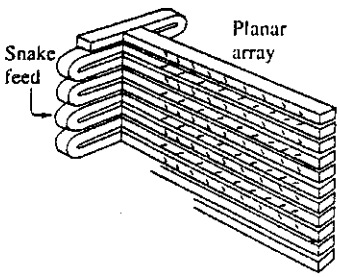


Figure 9.30 Sketch of a frequency-scan array showing, on the left, a folded waveguide delay line (snake feed) feeding a set of waveguides with radiating slots cut in their narrow wall.

Transmission Lines for Frequency Scan A popular form of transmission line for series-feeding a frequency-scan array has been the *snake feed*, shown at the left in Fig. 9.30. It has also been known as a *serpentine* or *sinuous feed*. Waveguide wrapped in the form of a helix has also been used, especially as the line feed for a parabolic cylinder reflector.

9.8 RADIATORS FOR PHASED ARRAYS

Types of Radiators Many different types of radiating elements (antennas) have been used in phased array radars, but the most popular have been various types of dipoles, slots cut into a wall of a waveguide, notch radiators, patch radiators, and open-ended waveguides, Fig. 9.31.[68,69] The pattern of a radiating element differs when located in the midst of a phased array than when it is all by itself in free space.[70] The *radiation impedance* (impedance which accounts for the power radiated) also can change; for example, the radiation impedance of a dipole in free space is 73 ohms; but in an infinite array with half-wavelength element-spacing and a back screen of quarter-wave separation, it is 153 ohms when the beam is directed broadside.[71] These changes are due to the effects of mutual coupling among neighboring elements. Furthermore, the impedance and antenna pattern

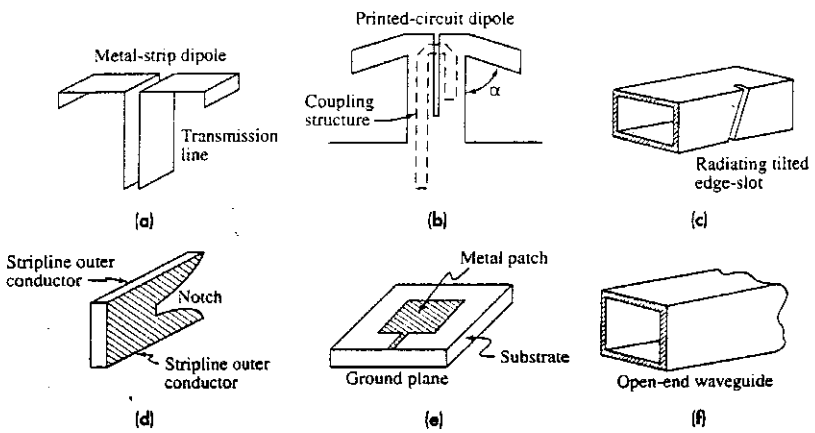


Figure 9.31 Sketches of single radiating elements for phased array antennas. (a) Metal-strip dipole with transmission line; (b) outline of printed-circuit dipole (solid lines) showing the coupling structure (dashed lines); (c) slot cut in the narrow wall of a waveguide; the tilt determines the amount of energy coupled from the slot; (d) notch radiator in stripline, radiation is to the right in this figure; (e) rectangular patch radiator; and (f) open-end waveguide.

near the edges. For example, it was mentioned previously that an element must be placed in the center of a 9 by 9 or an 11 by 11 array to accurately determine the effect of the neighboring elements on its performance. With half-wave spacing this means that significant coupling will occur between a radiating element and all elements that are within $2\frac{1}{2}$ wavelengths (if the 11 by 11 array represents an infinite array). In a square array of 60 by 60 elements, approximately 30 percent are edge elements. With such a large fraction being edge elements, one has to use caution when applying the results of infinite array theory.

The energy applied to a radiating element can appear in the main beam, the sidelobes, or be returned to the transmitter. The infinite array theory generally *assumes* that as the antenna scans off broadside, the reduction in radiated power due to the reduction in gain of the radiating element (the element factor) causes energy to be returned to the transmitter rather than be radiated elsewhere in space. The energy that returns to the transmitter results in an increase in the VSWR seen at each element. When the beam is steered, therefore, to a null of the element factor, some theorists assume that all the power is returned to the transmitter, which can be catastrophic. The situation has been called *blindness*[81] or *lost beams*.[82] This is a well-accepted concept by some antenna theorists, but there remain unanswered questions about its applicability to real antennas. When the beam is steered to the null of the element factor, the main beam will be distorted or even disappear, but one can argue that most, if not all, of the energy of the main beam radiates into space in other directions (producing a weird radiation pattern) rather than return to the transmitter.

Experimentally, the VSWR has been observed to increase at some of the many elements of a phased array when the beam is scanned; but there seems to be little experimental evidence for all or a large part of the transmitter power being returned from the array when the main beam is scanned to the direction of an element factor null. One might be skeptical about theories based on the infinite array model that assert that *all* the power returns to the source at certain beam-steering angles. If the power is not returned to the transmitter it seems plausible to assert that it is radiated into space. One should not expect a neat main-beam pattern or low sidelobes when this happens. An example of an experimental measurement in which the power was found to radiate in space rather than be returned to the transmitter when the beam is scanned is described next.

Forward-Wave Interaction in an Array Before leaving the subject of mutual coupling, an experimental measurement will be mentioned from the early 1960s which illustrates that further understanding of this subject is still required. Donald King and Harry Peters[83] reported measurements of small phased arrays using relatively high-gain closely spaced polyrod elements. Polyrods have seldom been used in phased arrays in the United States, but they illustrate the effects of mutual coupling that involve forward-traveling waves as well as diffraction effects and blockage. Measurements were made of a five-element array with polyrod radiators 6λ in length, spaced 3/4 wavelength apart. This is a rather close spacing for such high-gain endfire antenna elements, so it should not be expected that conventional array theory will apply. The E-plane pattern of a single 6λ polyrod in free space is shown in Fig. 9.32a. Its gain was 15.7 dB. The pattern of the center element of the five-element array is shown in Fig. 9.32b when the other four elements were terminated in a matched load. The pattern in the array is seen to be quite different from that in free space. It is broadened and becomes more rectangular in shape. The pattern of one

A *flared notch* antenna in stripline is indicated in Fig. 9.31d. These might be envisioned as starting with a dipole, tilting its two arms into a V shape, and then curving the arms of the V and smoothing the normally abrupt transition at the input. Such antennas might have a bandwidth of from 2 to 1 to 6 to 1.[74]

A *patch antenna*[75] consists of a thin metallic film bonded to a grounded dielectric substrate, Fig. 9.31e. Its shape is usually rectangular or circular, and it can be excited with microstrip. It has the advantage of being low profile, lightweight, and is easy and economical to manufacture. It can be mechanically robust and is readily employed with solid-state modules. Patch antennas, however, usually are not as broad band as other radiating elements.

Open-ended waveguides, Fig. 9.31f, are an extension of the waveguide sections in which the phase shifters are located. Their performance can be calculated or measured in a simple phased array waveguide simulator.[76] The waveguide may be loaded with dielectric to reduce its physical size in order to fit the element within the available space. If wide-angle scan is not required, the open-ended waveguides may be flared to form a horn radiator with greater directivity. An array of open-ended waveguides might be covered with a thin sheet of dielectric to better match the array to free space, as well as act as a radome to protect the array from the weather. A dielectric sheet across the face of the array, however, can result in coupling effects that modify the expected performance of the antenna.[77]

Other radiators that have been used with phased arrays include polyrods, Yagis, log-periodic antennas, spirals, and helices. Almost any type of radiator can be considered for application in a phased array; but the dipole, or its equivalent, probably has been the most popular.

Mutual Coupling The analysis of the phased array used in this chapter, as well in many other books and publications on antennas, is based on a relatively simple model, as was used in Sec. 9.5. It simply combines in space the radiation from the individual antenna elements, taking account of their relative amplitude and phase. Maxwell's equations are not involved, which is why it should not be surprising that the simple theory is not adequate for predicting the performance of actual phased arrays. In particular, the simple theory does not account for interaction among the radiating elements. The current at a particular element depends on the amplitude and phase of the currents in many of its neighboring elements, as well as the original current applied by the antenna feed network. The effect of one element on the other is expressed by the term *mutual coupling*. When the antenna is scanned from broadside, mutual coupling can cause a change in antenna gain, shape of the antenna pattern, shape of the individual element patterns, sidelobe levels, and the radiation impedance.

A major effect of mutual coupling is the change in the impedance seen at the element due to the presence of nearby elements. This is important for properly matching the element as the beam is scanned. The purpose of matching is to avoid high voltage standing-wave ratios (VSWR) that can result at certain scan angles.

Much of the classical theory of mutual coupling[78–80] has been based on modeling the antenna as an infinite array so that all elements within the array see the same environment. This has been a widely applied model for predicting phased array antenna behavior even though an infinite array is not realistic, and it can led to questionable results at times. In most practical arrays a large fraction of the elements can be considered *edge elements*, or elements

of the end elements was not given, but it would be different and would not be symmetrical in angle. Figures 9.32c and d show the beam scanned to 15 and 37.5°, respectively. The VSWR as a function of scan angle is given in Fig. 9.32e. (VSWR, or voltage standing-wave ratio, is a measure of how much signal is reflected by the antenna element back towards the transmitter.) Similar results were obtained when the element spacing was 1.5λ.

There are several interesting observations from these measurements. The VSWR did not increase as the beam was scanned from broadside, as would be expected from simple theory. The VSWR actually decreased, which is opposite to what is normally believed should happen when the beam of an array is scanned from broadside. There was no significant change in VSWR when the antenna was scanned to 37.5°. The antenna radiation pattern deteriorated considerably at this wide scanning angle and there was no main beam (the beam was "lost"), but there was energy radiated in other directions in space rather than reflected back to the transmitter, contrary to what is usually assumed in studies of mutual coupling. Thus there can be serious mutual coupling that affects the nature of the radiated pattern *without* energy being reflected back to increase the VSWR. Arrays of endfire elements such as polyrods, Yagis, log periodics, and axial-mode helices can experience similar mutual coupling due to forward-wave interaction, blockage, and diffraction rather than only the effects that appear at the terminals of the antenna elements. It might be expected that some of this behavior could also occur in an array of dipoles.

Closely spaced endfire elements have been used in radar antennas, but not often. The reason for mentioning this is that it gives evidence that what happens in the external near-region of the array and the interaction of the elements on the radiating aperture might be as important to understanding what is happening in an array as is the more familiar backward coupling that affects the circuit impedance, or VSWR, of the antenna. Further evidence of the importance of what takes place in front of the antenna is the effect of a dielectric slab or a periodic structure of baffles in front of the array, or even an array of open-ended waveguides, all of which can cause surface waves to travel across the array and contribute to mutual coupling effects.

Mutual Coupling and the Radar Systems Engineer The existing theoretical analyses of mutual coupling for finite (realistic) phased arrays do not seem complete since they do not account sufficiently for the forward-wave coupling among the elements. In spite of this deficiency, there is sufficient theoretical basis to allow the phased array antenna engineer to approach the design of real systems. As is common with many antenna design problems, the experimental skill and ingenuity of antenna engineers in dealing with real-world design have allowed phased array radars to be successfully built in spite of the shortcomings of existing theory.

9.9 ARCHITECTURES FOR PHASED ARRAYS

The term "architecture" is often used in the discussion of phased array radars but it has not been officially defined and there does not seem to be universal agreement as to what it encompasses. Here we use the term to include the various ways that phased array radars can be configured or structured.

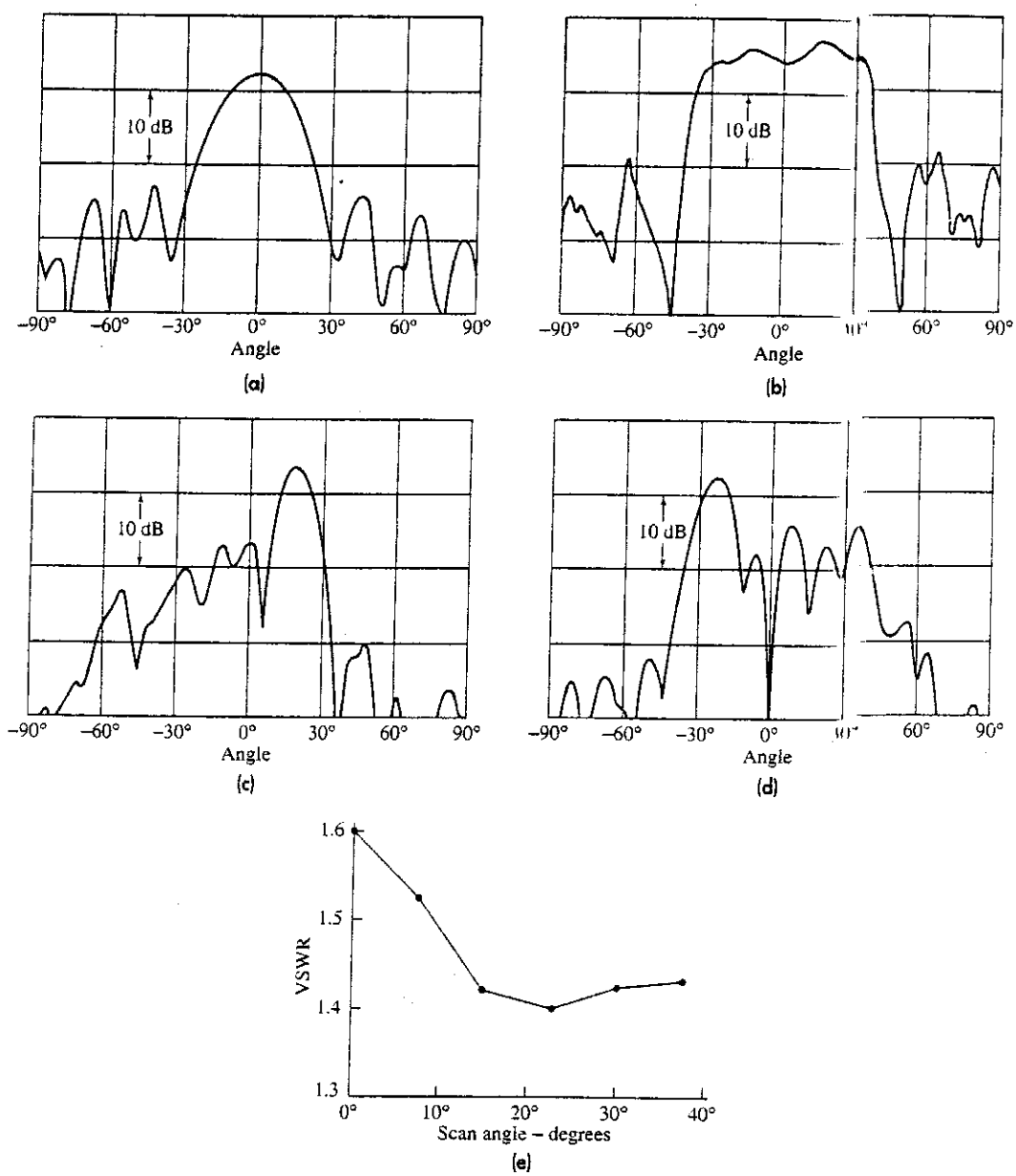


Fig. 9.32 Illustration of forward-wave mutual coupling in a five-element array with polyrod elements each six wavelengths long and spaced 3/4 wavelength apart. (a) measured *E*-plane radiation pattern of a single isolated polyrod element; (b) Measured *E*-plane radiation pattern of the center element with each of the other four elements terminated in a matched load; (c) pattern with the beam scanned to 15°; (d) pattern with the beam scanned to 37.5° (the main beam is not seen at its scan angle of 37.5°, but is "lost"); and (e) VSWR of the feed line of the center element as a function of the scan angle. Note there is no significant change in VSWR even for a scan angle of 37.5°.
(From King and Peters,[83] Courtesy Microwave J.)

feed to the vertical elements. This is called a *parallel-series* feed. It the columns were fed with a parallel feed, it would be a *parallel-parallel* feed. There can also be a *series-series* feed. Series feeds may sometimes be easier to implement, especially in an active-aperture array where the transmitters and receivers are located between the feed and the radiators; but they have narrower bandwidth than the parallel feed.

Losses in constrained feeds are due to attenuation in the transmission lines as well as reflections from the junctions. Mailloux[84] states that the total length l of transmission line from an antenna element to the single array input is at least

$$l = (N^{1/2} - 1)d \qquad [9.43a]$$

This expression applies to a square array with total number of elements N and element spacing d. (The total length l can be longer than given by this equation, depending on the specific architecture.) The longer the length of line between antenna element and input, the greater will be the loss due to transmission line attenuation. Mailloux also states that the total number of power dividers (which also introduce loss) in series with each element is

$$\text{Total number of power dividers} = \log_2 N = 3.32 \log_{10} N \qquad [9.43b]$$

When the power splitters in a parallel-feed are four-port hybrid junctions, or the equivalent, the feed is said to be *matched*. Theoretically, there are no spurious signals generated by internal reflections in a matched feed. It is not always convenient or practical to use four-port hybrid junctions. Three-port tee junctions are sometimes used instead, for economic reasons, to provide the power splitting. Such a network cannot be perfectly matched in theory, and internal reflections can occur which can appear as spurious sidelobes in the radiation pattern.

Brick and Tile Assemblies Brief mention should be made of two types of array architectures, one called *brick* and the other *tile*. They will not be described in sufficient detail to fully understand, but further information is available in the literature.[85,86] These terms relate to the manner in which the array is contructed in relatively larger sections to make assembly easier. They are used with a corporate feed. The array is often grouped into subarrays of rows, columns, or areas, with each subarray fed separately. In the brick construction, the array is assembled with circuits on boards that are mounted *perpendicular* to the array face. In tile construction, which has also been called *monolithic array construction*, the array is assembled with one or more layers *parallel* to the array face. The array face contains the radiating elements and semiconductor active subarrays. The phase shifter drivers might be mounted on a layer just below the radiating face, and another layer with the RF power dividers mounted below that. Brick construction utilizes greater depth than does tile construction, so it allows more room for circuits, better thermal dissipation, and more convenient maintenance. It is also compatible with dipole and flared-notch radiators which can have greater bandwidth than the flat microstrip patch-radiators of the tile type of construction. Tile construction has the advantage of being thin so that it can be made to conform to aircraft or missile surfaces. It can be folded and stowed for erection in space, and it can be compatible with robotic or other automatic means of fabrication. Generally, the brick and tile structures have been of more interest at the higher frequencies (X band and above) than at lower frequencies.

An extremely important part of any practical phased array system architecture is the means by which the power from the transmitter is efficiently divided and distributed to the radiating elements, and the reciprocal problem of combining the signals received at the elements and providing them to the receiver and signal processor. The structure that performs this function is called the *array feed.* Two major methods for this purpose are *constrained feeds* (a network of transmission lines, or power dividers) and *space feeds* (which resemble feeds for reflectors or lens antennas). The constrained feed is sometimes called a *corporate feed.* Analog multiple-beam-forming arrays, active-aperture arrays, and digital beam-forming also require special consideration of the array feed method. Array feeds should not introduce significant loss nor should they be of excessive weight and size. Since the array feed is not always one of the more "glamorous" or visible parts of an array, it is not always appreciated that any loss it introduces is equivalent to a loss in antenna power gain and has to be compensated by an increase in transmitter power or other, usually undesirable, means.

Constrained Feed Simple illustrations of a parallel-fed and a series-fed constrained feed for one-dimensional linear arrays were shown in Fig. 9.16. The constrained feed is basically a $1 \times N$ power divider, where N = total number of elements in the array. Figure 9.33 is an example of a constrained feed for an array in two dimensions. This particular example is a combination of a single parallel feed and a number of series feeds. In one dimension, the parallel feed is shown and in the other dimension series feeds are shown. Each element has its own phase shifter. The required phase shift to steer the beam in two dimensions can be determined for each element by the beam-steering computer and then distributed to each phase shifter. Alternatively, a small computation chip can be placed at each element to compute the phase required at that particular element, based on being told the azimuth and elevation angles to which the beam is to be steered. Feeds consisting of waveguide or coaxial transmission lines can handle high power with low loss and can be constructed with excellent precision. They can, on the other hand, be bulky and expensive. At the higher microwave frequencies (L band and above), strip lines which can be precisely fabricated with computer-aided manufacturing techniques are sometimes used.

The power distribution to the columns of the two-dimensional array of Fig. 9.33 is shown with a single parallel feed. The power in each column is distributed with a series

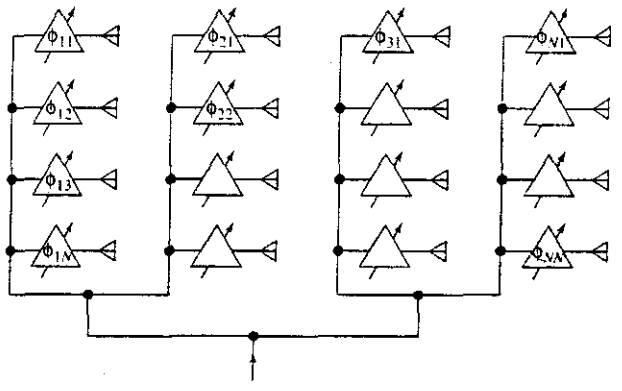


Figure 9.33 Planar array for scanning in two angular coordinates.

is reflected, and passes back through the phase shifters to be radiated as a plane wave in the desired direction. Because the energy passes through each phase shifter twice, a phase shifter need only be capable of half the phase shift needed for a lens array or a conventional array. The phase shifters, however, must be reciprocal. This can be a limitation since some ferrite phase shifters with excellent properties for use in phased arrays are nonreciprocal and therefore cannot be used in a reflectarray.

Comparison Spillover radiation from the feed of a space-fed array can result in higher sidelobes at angles far from broadside than would be obtained with a constrained feed, unless some means are taken to minimize the spillover. Both sets of antenna elements at the front and back of the lens array require matching. This increases the matching problem of the array and can result in lower antenna efficiency. It is relatively straightforward, however, for the space-fed array to generate a cluster of beams, as for monopulse angle tracking, by use of multiple feed horns similar to those used for monopulse tracking with a reflector antenna. Compared to the constrained feed, space-fed arrays have the advantage of lower loss.

The lens array allows more freedom than the reflectarray in designing the feed assembly since there is no aperture blocking. The presence of a back surface in the reflectarray, however, not only allows better mechanical support and heat removal than in a lens array, but it also makes it easier to provide the needed control signals to the phase shifters. Space-fed arrays are generally cheaper than conventional arrays because of the omission of transmission-line feed networks. The space-fed array with a single transmitter and receiver is usually cheaper than an active-aperture array (to be described later) whose transmitters (and receivers) are distributed along the aperture.

A space-fed array may be cheaper than an array with a constrained feed or an active-aperture array, but it might not have the ability to control the aperture illumination sufficiently well to obtain ultralow sidelobes; and it is not capable of the very large powers possible with a conventional array where a transmitter can be placed at each of the many elements of the array.

Parallel-Plate Feeds A folded pillbox antenna, a parallel-plate horn, or other similar microwave devices can be used to provide the power distribution to the antenna elements. These are called *reactive* feed systems. They are basically used with a linear array and can be considered as another form of space-fed array, but only for one dimension. They would have to be stacked to feed a planar array, and would thus be a heavy feed system.

Further information on the many methods for feeding a phased array can be found in Patton.[87]

Subarrays It is sometimes convenient to divide the array into subarrays. These can simplify the manufacture and assembly of the array, provide broader signal-bandwidth, and allow multiple transmitters to be used to obtain greater power. Each subarray could have its own transmitter and receiver, but it is not necessary to do so to utilize subarrays. It is also possible to give identical phase steering commands to similar elements in each subarray to allow simplification of the beam-steering unit (which generates the beam steering commands). Because of the discrete nature of the subarrays, the phase distribution

Monopulse Beams When multiple beams are needed for monopulse angle measurement with a constrained-feed phased array, the output of each receiving element can be split into three separate outputs that connect to three separate beam-forming networks. One output is used with a beam-forming system to provide the sum beam and the other two outputs are used with separate beam formers to generate the two angle beams. The two angle beams can have different aperture illuminations (weightings) than the sum beam so as to produce desirable difference patterns with low sidelobes and a good error-signal slope.

Space Feeds A space feed is similar to the feed of a reflector antenna. It enjoys the advantage of the relative simplicity that characterizes feeds for reflectors. There are two types of space feeds depending on whether the array is analogous to a lens or to a reflector.

Lens Arrays Although the lens array is considered first in this discussion, much of what is said about it is applicable to the reflectarray as well. The lens array, Fig. 9. 34a, is fed just as would a lens antenna. It is shown in Fig. 9.34a as a one-dimensional representation, but the space-fed array is almost always two-dimensional. The primary feed might be a single horn, a collection of horns, or a monopulse cluster of horns. (The space-fed array is described as a transmitting antenna, but an analogous description can be given as a receiving antenna.) An array of antenna elements collects the energy radiated by the feed and passes it through phase shifters which provide a phase correction to convert the incident spherical wave to a plane wave. The phase shifters also apply the phase shifts required to steer the plane wave to some angle off of broadside. The antenna elements on the opposite side of the lens array then radiate the beam into space. The feed illuminating the space-fed array provides a natural amplitude taper to produce lower sidelobes than would a uniform illumination. The feed may be placed off-axis to avoid reflections from the lens returning to the feed and producing a large VSWR.

Reflectarrays A space-fed reflectarray with an offset feed is diagrammed in Fig. 9.34b. The energy from the feed enters the antenna elements, passes through the phase shifters,

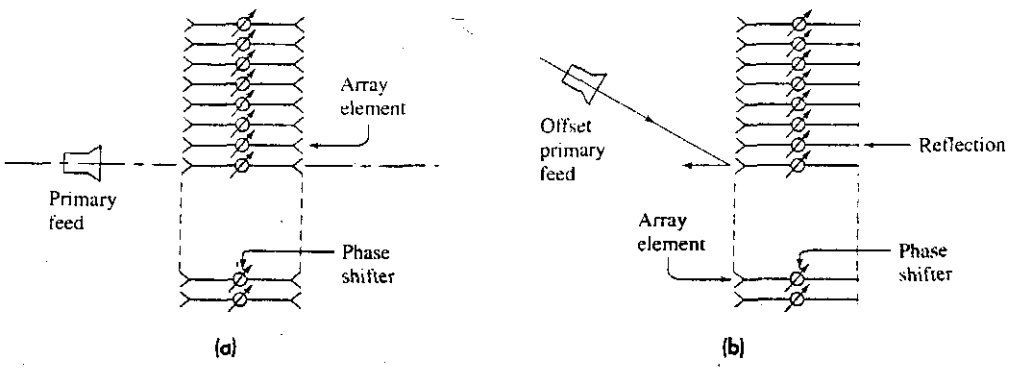


Figure 9.34 Space-fed arrays. (a) Lens array; (b) reflectarray.

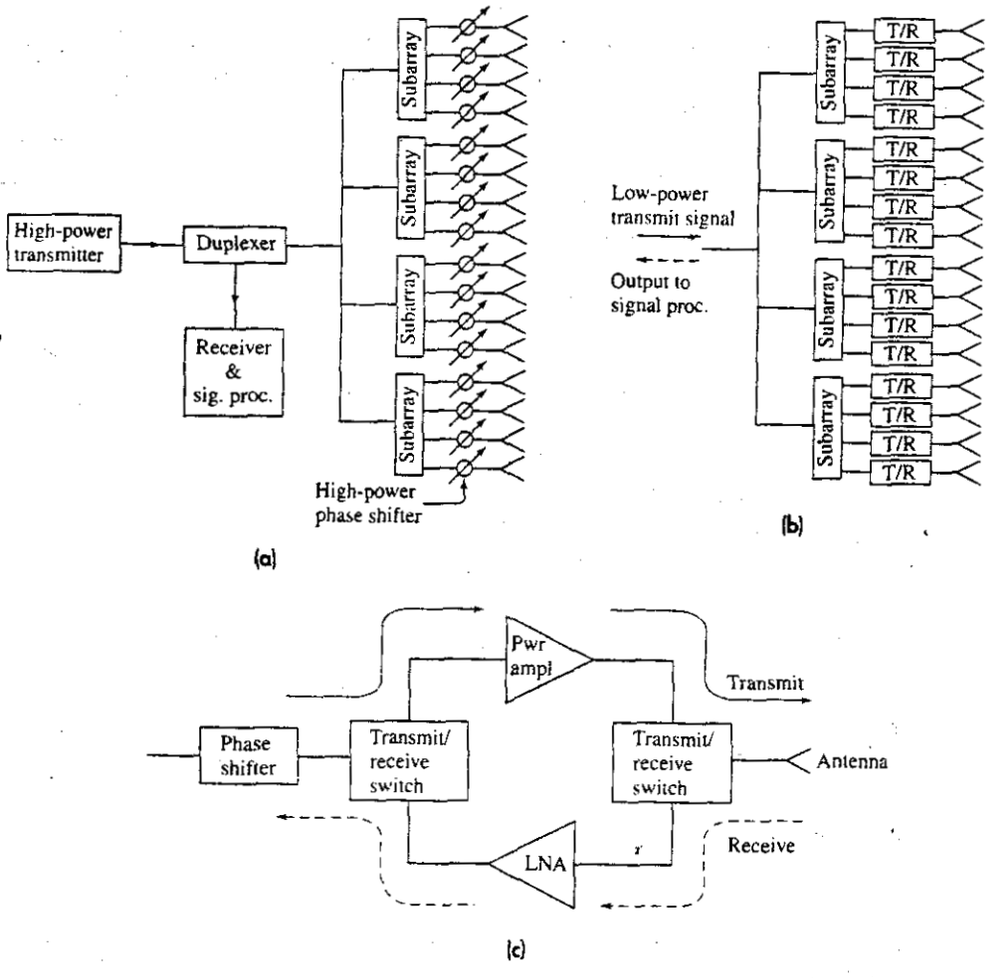


Figure 9.35 Comparison of the passive- and active-aperture array configurations, both shown with subarrays. (a) Passive-aperture phased array with a single high-power transmitter on the left and high-power phase shifters at each element. (b) Active-aperture array with T/R (transmit/receive) modules at each element. Sometimes there might be a booster amplifier on transmit at each subarray (that is bypassed on receive) and/or a time-delay element for increasing the signal bandwidth. (c) T/R module configuration, with a power amplifier on transmit and a low-noise amplifier (LNA) on receive.

active approach, and the choice as to which to use will depend on how the pros and cons balance for any particular application.

There are some generalities that can be cited regarding the relative costs of the two types of arrays. In the past it has been observed that it has often been cheaper to achieve a required total average power by employing a single high-power transmitter rather than obtain the same power by combining the outputs from a number of low-power sources. It has also been observed that high-power vacuum-tube transmitters usually have been

across the aperture has the character of a stair-step, with one step over each subaperture.[88] This can result in what are called *quantization lobes.* Such lobes can be reduced by either overlapping the subapertures or inserting a small amount of randomization[89] to the phases of the subapertures. The term subarray also has, at times in the past, been applied to the array feed networks of a constrained feed system that produce sum and difference *patterns for monopulse angle measurement.*[90]

Subarrays can achieve wide signal-bandwidth by employing a variable time-delay element at each subarray.[91] Although time-delay elements allow wideband operation, they have not been economically feasible to use at each radiating element of a large phased array. A compromise is to utilize them at each subarray.

The original U.S. Navy Aegis (AN/SPY-1) phased array radar system for ship air defense employed 32 transmitting and 68 receiving subarrays of different sizes.[92] Each of the 32 transmit subarrays had its own CFA (crossed-field amplifier) power amplifier.

Triangular Element Spacing[93,94] For the most part, it has been assumed here that the radiating elements of a phased array are laid out in a square grid. A triangular, rather than a square, arrangement of elements, however, permits a savings in the total amount of elements needed in an array when the spacings are determined to avoid grating lobes (spurious beams comparable to the main beam that appear in the radiation pattern when the element spacings are too large). The reduction in the number of elements depends on the solid angle over which the main beam is scanned. For example, if the beam is to be scanned anywhere within a cone defined by a half-angle of 45°, the number of elements required with equilateral triangular spacing is 13.4 percent less than with square spacing. (The altitude of the equilateral triangle in this case is equal to the element spacing of a square grid.) The smaller the angular region to be covered, the less is the saving. Triangular spacings are more likely to produce higher sidelobes in some directions because of the phase quantization of digital phase shifters.[95] For most applications, however, these quantization lobes do not significantly limit system performance.

Active-Aperture Phased Arrays There are several different ways to configure a phased array radar system. One traditional method is to use a high-power phase shifter at each antenna element, with a single high-power transmitter and a single receiver for the entire radar. This has sometimes been called a *passive aperture,* or passive array, in contrast to what is known as an *active aperture,* or active array. They are illustrated in Fig. 9.35a and b respectively. An active aperture has a transmitter (low or modest power) at each antenna element. There is also included at each element an individual receiver, phase shifter, duplexer, and control, as well as the RF power source. Thus an active-aperture phased array implies there is a miniature radar system at each of the array elements. The construction of the electronics at each element of an active-aperture array radar can be highly integrated as a module or with MMIC (monolithic microwave integrated circuitry) construction. The active-aperture module, Fig. 9.35c, is called the T/R (transmit/receive) module or transceiver module.

The passive aperture has had the advantage of usually being cheaper than an active aperture, but cost has not been the only criterion used in selecting a particular array architecture. There are factors in favor of the passive approach and factors in favor of the

Eglin Air-Force Base, Florida, in 1969.[98,99] In some respects, it can be said to have been the first active aperture phased array, in that it employed 5184 individual transmitter units, one at each of the radiating elements. It operated at UHF (centered at 442 MHz). Separate receiving and transmitting arrays were used since it was cheaper to employ two arrays rather than one array with duplexers. The receiving aperture was larger than the transmitting aperture and employed 19,500 receiving elements. Only 4660 of the elements in the receiving array were active (had receivers connected to them), the rest were inactive and were terminated. The receiving elements were arranged in a thinned, space-taper manner to reduce the number of receivers required while maintaining a suitable sidelobe level. The transmitters used a highly reliable tetrode as the final amplifier stage to produce a peak power output of 10 kW at each element. The total peak radiated power was 32 MW and the beamwidth was 1.4°. This radar was considered a success. It has been upgraded, but for a long time it continued to use a vacuum tube as the final stage at each transmitting element since it was cheaper to do so than convert to solid state. The radar has performed well its role in detecting, tracking, identifying, and cataloging earth-orbiting objects.

The first all solid-state active aperture phased array was the AN/FPS-115, more commonly known as Pave Paws.[100] It operated at UHF and was designed to detect subma-

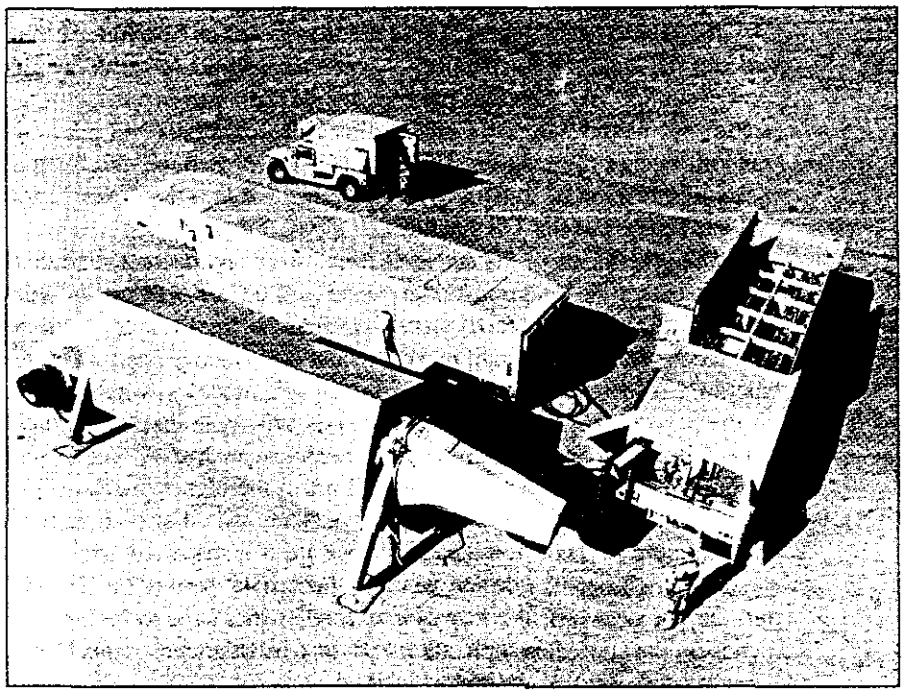


Figure 9.36 (a) The U.S. Army's THAAD X-band Ground-Based Radar (GBR) active-aperture radar for tactical ballistic defense. The 25,000-element array antenna equipment is shown at the lower left. Just behind it is the electronics equipment unit that houses signal and data processing, uninterruptable power supplies, receiver/exciter, and waveform generator. To the right is the cooling unit. At the upper left-center is the operator control unit. Not shown is the 1.1-MW diesel generator prime power unit.
(Courtesy, Raytheon, Inc.)

more efficient than the solid-state transmitters employed in active-aperture arrays at the higher microwave frequencies. These advantages of a single transmitter might be balanced by the fact that in an active aperture the power from the individual, low-power sources is combined in space so there is no loss in distributing the power as there is in a passive aperture that uses a constrained feed. If the passive array is space-fed, the loss is less than would be experienced by an array with a constrained feed. In a passive aperture the phase shifters must be capable of handling higher power than the phase shifters in the active aperture. The phase shifter loss in a passive aperture is often some fraction of a dB, which is low. The phase shifter loss in an active aperture, however, can be much larger than the phase shifter loss in a passive array since it occurs at low power levels and can be made up by increasing the gain of the power amplifiers that are located between the phase shifter and the antenna element. Thus there is little effect of phase shifter loss on the performance of an active aperture. Similarly on receive, the loss introduced by the phase shifter in the passive aperture can degrade the receiver noise figure. Loss in the phase shifters of an active aperture on receive is less important since the phase shifters are preceded by a low-noise amplifier that determines the noise figure.

There are various corporate-fed beamformer architectures that can be used with active-aperture arrays, depending on (1) whether the amplitude taper is applied in the beamformer network or in the T/R modules, (2) the degree of reliability (mean time between failures) required, and (3) whether the array is narrow or wide band.[96]

It has also been claimed[97] that the distributed architecture of the active aperture can smooth the effect of pulse-to-pulse amplitude and phase variations introduced by the RF power source, and therefore increase the MTI improvement factor and obtain better detection of moving targets in clutter. Since the amplitude and phase variations tend to be random among the many modules of the active aperture, the fluctuations combine in a noise-like fashion to smooth the effect. This assumes that a single large power supply is not used to power the T/R modules.

The proponents of the active-aperture array architecture state that one of its chief attributes is that the total transmitter power will be less than that of a passive aperture (that employs a constrained feed) since it avoids the losses of the high-power phase shifters and feed system. This might make the cost of the active aperture radar less than a passive aperture, if the cost of the active aperture T/R modules is not too large. The proponents of the passive aperture, on the other hand, will argue that the high cost of T/R modules, especially at the higher microwave frequencies (such as *X* band), as well as the lower efficiency of solid-state transmitters, will offset the higher losses of the passive aperture to make the active aperture more expensive. They would also argue that the cost advantage is even more in favor of the passive aperture if a space-fed array is used.

Although one can debate whether the active aperture or the passive aperture is better, the choice—just like many other choices that have to be made in engineering—depends on the particular application and the particular constraints imposed. It is not always obvious without full analysis which approach results in a more cost-effective phased array radar.

Examples of Active-Aperture Phased Array Radars The first "modern" phased array radar was the AN/FPS-85 satellite surveillance radar, which became operational at

radiating elements, each with its own gallium arsenide T/R module. In order to be able to operate with a wide-bandwidth signal, the aperture is divided into 72 subapertures, each containing 352 active elements. There is a time-delay steering element at each subarray to permit the use of wideband waveforms without distortion. The array aperture is 9.2 m^2 (almost 100 ft^2), which is quite a large aperture for an *X*-band phased array. Because there are so many of them, the T/R modules are a very important part of this radar (or any active-aperture array) and are the largest cost element of the array. It was said that "every $100 saved in the T/R module cost corresponds to $2.5M for the complete array." The entire array weighed over 46,000 pounds.

A different solid-state active-aperture phased array radar for theater ballistic missile defense is the Israeli *L*-band EL/M-2080 shown in Fig. 9.36b. It performs search, acquisition, and fire control as part of the stand-alone Arrow weapon system.[102] It is said to have detection ranges of hundreds of kilometers and can simultaneously track and engage many tens of missiles.

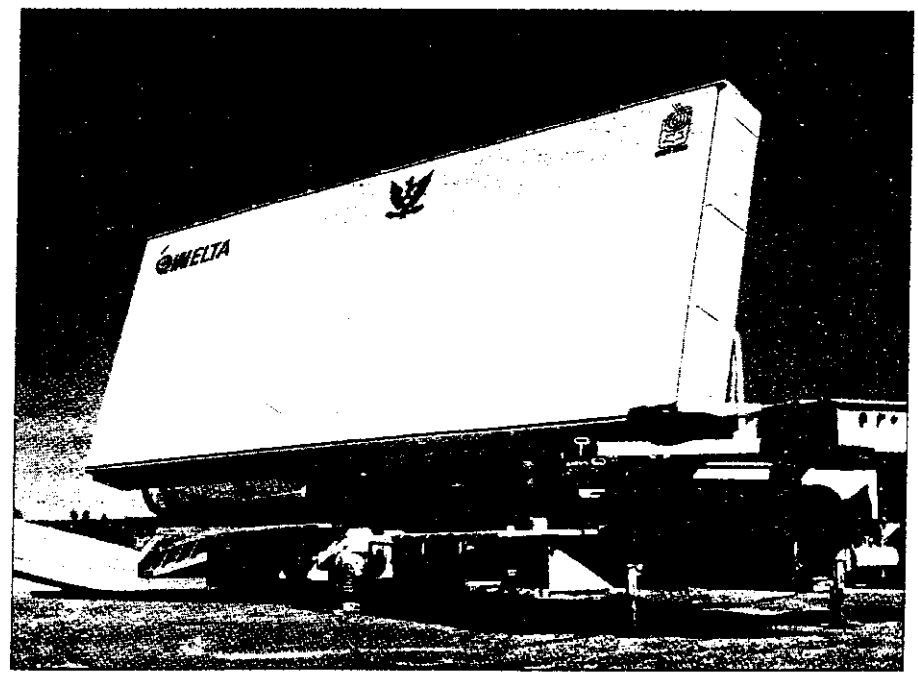
The active-aperture phased array has also been considered for airborne (fighter/attack) radar and for ship self-defense radar. In the airborne application the number of modules (and radiating elements) might be from 1000 to 2000, and for shipboard air defense there might be from 4000 to 8000 modules. A serious limitation of any fixed electronically steered phased array radar mounted in the nose of a fighter aircraft is its limited angle coverage. Although a phased array is usually said to be able to steer $\pm 60°$ coverage in angle, the main-beam gain decreases, the beam broadens, and the sidelobes rise significantly even before the beam approaches 60° from boresight. A fighter aircraft, however, requires its antenna beam to steer to even greater angles than 60°. This is practical to do with a mechanically steered antenna, but not with an electronically steered phased array. In the AN/APG-77 radar for the F-22 fighter aircraft, provision was made as a growth feature to allow installation of a "side array" on each side of the aircraft to allow coverage at angles beyond that available with a fixed electronically steered phased array antenna.[103] The addition of the side arrays, of course, increases the cost of an already expensive active-aperture radar.

Example of a Russian Phased Array Radar Architecture The Russians have generally employed a different approach from the United States to the design of their phased array air-defense radar systems. The U.S. Army's C-band Patriot and the U.S. Navy's *S*-band Aegis systems use multifunction phased array radars that perform the various radar functions required for air defense with a single system operating within a single frequency band. As discussed later in Sec. 9.14, this represents a compromise since the optimum frequencies for search and for track of aircraft are different. The Russians, on the other hand, use separate radars for each function, and the radars operate at frequencies more suitable for their particular function. Since Russia is a vast country requiring many air-defense systems, they emphasized a low-cost approach to radar design.

The Russian air-defense system S300V (NATO designation SA-12), used a 10,000-element *X*-band lens-array radar for multiple-target tracking and weapon guidance. The NATO designation of this *X*-band radar is *Grill Pan*. (The information in this subsection is taken from a paper by David Barton.[104]) The low cost and low RF loss experienced by this system was due, in part, to the separation of the surveillance and tracking functions

rine-launched ballistic missiles fired at the United States. It had a secondary mission to perform space surveillance. Pave Paws employed 1792 active elements arranged in a circular aperture 22.1 m (72.5 ft) in diameter, plus 885 dummy elements. The peak power per T/R module was 335 W, which produced a total peak power of 600 kW and an average power of 150 kW per face. A Pave Paws radar consisted of two faces to cover 240° in azimuth. There were 56 subarrays, each containing 32 modules feeding 32 radiating elements. Its range was said to be 3000 nmi for a 10 m^2 target. There were four operational Pave Paws radars in the United States. One of these, located in Georgia, was expected to be increased in capability by 10 dB (by employing more elements) and replace the AN/FPS-85. A larger version of Pave Paws has also replaced the parabolic torus reflector antennas in the Ballistic Missile Early Warning System (BMEWS).

The THAAD (Theater High Altitude Area Defense) radar, Fig. 9.36a, is an active aperture radar designed for ballistic missile defense.[101] It has also been known as the Ground based Radar, or GBR. Radars for ballistic missile defense have to perform target detection, acquisition, track, identification (recognition), discrimination (of reentry vehicles from decoys and chaff), and assessment of target kill as well as in-flight communication to the defensive missile. The THAAD GBR is an *X*-band radar with 25,344



(b) EL/M-2080 *L*-band active aperture radar for the Israeli Arrow Tactical Ballistic Missile Defense System.
(Courtesy IAI/ELTA Electronics Industries, Ltd.)

radiating elements, each with its own gallium arsenide T/R module. In order to be able to operate with a wide-bandwidth signal, the aperture is divided into 72 subapertures, each containing 352 active elements. There is a time-delay steering element at each subarray to permit the use of wideband waveforms without distortion. The array aperture is 9.2 m^2 (almost 100 ft^2), which is quite a large aperture for an *X*-band phased array. Because there are so many of them, the T/R modules are a very important part of this radar (or any active-aperture array) and are the largest cost element of the array. It was said that "every $100 saved in the T/R module cost corresponds to $2.5M for the complete array." The entire array weighed over 46,000 pounds.

A different solid-state active-aperture phased array radar for theater ballistic missile defense is the Israeli *L*-band EL/M-2080 shown in Fig. 9.36b. It performs search, acquisition, and fire control as part of the stand-alone Arrow weapon system.[102] It is said to have detection ranges of hundreds of kilometers and can simultaneously track and engage many tens of missiles.

The active-aperture phased array has also been considered for airborne (fighter/attack) radar and for ship self-defense radar. In the airborne application the number of modules (and radiating elements) might be from 1000 to 2000, and for shipboard air defense there might be from 4000 to 8000 modules. A serious limitation of any fixed electronically steered phased array radar mounted in the nose of a fighter aircraft is its limited angle coverage. Although a phased array is usually said to be able to steer $\pm 60°$ coverage in angle, the main-beam gain decreases, the beam broadens, and the sidelobes rise significantly even before the beam approaches 60° from boresight. A fighter aircraft, however, requires its antenna beam to steer to even greater angles than 60°. This is practical to do with a mechanically steered antenna, but not with an electronically steered phased array. In the AN/APG-77 radar for the F-22 fighter aircraft, provision was made as a growth feature to allow installation of a "side array" on each side of the aircraft to allow coverage at angles beyond that available with a fixed electronically steered phased array antenna.[103] The addition of the side arrays, of course, increases the cost of an already expensive active-aperture radar.

Example of a Russian Phased Array Radar Architecture The Russians have generally employed a different approach from the United States to the design of their phased array air-defense radar systems. The U.S. Army's C-band Patriot and the U.S. Navy's *S*-band Aegis systems use multifunction phased array radars that perform the various radar functions required for air defense with a single system operating within a single frequency band. As discussed later in Sec. 9.14, this represents a compromise since the optimum frequencies for search and for track of aircraft are different. The Russians, on the other hand, use separate radars for each function, and the radars operate at frequencies more suitable for their particular function. Since Russia is a vast country requiring many air-defense systems, they emphasized a low-cost approach to radar design.

The Russian air-defense system S300V (NATO designation SA-12), used a 10,000-element *X*-band lens-array radar for multiple-target tracking and weapon guidance. The NATO designation of this *X*-band radar is *Grill Pan*. (The information in this subsection is taken from a paper by David Barton.[104]) The low cost and low RF loss experienced by this system was due, in part, to the separation of the surveillance and tracking functions

rather than to their being combined in one multifunction array. The space-fed lens array utilized multimode monopulse horn-feeds, so it did not experience the larger loss that a constrained-feed system might have. Faraday rotation dual-mode ferrite phase shifters were used which operated with circular polarization. Rather than convert the normal linear polarization to circular for operation in the phase shifter, and back again to linear (which is normally done with the dual-mode ferrite phase shifter), the two polarization transformations characteristic of U.S. dual-mode ferrite phase shifters were omitted by having the radar transmit and receive circular polarization. The array received the orthogonal circular polarization; that is, it received left-hand circular if right-hand circular was transmitted. (The polarization of the echo from an aircraft when illuminated by one sense of circular polarization contains both right and left circular polarizations in roughly equal amounts.) The ferrite phase shifters were nonreciprocal, but the phase shifters did not have to be reset after transmission in order to receive since the same phase settings were applicable when the received signal was of a circular polarization orthogonal to that transmitted. Since the polarization on receive was different from that on transmit, the receiver was partially isolated from the transmitter. The isolation due to the orthogonal polarizations, plus the use of a rugged cyclotron-wave electrostatic amplifier as the receiver front-end, eliminated the need for a duplexer or solid-state receiver protector, further reducing the loss. (The electrostatic amplifier had a 3-dB noise figure and could withstand average leakage power of several hundred watts and much higher peak power.)

The Russian ferrite phase shifters had two sections in series to provide 720° of phase shift (instead of the more usual 360°). Each section had its own control coil. One coil was for setting the phase required to steer in elevation and the other for the phase to steer in azimuth. All the row coils were in series with each other to provide azimuth steering and all the column coils were in series to provide elevation steering. The 10,000 element array (100 × 100) required only 100 row drivers and 100 column drivers. There were no driver or logic circuits, data busses, or d-c power busses needed in the aperture for determining the combined phase shift for steering in two angles. The total two-way RF loss from transmitter to receiver, excluding propagation loss, was about 3 dB for the Grill Pan. This compares to the 7 to 12 dB losses found in comparable Western systems.

The lack of individual control of each phase shifter, however, can cause the phase errors at the elements to be correlated over entire rows or columns of elements. Loss of a driver can result in the loss of an entire row or column of elements. The simplicity and low cost of this method for setting phase shifters make it more difficult to achieve low sidelobes.[105]

In addition to lower system cost and more optimum frequency usage, another advantage in using multiple radars in an air-defense system rather than a single multifunction phased array is that the individual radars can employ long-dwell medium-prf and high-prf pulse doppler waveforms that are needed to detect moving aircraft and missile targets in heavy clutter.

The *X*-band Grill Pan described above was the target tracking and guidance radar for the SA-12 air-defense system. Air surveillance was performed with the *S*-band Bill Board radar, which is a scanning beam 3D radar using a phase-scanned planar array with slotted waveguide radiators. The array could be stowed for transport in the short time of one minute. The SA-12 also employed a separate sector-search radar for detection of tactical

ballistic missiles. The Russian SA-10, a similar air-defense radar system, also deployed on a tower a horizon search radar called Clam Shell for detection of low-altitude targets.

According to Barton, this approach to air defense seems to reflect "the Russian military's insistence on high performance against targets of low cross section in environments containing rain, chaff, and other sources of clutter, an almost insoluble problem when the multifunction approach is adopted."

Simultaneous Multiple Beams (Analog) As has been noted, the phased array can form a number of multiple simultaneous beams. This is important for monopulse angle tracking; but in this subsection multiple beams are considered to be more than what is normally needed for tracking. In principle, an N-element array can generate N independent beams. Multiple beams allow parallel operation and a higher data rate than with a single beam. The multiple beams may be fixed in space, steered independently, or steered as a group. They can be generated on transmit as well as on receive. When multiple beams are generated on receive, the transmit beam can have a wide radiation pattern that encompasses the coverage of the multiple receive beams. In the past, multiple beams were generated by analog components, but it is now advantageous to employ digital methods for beam-forming. Digital beam-forming is not appropriate for transmit, but this is not necessarily a limitation since the transmitting antenna is relatively simple and employs a broad radiation-pattern (omnidirectional in some cases).

The simple linear array that generates a single beam can be converted to a multiple-beam array by attaching additional fixed phase shifts to the output of each element. Each beam to be formed requires one additional phase shift per element, as in Fig. 9.37. For

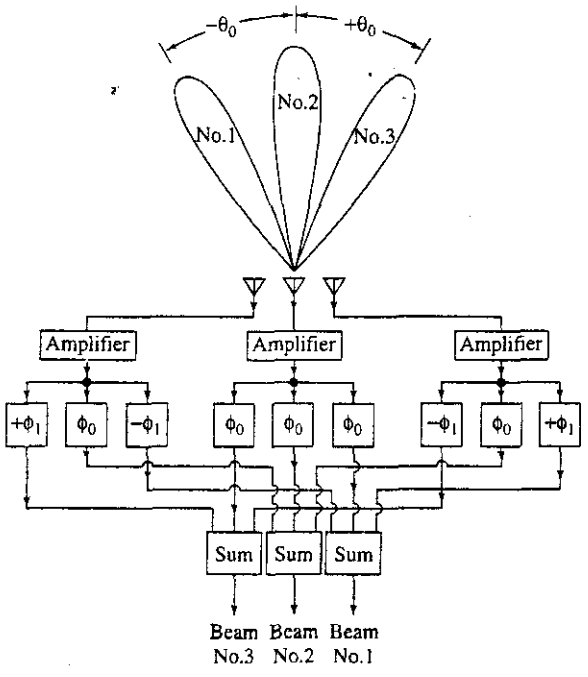


Figure 9.37 Simultaneous beam formation on receive (three beams shown). ϕ_0 = constant phase; $|\phi_1 - \phi_0| = |\Delta\phi| = |2\pi(d/\lambda) \sin \theta_0|$.

simplicity, the array in this figure is shown with but three elements, each with three sets of phase shifters. One set produces a beam broadside to the array ($\theta = 0$). Another set of three phase shifts generates a beam in the direction $\theta = +\theta_0$. The angle the beam points is determined by the relationship $\theta_0 = \sin^{-1}(\Delta\phi\ \lambda/2\pi d)$, where $\Delta\phi$ is the phase difference inserted between adjacent elements. Similarly, there is a set of phase shifters to produce a beam in the $-\theta_0$ direction. The receiving beam-forming networks may be at RF or at IF. At IF, tapped delay lines have been a convenient method to obtain the necessary phase shifts.

Butler Beam-Forming Array, or Butler Matrix[106] This is an analog RF beam-forming network, an example of which is shown in Fig. 9.38. It consists of 3-dB directional couplers (or hybrid junctions) and fixed phase shifts to form N contiguous beams with an N-element linear array. The number N is an integer expressed as some power of 2, that is, $N = 2^p$. The 3-dB directional coupler is a four-port junction. A signal fed into one port will divide its power equally between two other ports and no power will appear at the fourth port. In the process, a 90° phase difference is introduced between the phases of the two equally divided signals. Likewise, a signal inserted at the fourth port will divide its power equally between the same two ports with a 90° relative phase difference, and no power will appear at the first port. The relative phase difference in this case is of opposite sign compared to the phase difference resulting from a signal introduced at the first port.

The example of a *Butler matrix* depicted in Fig. 9.38 is an eight-element array that produces eight independent beams. It utilizes 12 directional couplers and eight fixed phase

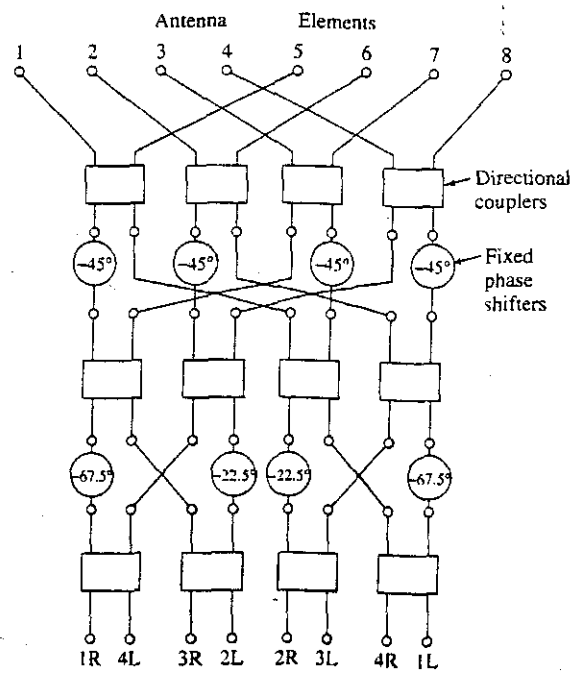


Figure 9.38 Eight-element Butler beam-forming matrix.

shifts. A Butler matrix has 2^p inputs and 2^p outputs. The number of directional couplers or hybrid junctions required for an N-element array is equal to $(N/2) \log_2 N$, and the number of fixed phase shifts is $(N/2) \log_2 (N - 1)$.

The Butler matrix is theoretically lossless in that no power is intentionally dissipated in terminations. There will always be, however, a finite insertion loss due to the inherent losses in the directional couplers, phase shifts, and transmission lines that make up the network. In a theoretically lossless, passive antenna radiating multiple beams, the radiation pattern and the crossover level of adjacent beams cannot be specified independently. With uniform illumination, as in the Butler matrix, the crossover level is 3.9 dB ($2/\pi$ in voltage) below the peak value of the beam. Crossover is independent of the beam position, element spacing, and frequency. The pattern is of the form $(\sin u)/u$, with -13.2 dB peak sidelobe.

The low crossover level of a Butler matrix is one of its disadvantages. If a lossless network could be achieved with a cosine illumination so as to reduce the peak sidelobe compared to that obtained with uniform illumination, the crossover level would be even worse (at a value of -9.5 dB); but the peak sidelobe would be -23 dB instead of -13.2 dB obtained with a uniform aperture illumination. By combining the output beams of the networks with additional circuitry, the Butler matrix can be modified to obtain an aperture illumination with lower sidelobes. The beamwidth will be widened, the gain lowered, and the network will no longer be theoretically lossless.

The flow diagram of the well-known fast Fourier transform (FFT) is similar to the circuit diagram of the Butler matrix.[107] The Butler matrix, however, was known before the FFT. Some of those familiar with the Butler matrix were surprised when they learned they were unknowingly using what would become an important procedure for the mathematical calculation of the Fourier transform. The FFT uses $(N/2) \log_2 N$ computations for an N-point transform, the same as the number of junctions needed for a Butler matrix.

The antenna equivalent of the conventional Fourier transform is called the *Blass beamforming network*. It requires N^2 couplers for N inputs and N outputs, just as the conventional Fourier transform requires N^2 computations for an N-point transform.

There have been other analog beam-forming methods considered for generating multiple beams with a phased array, as discussed in Sec. 8.7 of the second edition of this text. Most of them, as well as those mentioned above, are more suited for generating beams in one angular coordinate than in two coordinates.

Pattern Limitations in (Analog) Multiple-Beam Antennas In the above discussion of the Butler matrix it was said that it was not possible to arbitrarily select the crossover level of adjacent beams. The crossover of 3.9 dB down from the peak of the beam is determined by the theoretical lossless nature of the Butler matrix beamformer. This crossover level, however, is a characteristic of any passive, lossless antenna that forms multiple independent beams when the aperture illumination is uniform, as was indicated by Warren White.[108] He showed that a passive lossless beam-forming antenna requires that the individual beam patterns be orthogonal in space, which means that

$$\int_0^{2\pi} d\theta \int_{-\pi/2}^{+\pi/2} E_j(\theta,\phi)\, E_k^*(\theta,\phi) \cos\phi \, d\phi = 0 \qquad [9.44]$$

where θ = the longitude angle on the unit sphere, ϕ = the latitude angle, $E_j(\theta,\phi)$ = the radiation pattern associated with the jth input terminal, and $E_k^*(\theta,\phi)$ = the complex conjugate of $E_k(\theta,\phi)$. Independent orthogonal beams mean that when two or more beam input ports are simultaneously excited, the resulting radiation is a linear superposition of the radiations that would be obtained when the ports are excited separately. In addition, when a signal is applied to one port it should have no output at the other ports. An antenna which is lossless and passive means that the radiated power is the same as the input power.

As has been mentioned, the $(\sin u)/u$ pattern produced by a uniform illumination is an example of a pattern with orthogonal properties and produces a crossover of 3.9 dB between adjacent beams. With a cosine-squared illumination, however, White stated that the crossover with a lossless passive multibeam antenna is 15.4 dB down, and for a Hamming distribution it is 18.4 dB down. These are unacceptably high values. They apply to one-dimensional antennas. For two-dimensional antennas with pencil beams, the crossover in dB is double.

White suggested that these poor values of beam crossover can be avoided by use of active elements (amplifiers) inserted between the antenna radiating elements and the resistive beam-forming network. The sensitivity of the receiving antenna is established by the active circuits (low-noise amplifiers) at each element. Any losses that follow the amplifier do not affect the overall receiving system sensitivity. This method for generating efficient multiple-beam array antennas is similar to what is done in a digital beam-forming array, as will be described later in this section.

Systems Considerations for Multiple Beam-Forming Arrays A surveillance radar with a broad-beamwidth transmitting antenna (beamwidth θ_T) and with N fixed, narrow receiving beams (each of beamwidth θ_R) that cover the same angular region (so that $N\theta_R = \theta_T$) can have performance equivalent to a conventional radar that uses a single rotating transmit-receive beam of width θ_R. Equivalent performance means, in this case, that the two different types of radars will be able to detect the same size targets at the same range and have the same revisit time, which, in the case of a mechanically rotating antenna, is the antenna scan time, or the time to rotate 360° in azimuth. This requires that the signal integration time for the radar with fixed multiple beams also be the same as the time for the rotating antenna to make one revolution. The transmitting antenna gain of the multiple-beam radar is $1/N$th that of the scanning single-beam system. However, the observation time available to each of the fixed beams of the multiple-beam system is increased by $\theta_T/\theta_R = N$. If the signal is integrated without loss over this time, the reduction in transmitting antenna gain in the multiple-beam system is just compensated by the increased signal energy obtained because of its much longer integration time. The assumption of no integration loss is correct for perfect pre-detection (coherent) integration. When post-detection integration is used, however, there will be a theoretical integration loss so that the two radar systems will not be of exactly equivalent performance.

In a radar with fixed multiple receiving beams and a single broad transmitting beam, the two-way sidelobes are not as low as with a single-beam scanning radar. In some applications where low two-way sidelobes are important, it may be desirable to decrease the antenna sidelobes of the multiple receiving beams more than normal.

Multiple beam-forming array radars that use analog beam-forming are usually more expensive and require more equipment than a conventional scanning surveillance radar. They also do not seem to offer significant advantages in performance or capabilities over a conventional scanning radar. For these reasons they have seldom been considered for operational radar system applications. The production of multiple beams using digital beam-forming, to be discussed next, can also be expensive, but the use of digital technology offers advantages not readily available otherwise. Digital beam-forming, therefore, is of much more interest for multiple-beam radar applications than are analog beam-forming methods.

Digital Beam-Forming (DBF)[109–113] Almost all modern phased array radars utilize some sort of digital phase shifting to form a beam; but the term *digital beam-forming* (DBF) usually means something different—the formation of multiple receive beams by digitizing the outputs of the receiving array elements and forming beams by means of a digital processor. Although a single beam can be formed in this manner, digital beam-forming has more often been considered when multiple, simultaneous, directive beams are wanted. When digital beam-forming is employed to generate multiple receiving beams, the accompanying transmitting antenna has a broad beamwidth covering the same total angular region as the multiple receiving beams. The outputs from the multiple beams can be processed in parallel by the radar.

The basic configuration of a digital beam-forming receiving array might be as shown in Fig. 9.39. At each antenna element there is a receiver whose analog signal is digitized by the A/D converter. The A/D converter is a critical component since it can place a limit

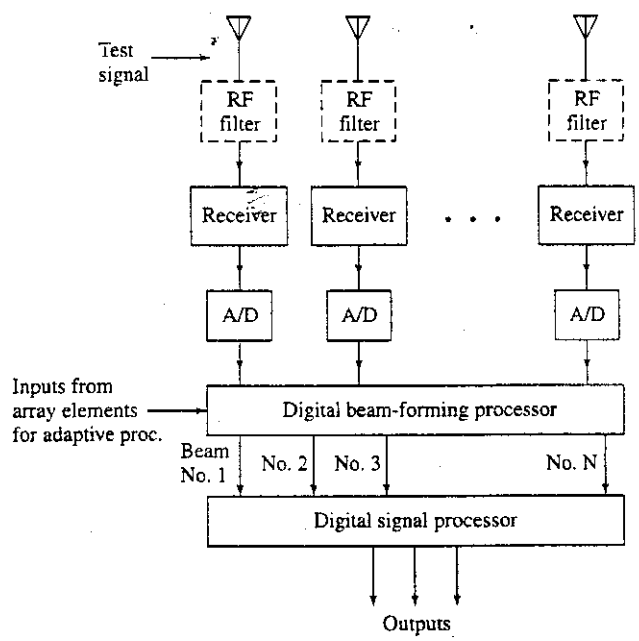


Figure 9.39 Basic configuration of a digital beam-forming receiving array.

on the system bandwidth and dynamic range. The lower the frequency at which the A/D converter operates, the greater can be its dynamic range and bandwidth. For this reason the input to the receiver is usually heterodyned (downconverted) to a lower frequency, filtered, and amplified to a power level suitable for the A/D converter. The receivers must be closely matched to one another so that the relative amplitude and phase of the signals at each element are preserved. If other than a uniform aperture illumination is desired (for example, to reduce the antenna sidelobes), amplitude weights can be applied to the digital output of each element and the weighted signals combined in the processor to produce the antenna beam. If a linear phase weight is applied to the digitized signal at each element, the antenna beam can be made to appear as if steered to a different angular direction. In this manner the digital processor can produce a multitude of receiving beams, each pointing in a different direction. An antenna beam, of course, is not actually formed in space by the processor. The pattern is evident as the variation of the output response of the digital signal processor as a function of angle.

Since the signal-to-noise ratio is established at the digital output of each receiver, there is no loss in signal-to-noise ratio when manipulating the digital outputs to form multiple beams. There can be any number of closely spaced antenna beams formed without degradation in SNR. Digital beam-forming, therefore, does not suffer the limitations inherent in analog beam forming, which are the low crossover of adjacent beams and the loss that occurs when trying to obtain lower peak sidelobes than the -13.2 dB of a uniform illumination.

The operation of the beam former is to take the outputs from each element, apply a complex weight (an amplitude and/or phase) to each, and then sum them to provide the output signal. The form of this output is similar to a digital Fourier transform (or inverse digital Fourier transform). When the phase weights are expressed in terms of the beam pointing angle θ_0, the relative response of the weighted outputs of the N-array elements as a function of the angle θ is represented as

$$g(\theta,\theta_0) = \sum_{n=1}^{N} a_n s_n \exp\left[j2\pi n(d/\lambda)(\sin\theta - \sin\theta_0)\right] \qquad [9.45]$$

where a_n = weight (amplitude for sidelobe control), s_n = output of the nth element, d = element spacing, λ = wavelength, and θ_0 = direction of the maximum beam response. In this manner, $M \leq N$ multiple beams can be formed. For each of the M beams there is a temporal sequence of digital numbers (data) that represent the temporal signals received at each beam position.

The computation of the Fourier transform in real time by the signal processor when there are many elements in the array can require a large number of operations per second. The computation requirements can be significantly decreased if the fast Fourier transform (FFT) is used. The FFT requires that the number of elements be some integer power of 2 (i.e., $N = 2^p$ and p = integer). There is less control of the beam patterns, however, with the FFT than if the conventional digital Fourier transform is used. The patterns of each FFT beam (in $\sin\theta$ space) are the same, the peak sidelobes will be high, and the crossover of adjacent beams is predetermined (by the orthogonality constraint) and may be lower than desired. Thus with the FFT one trades control of the radiation pattern for ease of computation.

Two-Dimensional Beam-Forming The above description of digital beam-forming assumed a linear array forming beams in one angular dimension. With a two-dimensional rectangular array of $M \times N$ elements, beam-forming in two angular dimensions can be accomplished with the two-dimensional FFT. An N-point FFT is taken on each of the M rows to obtain $N \times M$ outputs. This is followed by an M-point FFT on each of the N columns to provide $M \times N$ beam outputs.[112]

Baseband versus IF Digitizing In any radar the digitizing of the received signal can be done either in the IF stage of the receiver or at baseband (zero IF) with I (in-phase) and Q (quadrature) channels. The I and Q channels at baseband are obtained in a similar manner to that described in Chapter 3 for an MTI radar (Fig. 3.29). It will be recalled that the received signal is divided into two baseband channels. The two channels are identical except that the phase of the reference oscillator in one differs from the reference phase in the other by 90°. The bandwidth of the signals in each of the two baseband channels is half that of the signal bandwidth; but two A/D converters are necessary, one in each channel. The smaller bandwidth of each A/D can sometimes be of advantage as opposed to requiring digitizing with the full signal bandwidth in the IF stage. To obtain good performance, the phase difference between the I and the Q channels should not deviate significantly from 90°. This requires careful design or some form of feedback control to maintain the precise phase relationship between the two channels. The signals in the two channels must also be well matched in amplitude.

Digitizing of the received signal may also be done directly in the IF of the receiver. There is only one channel and one A/D converter, so that the problem of balancing the phases and amplitudes of the I and Q channels is not present. The A/D converter has to be capable of greater bandwidth (sampling rate) than with a baseband I,Q arrangement. P. Barton[109] states that the minimum sampling rate of an A/D converter in the IF has to be 5.4 times the signal (half-power) bandwidth. (The sampling rate has to be greater than the theoretical Nyquist rate of 2 × bandwidth because of the need to avoid distortion of the signal spectrum caused by the folding of the spectrum around zero frequency.) This compares with a minimum sampling rate of 1.4 times the signal bandwidth required for each of the two A/D converters of the equivalent baseband implementation.

A/D Converters The sampling rate and the dynamic range of an A/D converter can set limits on what might be achieved in a digital processor. In practical A/D converters the greater the sampling rate (bandwidth), the smaller will be the number of bits into which the A/D converter can digitize a signal. In addition, as the bandwidth or number of bits increase, the size and cost of the converter can also increase. The limitations of A/D converters are practical ones, but their capabilities have been continually improved over the years. Section 3.5 gives values of the performance of A/D converters, but it is always risky to state in a text such as this what the state of the art of such devices might be. The prudent engineer, of course, should always consult current manufacturers catalogs for up-to-date information.

Other Characteristics of Digital Beam-Forming In addition to the attractive aspects of digital beam-forming (DBF) array radars mentioned previously in this subsection, the following are further favorable attributes of such radars.

Self-calibration and error correction—Errors in the phase and amplitude that occur in the analog portion of the receivers of a DBF can be compensated with relative ease in the digital portion. This may be obtained by injecting a precise RF signal at the front-ends of each receiver, or by placing an external RF source at a known position in the near or far field of the antenna, or by use of the echo signal from a well-defined scatterer of the transmitter signal.[110] This precise RF signal is used as the standard to adjust the phase and amplitude at each element.

Low antenna sidelobes—The ability to digitally self-calibrate the array allows the potential for achieving low or even ultralow receiving antenna sidelobes after digital processing. The effect of mutual coupling can also be compensated. It has been said:[114] "Mathematically, the compensation consists of a matrix multiplication performed on the received signal vector. This, in effect, restores the signals as received by the isolated elements in the absence of coupling. An attractive feature is that this matrix is fixed and thus is valid for all desired pattern shapes and scan directions. Although it may be difficult to realize in analog form, it can be readily implemented in a digital beam-forming antenna system."

Adaptive nulling—The flexibility of a DBF array allows nulls to be placed in the antenna radiation pattern in the direction of noise sources (jammers) so as to reduce the noise that enters the receiver. The placement of nulls in a receiving antenna pattern to cancel unwanted noise sources that enter via the antenna sidelobes is a well known technique, and can be done with much simpler systems than a DBF array.[115] Sidelobe cancellation, for example, can be readily achieved with a mechanically scanned reflector antenna and a relatively small number of auxiliary elements. But when one has a DBF array radar, it can be implemented to cancel noise sources that appear in the antenna sidelobes in a manner different from the conventional sidelobe canceler or a fully adaptive array, and with advantages not found in radars that do not have multiple beams. Nulls are achieved in "beamspace" by using one or more formed beams properly attenuated to adaptively cancel noise sources rather than using one or more omnidirectional elements.[116,117,110] If there are J noise sources to be nulled, the angular location of each source is found in a conventional manner and J directive beams are then formed to adaptively cancel each source. Use of beamspace allows cancellation without disturbance to the main beam or the mainbeam sidelobes, except in the immediate vicinity of the noise sources. The use of beamspace for sidelobe cancellation has not normally been used in the past with conventional radar because of the complexity in forming multiple independent beams by other methods. In a DBF array, however, this is not a consideration since the forming of multiple independent beams is what such a system normally does anyway.

Adaptive nulling of clutter as a function of range—Nulls can be formed adaptively in those directions where there are large clutter echoes as well as in those directions in which there are noise sources. Unlike noise, however, clutter might be limited in range (for example, a large mountain or a patch of chaff). In such cases, range-dependent antenna pattern nulls can be formed only around those individual range cells containing localized clutter or chaff.

Correction for failed elements[111]—The complete failure of a sufficient number of antenna elements can seriously degrade the performance of a low sidelobe array antenna. It has been said[118] that it is possible to compensate, however, for the loss of antenna elements in a digital beam-forming receive array by using simple linear operations with the

outputs of a small group of good elements within the array. By properly utilizing the signals received at P elements of the array when P signals are received from different incident directions, it is possible to reconstruct the signal that would have appeared at the failed elements, if all of the array antenna elements have the same radiation pattern and some other restrictions apply. The technique has been said to work when the incident signal directions are not precisely known or even when they are only known to be within a broad angular sector.

Flexible data rates—As has been mentioned, the digital beam-forming array which looks everywhere can have a data rate that varies with the operational situation. This is unlike a conventional rotating radar whose data rate is fixed by the antenna rotation rate. At long ranges, a high data rate is not as important for an air-defense surveillance radar as it is at short ranges where weapons are engaged. Thus a lower data rata can be employed at long range which means that the integration times can be increased and improved target detection can be obtained without an increase in transmitter power. The flexibility of a DBF array to provide unrestricted data rates is important to a military weapon-control radar that should operate with high data rates during an engagement or when the target is seen to maneuver.

Simultaneous multiple functions—In an ubiquitous radar (one which looks everywhere all the time with fixed multiple receive beams) that uses digital beam-forming, it is possible to perform multiple functions simultaneously rather than sequentially. A conventional phased array such as Patriot or Aegis, on the other hand, has to time-share its various functions. Sometimes such radars run out of time to perform all the various functions required of them, so that some functions with lower priority have to be neglected in favor of more important ones. In an air-defense system, for example, priority will go to targets actually being engaged by missiles or to the searching for low-altitude pop-up targets, rather than to long-range surveillance. The ubiquitous radar with DBF, however, can perform its various functions simultaneously so long as it doesn't run out of computer capability.

Improved noncooperative target recognition—The ability to see everywhere all the time with whatever data rate is required is of benefit for those methods of noncooperative target recognition that depend on an observation time longer than normally needed for detection. The imaging of a ship or an aircraft by use of inverse synthetic aperture radar, for example, requires that the aspect of the target change sufficiently so that recognizable doppler-frequency shifts from different parts of the target can be isolated (resolved). Time is needed to allow for the target aspect to change, and time is something available with a DBF array. Recognition of helicopter targets based on the transient "flash" of the rotating blades when they are briefly oriented perpendicular to the radar line of site requires that the radar dwell long enough on the target to detect this phenomena.[119] The fixed beams of DBF provide this flexibility in target observation time.

Lower probability of intercept—It is relatively easy for an intercept receiver to detect the radiated signals of conventional radars at long ranges. To reduce a radar's detectability to a hostile intercept receiver, its peak power should be made as low as possible. The radiated energy should be spread over a wide angular region, over a long time interval, and over a wide frequency band. The ubiquitous DBF array offers the ability to spread the radiated energy over a wide angular region, something not possible with a scanning directive transmitting beam.

A beam-forming array that produces many multiple beams has capabilities not readily obtained with a conventional radar that employs a mechanically scanning antenna. Digital beam-forming offers advantages over analog beam-forming in generating multiple beams in that the same digital outputs from each array element are reused to readily generate multiple beams as well as perform other types of spatial and temporal processing. Digital processing has increased the feasibility and capability of the ubiquitous radar.

Examples of Digital Beam-Forming Arrays Digital beam-forming has been employed in both HF over-the-horizon radar and in 3D air-surveillance radar. Neither of these applications employ the full benefits of digital beam-forming arrays as described above, so they are not indicative of what might be done with this radar concept.

The U.S. Navy's Relocatable Over-the-Horizon Radar (ROTHR), AN/TPS-71, is designed to detect aircraft and ships at ranges from 500 to 2000 nmi.[120] It employs a receiving linear-array antenna 2.7 km in length with a total of 372 monopole antenna elements. At each antenna element there is a receiver that converts the signal to an IF frequency. The digital outputs of these 372 antenna elements are used to form 16 contiguous receiving beams that can be placed anywhere within the radar's angular coverage. Both spatial processing (beam forming) and temporal processing (doppler filtering and matched filtering) are performed with the same digital data from each receiving antenna element.

The *SMART 3D radar,* developed by Signaal of the Netherlands, is a 3D air-surveillance radar with an azimuth-rotating antenna which generates a number of simultaneous multiple-beams in elevation to provide a measurement of elevation angle using digital beam-forming. The original version of this radar was at *S* band, a later version was at *L* band. The *L* band radar formed 16 beams and the *S*-band version 14 beams.

9.10 MECHANICALLY STEERED PLANAR ARRAY ANTENNAS

Mechanically steered planar arrays offer important advantages not available with conventional reflector antennas in some radar applications. They have been employed at the lower radar frequencies (VHF) for air-surveillance applications, at microwave frequencies for 3D radars which need to obtain a measurement of target elevation angle, in airborne radars that operate from the nose of the aircraft, in missiles, civil marine radars, and in low-sidelobe antennas such as used for AWACS. The large number of elements in an array aperture allows better control of the aperture illumination and therefore, better control of the antenna radiation pattern.

Mechanically Rotating Arrays for Air Surveillance The first radar antennas developed by the United States in the 1930s, such as the Army's 100-MHz SCR-270 and the Navy's 200 MHz CXAM were mechanically steered phased arrays. At these low radar frequencies, it was natural to use planar arrays of dipoles since they were consistent with the communications antenna technology of that time. Also, they were well suited for the VHF frequencies at which the early radars operated. As radar frequencies increased, the parabolic

reflector was introduced since it was simpler than a planar array when there had to be a large number of dipoles. The German World War II Wurtzberg radar that operated at 550 MHz employed a parabola. The first U.S. microwave air-defense radar, the S-band SCR-584, also used a parabolic reflector. The parabolic reflector was well-known in optics, and it was not difficult to translate its technology and theory from optics to microwave frequencies.

The early air-surveillance radars were 2D in that they measured azimuth angle and range, but not elevation angle (or height). When pencil-beam 3D radars (such as the AN/SPS-39) first appeared in the 1950s to obtain elevation angle, they originally used a parabolic cylinder antenna with a line feed. These were followed by planar arrays (as in the AN/SPS-48) consisting of rows of slotted waveguides to obtain multiple beams in elevation. These original 3D systems utilized frequency-scanned phased arrays to electronically steer one or more beams in one angular coordinate (elevation). The other angle coordinate was obtained by mechanically rotating the entire antenna 360° in azimuth.

Almost all modern 3D air-surveillance radars employ a planar array antenna of some type that is mechanically steered in azimuth and which uses some form of electronic steering or beam-forming in elevation. Mechanical rotation is satisfactory for the air-surveillance application since it is not necessary to have rapid beam switching among many targets as it is in the weapons control application. The mechanically rotated planar array not only provides a convenient means for obtaining electronic steering in elevation, but the availability of many elements in the antenna also provides more flexibility in achieving the desired radiation pattern.

Mechanically Steered Slotted Planar Arrays[121–123] This type of planar array antenna is widely used for radars mounted in the nose of an aircraft. An example is shown in Fig. 9.40. Such antennas are common for military radars used in fighter and attack aircraft, airborne weather radar, and for missiles. An important advantage of this type of antenna for application in the nose radome of an aircraft is that it can be made relatively thin. This allows a larger diameter antenna to be mechanically scanned within an aircraft's nose radome than is possible with a relatively thick parabolic antenna and a feed that projects a distance from the reflector surface.

Microwave planar array antennas generally employ radiators that are slots cut into waveguide, a simple example of which was shown in Fig. 9.31c. They may be edge slots cut into the narrow wall or slots cut into the broad wall. The use of slots cut into the narrow wall allows the rows of an array to be spaced closer than if they were cut into the broad wall. The slot is the so-called *Babinet equivalent* of the dipole. Its pattern resembles that of a dipole, except that a vertical slot radiates horizontal polarization and a vertical dipole radiates vertical polarization. The waveguide feed structure lies directly behind the radiating slots of the array antenna.

There are two basic methods for structuring a series-fed mechanically scanned slotted waveguide radiator. One is called the standing wave, or resonant, configuration. The other is the traveling wave, or nonresonant, configuration. The spacings between elements in a resonant array is half a guide wavelength. (The guide wavelength is the wavelength measured in the waveguide rather than in free space.) The beam is broadside to the aperture, but the half-wavelength spacing between elements means that impedance mismatches

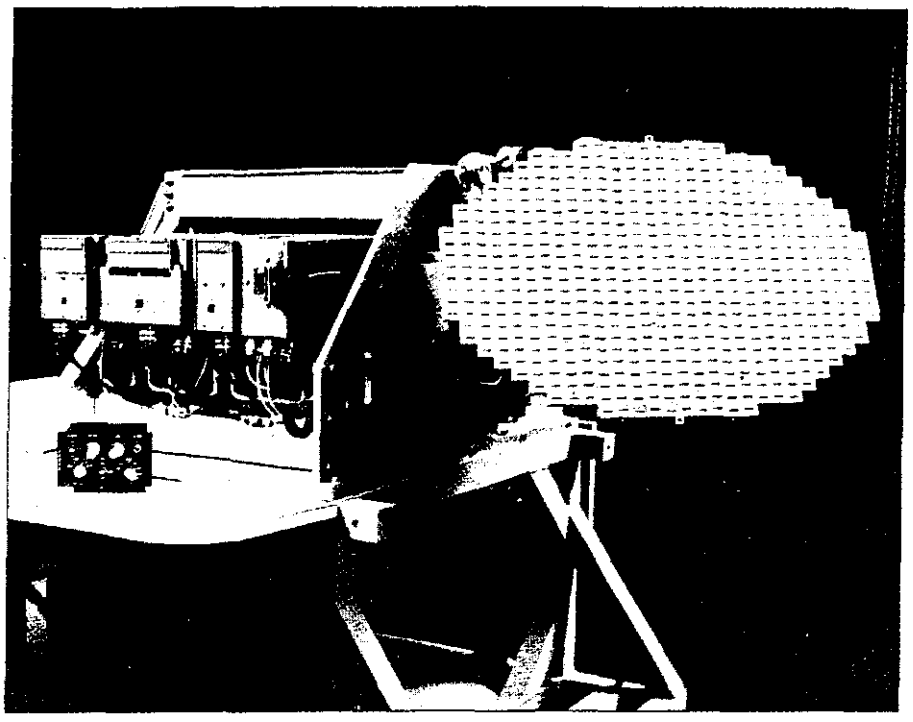


Figure 9.40 Mechanically steered planar array antenna for the AN/APG-68 airborne radar found on the F-16 aircraft.
(Courtesy Northrop Grumman Corp.)

at the elements can accumulate and cause a high VSWR (large mismatch). This results in restricting the number of elements in a linear array to less than 20.[124] which might be too low for most radar applications. Resonant arrays also have narrow bandwidths. The nonresonant array usually has an element spacing greater than half a guide wavelength and so does not experience the high VSWR or the narrow bandwidth of the resonant array. The waveguide of the nonresonant array, however, has to be terminated with a matched load to absorb the fraction of the input power that is not coupled to the other elements. The amount of power lost because of the matched load is about 5 to 10 percent of the power at midband.[121] Thus the array efficiency will be less than unity. (If the array is not well matched, a portion of the signal can be reflected from the termination and radiate as a high sidelobe in some spurious direction.) The direction of the peak of the beam radiated by a series-fed nonresonant array varies with frequency; similar to a frequency-scan array, but much less dramatic. With a wavelength λ, the beam is directed at an angle θ given by

$$\sin \theta = \lambda/\lambda_g - \lambda/2d \qquad [9.46]$$

where λ_g = guide wavelength, and d = element spacing. Yee and Voges[123] state that in most nonresonant slot arrays, $2d$ is selected to be greater than λ_g so that the angle of the

beam will move toward the load-end of the array as the frequency is increased. The frequency dependence of the beam direction might need to be taken into account in some applications.

The mechanically scanned planar array can be designed to provide monopulse beams in two angular coordinates.

Endfire Arrays In the above we have considered the array that radiates its main beam perpendicular (broadside) to the aperture. It is also possible to radiate a beam parallel to the aperture. These are known as *endfire array antennas.* If we consider the dipole as the radiating element, it must be oriented so that its element pattern allows radiation in the endfire direction, and the spacing between elements and the phase shift at each element allow propagation in the endfire direction. For example, the spacing between elements in an equal-amplitude linear endfire array fed from one end might be a quarter-wavelength, with phase shifts of $\pi/2$ radians between elements, to give an antenna pattern with most of its energy oriented in one endfire direction. In this example, the phase is retarded progressively by the same amount as that experienced by the traveling wave from one element to the next. (A phase retardation of 0.6π radians actually produces a higher directivity for this endfire antenna, according to the Hansen-Woodyard criterion, as mentioned by Kraus.[125])

The Yagi-Uda antenna, which originated in Japan, is a simple and inexpensive example of an endfire array. It consists of a single driven dipole plus several spaced parallel rods that form an endfire array, Fig. 9.41. Each rod may be thought of as a short-circuited dipole. The rod located behind (to the left of) the driven dipole acts to reflect the energy to the forward direction. One or more spaced rods in front of the driven dipole direct the energy forward. The rods are known as either *reflectors* or *directors.* The reflector might have a length of about a half-wavelength and be spaced a quarter-wavelength behind the driven dipole. The directors are slightly smaller (by about 10 percent) with spacings about a third of a wavelength.

Endfire antennas can be arrayed and have been used in radars, especially at VHF and UHF. An example is the antenna for the E2C Airborne Early Warning radar (that was shown in Fig. 3.45a). An advantage of the endfire antenna for this application is that a narrower beamwidth can be obtained in the vertical dimension without the need for a large vertical aperture. The beamwidth of an endfire antenna in the dimension orthogonal to its

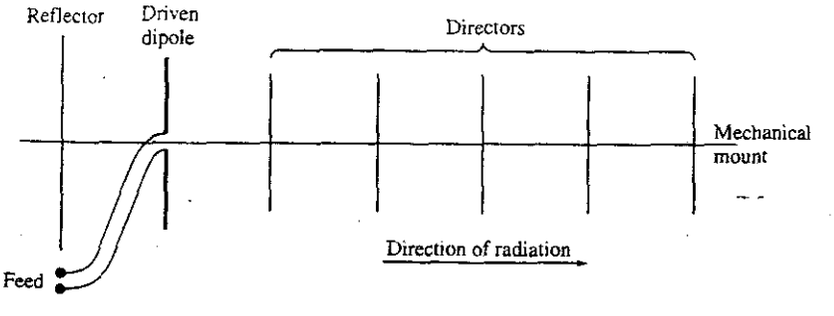


Figure 9.41 Sketch of a Yagi-Uda endfire antenna. (The mechanical mount is insulated from the dipoles.) For radar application they have been used in a linear array configuration; one example is that of the E2C AEW radar shown in Fig. 3.45a.

longitudinal axis is proportional to the square root of the antenna length, as compared to a conventional broadside antenna where the beamwidth is directly proportional to the antenna size.

Rotating Electronically Steered Phased Arrays At times it has been suggested[126,127] that a single face of an electronically steered phased array be mechanically rotated in azimuth but electronically steered in azimuth and elevation. Whether the single rotating phased array is an attractive alternative to a four-faced fixed phased array depends on the particular application and the assumptions that are made. Caution should be exercised, however, when considering this approach since in some cases it is conceivable that a rotating single phased array face might provide the worst attributes of both the phased array and the mechanically rotated antenna rather than the best of both. The mechanically trainable phased array, described next, is different and has some important operational advantages.

Trainable Phased Arrays A trainable phased array is one which is electronically steered in both azimuth and elevation over a wide angular sector, but which is mounted so as to be mechanically positioned to cover a desired sector. (Once in position it remains fixed rather than continuously rotated.) It is a convenient method for using an array for missile-range instrumentation. An example is the transportable *C*-band AN/MPS-39 MOTR (multiple object tracking instrumentation radar) developed by Lockheed-Martin at Moorestown, N. J. (shown in Fig. 4.1b). The 12-ft-diameter space-fed lens array with 8359 elements is mounted on a precision elevation-over-azimuth tracking pedestal so as to achieve coverage of a 60° cone anywhere within the full hemisphere. Its beamwidth is one degree and its measured antenna gain is just under 46 dB. In missile range applications, many targets might have to be tracked simultaneously; including the firing aircraft, the target missile or drone, surface-to-air missiles, air-to-air missiles, and other aircraft that might be on the range (range safety). The MOTR can simultaneously track up to 10 targets with an absolute angle accuracy of 0.2 mils rms and a range accuracy of 1 yd rms. Prior to such radars, a separate air-surveillance radar and more than one mechanical tracking radars had to be used.

The trainable phased array is also of advantage for ship air-defense when the agility of a phased array is needed. Traditionally, when a phased array is used on board a naval ship (for example, the *S*-band Aegis system or the U.S. Navy's original *S*-band AN/SPS-33 phased array radar system), there are four planar array faces mounted around the ship to provide 360 degrees of all-around coverage. Although the designer might want to configure the phased array so that all four apertures can operate simultaneously, the need to reduce the cost of such radars has sometimes resulted in having only one or two transmitters and receivers time-shared among the four array apertures. This might result in acceptable performance, but it is less than what could be obtained if the four faces were able to simultaneously operate all the time.

An alternative method for configuring an air-defense phased array radar system when four complete phased array faces are too costly (or even when they are not), is to employ two trainable arrays instead. An example used by the Russian navy on Cruisers is the trainable array radar shown in Fig. 9.42 which is known by its NATO nomenclature as the Top Dome SAN-6, or by its Russian/Soviet name of RIF. In a Russian Navy Cruiser,

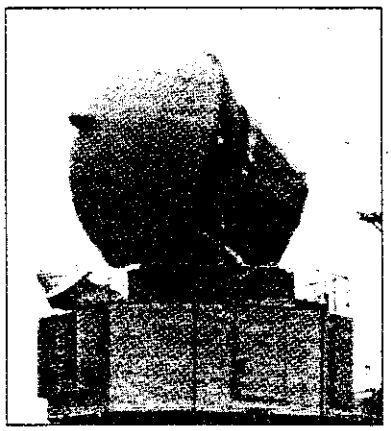


Figure 9.42 Soviet/Russian electronically scanned trainable phased array pulse doppler radar whose Russian name is RIF. Its NATO designation is Top Dome and is part of the SAN-6 shipboard surface-to-air missile system. It is said to employ row-and-column beam steering. The antenna assembly includes a hemispherical radome. In addition to the main phased array there is a wide-angle phased array for acquisition of the surface-to-air missiles just after launch.

one Top Dome radar is located fore and a second one is located aft. (A truism is that if something is important to have on a naval ship—such as a gun, missile system, or radar—there ought to be at least two of them to insure that one is available when needed.) The two trainable array radars can cover attacks simultaneously within any two 90° sectors. An important advantage with this configuration is that one or more attacks from within a single 90° sector can be engaged by both radars simultaneously. A four-face phased array, on the other hand, can bring only one array face to bear to defend against such an attack.

Two trainable phased arrays cannot engage a multiple simultaneous attack from more than two 90° sectors. The likelihood of this being a serious concern is small. The ability to mount a simultaneous attack from three different directions over 360° in azimuth is certainly possible, but having all targets appear simultaneously at the target ship from different directions is very difficult to do. If they are not simultaneous, the ship's air-surveillance radar with 360° of azimuth coverage can be expected to detect and recognize such attacks and the two trainable arrays can be scheduled to engage them without overlap. Also, it is seldom that a major naval ship that carries an expensive phased array radar for air defense operates by itself; so if there is a large multiple attack, one ship does not have to handle the total attack all by itself. Thus employing trainable phased arrays offers advantages not found with the traditional four-faced phased array system.

9.11 RADIATION-PATTERN SYNTHESIS[128]

A radar antenna radiation-pattern is required to have a specified beamwidth along with acceptably low sidelobe radiation. In some cases, the antenna radiation pattern must provide a desired contour, or shape, over a specified angular region. An example is the cosecant-squared elevation pattern of an air-surveillance radar. The aperture illumination for the squinted beams of an amplitude-comparison monopulse tracking radar, on the other hand, must have a suitable sum pattern with low sidelobes, as well as a suitable difference

pattern with low sidelobes and a large slope at beam crossover. This section reviews some of the methods available to the radar antenna designer to achieve the radiation patterns necessary to produce the desired radar performance.

As was mentioned previously in Sec. 9.3, the radiation pattern is determined by the distribution of current across the aperture. We have called the distribution of current the aperture illumination. Equation (9.10) gave the (electric field strength) radiation pattern $E(\phi)$ of a linear one-dimensional antenna in one angle-coordinate ϕ as the inverse Fourier transform of the aperture illumination $A(z)$. Similarly the radiation pattern of a two-dimensional planar antenna is given as a two-dimensional inverse Fourier transform of its aperture illumination. We shall first consider the problem of obtaining a desired main-beam pattern with acceptable low sidelobes and then the problem of obtaining shaped radiation patterns.

Obtaining a desired antenna pattern is slightly different for a continuous aperture (such as a reflector) than a phased array that consists of many individual elements. One can sometimes approximate the continuous (reflector) aperture illumination with a discrete (array) aperture illumination, and vice versa. The discussion of antenna patterns in this section is done chiefly for a linear one-dimensional aperture or for rectangular apertures where the illumination is separable; that is, $A(x,z) = A(x)A(z)$. When the illumination for an array is considered, the array is assumed to have uniformly spaced isotropic elements with element spacing generally taken to be half-wavelength. The radiating elements of a real array are not isotropic but have some element pattern $E_e(\theta)$. If the desired array antenna pattern is $E_d(\theta)$, the pattern to be found is $E_d(\theta)/E_e(\theta)$ when using a technique based on the assumption that the elements are isotropic.

Patterns with a Desired Beamwidth and Low Sidelobes Obtaining a pencil-beam or a fan-beam radiation pattern is not usually a synthesis problem. Instead, the patterns obtained from various aperture illuminations are calculated and a suitable one is selected. Table 9.1 in Sec. 9.3 lists a number of antenna patterns for several types of aperture illuminations that can be expressed in analytical form. These were considered in the past since their analytical form permitted the corresponding radiation patterns to be readily calculated. With modern computers, however, using aperture illuminations just because they are readily integrated is no longer necessary. Thus these aperture illuminations are not now generally used. The table is useful in that it illustrates how the maximum gain, beamwidth, and maximum sidelobe level vary with change in shape of the aperture illumination. The more tapered the aperture illumination (that is, the more rapidly its amplitude falls off as a function of the distance from the center of the aperture) the lower will be the sidelobe level, but the lower will be the antenna gain and the wider will be the beamwidth. In practice, other aperture illuminations, such as Taylor illuminations, are usually used rather than those given in Table 9.1.

Taylor Aperture Illumination For a specified maximum sidelobe level, the antenna pattern which has all of its sidelobes equal produces the narrowest beamwidth (where the beamwidth is measured by the angular distance between the first nulls that define the main beam). This is known as a Dolph-Chebyshev pattern since it was first shown by C. L. Dolph, a mathematician working at the U.S. Naval Research Laboratory during World

War II, who based it on equating the Chebyshev polynomial to the polynomial describing the pattern of an array antenna.[129] In spite of its desirable properties, the Dolph-Chebyshev pattern is seldom used for radar antennas since it is unrealizable with arrays containing other than a small number of elements. As the antenna size increases, the currents required at the ends of the aperture become nonmonotonic and large compared with the currents along the rest of the aperture. More of a restriction is the fact that these large currents are required to occupy a very narrow spatial region at the ends of the aperture, too narrow to be obtained with an actual antenna. This inability to achieve the theoretical aperture illumination sets an upper limit to the size of an antenna that can have a Dolph-Chebyshev pattern and therefore sets a lower limit to the width of the main beam that can be achieved.

A realizable approximation to the Dolph-Chebyshev aperture illumination was devised by T. T. Taylor.[130] The Taylor aperture illumination, as it is known, has been widely used for radar antennas. It produces a pattern with equal-amplitude sidelobes of a specified value, but only in the near vicinity of the main beam. Unlike the equal-sidelobe level of the Dolph-Chebyshev pattern, the sidelobes of the theoretical Taylor pattern are of uniform amplitude only within an angular region ϕ defined by

$$|(D/\lambda) \sin \phi| < \bar{n}$$

where $\bar{n}$ = integer, D = antenna dimension, and λ = wavelength. With a linear aperture, the sidelobes decrease as $1/\sin \theta$ with increasing angle θ outside this region (similar to the fall-off of a pattern with a uniform illumination). Hence $\bar{n}$ divides the sidelobes into a uniform region, which straddles the main beam, and a decreasing sidelobe region. The number of equal sidelobes on each side of the main beam is equal to $\bar{n} - 1$. The integer $\bar{n}$ is usually a small number. (Sometimes it is difficult to observe either on a calculated or an actual antenna pattern a region of equal sidelobes in the vicinity of the main beam, yet they are a part of the theoretical Taylor pattern.) The beamwidth of a Taylor pattern will be broader than that of a Dolph-Chebyshev. If the Taylor design sidelobe level is -25 dB, a value of $\bar{n} = 5$ gives a beamwidth almost 8 percent greater than the theoretical Dolph-Chebyshev. With $\bar{n} = 8$, the beamwidth is 5.5 percent greater.

The Taylor pattern is specified by (1) the peak design sidelobe level and (2) the value of $\bar{n}$. The integer $\bar{n}$ can take on only a small range of values for a given design sidelobe level. Taylor states that $\bar{n}$ must be at least 3 for a design sidelobe of -25 dB and at least 6 for a design sidelobe of -40 dB. The larger the value of $\bar{n}$ the sharper will be the beam. On the other hand, it cannot be too large since the same realizability difficulties will arise as with the Dolph-Chebyshev. A suitable criterion for obtaining a realizable Taylor pattern is to choose an illumination that decreases monotonically from the center out to the ends of the aperture and has a zero derivative at the ends of the aperture. The illumination need not be zero amplitude at the ends but can have a finite value (a pedestal). Taylor showed that a distribution with a pedestal, or nonzero value at the edges, is more effective in producing low sidelobes.

A rough guide to the selection of the parameter $\bar{n}$ has been given by Hansen.[131] For example, when the aperture illumination is monotonic, he states that the value of $\bar{n}$ equals 5 for a peak sidelobe of -25 dB, 9 for a peak sidelobe of -35 dB, and 11 for a peak sidelobe of -40 dB. The aperture efficiencies for these three examples are,

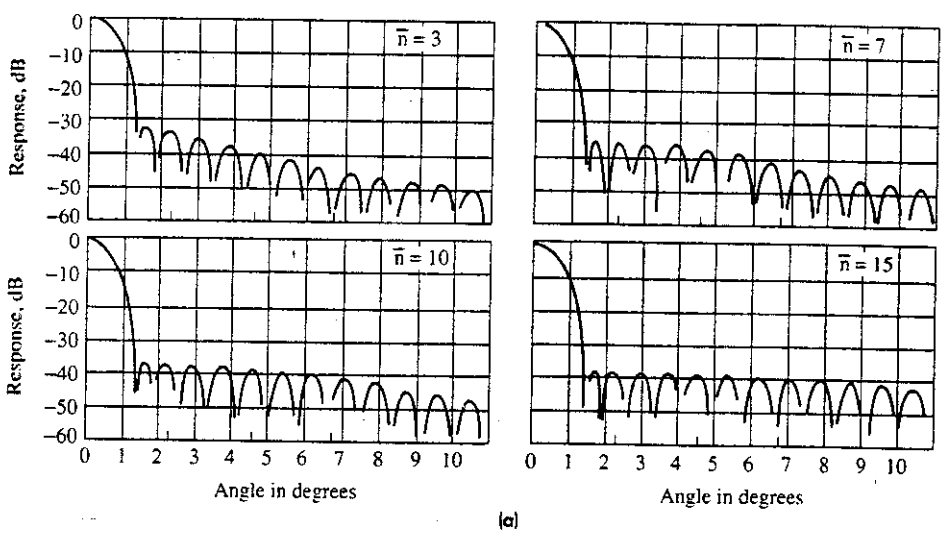
respectively, 0.91, 0.82, and 0.77. (Hansen later added that the maximum value of $\bar{n}$ for a monotonic illumination is 17 for a peak sidelobe of -50 dB and 23 for a peak sidelobe of -60 dB.[132]) Thus care must be exercised in the selection of the sidelobe level and the value of $\bar{n}$ for a Taylor pattern. Although the Taylor pattern was developed as a realizable approximation to the Dolph-Chebyshev, in practice it seldom resembles the theoretical equal-sidelobe pattern.

Taylor aperture illuminations can also be obtained for a circular aperture.[133] Figure 9.43 illustrates the nature of the Taylor circular aperture illumination and their corresponding radiation patterns. (This is from an Institute for Defense Analyses report[134] that was not widely circulated, but some of the information appears in the *IEEE AP-S Transactions.*[135]) Shown in Fig. 9.43a are four circular-aperture antenna patterns with the same half-power beamwidth, $70\lambda/D$, but with different values of the Taylor $\bar{n}$ parameter varying from 3 to 15. As $\bar{n}$ increases, the peak sidelobe decreases. From these patterns it would appear that one would want to select a large value of $\bar{n}$. There is a problem with this, however, as can be seen from their corresponding aperture illuminations in Fig. 9.43b. One might be able to realize the required aperture illumination for $\bar{n} = 3$, and might even be able to roughly approximate the illumination for $\bar{n} = 7$. But at the higher values of $\bar{n} = 10$ and 15, it is not likely that one can achieve the necessary aperture illuminations at the edge of the aperture.

There have been several other methods for selecting the aperture illumination for a conventional antenna pattern, as can be found in many of the texts[74,136] on antennas; but the Taylor seems to have been the most popular for radar applications.

Bayliss Illumination[137,138] Difference patterns are used in monopulse tracking radars along with the sum pattern, as was mentioned in Chap. 4. It has been said before in this text that the sum pattern which produces maximum gain is of the form $(\sin u)/u$, and is obtained with a uniform aperture illumination. The symbol $u = \pi(D/\lambda)\sin\theta$, where D = aperture dimension and λ = wavelength. If one forms a difference pattern by starting with a uniform aperture illumination and subtracting one half of the aperture from the other half (that is, $A(x) = -1$ for $-D/2 < x < 0$, and $+1$ for $0 < x < +D/2$), then the difference pattern is proportional to $(1 - \cos u)/u$. The peak sidelobe is 10.6 dB below the peak of the beam, which is relatively high. The optimum illumination for a difference pattern (one that produces maximum slope, or minimum error) is linear-odd over the aperture. That is, the aperture illumination is a straight line that passes through zero at the center of the aperture and has a maximum (say, for example, $+1$) at one edge and a minimum (-1) at the other edge. Its peak sidelobe of -8.3 dB is even worse than that of the uniform-illumination difference pattern. When a single aperture illumination is used to obtain both the sum and difference patterns, an aperture illumination has to be found which represents a suitable compromise for the gain of the sum pattern, the peak sidelobes of both patterns, and the slope of the difference pattern. A better approach, when permitted, is to use an antenna which can support independent sum and difference patterns.

When the difference pattern of a monopulse antenna can be selected independently of the sum pattern, as in a phased array, the criterion is to obtain the maximum angle accuracy commensurate with a desired sidelobe level. The Bayliss illumination has been popular for this purpose. It is based on the same principles as the Taylor illuminations. As with the Taylor, the Bayliss illumination[137] depends on the peak sidelobe level and the



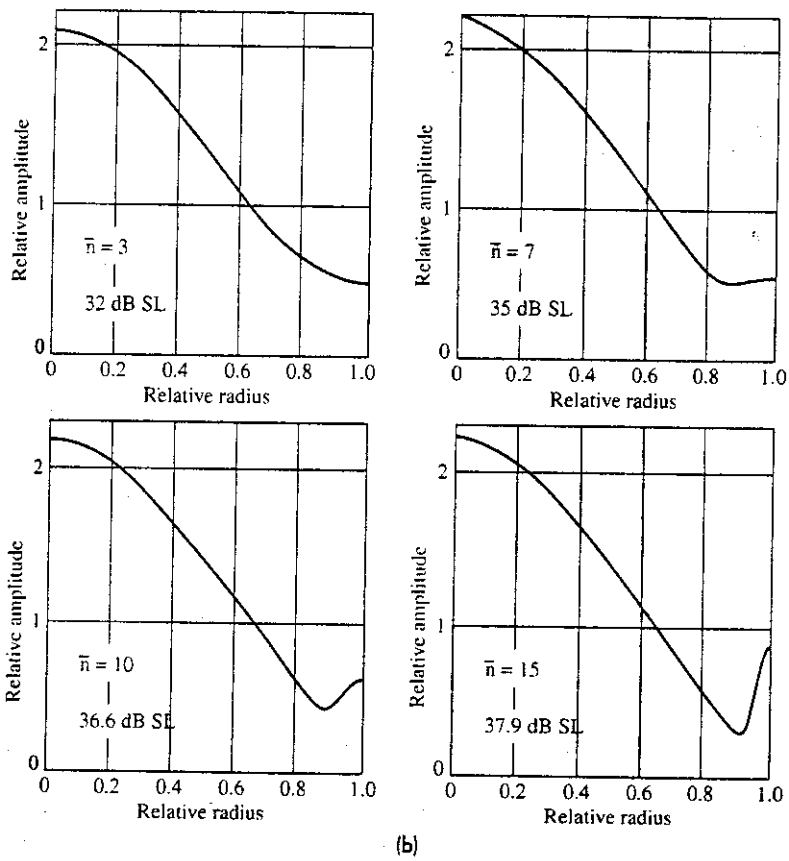


Figure 9.43
(a) Taylor circular aperture radiation patterns each having a beamwidth in degrees of $70\lambda/D$, but with different values of $\bar{n}$. (b) The corresponding aperture illuminations for the patterns of (a).
(From W. White[134] Courtesy Institute for Defense Analyses.)

number $(\bar{n} - 1)$ of equal sidelobes. Similar restrictions apply on the selection of these two parameters as with the Taylor illuminations.

The Bayliss difference pattern has also been applied to circular apertures, and can be found in any one of several antenna books, including those authored by Mailloux,[74] Hansen,[69] and by Elliott.[136]

Shaped Antenna Patterns In the above, we have considered aperture illuminations for obtaining suitable radiation patterns when a pencil beam or a simple fan beam was required. Sometimes it is necessary to form shaped antenna patterns where the patterns are wide compared to the minimum beamwidth (approximately λ/D radians) that can be obtained with an aperture of dimension D. An important example is the cosecant-squared beam discussed in Sec. 2.11. Other examples are a "square-top" pattern and an elevation pattern with a sharp cutoff at the horizon to minimize surface reflections.[139] In this subsection, we briefly describe two methods for synthesizing such patterns.[128] The methods for finding the aperture illumination to achieve some desired pattern are similar to the methods for finding the filter frequency response function to produce a desired time waveform.

Fourier Synthesis Since the antenna pattern is, in theory, given by the inverse Fourier transform of the aperture illumination [as in Eq. (9.10)], the aperture illumination required to achieve a desired antenna pattern can be found by taking the Fourier transform of the desired pattern, which is

$$A(z) = \frac{1}{\lambda} \int_{-\infty}^{+\infty} E(\phi) \exp\left(-j2\pi \frac{z}{\lambda} \sin \phi\right) d(\sin \phi) \qquad [9.47]$$

where the symbols are the same as described for Eq. (9.10). The limits of the integration actually are finite since $|\sin \phi| \leq 1$. The Fourier transform is such that when the angular region is finite the Fourier transform of $A(z)$ is infinite. Since we have only a finite aperture, the actual radiation pattern will only be approximate, and can be shown to be[140]

$$E_a(\phi) = \frac{D}{\lambda} \int_{-\infty}^{+\infty} E(\xi) \frac{\sin\left[\pi(D/\lambda)(\sin \phi - \sin \xi)\right]}{\pi(D/\lambda)(\sin \phi - \sin \xi)} d(\sin \xi) \qquad [9.48]$$

where $E_a(\phi)$ is the Fourier-integral pattern which approximates the desired pattern $E(\phi)$ when $A(z)$ is restricted to a finite aperture of dimension D. The angle ξ is the variable of integration. The approximation to the antenna pattern obtained on the basis of the Fourier integral for continuous apertures (or the Fourier-series method for discrete array antennas) has the property that the mean-square deviation between the desired pattern $E(\phi)$ and the approximate pattern $E_a(\phi)$ is a minimum. The larger the aperture, the better will be the approximation.

The Fourier series may be used to approximate the pattern of a discrete array, just as the Fourier integral may be used to approximate the pattern of a continuous aperture. The Fourier series method is restricted in practice to arrays with element spacing in the vicinity of a half-wavelength. Spacings larger than a wavelength produce undesired grating lobes. Spacings much smaller than a half-wavelength result in so-called "supergain" radiation patterns (beamwidths much smaller than λ/D radians) that are not realizable since they are a consequence of an overly simplified model of radiation.

Woodward-Levinson Method This is the spatial-domain analogy to the well-known sampling theorem for temporal signals. The classical sampling theorem for time waveforms states: A band-limited signal $s(t)$ with no frequency components greater than B Hz is determined by its amplitude at a series of points spaced $1/2B$ apart in time. The analogous sampling process applied to an antenna is that the radiation pattern $E(\phi)$ from an antenna with a finite aperture D is determined by a series of amplitudes spaced in angle λ/D apart. Figure 9.44a shows a pattern $E(\phi)$ and the sampled points λ/D radians apart. The sampled values $E_s(n\lambda/D)$, which determine the antenna pattern are shown in *b*. An antenna pattern $E_a(\phi)$ can be constructed from the sample values $E_s(n\lambda/D)$ using a pattern of the form $(\sin u)/u$ about each of the sample values, where $u = \pi(D/\lambda)\sin\phi$. The $(\sin u)/u$ pattern is called the composing function. The antenna pattern is given by

$$E_a(\phi) = \sum_{n=-(N-1)/2}^{(N-1)/2} E_s(n\lambda/D)\frac{\sin[\pi(D/\lambda)(\sin\phi - n\lambda/D)]}{\pi(D/\lambda)(\sin\phi - n\lambda/D)} \qquad [9.49]$$

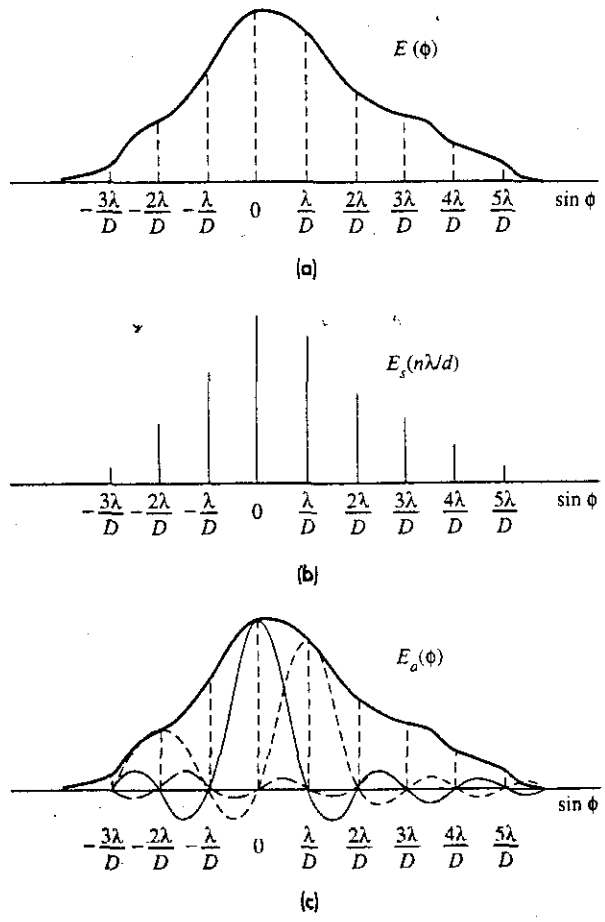


Figure 9.44 (a) Radiation pattern $E(\phi)$ with sampled values λ/D radians apart in angle, where λ = wavelength and D = aperture dimension; (b) sampled values $E_s(n\lambda/D)$, which specify the antenna pattern of (a); (c) reconstructed pattern $E_a(\phi)$ using $(\sin u)/u$ composing function to approximate the desired radiation pattern $E(\phi)$.

where N is the total number of samples, assumed to be odd. Thus the antenna pattern is constructed from a sum of individual $(\sin u)/u$ patterns spaced λ/D radians apart, each weighted in amplitude according to the sample values $E_a(n\lambda/D)$, as illustrated in Fig. 9.44c. The aperture illumination corresponding to the radiation pattern of Eq. (9.49) is

$$A(z) = \frac{1}{D} \sum_{n=-(N-1)/2}^{(N-2)/2} E_a(n\lambda/D) \exp(-j2\pi nz/D) \qquad [9.50]$$

The difference between the Woodward-Levinson method and the Fourier-integral synthesis method is that the former gives an antenna pattern which exactly fits the desired pattern at a finite number of points and the latter gives a radiation pattern whose mean-square deviation from the desired pattern is a minimum.

Cosecant-Squared Antenna Pattern This is an antenna with its beam shape proportional to the cosecant-squared of the angle in one plane (usually in elevation) and with a conventional narrow beam in the orthogonal plane. The original reason for using an antenna pattern whose elevation pattern was proportional to $\csc^2 \theta_e$ was to obtain an echo signal that did not vary with range so long as the target flew at a constant altitude and the earth could be assumed to be flat (Sec. 2.11). Many air-surveillance radars have been designed with such an elevation pattern. An even more important reason for using a cosecant-squared elevation pattern is that it allows the radar to provide coverage of targets at high altitudes and shorter ranges much more efficiently than would a conventional fan-beam antenna pattern. In most modern air-surveillance radars, the coverage is modified even further to allow for the curvature of the earth and to provide more radiation at high elevation angles and short range than would be obtained with a cosecant-squared pattern.[141] This additional coverage at high angles is necessary to compensate for the reduction in gain at short ranges caused by the use of sensitivity time control (STC). With STC the receiver gain is reduced at short ranges to attenuate the large clutter echoes that might appear. The cosecant-squared pattern, or a slight variation, has also been important for airborne surface-surveillance radars that must provide relatively uniform coverage of the surface.

The methods available for designing shaped beams, such as the Woodward-Levinson or the Fourier integral, can be applied to obtain the aperture illumination required for a cosecant-squared pattern or its variations. The cosecant-squared pattern may be approximated with a reflector antenna by modifying the shape of the reflecting surface. For example, the upper half of a parabolic reflector can be a parabola that reflects energy from the feed at its focus in a direction parallel to the axis of the parabola axis—as in any other parabolic reflector. The lower half, however, is distorted (tilted slightly forward) to direct a portion of the radiated energy in the upward direction. This is a very simple method, which can be adequate for some applications.

A cosecant-squared antenna pattern may also be obtained by (1) feeding a parabolic surface with two or more horns, (2) a linear array, (3) a parabolic cylinder fed by a linear array or a line source, or (4) a point-source feeding a reflector surface with double curvature. An example is the antenna for the U.S. Air Force's AN/TPN-19 S-band Airport Surveillance Radar which employed an offset paraboloid reflector fed from 12 feedhorns in a vertical line-feed with the uppermost feedhorn located at the focal point of the paraboloid.

Loss in Gain The gain of a cosecant-squared antenna will be less than the gain of a conventional antenna from which it was derived. A very approximate estimate of the reduction in gain for such an antenna is given as

$$\frac{G_{csc}}{G} = \frac{\theta_0}{\theta_0 + \sin^2 \theta_0 (\cot \theta_0 - \cot \theta_m)} \approx \frac{1}{2 - \theta_0 \cot \theta_m} \quad [9.51]$$

where G = the gain of a rectangular antenna pattern of width θ_0 that radiates uniformly from $\theta = 0$ to $\theta = \theta_0$, and G_{csc} = gain of the cosecant-squared antenna. The radiation decreases proportional to $\csc^2 \theta$ from θ_0 to θ_m, the maximum angle at which the cosecant-squared pattern is applied. (There is no radiation beyond θ_m.) The approximate expression on the right of Eq. (9.51) applies for small θ_0 and large θ_m. (In addition, the assumption that the beamwidth θ_0 is that of a rectangular-shaped beam can also affect the accuracy of this expression.) For example, if $\theta_0 = 6°$ and $\theta_m = 20°$, the gain of the cosecant-squared antenna is decreased by 2.2 dB compared with a rectangular beam 6° in width. A modified cosecant-squared pattern to allow for the coverage of the earth and to account for the action of the STC will have even lower gain.

Theoretical and Actual Antenna Patterns The discussion of theoretical antenna patterns in this section was based on the Fourier transform relationship between the antenna pattern and the aperture illumination. It has been widely accepted and widely used, but it has limitations. The Fourier transform relationship applies fairly well to the main beam of the antenna and to the region near the main beam. It is less accurate the farther one goes in angle from broadside, and it does not faithfully account for energy radiated near or beyond ±90°.

In a reflector antenna the spillover radiation from the feed, blockage by the feed or feed supports, the radiation diffracted by the reflector, and any leakage through a mesh reflector surface are not accounted for by the Fourier transform. The far-field pattern of any antenna also is affected by nearby structures or other obstructions that can block or diffract the energy radiated by the antenna. The blockage of an antenna by a mast on a ship, for example, might result in a −40 dB peak sidelobe being increased to as high as −15 to −20 dB or greater. Errors in the phase and amplitude across an antenna aperture due to either mechanical or electrical inaccuracies will increase the far-out sidelobe levels over what is predicted from the classical Fourier transform relationship. This is a subject that will be described next.

9.12 EFFECT OF ERRORS ON RADIATION PATTERNS

Experimentally measured patterns of actual antennas often deviate from the theoretically calculated pattern, especially in the region of the far-out sidelobes. Generally, the fault does not lie with the theory, but in the fact that it is not possible to reproduce precisely the necessary aperture illumination specified by synthesis theory. There are small, but ever-present, errors in the fabrication of the antenna and in how it is fed. These contribute unavoidable perturbations in the aperture illumination and result in a pattern different in detail from the one expected.

Errors in the aperture illumination may be either *systematic* or *random*. The former are predictable (usually), but the latter are not and can only be described in statistical terms. Examples of systematic errors include (1) aperture blocking by the feed and its supports in a reflector antenna, (2) mutual coupling among the elements of an array, (3) quantization lobes due to the discrete value of the phase in a digital phase shifter, (4) spurious lobes due to mismatch in a constrained feed for an array, and (5) periodicities introduced in an antenna during the manufacturing process. Random errors include (1) errors in the machining or manufacture of the antenna due to the finite precision of construction techniques, (2) the precision with which a phase shifter can be set to its required phase, (3) errors incurred in adjusting an array, (4) random distortions of the antenna surface, and (5) mechanical and electrical (phase) variations caused by temperature gradients or wind (and in some cases gravity) across the antenna. Although random errors throughout the antenna may be relatively small, their effect on the sidelobes (which are also small) may be relatively large compared to the small levels of the sidelobes. Systematic errors do not differ much from one antenna to another in any particular design constructed by similar methods. Random errors, on the other hand, can differ from one antenna to the next even though they may be of the same design and are constructed similarly. The effect of random errors, therefore, must be discussed in terms of statistical averages found over many similar antennas.

When no specific guidance is available, the antenna designer often assumes that the antenna should radiate a wavefront that differs in phase from the desired wavefront by no more than $\pm\lambda/16$, where λ = wavelength. Because of the two-way propagation from a reflector surface, the mechanical accuracy of a reflector antenna surface must be held within $\pm\lambda/32$. As we shall see, it is possible to obtain more precise criteria for antenna errors, especially when low or ultralow sidelobes are desired. Most of the discussion of errors in this section will concern random errors rather than systematic errors. (If the systematic errors are known, their effect on the antenna pattern can be ascertained by taking the Fourier transform of the actual aperture illumination, including the effects of systematic errors.) The discussion of the effects of random errors in reflectors is separated from the effects of errors in phased arrays.

Random Errors in Reflector Antennas The classical work on the effects of random errors on antenna patterns was due to the pioneering efforts of John Ruze.[142] For small phase errors he showed that the gain G of a circular aperture with mean-square phase-error $\overline{\delta^2}$ is approximately

$$G = G_0 \exp\left[-\overline{\delta^2}\right] = \rho_a(\pi D/\lambda)^2 \exp\left[-(4\pi\epsilon/\lambda)^2\right] \qquad [9.52]$$

where G_0 is the gain of the antenna without errors; the phase error δ, in radians, is with respect to the mean phase plane; ρ_a is the aperture efficiency; D is the diameter of the circular antenna; and ϵ is the rms error of the reflector surface in the same units as the wavelength λ. In the above, the expression $4\pi A_e/\lambda^2$ (discussed early in this chapter) was substituted for the gain G_0. For a given reflector size D, the gain increases as the square of the frequency when the errors are small, until the exponential term becomes significant. Differentiating Eq. (9.52), setting it equal to zero, and solving for wavelength gives the wavelength at which the maximum gain is obtained for an rms error ϵ, which is

$$\lambda_m = 4\pi\epsilon \qquad [9.53]$$

At this wavelength, the gain will be 4.3 dB below what it would have been in the absence of errors. The maximum gain of an antenna due to phase errors is then

$$G_{\max} = \frac{\rho_a}{43}\left(\frac{D}{\epsilon}\right)^2 \qquad [9.54]$$

For wavelengths shorter than λ_m, the gain drops off rapidly with decreasing wavelength (increasing frequency).

The gain of a reflector antenna is limited by the mechanical tolerance to which its surface can be constructed and maintained when in operation. The most precise antennas, under benign, controlled conditions, seem to be limited in practice to a precision of about one part in 20,000. From Eq. (9.54) the diameter of such an antenna is about 1600 wavelengths for maximum gain. Its beamwidth would be 0.04° and it would have a gain of about 68 dB. Special purpose nonradar antennas have been constructed with slightly better tolerances, but these generally have some means for measuring the antenna surface and correcting the surface automatically with feedback control while the antenna is operating. Such antennas are operated in controlled environments that may not be suitable for operational radar applications. In practice, therefore, radar antennas are seldom larger in dimension than approximately 300 wavelengths, which corresponds to a beamwidth of about 0.2°.

The construction tolerance of a reflector antenna is often described by its "peak" error, rather than its rms error. The ratio of the peak to the rms error is found in practice to be about 3:1. This truncation of errors occurs since large errors are usually corrected during manufacture.

With small phase errors the exponential factor in the gain expression of Eq. (9.52) can be approximated by

$$G \approx G_0\left(1 - \overline{\delta^2}\right) \qquad [9.55]$$

If the loss in antenna gain is to be less than 1 dB, this simple expression says that the rms phase variation about the mean phase surface should be less than 0.45 radian, or 26°. This is equivalent to an rms distance error of $\lambda/14$. For shallow reflector antennas, however, the two-way propagation path requires that the rms deviation of the surface from its true value be one half this value, or $\lambda/28$, for a 1 dB reduction of gain.

Ruze showed that under certain conditions the radiation pattern of a reflector antenna which is distorted by a large number of gaussian-shaped "bumps" can be expressed as

$$G(\theta,\phi) = G_0(\theta,\phi)e^{-\overline{\delta^2}} + (2\pi C/\lambda)^2 e^{-\overline{\delta^2}} \sum_{n=1}^{\infty} \frac{\left(\overline{\delta^2}\right)^n}{n!n} e^{-(\pi C u/\lambda)^2/n} \qquad [9.56a]$$

where C is the correlation distance of the error (the size of the region on the aperture where the errors cannot be considered independent) and u in this case is sin θ. The coordinate system for this equation is the classical coordinates of antenna theory, as was shown in Fig. 9.4, with the antenna lying in the x,y plane. The error current in one region of the antenna (the correlation distance) is assumed independent of error currents in other regions. The size of the correlation distance affects both the magnitude and the direction of the spurious radiation that results from the presence of errors. For small error, when only the first term of the series ($n = 1$) need be considered, Eq. (9.56a) becomes

$$G(\theta,\phi) = G_0(\theta,\phi)e^{-\overline{\delta^2}} + (2\pi C/\lambda)^2\overline{\delta^2}e^{-(\pi Cw/\lambda)^2} \qquad [9.56b]$$

The first term of the above equation [as well as Eq. (9.56a)] represents the no-error pattern reduced by a factor dependent on the mean-square phase error. The second term represents the average value of the sidelobes that are generated by the phase errors (not the average of the peaks, but the *average*). Near the main beam the sidelobes are determined mainly by the inverse Fourier transform of the aperture illumination [Eq. (9.10)], but eventually these drop below the error sidelobes and at angles far from the main beam the errors determine the sidelobe level. When the error sidelobes are dominant, the average sidelobe level is independent of angle.

Other observations about errors in reflector antennas are:

1. Ruze's original analysis[143] showed that the error sidelobes are proportional to the mean-square error and to the square of the correlation distance measured in wavelengths.
2. If errors are unavoidable, they should be kept small in extent; that is, for the same mechanical tolerance, the antenna with the smaller correlation distance will give lower sidelobes than an antenna with a larger correlation distance. An error stretching most of the length of the antenna is likely to have a worse effect than a localized bump or dent of much greater amplitude. Thus small disturbances such as screws and rivets on the reflector surface will have relatively little effect on the antenna radiation pattern.
3. An increase in frequency increases both the phase errors and the correlation distance (measured in wavelengths). Therefore the gain of a constant-area antenna does not increase as rapidly as the square of the frequency when errors are a factor.

Since the radiation pattern in the far-out sidelobe region is more likely to depend on the accuracy with which the antenna is constructed rather than the particular aperture illumination selected, the mechanical engineer, the skilled machinist, and technician are very important in realizing in practice a satisfactory antenna pattern.

Errors in Arrays In the above analysis of reflector-antenna errors, only the effect of the phase errors were considered. In an array antenna, however, other factors may enter to cause degradation of the radiation pattern. These include errors in the amplitude and phase of the current at each element of the array, missing or inoperative elements, rotation or translation of the element from its correct position, errors in the phase provided by the phase shifter, effects of a quantized phase shift, and variations in the individual element patterns because of mutual coupling. These errors can result in a decrease in antenna gain, increase in sidelobe level, generation of spurious sidelobes, and a shift in the location of the main beam.

It is not possible to predict the pattern of an antenna unless the actual errors experienced by that particular antenna are known. The average, or expected, value of a radiation pattern of an *ensemble* of antennas of the same type can be computed based on the rms values of the random errors. The statistical description of the radiation pattern cannot be applied to any particular antenna of the ensemble, but applies to the entire collection of similar antennas whose errors are described by the same statistics. Usually the average pattern is computed, but other statistical descriptions can be obtained if desired.

Average Radiation Pattern Due to Errors The ensemble-average radiation intensity pattern of a uniform array of M by N isotropic elements arranged on a rectangular grid with equal spacing between elements is given by[144]

$$\overline{|f(\theta,\phi)|^2} = P_e^2 e^{-\overline{\delta^2}} |f_0(\theta,\phi)|^2 + \left[\left(1 + \overline{\Delta^2}\right)P_e - P_e^2 e^{-\overline{\delta^2}}\right] \sum_{m=1}^{M} \sum_{n=1}^{N} i_{mn}^2 \qquad [9.57]$$

where

P_e = probability of an element being operative (or the fraction of elements that remain operating)

δ = phase error, radians (described by a gaussian probability density function)

$|f_0(\theta,\phi)|^2$ = no-error pattern

Δ = relative amplitude error (as a fraction of i_{mn})

i_{mn} = no-error current at the mnth element

Similar to the error pattern of a reflector antenna discussed earlier [Eq. (9.56)], the first term is the no-error pattern reduced by a factor which depends on the phase errors and the fraction of operative elements. The second term represents a statistical average sidelobe level due to the phase and amplitude errors and the fraction of elements that are operating. It also depends on the aperture illumination as given by the currents i_{mn}. This second term is independent of angle, and can be thought of as a statistical omnidirectional pattern which we shall call the error sidelobes. This second term causes the far-out sidelobes of the radiation pattern to be higher in the presence of errors as compared to the no-error pattern; but the shape of the main beam and the near-in sidelobes are not significantly affected by these errors, other then by the exponential term which is usually small. [Sometimes the rms error in amplitude is expressed in dB. When in dB it is not the mean square value which is in dB, but the value of $[1 - (\overline{\Delta^2})^{1/2}]^2$. For example, an rms amplitude error of 0.1 is equivalent to an error of 0.9 dB.]

If $P_e = 1$ (no missing elements) and if the errors are small, the normalized radiation intensity obtained by dividing Eq. (9.57) by the maximum radiation intensity at the center of the main beam, $|f_0(0,0)|^2$, is

$$\overline{|f(\theta,\phi)|^2} \approx |f_{0n}(\theta,\phi)|^2 + \left(\overline{\Delta^2} + \overline{\delta^2}\right) \frac{\sum_m \sum_n i_{mn}^2}{\left(\sum_m \sum_n i_{mn}\right)^2} \qquad [9.58]$$

The second term indicates that the larger the number of elements, the smaller will be the error-sidelobe level. The main-beam intensity, being coherent, increases as the square of the number of elements, while the error sidelobes, being noncoherent, increase only directly with the number of elements. The gain of a broadside array of isotropic elements is

$$G_0 = \frac{\left(\sum_m \sum_n i_{mn}\right)^2}{\sum_m \sum_n i_{mn}^2} \qquad [9.59]$$

When i_{mn} = constant, $G_0 = MN$, which states that the gain of an array of isotropic elements with uniform illumination is equal to the total number of elements. The normalized pattern of Eq. (9.58) then can be expressed as

$$\overline{|f_n(\theta,\phi)|^2} \approx |f_{0n}(\theta,\phi)|^2 + \frac{\overline{\Delta^2} + \overline{\delta^2}}{G_0} \qquad [9.60]$$

The average error-sidelobes is given by the second term. The greater the antenna gain, the less will be the effect of errors on the sidelobes. [Sometimes the denominator of the second term of the above equation is given as πG_0, where π is the gain of the so-called "ideal" element factor.]

The component parts of the normalized ensemble-average pattern as given by Eq. (9.60) are sketched in Fig. 9.45. The ordinate is in dB. The horizontal line shown G_0 (in dB) below the peak of the main beam is the radiation that would be produced by an isotropic antenna with the same power output as the directive antenna. The horizontal line shown $G_0/(\overline{\Delta^2} + \overline{\delta^2})$ (in dB) below the peak is the average value of the error sidelobes, which is independent of angle. The error sidelobes have little effect near the main beam, but they are the dominant factor affecting the far-out portion of the radiation pattern. The measured sidelobes of any actual array antenna, of course, would not be constant as shown in this figure. Instead they would have the usual shape expected of sidelobe radiation, but their ensemble-average value would be constant with angle.

Gain Reduction By substituting the radiation intensity of Eq. (9.57) into the definition of gain (or directivity) of Eq. (9.3), it can be shown that

$$G/G_0 = \frac{P_e}{(1 + \overline{\Delta^2}) \exp(\overline{\delta^2})} \approx \frac{P_e}{1 + \overline{\Delta^2} + \overline{\delta^2}} \qquad [9.61]$$

This states that the relative reduction in gain is independent of the number of elements and depends only on the fraction of elements that are operative and the mean-square value

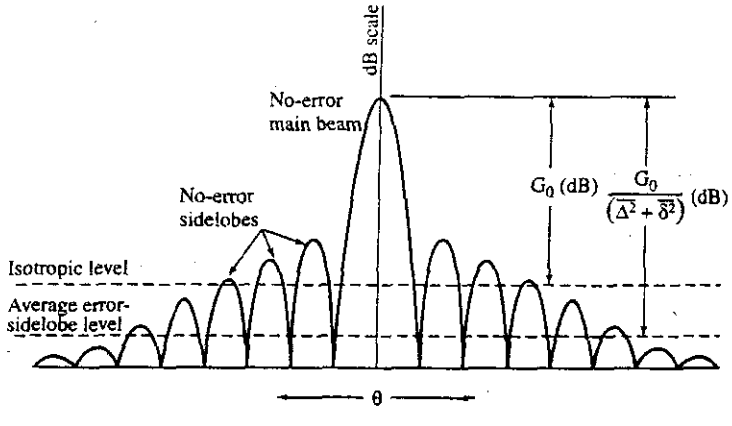


Figure 9.45 Qualitative sketch of the no-error radiation pattern and the average error-sidelobe level, as indicated by Eq. (9.60).

of the relative-amplitude and phase errors. When $P_e = 1$ and $\Delta = 0$, this expression, which applies for small phase errors, is similar to that of Eq. (9.55) for the reflector antenna.

Pointing Error Random phase and amplitude errors in the aperture illumination can give rise to an error in the pointing of the main beam.[145,146] If the aperture illumination is uniform across an M by M square array, the statistical rms beam pointing error in radians is

$$\delta\theta_0 = \frac{\sqrt{3}\sigma}{2\pi(d/\lambda)M^2} \qquad [9.62]$$

where σ = rms value of the normalized error current assuming Rayleigh distributed errors, d = element spacing, λ = wavelength, and M = number of elements along one dimension of a square array. According to this expression, the effect of errors on the beam-pointing accuracy generally is small.

Error-Sidelobe Statistics In the above we have considered the average value of the error sidelobes. The radar system engineer, however, is often more concerned about the peak sidelobe level rather than the average. The actual peak value cannot be predicted, but it can be described on a statistical basis. Usually one wants to determine (or specify) the probability that a sidelobe will not exceed a particular desired value. The approach outlined here is taken from James K. Hsiao.[147,148] He described three methods for finding the effect of errors on the peak sidelobe. These assume that the statistics of the radiation pattern with errors can be described by the Rice probability density function, something we have used in Sec. 2.5 when discussing the statistical detection of a signal in noise. Although all three methods are similar, we shall consider here only the second that he describes in his paper.

From the Rice cumulative probability of sidelobe level, Hsiao obtains a set of curves, shown in Fig. 9.46, for various values of the cumulative probability P that range from 0.90 to 0.98. This figure relates the peak sidelobe level (abscissa) and the amplitude and phase errors of the array (ordinate). The abscissa is actually the ratio of the design sidelobe level to the desired sidelobe level in dB. (The design sidelobe level is always less

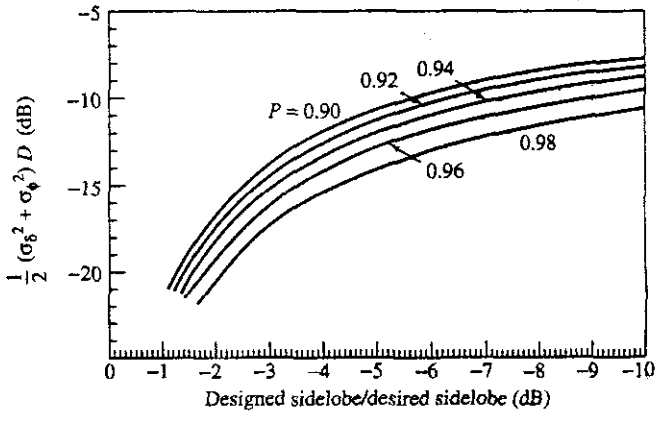


Figure 9.46 Curves for determining the rms amplitude error σ_δ and the rms phase error σ_ϕ as a function of the ratio of the *design* sidelobe level to the *desired* sidelobe level (abscissa) for various values of the cumulative probability P that the sidelobes will be less than the design sidelobe level.

The parameter $D = \sum i_{mn}^2/$(desired sidelobe level), where i_{mn} is the amplitude of the current at the mnth element.

(Due to James Hsiao, taken from refs. 147 and 148.)

than the desired sidelobe level.) The ordinate in Fig. 9.46 includes a factor D, as well as the phase and amplitude rms errors. Hsiao defines D as

$$D = \sum_n \sum_m i_{nm}^2/(\text{desired sidelobe level}) \qquad [9.63a]$$

where i_{nm} is the amplitude of the current at the nmth element, as determined by the desired aperture illumination. It is assumed that the illumination function i_{nm} is normalized such that $\sum_n \sum_m i_{nm} = 1$. Following Hsiao, we consider the example of a 100-element linear array designed with a Chebyshev aperture illumination to provide equal sidelobes -40 dB below the maximum value of the main beam. (As said before in this chapter, the Chebyshev pattern with such low sidelobes is likely to be unrealizable. It is used here to illustrate the procedure.) The design sidelobe level is then -40 dB. In this example we wish to keep the peak sidelobe of the array to a value no greater than -37 dB (the desired sidelobe) with a probability of 0.90. The ratio of the design sidelobe level to the desired sidelobe level (the abscissa of Fig. 9.46) is -3 dB. With an abscissa of -3 dB and a probability $P = 0.90$, we find the ordinate of Fig. 9.46 to be -14 dB. The aperture illumination for the 40 dB sidelobe Chebyshev pattern results in $\sum_n i_n^2 = -19$ dB. The value of D is then $-19 + 37$ dB $= 18$ dB and the sum of the mean square amplitude and phase errors is -29 dB. If the phase and relative amplitude errors are made equal, then the rms phase tolerance is 1.44° and the rms (relative) amplitude tolerance is 0.025, or 0.22 dB. Such tolerances are quite demanding and not easy to achieve. A plot is shown in Fig. 9.47 of the no-error -40 dB Chebyshev pattern along with a pattern where there are errors that give sidelobes with a cumulative probability of 90 percent that they will not exceed -37 dB.

If the design sidelobe level were -45 dB instead of -40 dB, and the desired sidelobe level were still -37 dB, one would find that the tolerance would be 2.6° for phase and 0.045 for amplitude. These are also quite demanding, but not as much as with the higher value of design sidelobe.

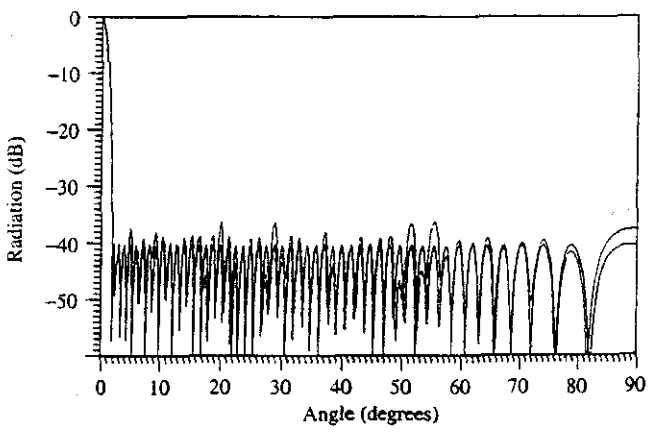


Figure 9.47 The design no-error Chebyshev pattern for a 100-element linear array with a -40 dB design sidelobe is shown along with the resulting desired pattern due to errors when $P = 0.9$, desired sidelobe level $= -37$ dB, rms amplitude error $\sigma_\delta = 0.025$, and rms phase error $\sigma_\phi = 1.44°$.
[Due to James Hsiao, taken from refs. 147 and 148.]

These are relatively tight tolerances. They result from not only requiring a low sidelobe level, but also because there are only a relatively few antenna elements (100 in this example of a linear array). The tolerances for a large planar array would be easier to achieve. As an illustration, Hsiao shows that a planar array with a directive gain of 40 dB, desired sidelobe of −40 dB, design sidelobe of −43 dB, and a probability of 0.9 that the sidelobes will not exceed the desired value results in an rms phase tolerance of 6.6° and an rms amplitude tolerance of 0.12. If the design sidelobe were reduced to −48 dB, the rms phase tolerance becomes 12.3° and the rms amplitude becomes 0.21.

When using a lower design sidelobe to ease the error tolerances, a price has to be paid. The lower design sidelobes result in an increase in the antenna half-power beamwidth and an even larger increase in the width between the first nulls of the antenna pattern. Alternatively, a larger aperture is required if the beamwidth is to remain constant.

A slightly different approach to determining the error tolerances required for keeping the sidelobes below a specified value was given by Cheston and Frank.[149]

Effect of Digital Phase Shifter Quantization[150,151] Phase shifters, whether analog or digital, will usually have a phase shift that is not exactly what one thinks has been set. There will always be some error, and its effects on the antenna pattern can be determined by the equations given in the above subsection. The deliberate quantization of phase that results with the use of digital phase shifters, however, introduces a different type of "error" in the desired aperture illumination, and produces pattern degradation similar to that produced by a random error.

Reduction in Gain The gain of an array antenna due to phase errors can be obtained from Eq. (9.61) by setting $P_e = 1$, $\Delta^2 = 0$, and assuming $\overline{\delta^2}$ is small, which results in

$$G = G_0\left(1 - \overline{\delta^2}\right) \doteq \left(1 - \frac{\pi^2}{3 \times 2^{2B}}\right) \qquad [9.63b]$$

The right-hand portion is obtained by assuming that the phase error of a digital phase shifter of B bits is described by a uniform probability density function (Sec. 2.4) that extends over an interval $\pm\pi/2^B$. From Eq. (9.63b) a three-bit phase shifter causes a reduction in gain of 0.23 dB and a four-bit phase shifter has a gain reduction of 0.06 dB. Thus, on the basis of the loss in antenna gain caused by digital phase quantization, a three- or four-bit phase shifter should be satisfactory for most purposes.

Increase in Sidelobes In addition to a reduction in main-beam gain, the quantization of the phase in digital phase shifters can result in an increase in the rms sidelobe level. With the assumptions that (1) the energy lost by the reduction in main-beam gain shows up as an increase in the rms sidelobe level, (2) the element gain is the same for the main beam and the sidelobes (within the region of space scanned by the array, (3) an allowance of one dB for the reduction in gain due to the aperture illumination, and (4) one dB for scanning degradation; then the sidelobe level due to phase quantization is

$$\text{rms sidelobe level} \approx \frac{5}{2^{2B}N} \qquad [9.64]$$

where N = total number of elements in the array. If an array has 4000 elements, a three-bit phase shifter would give rms sidelobes of 47 dB below the main beam, and a four-bit phase shifter gives −53 dB sidelobes. Thus three or four bits should be sufficient for most large arrays, except when very low sidelobes are desired.

Peak Quantization Sidelobe The above assumed a random distribution of phase error across the aperture for the purpose of computing the rms sidelobe level. The actual phase distribution with digital phase shifters, however, is likely to be periodic which gives rise to spurious quantization lobes, similar to grating lobes but with smaller amplitude. Peak sidelobes sometimes are of more concern to the radar system engineer than are the rms sidelobes. The peak quantization lobe relative to the main beam when the phase error has a triangular repetitive distribution is

$$\text{peak quantization lobe} = 1/2^{2B} \tag{9.65}$$

This applies when the main beam points close to broadside and there are many radiator elements within the period of quantized phase error. The position θ_q of the quantization lobe in this case is

$$\sin \theta_q \approx (1 - 2^B)\, \theta_0 \tag{9.66}$$

where θ_0 is the angle to which the main beam is steered.

Equation (9.65) is an optimistic estimate for the peak lobe. The greatest phase quantization lobe is said to occur when the element spacing is exactly one half the phase quantization period or an exact multiple thereof. With an element spacing of one-half wavelength, the quantization lobe will appear at $\sin \theta_q \approx \sin \theta_0 - 1$, and will have a value of

$$\text{peak quantization lobe} \approx \frac{\pi^2}{4} \frac{1}{2^{2B}} \frac{\cos \theta_q}{\cos \theta_0} \tag{9.67}$$

The peak sidelobes due to the phase quantization of the digital phase shifter can be significant. Attempts should be made to reduce them if their presence is objectionable. One method for reducing the peak sidelobe is to randomize the phase quantization. A constant phase shift can be inserted in the path to each element, with a value that differs from element to element by amounts that are unrelated to the bit size. The added phase shift is then subtracted in the phase command sent to the phase shifter. (With a space-fed array, such as the lens array or the reflectarray, decorrelation is inherent in the array geometry.)

Beam-Pointing Error The maximum pointing error $\Delta\theta_0$ due to quantization, according to C. J. Miller[151] is

$$\Delta\theta_0 = \theta_B \frac{\pi}{4} \frac{1}{2^B} \tag{9.68}$$

where θ_B is the beamwidth. A four-bit phase shifter, for example, allows an angle error of $\Delta\theta_0/\theta_B = 0.05$. Small steering increments are possible with quantized phase shifters. A linear array of 100 elements, for instance, can be steered in increments of about 0.01 beamwidth with three-bit phase shifters.[149]

9.13 LOW-SIDELOBE ANTENNAS

The highest sidelobe of an antenna pattern is usually, but not always, the first sidelobe adjacent to the main beam. A conventional reflector antenna might have a maximum sidelobe of about 23 to 28 dB below the peak of the main beam. Sidelobes much lower than −35 to −40 dB with conventional reflector antennas are difficult to obtain by normal methods. There are some radar applications, however, that require much lower sidelobes. An example is the airborne high-prf pulse doppler radar, such as AWACS that was discussed in Sec. 3.9. A high-prf radar sees many multiple-time-around clutter echoes that enter the radar receiver through the antenna sidelobes. Such sidelobe clutter can be large enough to limit the performance of an airborne doppler radar. It was the need for a low-sidelobe antenna for AWACS that led Westinghouse (now Northrop Grumman) radar antenna engineers to be the first to successfully demonstrate in the mid-1960s sidelobes almost three orders of magnitude lower than was the practice with conventional antennas. A low-sidelobe antenna is also helpful in combating hostile noise jamming that enters the receiver via the sidelobes. It also aids in combating antiradiation missiles (ARM) that home on the radar's radiation, and in making more difficult the task of a hostile intercept receiver. Low sidelobes, however, come with a price. Such antennas need to be more precise, are more complex, their beamwidth is widened, and they have to operate in a clear environment.

Table 9.2 lists typical performance of low-sidelobe antennas as given by Evans and Schrank.[152] (In their paper, the authors state that "it is probably possible to do 5 dB better with tuning of phase and amplitude during the test." I have taken the liberty to add this 5 dB to the numbers that originally appeared in their paper.) The phased arrays in this table are not scanned in angle.

There has been no generally accepted definition of what are low sidelobes. Schrank[153] proposed that low sidelobes be defined as from −30 to −40 dB and that ultralow sidelobes be below −40 dB. These values probably should be lower. One might consider antennas with sidelobes of the order of −40 dB to be low, and antennas with −50 dB or lower to be ultralow sidelobe antennas.

The peak sidelobe is not always a good measure of the difficulty involved in obtaining low sidelobe levels. The larger the antenna gain, the easier it is (relatively) to obtain low sidelobes. A better measure of the difficulty for an antenna engineer to achieve low sidelobes is how far the peak sidelobe level is below the isotropic value, which we take

Table 9.2 Performance of Low-Sidelobe Antennas

Type of Antenna	Peak Sidelobe	RMS Sidelobe	Bandwidth
Slotted waveguide	−50 dB	−60 dB	10%
Corporate-fed array	−45 dB	−55 dB	60%
Reflector	−45 dB	−55 dB	60%

[1] Adapted from Evans and Schrank, Ref. No. 152, *Microwave Journal*.

here to be the level which is $1/G_0$ down from the peak, where G_0 = antenna gain. Low or ultralow sidelobe antennas might have peak sidelobes from 10 to 20 dB below isotropic. Thus if an antenna has a gain of 30 dB, its sidelobes might be made as low as -40 to -50 dB. From the second term of Eq. (9.60) it can be seen that the mean square error $\overline{\Delta^2} + \overline{\delta^2}$ must be small compared to the desired peak sidelobe level times the antenna gain G_0. Thus the larger the gain the larger can be the errors for a given sidelobe level relative to the main beam.

To achieve low sidelobes, an antenna must (1) employ an aperture illumination that will theoretically provide the required sidelobe level and be practical to implement, (2) be constructed and maintained with high precision, and (3) have no blocking of the aperture by nearby structures. Both array antennas and reflector antennas can be made to have low sidelobes, but it is easier to do with an array than with a reflector.

Low-Sidelobe Aperture Illuminations The Taylor aperture illumination, discussed in Sec. 9.11, has often been used for obtaining low sidelobe antenna patterns. It will be recalled that the Taylor illumination is characterized by a parameter $\bar{n}$, such that the first $\bar{n} - 1$ sidelobes closest to the main beam are equal. Beyond the $\bar{n}$th sidelobe, the sidelobe level decreases. For a circular aperture, the Taylor far-out sidelobes fall off as $(\sin \theta)^{-3/2}$. As was mentioned in Sec. 9.11, the parameter $\bar{n}$ must be chosen so that the aperture illumination is realizable.

Figure 9.48, from a report by Warren White,[154] is an example of how the normalized beamwidth β of a Taylor pattern varies with the peak sidelobe level and the parameter $\bar{n}$

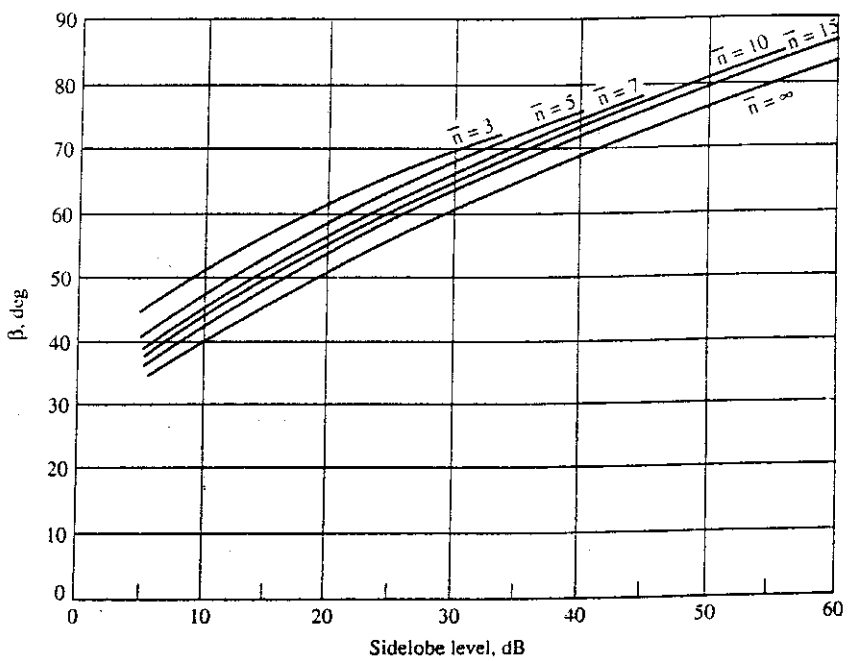


Figure 9.48 Normalized beamwidth β versus the sidelobe level and the value of $\bar{n}$, where the half-power beamwidth $\theta_B = \beta\lambda/D$. (From White.[154] Courtesy Institute for Defense Analyses.)

for a continuous circular aperture. The half-power beamwidth is $\theta_B = \beta\,(\lambda/D)$. The curve for $\bar{n} = \infty$ applies for the Chebyshev illumination, which is not realizable. Comparing a pattern with $\bar{n} = 5$ and a -25 dB sidelobe level with a pattern having $\bar{n} = 15$ and a -60 dB sidelobe level, the half-power beamwidth increases by a factor of almost 1.4 when lowering the sidelobes level from -25 dB to the level of -60 dB. The null width of the Taylor distribution increases even faster; over the same range of sidelobe levels, the null width increases by a factor of about 1.9. Thus one of the costs of low sidelobes is that the beamwidth increases as the sidelobe level is decreased. The gain also decreases with decreasing sidelobes. If the same beamwidth is to be maintained as the sidelobes are lowered, the antenna aperture must be increased in size.

Previously, Fig. 9.43 in Sec. 9.11 illustrated that the Taylor aperture illuminations are not always realizable because of the difficulty in obtaining the required shape of the currents at the edges of the aperture. The value of $\bar{n}$ must be chosen appropriately in order to achieve a pattern that is practical. A suitable criterion for selecting a realizable Taylor illumination is that it have a monotonically decreasing illumination, the slope at the edge of the aperture should be zero, and it should not turn positive. Ludwig[155] states that for a circular aperture the value of $\bar{n}$ should be no greater than 7 for -40-dB sidelobes, 11 for -50-dB sidelobes, and 16 for -60-dB sidelobes. The value of $\bar{n}$ should also not be too small. It should be at least 3 for -30-dB sidelobes and 4 for -40-dB sidelobes.[156] [Note these values of $\bar{n}$ for a circular Taylor aperture-illumination differ slightly from those given in Sec. 9.11 for a line-source Taylor illumination.]

Achieving the Low-Sidelobe Pattern In the above we indicated it is necessary to obtain an aperture illumination that can be implemented. In addition, the desired illumination must be maintained. The phase and amplitude tolerance on the aperture illumination must be determined, as in the previous section. As was indicated with the discussion of Fig. 9.47 in the previous section, the required error tolerances can be quite demanding. Schrank[153] points out that for a low-sidelobe array antenna, one must control the systematic errors and the mutual coupling in addition to the random errors. Mutual coupling can be compensated by appropriate computer-aided design. Systematic errors must be controlled by careful fabrication and serious attention to tolerances. Systematic errors generally affect the near-in sidelobes and can generate spurious lobes. Random errors, as we have seen, affect the far-out sidelobes. According to Schrank, it is the random errors that ultimately limit the ability to obtain low sidelobes.

Blocking or Masking of the Aperture The low-sidelobe antenna must be located in a clear environment in order to maintain the low sidelobes. Any obstruction in front of the antenna can alter the radiation pattern and result in an increase in sidelobes. Obstructions can include nearby buildings and trees. The need to operate in a clear environment is why the AWACS antenna (Fig. 3.45b) is located well above the fuselage of the aircraft that carries it. It is difficult to avoid the blockage caused by the tail, but the antenna is high enough to minimize the effects of the aircraft structure and the engines. The antenna is mounted in a rotodome that rotates in synchronism with the antenna, so that the antenna always sees the same radome environment.

A classic example of the effect of aperture blocking on antenna sidelobes is that caused by the masts and other superstructure on a ship.[157] No matter how low the sidelobes of an antenna might be in free space, when a mast is situated so as to block a portion of its radiation, the sidelobes can increase considerably. Blockage of a shipboard radar antenna can be avoided by mounting the antenna at the top of the mast. This is something that can be accommodated during the design of a new ship, but it is difficult to mount a heavy antenna at the top of a mast on an old ship which very likely has used up all of its margin for topside weight and moment.

There are two factors that cause the sidelobes of an antenna to increase when blockage occurs. One is that a part of the beam has been masked, which is equivalent to having a part of the aperture illumination excised. In other words the effective aperture illumination has been modified. The other effect is that the obstruction can scatter the radiated energy in new directions, so that target echoes or clutter echoes might appear as false targets in an erroneous direction. The radar thinks the scattered echo is in the direction at which the main beam points at the time, when the energy is from some other direction because it was scattered by the mast. Shaping of the mast or covering it with absorbing material can reduce the scattering that produces false echoes, but it does not reduce the distortion of the pattern caused by excising part of the radiated energy.

It is usually difficult to avoid degradation to the sidelobes and the main beam by blockage of a mast. Masts are often of steel, but masts made of dielectric will also cause similar blockage effects. The mast is often modeled as a cylinder for purposes of calculating its scattering effects, but actual shipboard masts are much more complex than simple cylinders. One method to avoid aperture blocking is to employ a four-face phased array antenna distributed around the ship so that none of its faces look into a mast. Even this is not perfect, however, since the ship's superstructure or deck might intercept part of the beam when the ship encounters large pitch or roll angles because of sea conditions.[158,159]

Effect of a Radome on Sidelobes The phase and amplitude of a signal can change when propagating through a radome, and can limit how low the antenna sidelobes can be. The periodicity of a metal space-frame geodesic dome precludes its use with very low sidelobe antennas. A air-inflated "bag" of thin dielectric material will allow lower sidelobes than can the space-frame geodesic dome. An airborne radar antenna located in the nose of an aircraft behind an aerodynamically shaped radome also can inhibit achieving low sidelobes. The radar designer does not have much control over this type of airborne radome since its first priority is to maintain structural integrity and conform to aerodynamical requirements. In addition, the radome must be able to withstand bird strikes, rain erosion, and lightning strikes. For lighting protection the antenna might be enclosed in a cage of metal rods. A radome can cause phase and amplitude changes in the radiated wave that result in loss and distortion. Energy that is scattered by the inner surface of the radome can result in a sidelobe that is known as the *radome flash.* In some cases[160] the peak value of the radome flash was found to be of the order of 40 dB below the main beam. Thus it can be more difficult to achieve low sidelobes with an antenna inside a radome. The effect the radome can have on an antenna pattern must be determined and corrected if low sidelobes are to be achieved. It should be kept in mind that the effects of a radome mounted

above the fuselage can be successfully handled, as it was with the impressive low-sidelobe antenna and radome employed on AWACS.

No-Sidelobe Array In the quest for a low-sidelobe antenna, it might be asked what would an antenna be like if there were no sidelobes whatsoever. Such an antenna pattern can be produced by a linear array with half-wave spacing and with element currents given by the coefficients of the binomial expansion. This no-sidelobe pattern is attributed to J. S. Stone who obtained patents for it in the late 1920s. Leon Ricardi[161] gives its beamwidth as $0.975/(N-1)^{1/2}$ rad and its directive gain as $1.77N^{1/2}$, where N is the number of elements. Although this antenna has no sidelobes it has a very fat main beam (the radiated energy has to go somewhere), the ratio of the current at the center of the array aperture to the current at the edge element is quite large for values of N that may be of interest, and its gain increases only as the square root of the aperture size. The currents at the elements of a nine-element binominal array, for example, are proportional to 1, 8, 28, 56, 70, 56, 28, 8, 1. The ratio of the current at the center to that at the edge is 70 to 1. For many reasons, the antenna with absolutely no sidelobes produces poor patterns and is not something that a radar engineer should aspire to use as a design goal.

It has been said that the gaussian antenna radiation pattern can be used to approximate the pattern of a no-sidelobe binomial illumination when the number of elements in the binomial array is greater than five and if the standard deviation of the gaussian pattern is made equal to $(L\lambda/8)^{1/2}$, where L is the antenna dimension.[162]

Examples of Low-Sidelobe Antennas The first low-sidelobe antenna was for the AN/APY-1 S-band radar that was the basis for the AWACS (E3A) airborne warning and control system.[152] As was mentioned, an antenna with exceptionally low sidelobes was needed for such an application in order to limit the large clutter echoes that might enter the receiver via the antenna sidelobes. The antenna, Fig. 9.49, consisted of 30 slotted waveguides, called *sticks*. These were fed from one end (series fed). The antenna was 24 ft (7.3 m) in width and 5 ft (1.5 m) in height, and was enclosed in a rotodome that was mechanically rotated in azimuth at 6 rpm. It is a rugged antenna, as is needed to maintain the mechanical tolerances.

Slotted Array The series-fed slotted array, of which the AWACS end-fed array was an example, is well suited as a low-sidelobe antenna since the slots provide a convenient means for achieving the necessary aperture illumination. The slots can be milled with precision computer-controlled machines, the structure can be made mechanically sturdy, and there is no blockage of the aperture as would be the case with a conventional reflector antenna. To achieve the desired low sidelobes, the mutual coupling between the slots in each stick and the coupling between sticks must be properly taken into account. The slots may be in the broadwall or the sidewall of the waveguide slots. (The AWACS antenna used sidewall slots.) Sidewall slots must be tilted at an angle in order to achieve coupling of energy from the waveguide. The tilt of the slots causes cross-polarized radiation, which must be surpressed in some system applications. Periodic errors due to the introduction of the necessary phase reversals in adjacent slot radiators (in order to provide the necessary phase in a traveling-wave waveguide array) must also be surpressed.

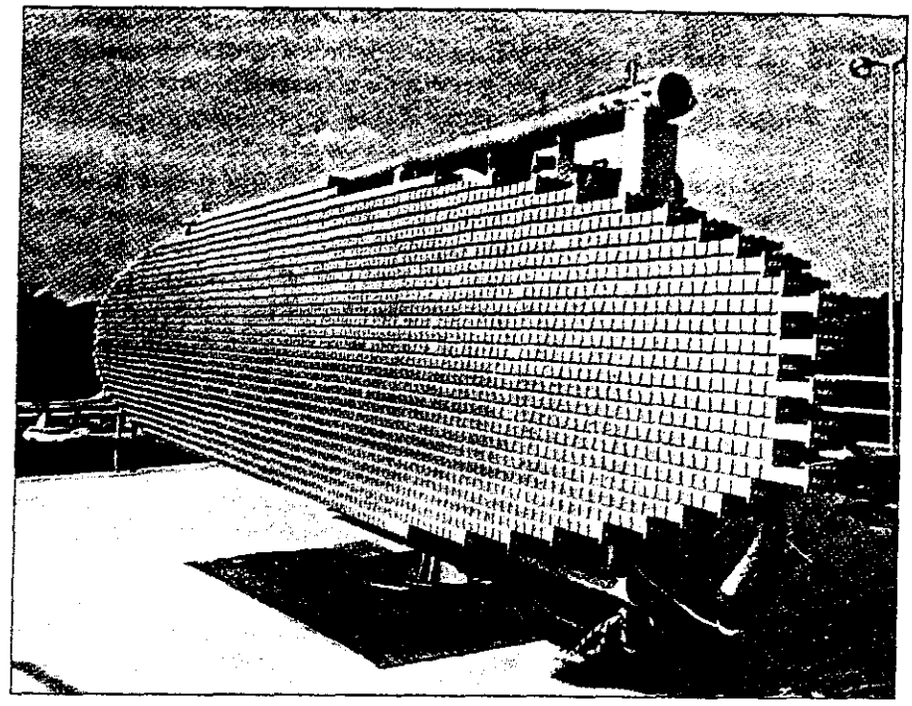


Figure 9.49 Slotted-array low-sidelobe antenna for the AWACS, AN/APY-1 radar. (Courtesy Northrup Grumman Corp.)

The end-fed, or any series-fed, slotted array is usually not of wide bandwidth since the direction of the radiated beam will squint (change angle) when the frequency is changed. With a narrowband signal this is not a serious problem since the direction the beam points relative to the antenna is known and can be compensated accordingly when extracting the angle measurement. When a wideband signal, however, is radiated by such an antenna, the beam will broaden, or smear. If a wide signal bandwidth is required in a low-sidelobe antenna, the corporate-fed array can be used.

Corporate-Fed Planar Array Operation over a wide bandwidth can be obtained with a corporate-fed array if equal path lengths are used connecting the antenna input to each element of the array. With equal path-lengths, changes in frequency do not cause changes in phase between the radiating elements, as they do in a series-fed array. The feed network of the corporate-fed array is not as simple as that of a series-fed array. There are couplers and branch lines in each path to the radiating elements, and they must be exceptionally precise in order to achieve the precision aperture illuminations required for low sidelobes. The use of dielectrics in such a feed system should be kept to a minimum since they can produce a phase change with a change in frequency. Strip transmission lines, with the strips supported by a minimum of dielectric, can be employed if their power handling is satisfactory. The precise tolerances and complexity of the corporate-fed array

can result in greater cost and complexity than the series-fed slotted array, but this is the price that has to be paid to have low sidelobes with wide bandwidth. Because of its greater complexity, Table 9.2 indicates that the corporate-fed array is more likely to have higher sidelobes than the slotted array.

Electronically Steered Low-Sidelobe Arrays It is more difficult to achieve low sidelobes when the array must be electronically steered. (The wide vertical beamwidth of the AWACS antenna can be steered electronically over a limited range of elevation angles, but this is different from an antenna that scans over wide angles in both azimuth and elevation.) Phase shifters introduce error and if they are digital, they must contain a large number of bits in order to suppress spurious sidelobes. The phase shifters might have to have six to eight bits rather than the three to four bits acceptable for antennas with conventional sidelobe levels. The effects of mutual coupling are more difficult to correct since mutual coupling will change when the antenna beam is electronically scanned. Arrays made up of subarrays also can produce periodic errors that can result in high sidelobes. Subarrays, especially large ones, can be a problem when low sidelobes are desired.

The additional cost of low-sidelobe phased arrays has been examined by W. Patton.[163] Based on analyzing three different arrays, each with 40 dB directive gain, but with sidelobes differing by steps of 6 dB, he concluded that the cost of building a phased array increased about 2.3 percent for each dB the sidelobe level is reduced when the size of the array is held constant and about 3.2 percent for each dB of sidelobe reduction when the antenna beamwidth is held constant.

FASR, an Electronically Steered, Low-Sidelobe Array Radar[164] This was one of the first low-sidelobe electronically steerable phased arrays. FASR (Fixed Array Surveillance Radar) was an experimental UHF radar developed by the Naval Research Laboratory to demonstrate how low-sidelobe antennas can be obtained in a shipboard environment by placing four fixed phased-array faces around a ship so as to avoid blockage by masts or superstructure. A single array antenna was 32 by 12.5 ft with 297 dipole radiators arranged in 27 columns and 11 rows. Although not small in physical size, the array contained a small number of elements and was not of high gain. Thus it was more difficult to achieve low sidelobes (tighter tolerances were required) than in an antenna with higher gain (and thus more elements). FASR was designed for −40 dB sidelobes and used six-bit digital diode phase-shifters. The 5 by 12° beam scanned 120° in azimuth and 90° in elevation. Monopulse sum and difference beams were generated. The desired sidelobes of −40 dB were achieved at broadside, but they increased slightly when the beam was scanned. The difference pattern also achieved −40 dB sidelobes with respect to the sum pattern.

Parabolic Reflector[165] There are many advantages of a parabolic reflector for radar applications; however, it is more difficult to achieve low sidelobes with a reflector. Among the things that need to be done to obtain low sidelobes are:

1. A solid rather than a mesh reflector surface should be used to avoid leakage in the back direction.

2. The feed system should illuminate the edges of the reflector with low energy, not only to obtain a highly tapered illumination, but also to minimize the sidelobe energy caused by spillover.
3. Spillover radiation from the edges of the reflector can be attenuated by the appropriate placement of absorbing materials or shields.
4. The feed system might have to consist of more than one horn or radiator in order to properly control the aperture illumination.
5. There can be no aperture blockage by the feed system, which leads to the use of an offset reflector.
6. The mechanical tolerances of the reflector surface have to be better than that of an array by a factor of two because of the two-way path on reflection from the surface.

Scudder[166] described the design of an *S*-band 3D air-surveillance radar using an offset reflector 20 ft wide by 12 ft high with a bandwidth of 600 MHz. The azimuth beamwidth was 1.4°. Seven corrugated feed horns were positioned to provide seven overlapping beams in elevation from 0° to 20°. More than one feed horn was used to form each elevation beam. Absorbers were placed at both the top and bottom edges of the reflector to suppress spillover radiation. Based on measurements of a one-tenth scale model operating at K_a band, the peak near-in sidelobes were -40 dB and decreased rapidly to less than -50 dB, with the wide-angle sidelobes below -60 dB.

Another example is that described by Williams et al.[167] Their elliptical shaped reflector had a major axis of 45 wavelengths and a minor axis of 15 wavelengths producing an azimuth beamwidth of 1.7° and an elevation beamwidth of 5°. It was feed from an offset array of four conical horns. The aperture illumination was designed to produce a peak sidelobe of -50 dB. The price paid for this low sidelobe design was that the beamwidth was broadened by a factor of 1.47 compared to the beamwidth that would have been produced by a uniformly illuminated circular aperture. An *X*-band model produced a peak sidelobe of -43 dB, which was attributed by the authors to spillover radiation from the feed support struts.

An offset-fed parabolic cylinder has some advantage over other reflector antennas for producing low sidelobes because the line source feed (which may be a linear array or a pill box) allows much better control of the aperture illumination than when a conventional parabolic reflector is used with one or several horn feeds.[168] If a corporate-fed linear array is used with equal lengths of lines to the radiators, the antenna can have much broader bandwidth than a series-fed array.

Measurement of Low-Sidelobe Radiation Patterns A good pattern range is needed to accurately measure the radiation pattern of a low-sidelobe antenna, especially if the depth of the nulls are of interest.[169] The pattern range must be sufficiently large so that the curvature of the wavefront due to the finite distance does not affect measurement accuracy. The rule of thumb used by antenna engineers for conventional antennas is that the distance between the antenna and the pattern-measuring source should be at least $2D^2/\lambda$, where D is the antenna dimension and λ is the wavelength. This is adequate for sidelobes

down to about −30 dB; but not for lower sidelobe levels. Hacker and Schrank[170] state, however, that the distance of $2D^2/\lambda$ can be satisfactory for the accurate measurement of the wide-angle (far-out) sidelobes, but it does not accurately measure the first one or two near-in sidelobes. If, on the other hand, the entire antenna pattern is to be determined to an accuracy of better than 0.5 dB, the pattern range must be $8D^2/\lambda$ when the pattern is a modified Taylor with first sidelobe of −50 dB. Using the Taylor $\bar{n}$ linear aperture illuminations, Hansen added to this by determining that that if an error of 1 dB or less is required when measuring a −40 dB sidelobe, the measurement distance should be $6D^2/\lambda$.[131] For a −60 dB sidelobe design, the distance should be $12D^2/\lambda$.

System Implications of Low Sidelobes Low antenna sidelobes are important for achieving the desired performance of pulse doppler radars in the face of heavy clutter. They can also be useful for reducing the effects of sidelobe jamming and making the job of hostile intercept receivers and antiradiation missiles more difficult. Low sidelobes, however, do not come without cost—both monetary and performance.

Compared to the conventional parabolic reflector, low-sidelobe antennas are more expensive, less rugged, more likely to be heavier, require better mechanical and electrical tolerances, and are harder to maintain. As has been mentioned, the lower the sidelobes the less the antenna gain and the wider will be the main beam. If a larger antenna cannot be used to maintain the beamwidth, the wider beam means poorer angular resolution and accuracy, greater susceptibility to main-beam jamming, and larger clutter echoes received in the main beam. In an airborne doppler radar that is concerned with detecting low velocity ground-moving vehicles with doppler frequencies near that of the main-beam clutter, a narrow antenna beamwidth with conventional sidelobe levels may be more important than a low sidelobe antenna with an increased beamwidth.

To avoid negating the benefits of low sidelobes, the antenna must be operated in a clear environment without obstructions or nearby objects that block or scatter the radiated energy. When a low-sidelobe antenna must be operated within a radome, special considerations need to be given to the design of the radome in order that the sidelobes are not degraded. Usually the radome will rotate along with the antenna (a rotodome) in order to maintain the same radome environment seen by the antenna as it rotates.

As with most things, low-sidelobe antennas have both good and bad effects. They are not universally applicable and should be used only when their desirable features outweigh their disadvantages.

9.14 COST OF PHASED ARRAY RADARS

A phased array radar generally costs more than a conventional radar that employs a mechanically scanned antenna but it also can provide unique capabilities not available with other antennas. It has been used chiefly in military applications where the unique features of a phased array may sometimes compensate for its higher cost. It has seldom been employed for civilian radar applications since there are few applications where its greater expense can be justified when competitive market forces make price an important

consideration. There has been much interest in reducing the costs of phased arrays to make them more competitive for both military and civilian applications.

Factors Affecting the Cost of a Phased Array Radar Tang and Brown[171] state that the cost problem of a phased array is attributable to the following:

- The large number of discrete components in a conventional phased array that have to be individually fabricated, assembled, tested, and installed.
- The production yield of components, especially solid-state amplifiers that use monolithic high-power chips.
- The labor involved.
- The limited production quantity of any particular radar system which does not justify large capital investments for a dedicated production line with special production tooling that could reduce manufacturing costs.

These four categories are only a part of the problem. Cost depends on the particular radar architecture, the degree to which multifunction operation is used, computer software for operating the radar, and the radar frequency. Each of these will be briefly reviewed.

Effect of Radar Architecture on Cost Section 9.9 discussed the various architectures for phased array radars, and some mention was made of the effect of the architecture on radar cost. It is difficult to be quantitatively accurate regarding predictions of radar cost, but some generalizations can be made. For example, a space-fed array using a single high-power vacuum tube transmitter has been less expensive than a corporate-fed array or an active aperture solid-state phased array. Also, the row-column control of phase shifters is usually cheaper than when each phase shifter is controlled individually.

There was a brief description in Sec. 9.9 of how Russian (Soviet) air-defense phased array systems achieved lower cost. Cory[105] summarized these as:

- Minimizing the total number of phase shifter modules.
- Simplifying the radar architecture (such as by using a space-fed array).
- Designing simple and inexpensive components.
- Minimizing the size and complexity of the control system.
- Simplifying the feed design.

In addition the Russians used several low-cost radars, each performing a single air-defense function rather than the more complex multifunction phased array that has been the more usual practice in the U.S.

Multifunction Radar and Cost Because of its flexibility and rapid beam steering, an electronically steered phased array antenna can be used to perform multiple radar functions including search, track, weapon control, missile guidance, target recognition, and perhaps others. The ability to perform multiple functions with one phased array radar has been a major selling point of those who market radar systems that employ phased arrays. One

should be cautious, however, since a single multifunction phased array radar might not be the best approach for all radar applications. This is especially true for air defense.

The basic problem with a multifunction electronically steered phased array radar for air defense is that compromises must be accepted when a single radar is used for both surveillance and weapon control. It is well known among radar system engineers that the lower microwave frequencies are more suited than the higher frequencies for long-range air-surveillance radars. The frequency of choice is usually L band (1.215–1.4 GHz). On the other hand, the higher microwave frequencies are more desirable for weapon control, with X band the usual choice (8.5–10 GHz). When a single phased array radar is required to perform both surveillance and weapon control, a compromise choice of a single frequency has to be made, generally somewhere between L and X bands. The U.S. Navy's Aegis air-defense system is at S band and the U.S. Army's Patriot air-defense system is at C band, yet they basically have the same mission (except one is on ships and the other is land mobile). When the beamwidths must be the same no matter what the frequency (which implies that the antenna gain is independent of frequency) the antenna aperture will be smaller at the higher frequencies. Thus performing the surveillance function at S band results in less range performance (for a given average transmitter power) than if it were at L band. Air surveillance at higher frequencies (such as C band) result in even less range performance. Also, the higher the frequency the less will be the available doppler space (to detect moving targets in clutter) because of blind speeds. When weapon control radars operate at C band or S band, the antenna beamwidths will be wider than they would be at X band, resulting in poorer angle accuracy and less ability to deal with multipath effects from surface reflections. Thus, even if cost were of no concern whatsoever, one usually has to accept lesser performance in both surveillance and weapon control when a multifunction radar is employed for air defense at a single frequency band.

It is not always necessarily true that an air-defense radar system with multiple radars at different frequencies has to be more expensive than a single-frequency multifunction system which has the same performance. Both Barton[104] and Cory[105] have written about the benefits of the Russian design approach to air-defense radar systems that employ several simple cost-effective phased array radars to perform the various functions of surveillance, weapon control, and low-altitude detection.

In some radar system applications the optimum frequencies for surveillance and for track might be the same, so that multifunction radars do not have the same limitations that are experienced with their use for air defense. Space surveillance for the detection and tracking of satellites is one example where both the surveillance and the tracking functions can be performed at the same frequency. (UHF is a good choice, as in the AN/FPS-85 and Pave Paws.) Thus no significant compromises in performance need be made when using a single multifunction radar for the space surveillance functions of detection and tracking.

Multifunction radars that use mechanically scanned planar array antennas have been well suited for military combat aircraft and, in the past, have been the norm. A modern radar for a fighter/attack aircraft has to perform a number of functions, maybe from 6 to 9, for air-to-air purposes and a similar number of different functions for air-to-ground purposes. There have been no significant limitations (other than having sufficient time) in using the same airborne radar to perform all the many functions at X band. The multifunction

electronically steered phased array can also be used for airborne fighter/attack radar application, but a single phased array face in the nose of a fighter/attack aircraft (usually limited to less than $\pm 60°$ in angle) cannot provide as large an angular coverage as can a mechanically steered planar array antenna. (Large angular coverage is especially important for the dog-fight role.) Two or more phased arrays might be used for increased coverage, but they result in increased system size and weight. Mechanical antennas thus can be competitive, and in some ways superior, to electronically steered phased arrays for military airborne applications since they can do what is required of an airborne radar antenna at less cost and less weight than an electronically steered phased array.

Offensive bomber aircraft, such as the B-1B, have also employed the multifunction electronically steered phased array to perform the many radar functions unique to the bomber. Such radars have been more expensive than conventional mechanically scanned radars for the same purpose.

Computer Software and Cost The computer can be a significant part of the cost of a versatile multifunction phased array radar. It has not been the computer hardware but the software that can be a sizable fraction of the total radar development cost. In early long-range multifunction phased array radar systems, software cost was about 30 to 40 percent of the total. Over time, computer software has become better and easier to obtain, but it is still a significant factor in achieving a successful phased array radar system. The design of the computer software for phased arrays must enter the radar system development process at an early stage, with sufficient time and funds allowed for it to be successfully completed. Without sophisticated computer control, a phased array can do very little.

Effect of Frequency on Phased Array Radar Cost In general, the lower the frequency of an air-surveillance radar that uses a phased array antenna, the lower the cost. The rationalization is as follows. A passive phased array is assumed, one with a single receiver and a single high-power transmitter, that requires a specified power-aperture product. (That is, the average transmitter power times the antenna area is a constant, Sec. 2.13.) The cost of an antenna element and a phase shifter is more or less independent of frequency, but the cost of an array is proportional to the number of elements. In this comparison it is further assumed that the antenna must have the same gain (same elevation and azimuth beamwidths) at whatever frequency is selected for its operation. Thus a 9000-MHz (X-band) phased array radar would have the same number of elements as a 450-MHz (UHF) radar. With the above assumptions, the array antennas at the two different frequencies would cost approximately the same. The aperture of the 450 MHz radar, however, is 20 times larger in linear dimension and 400 times larger in area than that of the X-band radar. Since the power-aperture product in this example is assumed to be the same at both frequencies for equal performance, the transmitter power for the X-band radar must be 400 times that of the UHF radar, with the result that the cost of the X-band radar will be many times greater than that of the radar at UHF. The effect of a higher frequency on cost is probably even greater for active-aperture phased array radars that employ a T/R module at each element containing its own solid-state transmitter, receiver, phase shifter, and duplexer. The cost of a T/R module is likely to be greater at the higher frequencies rather than be relatively independent of frequency as in the passive array. The above

argument assumed a surveillance radar where $P_{av}A$ is constant. In a tracking radar, generally the product $P_{av}A^2$ is a constant with frequency. This makes it even more likely that an equivalent radar at a lower frequency will cost significantly less than one at a higher frequency.

The above has been a very simplistic argument with some very gross assumptions. There may be other requirements the radar must meet which require that a phased array radar operate at higher frequencies in spite of higher cost (such as if it has to fit into the nose of an aircraft). Nevertheless, it is often true that the lower the frequency the more affordable will be the phased array radar.

Reducing Phased Array Cost This subsection summarizes some of the guidelines for lowering the cost of phased array radars.[172]

1. *Operating at as low a frequency as the application will permit.*
2. *Emphasizing low-loss design.* This might seem obvious or trivial, but it has not always been given sufficient attention.
3. *Time sharing a four-faced phased array with a single transmitter.* This is done as a cost-saving measure, but it can affect the overall performance of the system.
4. *Employing the active aperture architecture only when it is appropriate.* The losses in the active aperture array are less than the losses in a passive array with a constrained feed. It is not always obvious, however, that the total system cost of an active aperture radar will be lower or its performance better than other phased-array architectures.
5. *Single transmitter.* It has usually been true that the greater the RF average power from a single device, the lower will be its cost per watt.
6. *Space-fed arrays.* Lower loss and less complexity of the space-fed array can result in lower cost.
7. *Attention to computer issues.* The cost of the computer software for a phased array and the time required to generate it can be significant.
8. *Trainable arrays.* If an application requires 360° of coverage, one might not want to have the expense of four identical phased array radar systems to provide the total coverage. As mentioned in no. 3 above, one or two transmitters might be time-shared among the four apertures of an array radar system. The combination of two trainable arrays and a conventional 2D air-surveillance radar could be less expensive that a full four-face phased array.
9. *Avoiding a multifunction array.* There can be, in some applications, other less costly approaches that can perform more effectively the same mission.

It was mentioned that phased array radars have the advantage of being more readily hardened to withstand nuclear blast effects than a mechanically steered radar. Hardening only adds to the already large costs of phased arrays. In the U.S. development of intercontinental ballistic missile defense systems in the 1960s and early 1970s, the high cost of a fully hardened system was a factor in leading to the ABM (antiballistic missile) treaty between the U.S. and the Soviet Union.

It might be mentioned that the life-cycle costs of military systems, which were not considered in the above, include the cost of development, procurement, installation, training, operating, and maintenance. A further consideration is that the development cost of a radar usually is a small fraction of the total life-cycle costs of the system.

9.15 OTHER TOPICS CONCERNING PHASED ARRAYS

Bandwidth of a Phased Array Antenna Two different kinds of bandwidths need to be considered for phased arrays. One is the instantaneous, or signal, bandwidth, which is an indication of the maximum bandwidth of a signal that the array can handle without distortion. Usually, it is difficult to obtain an array signal bandwidth of more than a few percent. The other is the operating, or tunable, bandwidth over which a narrowband signal can be received (or transmitted) without distortion.

Signal Bandwidth Figure 9.15a showed a two element array receiving a signal that arrives at an angle θ_0 relative to broadside. The signal appears at element 2 before it appears at element 1. If a delay line of the proper length is inserted at element 2, the two signals will coincide and add without loss. There is no theoretical limitation to the signal bandwidth in this case when delay lines are used to bring signals from the various elements of the array into time coincidence. As has been said in the original discussion of Fig. 9.15, time delays inserted at each of the many elements of a large array have not been not practical. Instead, the delay line is replaced with a phase shifter that is limited in phase to the range 0 to 2π radians. Signals can be phase coherent so long as they overlap in time. But the signals do not overlap with 0 to 2π phase shifters during the transient build-up time. Thus phase shifters produce coherent addition only for narrowband (long time duration) signals.

The limitation on bandwidth when phase shifters are used in an array is dependent on the rise time, or build-up time, of the signal as it transits across the array. The transient response of the incident signal as it builds up across the array has the same effect on signal bandwidth as the transient response, or build-up time, of a conventional filter. The build-up time of a linear array of dimension D when a signal is incident on the array at an angle θ_0, is $(D \sin \theta_0)/c$, where c is the velocity of propagation. If, for example, the angle of arrival $\theta_0 = 45°$, and $D = 30$ ft, the transient build-up time is about 22 ns. The signal bandwidth is thus limited to the reciprocal of the transient build-up time, or about 45 MHz. There is zero build-up time for a signal that arrives from the broadside direction ($\theta_0 = 0$); hence, there is no theoretical bandwidth limitation. (This assumes that the signals from all the antenna elements are summed with equal-length transmission lines.)

Another aspect of signal bandwidth has to do with the change of phase with a change in frequency. The phase shift ϕ required to steer a beam to a given direction is assumed to be independent of frequency. If the value of ϕ is chosen so as to point the beam to a direction θ_0 when the frequency is f_1, the beam will point to a new direction when the frequency is f_2. If it is assumed that the signal's frequency spectral width must not cause the beam to scan more than $\pm$ one-fourth beamwidth, Cheston and Frank[173] show that

$$\text{signal relative-bandwidth in percent} = \text{broadside beamwidth in degrees} \quad [9.69]$$

where the signal relative-bandwidth in percent is 100 times the absolute bandwidth divided by the RF carrier frequency, or $(B/f_0) \times 100$. This expression is based on the array having an equal-path-length feed and the beam is scanned to an angle of 60°.

Although it has been impractical to employ delay lines at each element of an array, they have sometimes been used at the subarrays to increase bandwidth. Phase shifters are used at each element in addition to the delay lines at each subaperture. This reduces the complexity of an array compared to one with delay lines at every element, but it also has less bandwidth than a true time-delay array. Subarrays with delay lines increase the bandwidth of a linear array in proportion to the number of subarrays, compared to the bandwidth of an array that does not have subarrays.[174] The sidelobes will increase, however, with the use of subarrays, which may not be desirable for some applications.

Operating, or Tunable, Bandwidth Although the build-up time of an array limits the signal bandwidth that it can handle without distortion, it is possible to operate an array over a very wide band of frequencies by retuning; that is, by re-setting the phase shifters to new values when the frequency (of a narrowband signal) is changed. Such an array can radiate different narrowband signals one at a time at different frequencies by readjusting the phase shifters with each new frequency. The operating bandwidth of the array can be quite large and might be limited only by the onset of grating lobes. This assumes that the antenna elements and other components of the array are wideband.

It has been reported[175] that an array containing 4096 open-ended waveguide radiators with a triangular arrangement of elements was capable of operating over a 30 percent frequency band and over a scan volume of more than 120° in both azimuth and elevation. A wide-angle impedance matching dielectric sheet was placed in front of the array. (The frequency at which this array operated was not given.)

Computer Control of an Array Although a radar with a conventional mechanically scanned antenna can operate without computer control, the multifunction electronically scanned phased array must be controlled by a computer if it is to achieve its full potential. An important task of a computer is to generate the phase shifter commands for each element of the array to steer the beam in the desired direction. But this is only a small part of what an array computer must accomplish. A much more demanding task is to effectively manage the various radar functions required of the array. The computer hardware for the multifunction operation of an array usually is not as much a concern as is the computer software needed to generate and schedule the various waveforms, data rates, and processing without degradation of performance.

The demands on the computer that controls the phased array radar vary with the application. Generally, the various radar tasks that have to be performed by a phased array under computer control must be done sequentially rather than simultaneously. In a phased array radar for air defense, for example, it is the job of the computer to allow the radar to track a large number of targets as well as search a large volume of space within a specified time for the detection of targets. The tracking data rate depends on whether or not the target is considered hostile and is being engaged. The revisit time during search, for example, might need to be one or two seconds when looking for pop-up, low-altitude targets that first appear over the horizon at short ranges (perhaps 8 to 20 nmi). With long-range

targets (150 to 200 nmi), the revisit time can be greater (perhaps 10 s) since there is more time for an air-defense system to react to a long-range threat than to a short-range threat. (A sea-skimming missile flying at Mach 3 at very low altitude might first appear above the radar horizon at a range of 10 nmi from the radar. If the detection decision is made almost instantaneously, then there is then less than 20 s available to destroy the missile before it reaches its objective.) If the radar is to detect tactical ballistic missiles at long ranges, the search patterns and data rates will be different from those used for aircraft targets. Thus the computer must program different search procedures depending on the range and type of target expected.

Once a target is detected, the data is used to update an existing track or initiate a new track. When the radar is performing its search function, the scanning must be interrupted periodically in order to radiate one or more pulses in the direction of known targets already held in track. Since the target is in track, the radar will know the approximate time the echo is supposed to arrive back at the radar. At that time, the phased array beam can be pointed to the direction of the target so as to receive the echo signal. Search and track are therefore accomplished in an interleaved manner. The computer has to be programmed to be able to accomplish this efficiently; that is, to search the required volume with the required revisit times and to track a large number of targets without serious degradation.

In spite of best efforts, there will likely be radar functions that become overloaded when the number of targets is large. This occurs because the various functions of a conventional phased array are performed sequentially in time. Everything cannot be done at once, so priorities must be assigned to the various functions to be performed. Those with less importance are performed at a lower data rate—or maybe, not at all. Among the highest priority functions are those involved with the engagement of threatening hostile attacks. This includes tracking of the attacker and providing guidance information to the intercepting missile. Time critical functions, such as short-range horizon-search for low altitude threats are also of high priority. Next in priority might be the tracking of confirmed hostile threats. Of lower priority is the tracking of known friendly targets. Low priority is also given to above-horizon search for targets at long range. These are only a few of the many tasks or functions that a phased array radar has to perform as part of an air-defense system.[176,177]

As the number of targets increases or as the demands of the air-defense increase, the computer can become increasingly overloaded and it will be more difficult for the phased array radar to perform all of its tasks with equal effectiveness. The problem has sometimes been described as there not being enough microseconds in a second to do all that is required of the phased array radar. One author[178] described the problem as not being able "to get round the sky quickly enough." Among the many factors that can cause the phased array to be overloaded and result in compromised performance are (1) a relatively large number of angular resolution cells that the radar must examine, (2) the need for the radar to detect small moving targets in large clutter, which requires that it dwell in each direction long enough to obtain good doppler filtering to suppress the clutter, and (3) a relatively long time to acquire targets and to initiate tracks.

The overloading of the functions performed by a conventional phased array radar is less of a problem with a radar that uses digital beam forming to look everywhere all the time since the various functions can be performed simultaneously (in parallel) rather than sequentially (in series), as was discussed in Sec. 9.9.

Radant Phased Array[179,180] Radant, originally developed by Thomson-CSF of France, is a different method for employing diodes or microelectromechanical switches (MEMS) to produce agile beam steering. Instead of using diode phase shifters in the conventional manner, Radant employs strips of metalized diodes arranged in columns to produce phase shift in one angular dimension by changing the voltage applied to each strip of diodes. The principle of the Radant diode lens for scanning in one angular coordinate is illustrated in Fig. 9.50. The vertical strips of diodes that produce the phase shifts are illuminated from the back by a plane wave as might be generated by a simple planar array. A parabolic reflector or a lens also might be used to illuminate the Radant, as might space feeding (but not offset space feed).

Radant can be thought of as a lens whose index of refraction (or dielectric constant) can be varied by appropriately biasing each string of diodes.[181] Biasing the PIN diodes can provide the desired change in susceptance to change the "index of refraction" of the lens consisting of many diodes. A number of planes of diodes are used to obtain the total phase shift. In a Radant array antenna with the geometry of Fig. 9.50, there might be 20 planes of diodes having a total thickness of 4 inches at X band.

As described, beam steering occurs in only one plane. Figure 9.51 illustrates two-dimensional steering. In this arrangement the first lens steers the beam in the horizontal direction. The plane of polarization is then rotated 90°, and a second lens oriented 90° to lens no. 1 steers the beam in the vertical. Control of two-dimensional beam steering is done with row and column commands so that with a $N \times M$ element array there need be only $N + M$ commands rather than $N \times M$. An advantage claimed for the Radant antenna is that it can be of lower cost than other array configurations.[182]

A Radant lens electronically steered phased array radar, called the RBE2, was developed by Thomson-CSF for the French Rafale multirole combat aircraft built by Dassult Aviation.[183]

Ferroelectric Phased Arrays A change in phase can be had by a change in the dielectric constant (permittivity) of the material in which the electromagnetic signal propagates. Materials whose dielectric constant varies with the d-c voltage applied across it are known

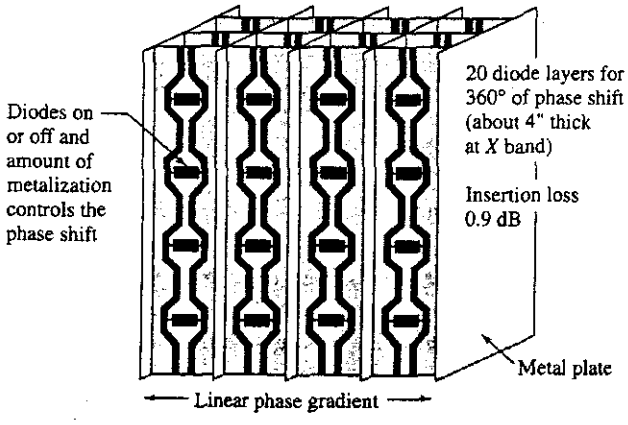


Figure 9.50 Principle of the Radant antenna for scanning in one angle coordinate, horizontal in this case.
(Courtesy of Jagannohan Rao, Naval Research Laboratory Radar Division.)

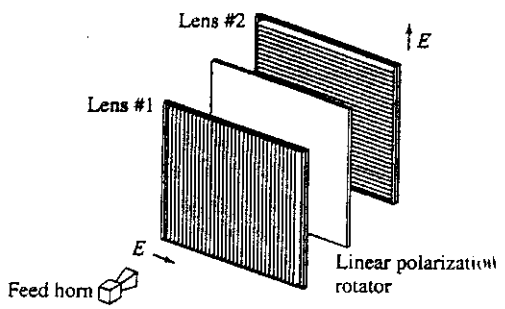


Figure 9.51 Method of obtaining a two-coordinate beam steering Radant antenna. (Courtesy of Jaganmohan Rao, Naval Research Laboratory Radar Division.)

as *ferroelectrics.* Ferroelectric phase shifters have been investigated in the past, but they were difficult to match because of their large dielectric constant and they have higher losses than might be desired. With time, ferroelectric materials have improved and they have shown more promise.

An interesting approach to a ferroelectric phased array based on a lens array with bulk phase shifting is that of Jay Rao and colleagues.[184] Figure 9.52 is a very simplified sketch of the operation of a ferroelectric lens array that steers in one plane. It is made up of ferroelectric slabs sandwiched between conducting plates that apply the d-c voltage that determines the dielectric constant and the phase shift. Changing the voltage on the conducting plates changes the direction of the beam. Beam steering in two orthogonal planes is obtained by rotating the plane of polarization 90° and using a second lens array oriented 90° to the first lens, as was illustrated in Fig. 9.51 for the Radant lens. Similarly the ferroelectric array can be illuminated with a conventional non-steerable planar array or by a horn feed, just as for Radant. The material used was a bulk oxide-ceramic composite of barium strontium titanate oxide (BSTO) with a typical dielectric constant between 90 and 120.

Theory and experimental measurements indicate that a ferroelectric array at *X* band can operate over the frequency range from 8 to 12 GHz with a voltage standing-wave

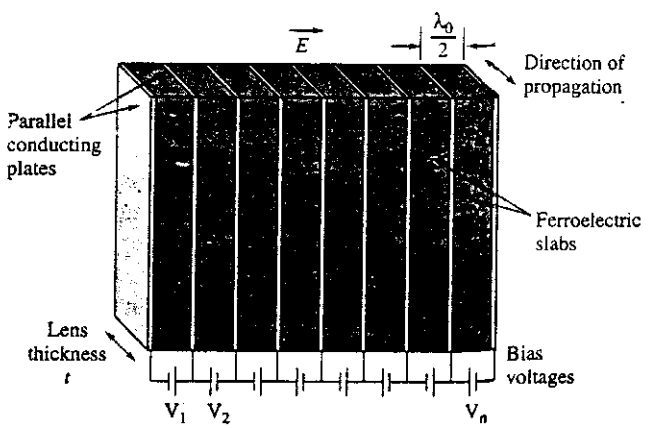


Figure 9.52 Basic configuration of a ferroelectric lens array for steering in the horizontal plane with horizontal polarization.[184] (Provided by J. Rao and D. Patel of the Naval Research Laboratory Radar Division.)

ratio less than two, and with a loss between one and two dB.[184] The advantage of this form of phased array compared to other phased arrays is its potential for low cost. It uses row and column steering that requires far fewer phase-shift control signals than an array which requires control signals for each element of the array. It is said to have smaller lens thickness, higher power capability, simpler beam steering controls, and use less power to control the phase shift than a Radant lens array. The use of row-column bulk phase steering instead of individual phase shifters at each element can make it more difficult, however, to achieve low sidelobe levels.

Conformal Array Antennas A long-sought, but difficult to achieve, desire of radar system engineers is to be able to place array elements anywhere on a surface of a relatively arbitrary shape and obtain a directive beam with good sidelobes, good efficiency, and which can be conveniently scanned electronically. An array on a nonplanar surface is called a *conformal array.* Such an array, if practical, might be configured to conform to the nose, wing, or fuselage of an aircraft or missile. Most of the work on conformal arrays has been for relatively simple shapes such as the cylinder, hemisphere, cone, or truncated cone.[185] It has been difficult to achieve a practical conformal array, except in simple geometries. There are good reasons why almost all operational phased arrays are planar.

In some conformal geometries, equal element-to-element spacing is not practical. Because of the nonplanar surface the element polarization can vary from one point on the array to another and with the beam pointing direction. The aperture illumination cannot be separated into two orthogonal patterns, as it can in a rectangular array. The boresight of a difference pattern might vary with angle. The calculation of mutual coupling is more difficult with a nonplanar surface. The issue is not whether a conformal array can be built, but whether some other solution is better.

If one wanted to have a cylindrical antenna, one way to achieve it would be to use four planar arrays arranged in a square and cover the structure with a cylindrical radome—and not let anyone look inside. If the change in beamwidth and gain with scan angle experienced by planar arrays were of concern, adjacent faces of a four-face phased array might be used cooperatively to maintain the beam relatively unchanged with scan. As the beam from one planar array is scanned off broadside, some of the transmitter power can be applied to an adjacent face and diverted to scan in the same direction as the original face.[186] Thus each face might work in conjunction with its two adjacent faces to maintain the beam shape almost independent of scan angle.

Thinned, or Unequally Spaced, Arrays[187,188] A normal phased array antenna has its elements spaced about a half-wavelength apart for good performance. When the elements are, on average, spaced much greater than half-wavelength the array is said to be thinned, or unequally spaced. (This assumes that a wide angular coverage is required; otherwise it would be known as a limited-scan array.) The beamwidth is determined by the electrical size (in wavelengths) of the array but its gain and sidelobe levels are determined by the number of elements that remain. A thinned array will have about the same beamwidth as a filled array, but its gain will be reduced in proportion to the number of elements

removed. Its peak and average sidelobes will increase. [The simple expression for antenna gain, $G = \pi^2/\theta_a\theta_b$ of Eq. (9.5b), doesn't apply to thinned arrays.] Many methods have been tried to determine element spacings that will result in acceptable antenna patterns. Two of the more successful will be described briefly.

Dynamic Programming One method is to have a computer calculate, for a fixed number of elements, the antenna patterns for all possible locations of the elements within the array and determine which is best. This method of total enumeration, is impractical because of the large number of combinations that are possible. A more practical approach is *dynamic programming*, an optimization method that can produce an equivalent result to total enumeration under certain conditions. Its advantage is that it is much less computer intensive. Dynamic programming determines the optimum solution to a multistage problem by optimizing each stage of the problem on the basis of the input to that stage. It is a good method for finding the spacings of a thinned array when the number of elements is not too large.[189]

Density Taper The other method, which is applicable to large linear or large planar arrays, is to employ a *density taper*.[190] Consider a uniform grid of possible element locations with equal spacing of one-half wavelength. The desired amplitude illumination for a conventional filled array is used as the model for determining the *density* of equal-amplitude elements. That is, the density of equal-amplitude elements is made to approximate the desired aperture illumination. The choice of whether or not to include an element in a possible location may be made statistically or deterministically. In one design, a one-degree beamwidth circular array which would have 7800 elements if completely filled, had 3773 elements when a density taper was employed based on a 30-dB Taylor amplitude illumination.[187] This represents a thinning of 52 percent, where the degree of thinning is defined as the ratio of the number of elements removed from a filled array divided by the original number of elements. Density taper was used in designing the receiving aperture of the AN/FPS-85 and the Cobra Dane (AN/FPS-108) space-surveillance phased array radars.

System Degradation Thinning of a phased array will produce serious undesirable characteristics for many radar applications when the degree of thinning is too high. The gain is significantly reduced and there will be high peak and average sidelobes.

A conventional filled phased array will have almost all of its radiated energy within its main beam. Only a few percent of the radiated energy will appear in the sidelobes. With a highly thinned array, however, the reverse is true. Too much energy is wasted in the sidelobes. An array with 90 percent thinning might have about 90 percent of its energy in the sidelobes. If clutter is a problem, as it is in high-prf pulse doppler radars, the high sidelobes throughout space can result in high levels of clutter entering via the sidelobes.

Thinning may look attractive at first glance because it seems to allow a narrow antenna beamwidth with a reduced number of elements, but one does not usually get something for nothing. It should be attempted only with one's eyes wide open to the consequences.

9.16 SYSTEMS ASPECTS OF PHASED ARRAY RADARS

Attractive Attributes of Phased Array Radar The electronically steered phased array antenna is of interest since it can provide capabilities not readily available with other types of antennas. Its advantages are summarized below.

Interialess, rapid beam-steering. The beam from an array can be scanned, or switched from one position to another, in a time determined by the switching speed of the phase shifters. A diode phase shifter, for example, allows the beam to be switched in several microseconds or less. Ferrite phase shifters provide switching speeds that are slightly longer.

Multiple, independent beams. A single array aperture can generate independent simultaneous beams on receive, as described for the digital beam-forming phased array discussed in Sec. 9.9. On the other hand, multiple simultaneous beams on transmit are difficult to obtain, which is why a broad transmitting beam sometimes is used in conjunction with a number of contiguous narrow receive beams. For almost simultaneous tracking of many targets, a simpler method is to rapidly switch a single transmitting beam through a sequence of positions by means of a time-sequenced burst of pulses, with each pulse steered to a different direction. Rapidly acting phase shifters are needed, as well as an application that does not require a short minimum range. Since the targets are in track, their directions are known. On reception the receive beam is switched at the proper time to the direction from which the echo is expected so as to update a target already in track.

Potential for large peak and/or average power. Each element of an array can have its own individual transmitter with the outputs combined in "space" to obtain a large total power. (The power per element will be limited by the need for each individual transmitter or T/R module to fit within the space available between adjacent elements.) The active-aperture radar, Sec. 9.9, is an example. In addition to being able to achieve a large radiated power, an array with a transmitter at each element avoids the loss that can occur when the power from a single high-power transmitter has to be divided and distributed to each radiating element.

Control of the aperture illumination. Since the aperture illumination is determined by the currents at a large number of individual radiating elements across the array, a particular antenna radiation pattern is much easier to obtain with an array than with other antennas. This is important when shaped beams or very low sidelobes are desired. Separate monopulse sum and difference patterns, each with its own optimum characteristics, can also be obtained with arrays.

Adaptive processing. Adaptive arrays are designed to automatically adjust the aperture illumination to place nulls in the antenna pattern in the direction of external noise sources and/or clutter echoes. Full array adaptivity, however, in which each element of a large array is part of the adaptivity process, has been too expensive in the past to implement. The sidelobe canceler is an example of a practical adaptive system that requires the use of only a few auxiliary low-gain antennas. When

sidelobe canceling is employed in a phased array, several of the array's elements can be used as the auxiliary antennas.

Lower radar cross section. Because of its flat surface, the phased array can have a lower radar cross section than conventional reflector antennas, when illuminated by a radar with a frequency lower than the radar frequency. The flat face of a phased array can produce a large specular reflection, but the array can be tilted so that its specular scattering is directed to an angle where it is less likely to be detected. The effect of the tilt on beam steering can be compensated in the commands to the individual phase shifters. On the other hand, the radar cross section of a planar array might not be low when illuminated by a radar at a frequency higher than that for which it was designed.

Flush aperture shape. The flat surface of an array permits it to be flush mounted and to be hardened to resist the effects of blast.

Multiple functions. The agile beam-steering offered by a phased array allows a single array radar to be time-shared (sequenced) among more than one radar function.

Electronic beam stabilization. The ability to steer the beam electronically allows stabilization of the beam-pointing direction when the radar is on a ship or aircraft that is subject to roll, pitch, and/or yaw. This avoids the need for heavy mechanical stabilization machinery, but it also requires that the array be able to steer the beam over wider angles than when it is mechanically stabilized.

One other advantage sometimes claimed for a phased array is that it degrades gracefully when failures occur in the system. Since there can be many elements in the array (several thousand to several tens of thousands, or more), the effect of the failure of a few individual elements is small. It has been said[191] that for an active-aperture phased array "typically, 5 percent or more [T/R] module failures can be tolerated while maintaining acceptable performance as a multimode radar." There are several reservations that need to be mentioned, since graceful degradation is not guaranteed. First, there can be failure modes in a phased array radar that can affect a large number or all of the elements, or even be catastrophic. Second, no matter how "graceful" the failure of an array might be, sooner or later there will come a time when failures finally have to be replaced. Third, if the buyer of the radar (especially if the radar is for the government) is told by the radar company's marketing department that the radar can operate satisfactorily when a significant fraction of its elements fail, it is likely that at some time during the development of the radar—when serious overruns in money or time occur—the margin that allows for graceful degradation might be quietly removed. In spite of difficulties, graceful degradation should be designed into a phased array radar and the radar systems engineer should try to make sure it is not removed.

Limitations of Phased Arrays As with most things in life, the desirable benefits of the phased array do not come without their price—and the price is sometimes measured in more than just dollars.

Complexity The phased array radar is much more complex than a radar with a reflector antenna. In addition to being made up of many thousands of individual elements, there

must be means for insuring that the phase and amplitude at each element are what they were designed to be; and, if not, there needs to be means to readjust them to their correct value. If the phased array is to perform the functions of multiple radars, there must be more equipment behind the aperture than if only a single radar function were being performed. The antenna aperture of a phased array radar is much like the "tip of an ice berg." There is a lot more than what is normally visible. The complexity of the array also increases the problems associated with its *maintainability*, achieving high *reliability*, and assuring that its *availability* is such that it will be able to operate when needed.

Software Intensive Everything that the phased array does is commanded by a computer. The cost of the computer software to control the beam steering of the array and the various radar functions can be a significant portion of the total radar system cost, especially if the radar must perform the multiple functions of search, track, and weapon control.

Cost The cost of phased arrays was discussed in Sec. 9.14, and nothing more need be said here other than the high cost of phased array radars has limited their use to applications where the customer has been willing to pay the higher costs in order to obtain the special attributes of an array radar. For this reason, the phased array that scans in two angular coordinates has seldom been used for other than military applications.

System Limitations An advantage of a phased array is that it can be time shared to perform multiple functions. Some of the functions performed by an array, however, might take more time than is available. Three radar tasks that usually require more time than usual to accomplish properly are doppler processing, target acquisition, and long-range surveillance.

Doppler processing (as in MTI or pulse doppler radar) is used to detect moving targets in heavy clutter. As has been seen in Chap. 3, the longer the dwell time, the better can moving targets be separated from clutter. When a target is detected by the surveillance waveform of the array and is then designated for tracking, the tracking beam must be accurately directed to the direction of the target so that the beam can rapidly acquire the target without having to search for it. (A weapon control radar using mechanical trackers, on the other hand, usually can take the time—a few seconds—to search a limited angular region to find and acquire the target to be tracked. The phased array that has to perform multiple functions generally doesn't have this luxury.) Long-range air-surveillance requires the radar beam to remain in a fixed direction until the potential echoes from all ranges can return to the radar. These and other radar tasks, such as some forms of noncooperative target recognition and burnthrough against ECM jamming, can require long dwell times that could overload the scheduling of the array and cause it to omit or delay tasks of lower priority.

More than Just Firepower One of the reasons the phased array has been used for military air-defense applications is that it has been said to provide increased *firepower*. Firepower has not always been a well-defined term, but it has been defined in Webster's Ninth New Collegiate Dictionary as "the capacity to deliver effective fire on a target." In addition to firepower, a balanced air-defense system must avoid *leakage* (the penetration of

the defense by an attacker when the kill probability is too low), *saturation* (which means that the attack is so large and occurs within such a short time that the defense becomes overwhelmed and cannot engage all targets), and *exhaustion* (when the defense runs out of missiles before it runs out of attackers). The phased array radar addresses only one of these three, which is saturation.

9.17 OTHER ANTENNA TOPICS

This section considers some miscellaneous antenna topics that did not seem to fit in other sections of the chapter. They are discussed in no particular order of importance.

Mirror-Scan Antenna, or Inverse Cassegrain[192] The radiated beam of the antenna configuration shown in Fig. 9.53 can be rapidly scanned over a wide angle by mechanical movement of a light weight planar mirror called a twist reflector. This has been known by many names, including mirror-scan antenna, mirror-track antenna, polarization-twist Cassegrain, flat-plate Cassegrain, parabolic reflector with planar auxiliary mirror, and inverse Cassegrain. The parabolic reflector shown on the left in Fig. 9.53 is made up of parallel wires spaced less than a half-wavelength apart. (The wires are usually supported by a low-loss dielectric material.) For purposes of discussion, assume they are oriented vertically. The thin parallel vertical wires of the parabolic reflector make it sensitive to vertical incident polarization (when the E field is vertical). The parabola will completely reflect linear vertical polarization and be transparent to linear horizontal polarization. If the energy radiated by the feed in the center of the figure is vertically polarized (E field parallel to the vertical wires of the parabolic reflector), it will be completely reflected and directed towards a planar reflector (a mirror) called the *twist reflector*. The twist reflector has the property that it imparts a 90° rotation of the plane of polarization to the energy reflected from it. The polarization of the energy reflected from the twist reflector will then be horizontal and will pass through the parabolic reflector with negligible attenuation. The

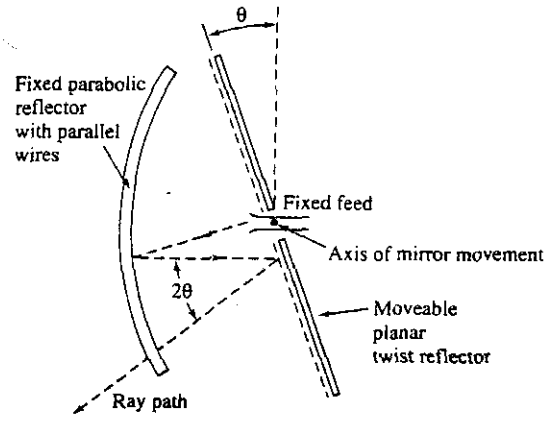


Figure 9.53 Geometry of the polarization-twist mirror-scan antenna, using a polarization sensitive parabolic reflector and a planar polarization-rotating twist-reflector. Rapid scanning of the beam in azimuth and elevation is accomplished by mechanical movement of the lightweight planar twist-reflector.

radiated beam is steered in angle by mechanically rotating the low inertia twist reflector. When the twist reflector is rotated by an angle θ, the radiated beam is rotated through an angle 2θ. The beam can be rapidly scanned over an angle of $\pm$ 90° without the need for microwave rotary joints.

One method of making a twist reflector is to orient a grating of thin wires 45° to the incident polarization and placed a quarter-wavelength in front of the planar reflecting surface. This type of construction is limited to a bandwidth of about 10 percent. Much broader bandwidth is possible with other constructions. A meanderline polarizer backed by a reflector surface, for example, can achieve an octave bandwidth.[193] A mirror-scan antenna with a twist reflector using a number of log-periodic layered structures demonstrated operation over a frequency range from 2 to 12 GHz.[194]

The mirror-scan antenna has been widely used by many countries, especially the former Soviet Union, for land-based, shipborne, and airborne radar applications.

Beam Steering of a Reflector Antenna by Movement of the Feed The beam of a parabolic reflector antenna can be scanned by laterally displacing the feed from the focus of the antenna. Generally the beam can only be scanned a few beamwidths off axis before the antenna pattern degrades significantly.[195,196] The antenna gain decreases and the sidelobes increase; and in radar applications it is often the increase in sidelobe level that determines how far the beam can be scanned rather than the decrease in gain. The larger the f/D ratio (f = focal length and D = antenna diameter) the greater in angle the beam can be scanned, but it is still quite limited.

A *spherical reflector* can produce a slightly larger scan angle when positioning the feed off the focal point, but spherical reflectors have aberrations that cause high sidelobes. The Arecibo antenna in Puerto Rico used for radar and radio astronomy employs a large spherical reflector 1000 ft in diameter. It has a specially designed feed that corrects for aberrations so that the UHF beam can be steered about 20° off axis.

A *parabolic torus* is a reflector antenna that is generated by rotating the parabolic section of Fig. 9.7 over an arc of a circle whose center is on the axis of the parabola.[197,198] Thus the vertical profile of the parabolic torus is a parabola and its horizontal profile (the plane in which the beam is scanned) is a circle of radius r. The radius r of the circular contour of the torus is made large enough so that the portion of the antenna surface illuminated by the feed does not appreciably differ from a true parabola. In other words, even though the horizontal contour of the torus is circular, the portion of the aperture illuminated by the feed approximates a parabolic shape, which is why the antenna generates satisfactory radiation patterns. Only a portion of the reflector surface is illuminated at any beam position. The main use of a parabolic torus is to rapidly scan a beam in a single plane by either mechanically moving a single feed along a circle whose radius is $r/2$, or by switching among many fixed feeds located on the circle of radius $r/2$. The latter method of scanning was used with the original antenna for the Ballistic Missile Early Warning System (BMEWS) radar, a high-power UHF radar for the detection of intercontinental ballistic missiles at ranges over 2000 nmi. Its beam was mechanically scanned 120° in azimuth in two seconds. The BMEWS antenna was 165 ft high by 400 ft wide, and although it is a large antenna, its cost was relatively low since it was a fixed structure and only the feeds were switched by an organ-pipe scanner.[199] The BMEWS parabolic torus

antennas have been replaced by a phased array based on the technology of the Pave Paws radar.

The parabolic torus also was used in the past for naval height-finder radars, such as the AN/SPS-30, which rapidly scanned a horizontal fan beam in the vertical plane to extract an aircraft target's elevation angle.

The torus reflector antenna also has been used to generate two spaced antenna beams separated in azimuth, in addition to the main antenna beam, for the AN/SPQ-9B shipboard radar which uses a single mechanically rotating reflector, Fig. 9.54. The purpose of the two additional azimuth beams is to allow confirmation looks on a target after detection by the main beam so as to more rapidly establish a track on a target. After a single pass by the target, this allows the air-defense system to begin to establish a track on the target.

It has been reported[200] that a torus-like reflector with an elliptical contour rather than a circular one (but with a parabolic contour in the orthogonal plane) can have wide-angle scanning with good performance and yet be much smaller than the conventional parabolic torus discussed in the above. With a modified contour in the plane of scan, this approach has also been demonstrated to scan the beam using an offset-fed reflector to avoid aperture blockage.[201]

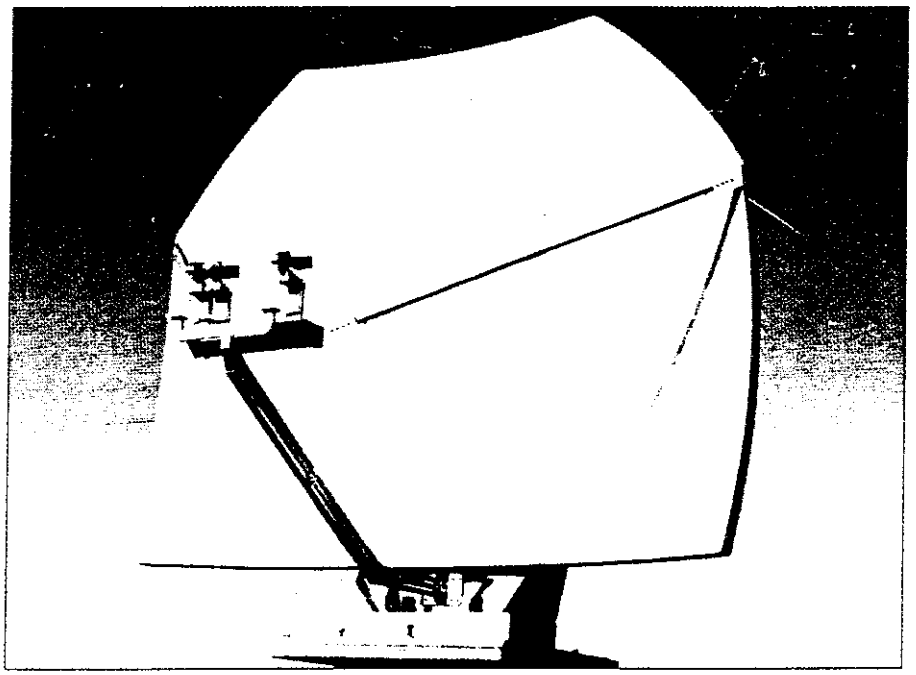


Figure 9.54 Parabolic torus reflector used with the AN/SPQ-9B shipboard radar to obtain three beams separated in azimuth.
(Provided by L. Leibowitz and B. Cantrell of the Naval Research Laboratory Radar Division.)

Lens Antennas[202] Microwave lenses may be constructed from dielectric materials, artificial dielectrics, or metal plates (waveguide media) to cause a focusing action similar to that of an optical lens. Dielectric lenses (the microwave analogy of the optical lens) are generally heavy and difficult to obtain with uniform properties. Lenses constructed from artificial dielectrics[203] can be made lighter but are usually poor conductors of heat so that it might be difficult to dissipate the heat generated in such materials when they are used for high-power radar transmitting antennas.

The metal-plate waveguide lens is constructed from side-by-side parallel-plate waveguides.[204] The phase velocity in a parallel-plate waveguide is greater than that in free space so that its index of refraction (and its dielectric constant) is less than unity, which is why they can be used to make lenses. They were used in early monopulse tracking radars to avoid aperture blockage caused by the large feed systems needed with reflector antennas.

The Luneburg lens[205] differs from optical lenses and other microwave lenses in that it is spherical and its index of refraction η is not uniform, but varies with distance from the center of the sphere as

$$\eta = \left[2 - (r/r_0)^2\right]^{1/2} \qquad [9.70]$$

where r is the radial distance and r_0 is the radius of the lens. It has the property that a plane wave incident on the sphere is brought to a focus on the surface at the diametrically opposite side. Likewise, a transmitting point source on the surface of the sphere emerges as a plane wave on passing through the lens. The beam of a Luneburg lens can be steered by moving the feed along the surface of the lens. With multiple feeds, it can form multiple simultaneous beams which was why it was seriously considered as the antenna for the Nike Zeus ballistic missile defense system, an early intercontinental ballistic missile defense system concept conceived in the late 1950s by Bell Telephone Laboratories for the U.S. Army (but never reached deployment).

The Fresnel zone-plate lens is an interesting form of lens that is simple, of small thickness, light weight, low loss, and low cost, especially for use at millimeter waves.[206] It has not had significant microwave radar application, however.

The lens antenna generally is less efficient than a reflector since unwanted reflections occur from both the front and rear surfaces of the lens. It is not as easy to support mechanically as is a reflector. Dielectric lenses have problems in dissipating heat when radiating high power. The lens antenna has interesting attributes, but it is not often used for radar application.

Radomes[207–209] Mechanical engineers can design a ground-based or shipborne antenna to be structurally strong enough to operate in high winds, icing, and other adverse weather conditions. It is often much cheaper and better, however, to enclose the antenna in an electromagnetically transparent protective shield called a *radome*. An antenna enclosed by a radome can be lighter and have a smaller drive motor than an antenna exposed to the elements.

Radomes for ground-based radars are often in the shape of a sphere (for example, a three-quarters sphere). The sphere is a good mechanical structure and offers aerodynamic advantages in high winds. Precipitation particles blow around a sphere rather than

impinge upon its surface, so that snow or other frozen particles are not readily deposited. Antennas mounted on aircraft must be housed within a radome that does not interfere with the aerodynamics of flight, be strong enough to be a part of the aircraft's structure, as well as not distort the antenna pattern.

A radome with good electromagnetic properties should be of low loss, have an adequate bandwidth, not raise the sidelobe level significantly, provide a low VSWR, a low antenna noise temperature, and not cause the boresight (pointing direction) to shift. In highly accurate tracking radars, the radome must also not increase the rate of change of the shift in the boresight.

There are two types of radomes that have been used for ground-based and shipboard radars: the rigid self-supporting radome and the air-supported radome.

Rigid Radomes An example of a rigid radome is the *space frame,* an example of which is shown in Fig. 9.55. This type of radome consists of a three-dimensional lattice of primary load-bearing structural members enclosed with thin dielectric panels (0.02–0.04 inch thick). The panels can be made of teflon-coated fiberglass and can be very thin, even for large radomes, since they do not carry the main loads or stresses. This type of construction whereby a spherical structure is constructed from flat panels of simple geometric shapes is sometimes called a *geodesic dome.* The supporting framework can be of steel, aluminum, or plastic. Metal structures are superior in electrical performance compared to plastic or fiberglass since they can be made smaller because of their increased strength. Smaller thickness means less aperture blocking. Metal space frames are generally cheaper

Figure 9.55 Rigid space-frame radome for the Argos-10 air-surveillance radar.
(Courtesy Alfonso Farina and Alenia Marconi Systems.)

and easier to fabricate, transport, and assemble, and can be used for larger diameter configurations.

Aluminum structural members are typically used for the space frame. They are larger than steel of equivalent strength, but they are lightweight, noncorrosive, and require no maintenance. It is important that the plastic panels be able to repel water (hydrophobic) rather than absorb water. A hydrophobic material will cause the water on the radome surface to form into beads rather than coat the surface in a sheet or a film. Water in the form of beads usually has dimensions small compared to the radar wavelength and do not cause as much of an effect as a water film.[210] In some radomes, the exterior surface is coated with a white radar-transparent paint, such as Hypalon, to reduce the interior temperature rise caused by solar radiation. A material known as Tedlar has also been used as a surfacing material.

A metal space-frame radome might be made up of individual triangular panels with a relatively uniform pattern. Instead of panels of uniform shape, there can be a quasi-random selection of different panel sizes and shapes to minimize the periodicity of the structure and avoid the generation of spurious sidelobes obtained with a periodic structure. The randomization of the space frame also makes it less sensitive to polarization. A metal space-frame radome might typically have a transmission loss of 0.5 dB and cause the antenna sidelobes to increase an average of 1 dB at the −25 dB level. The boresite might be shifted less than 0.1 mrad and the antenna noise temperature might increase less than 5 K. Space-frame radomes can be larger than 150 ft in diameter, and they can be used at any microwave radar frequency. Typically they can be designed for wind speeds of 150 mph, but they have also been designed to withstand winds as high as 300 mph.

Rigid radomes also have been made of fiberglass-reinforced solid plastic laminates and as a rigid shell sandwich, but these are usually of smaller size than the space frame. Solid laminate radomes can be up to 35 ft in diameter and operate at frequencies up to 3 GHz. Panels are sometimes arranged in an "orange peel" geometry. Large-diameter sandwich radomes, which can be as large as 80 ft, are usually constructed with an A-type sandwich (described later), but are often narrowband.[207] Some *L*-band sandwich radomes for long-range FAA enroute ARSR radars have been said to have transmission efficiencies of 98 percent (less than 0.1 dB).[211]

Weather Effects on Rigid Radomes An important advantage of the rigid radome is its ability to withstand the rigors of severe climate. Rime ice, the prevalent type of icing in the Arctic region, has little or no effect on most radomes. Although it tends to collect on many types of structures and can obtain large thickness, both theory and experiment show a lack of rime-ice formation on a spherical radome.[212] The trajectories of water droplets in the air stream flowing around a large spherical radome do not impinge upon the surface. Droplets of freezing rain, on the other hand, are large and almost 100 percent of freezing rain can collect on the radome's surface. Dry snow does not stick to cold surfaces and is generally not a problem. Wet snow can stick to the radome and affect its transmission properties. Removal of snow by thermal means is expensive; but on the smaller radomes snow can be removed mechanically by tying a rope at the top of the radome and having someone walk the rope around the radome to knock off the snow.

Liquid water can collect on a radome as a film due to condensation or rainfall if the surface is non-hydrophobic. The films can be very thin and still cause attenuations of

several dB or more depending on the frequency and the amount of water. Effenberger et al.[213] state that at *X* band a water film can result in 8 dB of added transmission loss. If the water is in the form of droplets instead of a film, the added transmission loss is reduced to 1 dB. These losses compare to a loss of a dry air-supported radome of 0.1 dB and a dry metal space frame of 0.7 dB. Thus it is important that materials used for a radome not absorb water and not allow water to form a film on its surface. Radomes should be checked periodically to insure that their hydrophobic properties do not weaken with time.

Air-Supported Radomes The first large radomes for ground-based radar antennas that appeared shortly after World War II were constructed of a strong, flexible rubberized airtight material supported by air pressure from within. Since the radome material is relatively thin and uniform, it approximates the electrically thin shell that provides very low loss (less than 0.1 dB) and very small boresight error. The air-supported radome is constructed of gore-shaped fabric sections with the gores or seams in the vertical direction. Reliable operation depends on the use of uninterruptable power supplies and redundant blower systems. Teflon fiberglass is a commonly used material.[210] Air-supported radomes can be folded into a small package which makes them suitable for transportable radars requiring mobility and quick assembly and disassembly times. At a prepared site, a 50-ft air-supported radome can be installed in about one or two hours.[214] They are also of interest when wideband operation is important.

The life of an air-supported radome is limited by exposure to ultraviolet light, surface erosion, and the constant flexing of the material in the wind. The life can be increased by the use of better materials such as neoprene-coated nylon and teflon-coated fiberglass. In high winds the material can be damaged by flying debris. The rotation of the antenna might have to be stopped to prevent the fabric from being blown against the antenna and torn. Nevertheless, these structures are designed to withstand winds of 100 mph, and in special applications it has been extended to 200 mph.[215] Maintaining the internal pressure in high winds can sometimes be difficult. Frequent and costly maintenance is another problem. The air-supported radome has superior electrical properties to the rigid spaceframe radome, but the latter is much more rugged.

Aircraft Radomes[208,216] The shape of the radome used in the nose of a military fighter/attack aircraft is determined primarily by aerodynamic requirements rather than by electromagnetics. It is often an ogive or some other similar conical shape. Because of the aerodynamic shape of the nose radome, the angle of incidence of the beam on the surface of the radome depends on the scan angle and might be from 0 to 80°, or more. Also, the incident polarization can vary with angle of scan. This is unlike the situation with spherical shaped ground-based radomes where the incidence angle and polarization are relatively independent of the scan angle. Thus there can be distortions in the antenna pattern, spurious sidelobes, and angle errors that depend on the angle of scan. Airborne radomes, especially those used at supersonic speeds, can be subjected to mechanical stress and aerodynamic heating so severe that the electromagnetic requirements of a radome made of dielectric materials might have to be sacrificed in order to obtain sufficient mechanical strength to survive. Rain impinging on an unprotected radome in flight can cause

structural damage within minutes due to erosion at high subsonic speeds. Rain erosion is reduced by coating the radome surface with a shock-absorbing rubber-like material such as neoprene or polyurethane elastomer. A lightning strike can puncture a hole in the radome wall and severely damage the radar equipment inside. The damaging effects of lightning are reduced by placing conductors on the external surface of the radome to divert the lightning to these conductors rather than puncture the radome wall. If a large number of conductors is necessary to insure that lightning is diverted, the performance of the radar might be affected. Other factors in radome design are avoidance of static charge build-up and damage due to bird strike or the impact of hail. The radome must, in addition, withstand high temperatures and not be too heavy. It has been said[216] "Few, if any, other components [of an airborne radar system] have such a variety of conflicting requirements as does a radome."

Radome Wall Construction The following are some wall constructions that have been used in radomes:

Thin wall. The wall is electrically thin compared to the radar wavelength. If the wall physical thickness is d, with a dielectric constant ϵ, and a wavelength λ, then a thin wall radome is characterized by $d < 0.05\lambda/\sqrt{\epsilon}$. A thin wall radome has good electrical properties, but it can be weak structurally.

Half wavelength. This is a solid dielectric surface whose electrical thickness is approximately a half-wavelength. Theoretically, the halfwave-thick surface is nonreflecting and has no loss other than ohmic losses of the material. It is of limited bandwidth, as well as limited in the range of incidence angles over which the electromagnetic energy can be transmitted with minimal reflection. There can be multiple half-wave surfaces.

A-sandwich. A three-layer wall consisting of two thin relatively high-dielectric-constant skins separated by a low-dielectric-constant core whose thickness is approximately one-quarter wavelength. The skins are glass reinforced plastic laminates and are thin compared to a wavelength. The core might be a honeycomb or a foam.

B-sandwich. The inverse of the A-sandwich with quarter-wavelength skins having a dielectric constant lower than that of the core.

C-sandwich. Two back-to-back A-sandwiches. It is used when the ordinary A-sandwich does not provide sufficient strength.

Multilayer. A general term for more layers than that of the C-sandwich.

Metallic Radomes A thin metal sheet with periodically spaced openings (such as slots) has a bandpass characteristic and can be used for radomes. The metallic structure overcomes the mechanical limitations of radomes made of dielectric materials, yet has good electrical properties.[217] It also is better able to distribute frictionally induced heating and better able to withstand the stresses caused by rain, hail, dust, and lightning. This type of radome is a *frequency selective surface* that reduces interference because of its bandpass characteristic. It also reduces the nose-on radar cross section of the aircraft when viewed by radar systems not within its own passband.[218] The penalty for its use, however, is

narrow bandwidth. A frequency selective surface can be more sophisticated and have better properties than a thin metallic sheet with simple periodic slots.[219]

Rotodomes In some cases the radome is made to rotate in synchronism with the antenna. It is then called a *rotodome*. An example is the AWACS AN/APY-1 antenna, Fig. 3.45b, where the antenna and radome are designed together so that the antenna can maintain the very low sidelobes required of a high-prf pulse doppler radar. A rotodome is also used with the ASDE-3, see Fig. 1.9, the Airport Surface Detection Equipment found at the top of the control tower of major airports for monitoring taxiing aircraft and ground traffic at airports.

Adaptive Antennas[220–224] An adaptive antenna, usually an array, is one that senses the received signals incident across its aperture and adjusts the phase and amplitude of the aperture illumination to maximize the signal-to-external-noise ratio or signal-to-clutter ratio. Adaptive arrays usually require some prior knowledge of the desired signal and the nature of the noise or clutter to be rejected. When one speaks of a *fully adaptive array* it means one where each element of the array is part of the adaptive process. There have been many investigations of the theory and algorithms to perform adaptive array processing, but the difficulty and cost of implementing fully adaptive processing at each element of a large array antenna has been prohibitive. Until large fully adaptive arrays become more practical, they will be mainly of academic interest. There are, however, two important radar applications where adaptive array technology has proven to be practical and important. These occur when only a relatively few adaptive elements are needed. One such application is the *sidelobe canceler;* the other is the AEW radar that employs *space-time adaptive processing.*

Sidelobe Cancellation[225] This adaptively places nulls in the antenna radiation pattern in the direction of a limited number of noise jammers so as to reject the noise before it enters the receiver. A small number of omnidirectional (or wide beamwidth) auxiliary antenna elements are placed on or in the near vicinity of the main radar antenna. Typically there might be from 3 to 6 auxiliary elements used for the sidelobe canceler. The main radar antenna can be a reflector or a phased array. In theory, one auxiliary element (one degree of freedom) can create one adaptive null; but, in practice, especially when multipath propagation occurs, two auxiliary elements might be needed in some instances to place a suitable null in the direction of one noise source. The sidelobe canceler has been a successful application of adaptive antennas and has been applied operationally to a number of radar systems.

Space-Time Adaptive Processing (STAP) Military airborne MTI radars, or Airborne Early Warning (AEW) radars, must be able to cancel ground clutter echoes which have a non-zero doppler velocity with respect to the moving radar, along with hostile jamming that enters the radar via the antenna sidelobes. According to J. Ward[226] "Space-time adaptive processing (STAP) refers to multidimensional adaptive filtering algorithms that simultaneously combine the signals from the elements of an array antenna and the multiple pulses of a coherent radar waveform, to suppress interference and provide target

detection." Fred Staudaher,[227] of the Naval Research Laboratory, described the differences between spatial processing, temporal processing, and space-time processing as follows: "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 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."

STAP and other forms of antenna adaptive processing are more subjects in circuit design and algorithm development than antennas.

The Quest for Superresolution The ability of a radar to resolve two targets in angle depends on their relative radar cross section, signal-to-noise ratio, antenna beamwidth, the phase difference between the two signals, and the criterion used to establish resolution. It has been generally accepted that two equal targets can be resolved in angle when they are separated by eight-tenths of a beamwidth, provided the signal-to-noise ratio is large enough for good detection. Resolution can be better than this with high signal-to-noise ratios and a criterion which acknowledges that there can be a phase difference between the two signals. Every now and then, however, there have been different proposals for obtaining better angular resolution with radar systems—all without true success thus far!

A technique that has had impressive claims made for it is known as *spectral estimation* or *superresolution*.[228] It is also known by some of the many algorithms that are used, such as the maximum entropy method, autoregression, Burg algorithm, and others. These angular resolution methods apply for noncoherent sources, such as independent noise radiators (jammers or radio stars). Superresolution methods are basically the same as adaptive antennas which place sharp nulls in the direction of noise sources. Superresolution and adaptive antennas use the same algorithms and the same hardware, and the plots of their outputs are the same except that one is plotted upside down with respect to the other. That is, the nulls of the adaptive antenna become the narrow spikes of superresolution when inverted.

Superresolution may resolve closely spaced noise-like sources, but it does not reliably resolve the echoes from multiple targets illuminated by the same radar. The echoes from targets illuminated by the same radar have a phase relationship among each other and are thus coherent. Superresolution, or spectral estimation, algorithms employ nonlinear mathematical operations. When multiple echo signals from the same radar are subject to nonlinear processing, they can produce spurious signals that do not allow good resolution capabilities. Thus superresolution does not provide improved angular resolution with radar echo signals. This was first stated in the radar literature by A. W. Rihaczek.[229]

Other "Superresolution" Concepts In the traditional antenna literature, one can often find discussed the concept of a "superdirective" array antenna (formerly called supergain).[230] This is defined as an array antenna with a directivity higher than that obtained when the same antenna has a uniform aperture illumination. (Its antenna illumination efficiency is said to significantly exceed 100 percent.) Many reasons have been offered why such an antenna is not practical (narrow bandwidth, high Q, large aperture currents, high loss, and

extremely precise tolerances), but supergain antennas require aperture illuminations that must change amplitude (spatially) across the aperture faster than can be expected of a signal operating at the given RF frequency. Superdirectivity appears to result from the simple algebraic models of an antenna pattern rather than as a solution to Maxwell's equations applied to a real radar antenna.

In the 1960s the *multiplicative array* was the "superresolution" technique that caused excitement, for a while. In a multiplicative array the outputs of the individual radiating elements were combined in a nonlinear manner rather than linearly. The nonlinear manipulation of the aperture illumination results in an apparently narrower antenna pattern. For example, the pattern of an N-element array can be expressed by a polynomial of degree N. If the output of one half of the array is multiplied by the output from the other half, the resulting expression is a polynomial of order $(N^2/4)$, which when plotted appears as a much narrower pattern of an array with $N^2/4$ elements rather than N. This narrower pattern looks exciting as a means for obtaining improved resolution by simple multiplicative processing of the array output. When two closely spaced targets are examined, however, with such multiplicative processing the resolution is not that of a larger array and, to make matters worse, there will be spurious signals generated because of the nonlinear mathematical operation. (The nonlinear operation of squaring or cubing the antenna pattern also will make it narrower and appear to provide better resolution than a conventional antenna, but when multiple targets are present, the resulting pattern is not the superposition of the individual patterns, but much worse.)

There have also been attempts to achieve improved resolution by what was called "data restoration," which smoothed the received aperture illumination and extrapolated it beyond the physical confines of the antenna. This also did not produce the significant improvement in resolution that was desired.

Thus one should not expect to obtain significantly better angular resolution with an antenna by some nonlinear form of antenna processing. There seems to be no magic radar resolution algorithm. The only method that has worked in the past for obtaining improved resolution when the electrical size of the aperture cannot be increased is to increase the signal-to-noise ratio and recognize in the resolution procedure that the echoes signals can be of different phase.

Microelectromechanical Switches in Phased Arrays[232,233] A microelectromechanical switch, or MEMS, is a small, low-inertia fast-acting switch activated by an electrostatic field. The switching mechanism may be in the form of a cantilever, rotary, or membrane configuration. In one example, the upper contact is a 0.3-μm aluminum membrane suspended across polymer posts. This suspended membrane is 4 μm above a substrate surface with the bottom contact of 0.7-μm gold or aluminum metal layer. On top of this metal layer is a thin dielectric layer, typically 0.1 μm (1000 Å) of silicon nitride. It is not a metallic contact switch, but switches by providing a change in capacitive impedance. The dielectric on the bottom part of the switch makes contact with the metallic portion of the suspended membrane, which eliminates the problem of striction that would occur if two metallic layers came into contact.

MEMS can have very wide bandwidth and can be made to operate with signals from a few MHz to 40 GHz. Measured insertion loss is less than 0.2 dB per switch. The switch

can activate in 2 to 5 μs and it can handle RF power up to 10 watts. The pull-in voltage is from 10 to 30 volts. Since a MEMS is activated by a d-c electrostatic field, no d-c current is required and power consumption is small. The energy required to activate a switch is on the order of 10 nJ.

The MEMS can be used in the same way that diode switches are used in the description of the digital phase shifter, shown in Fig. 9.17, to switch in and out fixed length of lines to obtain various phase shifts. In this type of *X*-band phase shifter the loss in a four-bit phase shifter is from 1.2 to 2 dB. Its size is approximately 6 mm by 9 mm. These phase shifters can be fabricated on silicon wafers. Hundreds of phase shifters can be built on a single 8-in. wafer, which makes them inherently low cost.

It has been said[234] that the MEMS for electronically scanned phased array radars has the potential to reduce the cost, weight, and power consumption for such systems when the array size exceeds 10,000 elements.

REFERENCES

1. Kraus, J. D. *Antennas,* 2nd ed. New York: McGraw-Hill, 1988, Sec. 10–12.
2. Probert-Jones, J. R. "The Radar Equation in Meteorology." *Quart. J. Roy. Meteor. Soc.* 88 (1962), pp. 485–495.
3. Stutzman, W. L. "Estimating Directivity and Gain of Antennas." *IEEE Antennas and Propagation Magazine* 40 (August 1998), pp. 7–11.
4. Evans, G. E. *Antenna Measurement Techniques,* Artech House, Boston, MA, 1990, p. 115.
5. Cutler, C. C., A. P. King, and W. E. Kock. "Microwave Antenna Measurements." *Proc. IRE* 35 (December 1947), pp. 1462–1471.
6. Sherman, J. W. "Aperture-Antenna Analysis." *Radar Handbook,* 1st ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1970, Chap. 9.
7. Mints, M. Ya., Ye. D. Prilepskiy, and V. M. Zaslonko. "Optimization of the Radiation Power Concentration Factor of an Antenna with a Circular Aperture and a Maximally Flat Radiation Pattern." *Soviet J. of Communications Technology and Electronics* 34 (May 1989), pp. 33–39.
8. Silver, S. *Microwave Antenna Theory and Design,* vol. 12 of the M. I. T. Radiation Laboratory Series. New York: McGraw-Hill, 1949, Chap. 6.
9. Silver, S. Ref. 8, Sec. 6.5.
10. Sherman, J. W. Ref. 6, Sec. 9.2
11. Bodnar, D. G. "Materials and Design Data." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, 1993, Chap. 46, Sec. 46–5.
12. Jasik, H. "Fundamentals of Antennas." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.), New York: McGraw-Hill, 1993, Chap. 2, Sec. 2.7.

13. Johnson, R. C. *Designer Notes for Microwave Antennas.* Boston: Artech House, 1991, Sec. A.12.

14. Cutler, C. C. "Parabolic Antenna Design for Microwaves." *Proc. IRE* 35 (November 1947), pp. 1284–1294.

15. Sciambi, A. F. "The Effect of the Aperture Illumination on the Circular Aperture Antenna Pattern Characateristics." *Microwave J.* 8 (August 1965), pp. 79–84.

16. Olver, A. D., P. J. B. Clarricoats, A. A. Kishk, and L. Shafai. *Microwave Horns and Feeds.* New York: IEEE Press, 1994.

17. Ruze, J. "Feed Support Blockage Loss in Parabolic Antennas." *Microwave J.* 11 (December 1968), pp. 76–80.

18. Kildal, P-S, E. Olson, and J. A. Aas. "Losses, Sidelobes, and Cross Polarization Caused by Feed-Support Struts in Reflector Antennas: Design Curves." *IEEE Trans.* AP-36 (February 1988), pp. 182–190.

19. Rudge, A. W., and N. A. Adatia. "Offset-Parabolic-Reflector Antennas: A Review." *Proc. IEEE* 66 (December 1978), pp. 1592–1618.

20. Cook, J. H., Jr. "Earth Station Antennas." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.), New York: McGraw-Hill, 1993, Chap. 36, pp. 36-8 to 36-10.

21. Terada, M. A., and W. L. Stutzman. "Design of Offset-Parabolic-Reflector Antennas for Low Cross-Pol and Low Sidelobes." *IEEE Antennas and Propagation Magazine* 35, no. 6 (December 1993), pp. 436–449.

22. Terada, M. A., and W. L. Stutzman. "Computer-aided Design of Reflector Antennas." *Microwave J.* 38 (August 1995), pp. 64–73.

23. Weiss, H. G. "The Haystack Microwave Research Facility." *IEEE Spectrum* 2 (February 1965), pp. 50–59.

24. Hannan, P. W. "Microwave Antennas Derived from the Cassegain Telescope." *IRE Trans.* AP-9 (March 1961), pp. 140–153.

25. Josefsson, L. G. "A Broad-Band Twist Reflector." *IEEE Trans.* AP-19 (July 1971), pp. 552–554.

26. Lewis, B. L., and J. P. Shelton. "Mirror Scan Antenna Technology." *Record of the IEEE 1980 International Radar Conf.,* Arlington, VA, pp. 279–283. IEEE Publication 80CH1493-6 AES.

27. Dwight, H. B. *Tables of Integrals and Other Mathematical Data.* New York: Macmillan, 1947, Equation No. 420.3.

28. Bickmore, R. W. "A Note on the Effective Aperture of Electronically Scanned Arrays." *IRE Trans.* AP-6 (April 1958), pp. 194–196.

29. Elliott, R.S. "The Theory of Antenna Arrays." *Microwave Scanning Antennas,* R. C. Hansen, (Ed.). New York: Academic, 1966, vol. II, Chap. 1.

30. Koul, S. K., and B. Bhat. *Microwave and Millimeter Wave Phase Shifters, Vol. II, Semiconductor and Delay Line Phase Shifters.* Boston: Artech House, 1991.

31. Garver, R. V. *Microwave Diode Control Devices.* Boston: Artech House, 1976.

32. Tang, R., and R. W. Burns. "Phased Arrays." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, 1993, Chap. 20, pp. 20-36 to 20-44.

33. White, J. F. *Microwave Semiconductor Engineering.* New Jersey: Van Nostrand, 1982.

34. Temme, D. H. "Diode and Ferrite Phaser Technology." *Phased Array Antennas,* A. A. Oliner and G. H. Knittel (Eds.). Boston: Artech House, 1972, pp. 212–218.

35. Koul, S. K., and B. Bhat. Ref. 30, Chap. 11.

36. Andricos, C., I. J. Bahl, and E. L. Griffin. "C-Band 6-Bit Monolithic Phase Shifter." *IEEE Trans.* MTT-33 (December 1985), pp. 1591–1596.

37. Koul, S. K., and B. Bhat. Ref. 30, Chap. 12.

38. Shenoy, R. P. "Phased Array Antennas." *Advanced Radar Techniques and Systems,* G. Galati (Ed.). Peter Peregrinus, 1993, Chap. 10, Sec. 10.27.

39. Stark, L. "Microwave Theory of Phased Array Antennas—A Review." *Proc. IEEE* 62 (December 1974), pp. 1661–1701.

40. Koul, S. K., and B. Bhat. *Microwave and Millimeter Wave Phase Shifters, Vol. I, Dielectric and Ferrite Phase Shifters.* Boston: Artech House, 1991.

41. Rodrique, G. P. "A Generation of Microwave Ferrite Devices." *Proc. IEEE* 76 (February 1988), pp.121–137.

42. Stark, L., R. W. Burns, and W. P. Clark. "Phase Shifters for Phased Arrays," *Radar Handbook,* 1st ed., M. Skolnik (Ed.), New York: McGraw-Hill, 1970, Chap. 12.

43. Whicker, L. R. (Ed.). *Ferrite Control Devices, Vol. 2, Ferrite Phasers and Ferrite MIC Components.* Boston: Artech House, 1974.

44. Wicker, L. R., and R. R. Jones. "A Digital Current Controlled Latching Ferrite Phase Shifter." *IEEE 1965 International Convention Record,* pt. V, pp. 217–223.

45. Cattgasrin, G., et al. "A Digital Ferrite Phase-Shifter for High Power S-Band Operation." *Rivista Tecnica, Selenia* 8, no. 2 (1982), pp. 29–34.

46. Hord, W. E. "Microwave and Millimeter-Wave Ferrite Phase Shifters." *Microwave J.* 1989 State of the Art Reference, pp. 81–93.

47. Ince, W. J., and E. Stern. "Nonreciprocal Remanence Phase Shifters in Rectangular Waveguide." *IEEE Trans.* MTT-15 (February 1967), pp. 87–95.

48. Koul, S. K., and B. Bhat. Ref. 40, Sec. 4.8.

49. Junding, W., et al. "Analysis of Twin Ferrite Toroidal Phase Shifter in Grooved Waveguide." *IEEE Trans.* MTT-42 (April 1994), pp. 616–621.

50. DiBartolo, J., W. J. Ince, and D. H. Temme. "A Solid State 'Flux Drive' Control Circuit for Latching-Ferrite-Phaser Applications." *Microwave J.* 15 (September 1972), pp. 59–64.

51. Koul, S. K., and B. Bhat. Ref. 40, Sec. 5.6.

52. Fox, G. A., S. E. Miller, and M. T. Weiss. "Behavior and Application of Ferrites in the Microwave Region." *Bell System Tech. J.* 34 (January 1955), pp. 5–103.

53. Fox, A. G. "An Adjustable Wave-Guide Phase Changer." *Proc. IRE* 35 (September 1947), pp. 1489–1498.

54. Yansheng, X., and J. Zhengchang. "Dual-Mode Latching Ferrite Devices, Part I." *Microwave J.* 29 (May 1986), pp. 277–280.

55. Whicker, L. R., and C. W. Young, Jr. "The Evolution of Ferrite Control Components." *Microwave J.* 21 (November 1978), pp. 33–37.

56. Ince, W. J. "Recent Advances in Diode and Ferrite Phaser Technology for Phased-Array Radars, Part II." *Microwave J.* 15 (October 1972), pp. 31–36.

57. Yansheng, X., and J. Zhengchang. "Dual-Mode Latching Ferrite Devices, Part II." *Microwave J.* 29 (May 1986), pp. 282–286.

58. Monaghan, S. R., and M. C. Mohr. "Polarization Insensitive Phase Shifter for Use in Phased-Array Antennas." *Microwave J.* 12 (December 1969), pp. 75–80.

59. Hord, W. E. "Design Considerations for Rotary-Field Ferrite Phase Shifters." *Microwave J.* 31 (November 1988), pp. 105–115.

60. Boyd, C. R., Jr. "Progress in Ferrite Phase Shifters," Microwave Applications Group brochure, Santa Maria, CA (no date).

61. Varadan, V. K. "A Novel Microwave Planar Phase Shifter." *Microwave J.* 38 (April 1995), pp. 244–254.

62. Ajioka, J. S. "Frequency-Scan Antennas." *Antenna Engineering Handbook*, R. C. Johnson (Ed.), New York: McGraw-Hill, 1993, Chap. 19.

63. Hammer, I. W. "Frequency-Scanned Arrays." *Radar Handbook*, 1st ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1970, Chap. 13.

64. Begovich, N. A. "Frequency Scanning." *Microwave Scanning Antennas, Vol. III*, R. C. Hansen (Ed.). New York: Academic, 1966, Chap. 2.

65. Johansson, F. S., L. G. Josefsson, and T. Lorentzon." A Novel Frequency-Scanned Reflector Antenna." *IEEE Trans.* AP-37 (August 1989), pp. 984–989.

66. Johansson, F. S. "Frequency-Scanned Gratings Consisting of Photo-Etched Arrays." *IEEE Trans.* AP-37 (August 1989), p. 996–1002.

67. Croney, J. "Doubly Dispersive Frequency Scanning Antenna." *Microwave J.* 6 (July 1963), pp. 76–80.

68. Tang, R. "Practical Aspects of Phased Array Design." *Antenna Handbook*, Y. T. Lo and S. W. Lee (Ed.). New York: Van Nostrand Reinhold, 1988, Chap. 18, pp. 18-6 to 18-11.

69. Hansen, R. C. *Phased Array Antennas*. New York: John Wiley, 1998, Chap. 5.

70. Hansen, R. C. Ref. 69, Sec. 8.2.

71. Stark, L. "Comparison of Array Element Types." *Phased Array Antennas, Proc. 1970 Phased Array Antenna Symp.* Dedham, MA: Artech House, 1972, pp. 51–67.

72. Tang, R., and R. W. Burns. "Phased Arrays." *Antenna Engineering Handbook*, 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, 1993, Chap. 20.

73. Edward, B., and D. Rees. "A Broadband Printed Dipole with Integrated Balun." *Microwave J.* 35 (May 1987), pp. 339ff.

74. Mailloux, R. J. *Phased Array Antenna Handbook.* Boston: Artech House, 1994.

75. Richard, W. F. "Microstrip Antennas." *Antenna Handbook,* Y. T. Lo and S. W. Lee (Eds.). New York: Van Nostrand Reinhold, 1988, Chap. 10.

76. Lewis, L. R. "Phased Array Elements—Part 2." In *Practical Phased Array Antenna Systems,* E. Brookner (Ed.). Boston: Artech House, 1991, Lecture 5.

77. Cheston, T. C., and J. Frank. "Phased Array Radar Antennas." *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 7, pp. 7.31–7.32.

78. Mailloux, R. J. Ref. 74, Chap. 6.

79. Hansen, R. C. Ref. 69, Chap. 7.

80. Hannan, P. W. "The Element-gain Paradox for a Phased-Array Antenna." *IEEE Trans.* AP-12 (July 1964), pp. 423–433.

81. Pozar, D. M., and D. H. Schaubert. "Scan Blindness in Infinite Arrays of Printed Dipoles." *IEEE Trans.* AP-32 (June 1984), pp. 602–610.

82. Byron, E. V., and J. Frank. "'Lost Beams' from a Dielectric Covered Phased-Array Aperture." *IEEE Trans.* AP-16 (July 1968), pp. 494–499.

83. King, D. D., and H. J. Peters. "Element Interaction in Steerable Arrays." *Microwave J.* 6 (February 1963), pp. 73–77.

84. Mailloux, R. J. Ref. 78, p. 314.

85. Kinsel, J., B. J. Edward, and D. E. Rees. "V-Band Space-Based Phased Arrays." *Microwave J.* 30 (January 1987), pp. 89–102.

86. Mailloux, R. J. "Phased Array Architecture." *Proc. IEEE* 80 (Jaunary 1992), pp. 163–172. See also Mailloux, Ref. 74, Sec. 5.3.1.

87. Patton, W. T. "Array Feeds." In *Practical Phased-Array Antenna Systems,* E. Brookner (Ed.). Boston: Artech House, 1991, Lecture 6.

88. Hansen. R. J. Ref. 69, Secs. 2.3.4 and 2.3.5.

89. Smith, M. S., and Y. C. Guo. "A Comparison of Methods for Randomizing Phase Quantization Errors in Phased Arrays." *IEEE Trans.* AP-31 (November 1983), pp. 821–828.

90. Cheston, T. C., and J. Frank. Ref. 77, Sec. 7.8.

91. Cheston, T. C., and J. Frank. Ref. 77, pp. 7.53–7.55.

92. Patton, W. T. "Compact, Constrained Feed Phased Array for the AN/SPY-1." In *Practical Phased-Array Antenna Systems,* E. Brookner (Ed.). Boston: Artech House, 1991, Lecture 8.

93. Sharp, E. D. "A Triangular Arrangement of Planar-Array Elements that Reduces the Number Needed." *IRE Trans.* AP-9 (March 1961), pp. 126–129.

94. Cheng, D. H. S. "Characteristics of Triangular Lattice Arrays." *Proc. IEEE* 56 (November 1968), pp. 1811–1817.

95. Nelson, E. A. "Quantization Sidelobes of a Phased Array with a Triangular Element Arrangement." *IEEE Trans.* AP-17 (May 1969), pp. 363–365.

96. Agrawal, A. K., and E. L. Holzman. "Beamformer Architectures for Active Phased-Array Radar Antennas." *IEEE Trans.* AP-47 (March 1999), pp. 432–442.

97. Holzman, E. L., A. K. Agrawal, and J. G. Ferrante. "Active Phased Array Design for High Clutter Improvement Factor." *1996 IEEE International Symposium on Phased Array Systems and Technology,* October 15–18, 1996, IEEE Catalog Number 96TH8175, pp. 44–47.

98. Reed, J. E. "The AN/FPS-85 Radar System." *Proc. IEEE* 57 (March 1969), pp. 324–335.

99. Grimes, M. D., J. M. Major, and T. J. Warnagiris. "Peak Power Tailoring and Phase Nulling of the AN/FPS-85 Radar." *SPIE* 2154, *Intense Microwave Pulses II,* pp. 241–246; 1994.

100. Brookner, E. *Aspects of Modern Radar.* Boston: Artech House, 1988, Sec. 2.2.1.2, p. 198, and pp. 279–281.

101. Sarcione, M., et al. "The Design, Development and Testing of the THAAD (Theater High Altitide Area Defense) Solid State Phased Array (formerly Ground Based Radar)." *1996 IEEE International Symp. on Phased Array Systems and Technology,* October 15–18, 1996, IEEE Catalog Number 96TH8175, pp. 260–265.

102. Dryer, et al. "EL/M 2080 ATBM Early Warning and Fire Control Radar System." *1996 IEEE International Symp. on Phased Array Systems and Technology,* October 15–18, 1996, IEEE Catalog Number 96TH8175, pp. 11–16.

103. Malas, J. A. "F-22 Radar Development." *Proc. IEEE 1997 NAECON* 2, pp. 831–839, IEEE Catalog no. CH36015-97.

104. Barton, D. K. "The 1993 Moscow Air Show." *Microwave J.* 37 (May 1994), pp. 24ff.

105. Corey, L. E. "A Survey of Russian Low Cost Phased-Array Technology." *1996 IEEE International Symp. on Phased Array Systems and Technology,* October 15–18, 1996, IEEE Catalog Number 96TH8175, pp. 255–259.

106. Ajioka, J. S., and J. L. McFarland. "Beam-Forming Feeds." Chap. 19, *Antenna Handbook,* Y. T. Lo and S. W. Lee (Ed.). New York: Van Nostrand Reinhold, 1988. See also references 86 to 100, Chap. 8. In *Introduction to Radar Systems,* 2nd ed. M. Skolnik.

107. Shelton, J. W. "Fast Fourier Transform and Butler Matrices." *Proc. IEEE* 56 (March 1968), p. 350.

108. White, W. D. "Pattern Limitations in Multiple-Beam Antennas." *IRE Trans.* AP-10 (July 1962), pp. 430–436.

109. Barton, P. "Digital Beam Forming of Radar." *IEE Proc.* 127, Pt. F., No. 4 (August 1980).

110. Steyskal, H., and J. F. Rose. "Digital Beamforming for Radar Systems." *Microwave J.* 32 (January 1989), pp. 121ff.

111. Steyskal, H. "Digital Beamforming at Rome Laboratory." *Microwave J.* 39 (February 1996), pp. 100ff.

112. Farina, A. *Antenna-Based Signal Processing Techniques for Radar Systems.* Boston: Artech House, 1992, Sec. 2.7.

113. Skolnik, M. "Improvements for Air-Surveillance Radar." *Proc. 1999 IEEE Radar Conf.* April 20–22, 1999, IEEE Catalog Number 99CH36249, pp. 18–21.

114. Steyskal, H., and J. S. Herd. "Mutual Coupling Compensation in Small Array Antennas." *IEEE Trans.* AP-38 (December 1990), pp. 1971–1975.

115. Lewis, B. L., F. F. Kretschmer, and W. W. Shelton. *Aspects of Radar Signal Processing.* Norwood, MA: Artech House, 1986, Chap. 3.

116. Brookner, E., and J. M. Howell, "Adaptive-Adpative Array Processing." *Proc. IEEE* 74 (April 1986), pp. 602–604.

117. Gabriel, W. F. "Using Spectral Estimation Techniques in Adaptive Processing Antenna Systems." *IEEE Trans.* AP-34 (April 1986), pp. 291–300.

118. Mailloux, R. J. "Array Failure Correction with a Digitally Beamformed Array." *IEEE Trans.* AP-44 (December 1996), pp. 1543–1550.

119. Wirth, W. D. "Long Term Integration for a Floodlight Radar." *1995 IEEE International Radar Conf.* Arlington, VA, pp. 698–703.

120. Headrick, J. M. "HF Over-the-Horizon Radar." *Radar Handbook,* M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 24.

121. Sparks, R. A. "Systems Applications of Mechanically Scanned Array Antennas." *Microwave J.* 31 (June 1988), pp. 26–48.

122. Richardson, P. N., and H. Y. Yee. "Design and Analysis of Slotted Waveguide Antenna Arrays." *Microwave J.* 31 (June 1988), pp. 109ff.

123. Yee, H. Y, and R. C. Voges. "Slot-Antenna Arrays." *Antenna Engineering Handbook.* 3rd ed., R. C. Johnson (Ed.), New York: McGraw-Hill, 1993, Chap. 9.

124. Watson, C. K., and K. Ringer. "Feed Network Design for Airborne Monopulse Slot-Array Antennas." *Microwave J.* 31 (June 1988), pp. 129ff.

125. Kraus, J. D. *Antennas,* 2nd ed., New York: McGraw-Hill, 1988, Sec. 4.6.

126. Butler, J. M., A. R. Moore, and H. D. Griffiths. "Resource Management for a Rotating Multi-Function Radar." *Radar-97, 14-16 October 1997,* IEE Publication No. 449, pp. 568–572.

127. Billam, E. R. "Rotating vs Fixed Active Arrays for Multifunction Radar." *Radar-97, 14-16 October 1997,* IEE Publication No. 449, pp. 573–575.

128. Mailloux, R. J. Ref. 74, Chap. 3, "Pattern Synthesis for Linear and Planar Arrays."

129. Dolph, C. L. "A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beamwidth and Side Lobe Level." *Proc. IRE* 34 (June 1946), pp. 335–348; also discussion by H. J. Riblet, vol. 35, pp. 489–492.

130. Taylor, T. T. "Design of Line-Source Antennas for Narrow Beamwidth and Low Side Lobes." *IRE Trans.* AP-3 (January 1955), pp. 16–28.

131. Hansen, R. C. "Linear Arrays." *Handbook of Antenna Design,* vol. 2, A. W. Rudge, K. Milne, A. D. Oliver, and P. Knight (Eds.). London: Peter Peregrinus, 1983, Chap. 9.

132. Hansen, R. C. "Measurement Distance Effects on Low Sidelobe Patterns." *IEEE Trans.* AP-32 (June 1984), pp. 591–594.

133. Taylor, T. T. "Design of Circular Apertures for Narrow Beamwidths and Low Sidelobes." *IRE Trans.* AP-8 (January 1960), pp. 17–22.

134. White, W. D. "Desirable Illuminations for Circular Aperture Arrays." Institute for Defense Analyses, Arlington, VA, Research paper P-351, IDA Log No. HQ 67–6476, December 1967.

135. White, W. D. "Circular Aperture Distribution Functions." *IEEE Trans.* AP-25 (September 1977), pp. 714–716.

136. Elliott, R. S. *Antenna Theory and Design.* Englewood Cliffs, NJ: Prentice-Hall, 1981.

137. Bayliss, E. T. "Design of Monopulse Antenna Difference Patterns with Low Sidelobes." *Bell System Tech. J.* 47 (May–June 1968), pp. 623–650.

138. Hansen, R. C. *Phased Array Antennas.* New York: John Wiley, 1998, Sec. 3.7.

139. Lopez, A. R. "Sharp Cutoff Radiation Patterns." *IEEE Trans.* AP-27 (November 1979), pp. 820–824.

140. Ruze, J. "Physical Limitations on Antennas." MIT Research Lab. Electronics Tech, Rept. 248, Oct. 20, 1952; or see p. 255 of the 2nd ed. of this text.

141. Shrader, W. W., and V. Gregers-Hansen. "MTI Radar." *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 15, Fig. 15.64.

142. Ruze, J. "Antenna Tolerance Theory—A Review." *Proc. IEEE* 54 (April 1966), pp. 633–640.

143. Ruze, J. "Physical Limitations on Antennas." MIT Research Lab. Electronics Tech. Rept. 248, Oct. 30, 1952.

144. Skolnik. M. I. "Nonuniform Arrays." *Antenna Theory,* Pt I, R. E. Collin and F. J. Zucker (Eds.). New York: McGraw-Hill, 1969, Chap. 6, Sec. 6.6.

145. Rondinelli, L. A. "Effects of Random Errors on the Performance of Antenna Arrays of Many Elements." *IRE Natl. Conv. Record* 7, pt. 1 (1959), pp. 174–187.

146. Lichter, M. "Beam-Pointing Errors of Long Line Sources." *IRE Trans.* AP-8 (May 1960), pp. 268–275.

147. Hsiao, J. K. "Design of Error Tolerance of a Phased Array." *Electronic Letters* 21, no. 19 (September 12, 1985), pp. 834–836.

148. Hsiao, J. K. "Array Sidelobes, Error Tolerance, Gain, and Beamwidth." Naval Research Laboratory, Washington, D.C., Report 8841, September 28, 1984.

149. Cheston, T. C., and J. Frank. "Phased Array Radar Antennas." Chap. 7, *Radar Handbook,* 2nd ed., M. I. Skolnik (Ed.), New York: McGraw-Hill, 1990, p. 7.41.

150. Cheston, T. C., and J. Frank. Ref. 148, Sec. 7.6.

151. Miller, C. J. "Minimizing the Effects of Phase Quantization in an Electronically Scanned Array." *Proc. of Symp. on Electrically Scanned Array Techniques and Applications,* Rome Air Development Center Technical Documentary Report RADC-TDR-64–225, vol. 1, pp. 17–38, July, 1964.

152. Evans, G. E., and H. E. Schrank. "Low-Sidelobe Radar Antennas." *Microwave J.* 26 (July 1983), pp. 109–117.

153. Schrank, H. E. "Low Sidelobe Phased Array Antennas." *IEEE APS Newsletter* (April 1983), pp. 5–9.

154. White, W. D. "Desirable Illuminations for Circular Aperture Antennas, Research Paper P-351." Institute for Defense Analyses, Arlington, VA, December 1967.

155. Ludwig, A. C. "Low Sidelobe Aperture Distribution for Blocked and Unblocked Circular Apertures." *IEEE Trans.* AP-30 (September 1982), pp. 933–946.

156. Taylor, T. T. "Design of Circular Apertures for Narrow Beamwidth and Low Sidelobes." *IRE Trans.* AP-8 (January 1960), pp. 17–22.

157. Green, T. J. "The Influence of Masts on Ship-Borne Radar Performance." *Radar-77, IEE (London) Conference Publication* no. 155 (October 1977), pp. 405–408.

158. Mangulis, V. "Effective Sidelobe Levels Due to Scatterers." *IEEE Trans.* AES-15 (May 1979), pp. 325–333.

159. Mangulis, V. "Antenna Sidelobes in the Presence of Flat Reflectors." *IEEE Trans.* AES-30 (October 1994), pp. 1122–1125.

160. Scorer, M. "The Calculaton of Radome Induced Sidelobes." *Radar-77,* IEE (London) Conference Publication No. 155 (October 1977), pp. 414–418.

161. Ricardi, L. J. "Radiation Properties of the Binomial Array." *Microwave J.* 15 (December 1972), p. 20.

162. Krall, A. D., D. G. Jablonski, and J. Coughlin. "Radiation Properties of a 'Gaussian' Antenna." *Microwave J.* 27 (May 1984), pp. 283 & 288.

163. Patton, W. T. "Low-Sidelobe Antennas for Tactical Phased-Array Radars." *RCA Engineer* 27 (Sept./Oct 1982), pp. 31–36.

164. Maine, E. E., Jr., and J. M. Willey. The Fixed Array Surveillance Radar, private communication.

165. Schrank, H. E. "Low Sidelobe Reflector Antennas." *IEEE APS Newsletter,* (April 1985), pp. 5–16.

166. Scudder, R. M. "Advanced Antenna Design Reduces Electronic Countermeasures Threat." *RCA Engineer* 23 (Feb./Mar. 1978), pp. 61–65.

167. Williams, N., P. Varnish, and D. J. Browning. "Reduced Cost Low Sidelobe Reflector Antenna Systems." *IEE (London) International Conf. Radar-82* (October 1982), pp. 351–354.

168. Fante, R. L., P. R. Franchi, N. R. Kerweis, and L. F. Dennett. "A Parabolic Cylinder Antenna with Very Low Sidelobes." *IEEE Trans.* AP-28 (January 1980).

169. Evans, G. E. *Antenna Measurement Techniques.* Boston: Artech House, 1990.

170. Hacker, P. S., and H. E. Schrank. "Range Distance Requirements for Measuring Low and Ultralow Sidelobe Antenna Patterns." *IEEE Trans.* AP-30 (September 1982), pp. 956–966.

171. Tang, R., and R. Brown. "Cost Reduction Techniques for Phased Arrays." *Microwave J.* 30 (January 1987), pp. 139–146.

172. Skolnik, M. "The Radar Antenna—Circa 1995." *J. Franklin Inst.* 332B, no. 5 (May 1995), pp. 503–519.

173. Cheston, T. C., and J. Frank. "Phased Array Radar Antennas." Chap. 7, *Radar Handbook,* M. Skolnik (ed.) New York: McGraw-Hill, 1990, Chap. 7, Sec. 7.7.

174. Mailloux, R. J. Ref. 74, Sec. 8.3.

175. Smolders, A. B. "Design and Construction of a Broadband Wide-Scan-Angle Phased-Array Antenna with 4096 Radiating Elements." *1996 IEEE International Symp. on Phased Array Systems and Technology,* Boston, MA, October 15–18, 1996, pp. 87–92.

176. Baugh, R. A. *Computer Control of Modern Radars.* published by RCA Corp., (now Lockheed Martin) Aegis Department, Moorestown, NJ, 1973.

177. Huizing, A. G., and A. A. F. Bloemen. "An Efficient Scheduling Algorithm for a Multifunction Radar." *1996 IEEE International Symp. on Phased Array Systems and Technology,* Boston, MA, October 15–18, 1996, pp. 359–364.

178. Billam, E. R. "The Problem of Time in Phased Array Radar." *Radar-97,* October 14–16, 1997, IEE Publication No. 449, pp. 563–567.

179. Bony, Gilbert. "Electrically Controlled Dielectric Panel Lens," United States Patent No. 3,708,796, Jan. 2, 1973.

180. Skolnik, M. I. "The Radar Antenna—Circa 1995." *J. Franklin Institute,* 332B, No. 5 (1995), pp. 503–519.

181. Chekroun, C., et al. "RADANT. New Method of Electronic Scanning." *Microwave J.* 24 (February 1981), pp. 45–53.

182. Rao, J. B. L., G. V. Trunk, and D. P. Patel. "Two Low-Cost Phased Arrays." *1996 IEEE International Symp. on Phased Array Systems and Technology,* Boston, MA, October 15–18, 1996, pp. 119–124.

183. Colin, Jean-Marie. "Phased Array Radars in France. Present & Future." *1996 IEEE International Symp. on Phased Array Systems and Technology,* Boston, MA, October 15–18, 1996, pp. 458–462.

184. Rao, J. B. L., D. P. Patel, and V. Krichevsky. "Voltage-Controlled Ferrolelectric Lens Phased Arrays." *IEEE Trans.* AP-47 (March 1999), pp. 458–468.

185. Mailloux, R. J. "Conformal and Low-Profile Arrays." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, Chap. 21, 1993.

186. Hsiao, J. K. "Approximation of a Conformal Array with Multiple, Simultaneously Excited Planar Arrays," Naval Research Laboratory, Washington, D. C., Report 7442, July 28, 1972.

187. Skolnik, M. I. "Nonuniform Arrays." *Antenna Theory,* pt. I, R. E. Collin and F. J. Zucker (Eds.). New York: McGraw-Hill, 1969, Chap. 6.

188. Mailloux, R. J. *Phased Array Antenna Handbook.* Boston: Artech House, 1994, Sec. 2.4.

189. Skolnik, M. I., G. Nemhauser, and J. W. Sherman. "Dynamic Programming Applied to Unequally Spaced Arrays." *IEEE Trans.* AP-12 (January 1964), pp. 35–43.

190. Skolnik, M. I., J. W. Sherman, and F. C. Ogg. "Statistically Designed Density-Tapered Arrays." *IEEE Trans.* AP-12 (July 1964), pp. 408–417.

191. Hall, W. P., Jr., and R. D. Nordmeyer. "Active-Element, Phased Array Radar: Affordable Performance for the 1990s." *IEEE National Telesystems Conf. Proc.*, Atlanta, GA, pp. 193–197, March 26 and 27, 1991.

192. Cross, D. C., D. D. Howard, and J. W. Titus. "Mirror-Antenna Radar Concept." *Microwave J.* 29 (May 1986), pp. 323–335.

193. Orleansky, E., C. Samson, and M. Havkin. "A Broadband Meanderline Twistreflector for the Inverse Cassegrain Antenna." *Microwave J.* 30 (October 1987), pp. 185–192.

194. Lewis, B. L., and J. P. Shelton. "Mirror Scan Antenna Technology." *1980 IEEE International Radar Conf.,* Washington, D.C., pp. 279–283, April 1980.

195. Ruze, J. "Lateral Feed Displacement of a Paraboloid." *IEEE Trans.* AP-13 (September 1965), pp. 660–665.

196. Imbriale, W. A., P. G. Ingerson, and W. C. Wong. "Large Lateral Feed Displacement in a Parabolic Reflector." *IEEE Trans.* AP-22 (November 1974), pp. 742–745.

197. Kelleher, K. S., and H. H. Hibbs. "A New Microwave Reflector." Naval Research Laboratory, Washington, D.C., Report 4141, 1953.

198. Kelleher, K. S., and G. Hyde. "Reflector Antennas." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, 1993, Chap. 17, pp. 17-46 to 17.52.

199. Kelleher, K. S. "Electromechanical Scanning Antennas." *Antenna Engineering Handbook.* 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, 1993, Chap. 18, pp. 18-23 to 18-24.

200. Rappaport, C. M., and W. P. Craig. "High Aperture Efficiency Symmetric Reflector Antennas with up to 60° Field of View." *IEEE Trans.* AP-39 (March 1991), pp. 336–344.

201. Craig, W. P., C. M. Rappaport, and J. S. Mason. "A High Aperture Efficiency, Wide-Angle Scanning Offset Reflector Antenna." *IEEE Trans.* AP-41 (November 1993), pp. 1481–1490.

202. Bodnar, D. G. "Lens Antennas." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.). New York: McGraw-Hill, 1993, Chap. 16.

203. Harvey, A. F. "Optical Techniques at Microwave Frequencies." *IEE Proc.* 106, pt. B (March 1959), pp. 141–157. Contains an extensive bibliography.

204. Kock, W. E. "Metal Lens Antennas." *Proc. IRE* 34 (November 1946), pp. 828–836.

205. Luneburg, R. K. *Mathematical Theory of Optics.* Berkeley: University of California, 1964. (Originally mimeographed lecture notes, Brown University Graduate School, Providence, R.I., 1944.)

206. Wiltse, J. C., and J. E. Garrett. "The Fresnel Zone Plate Antenna." *Microwave J.* 34 (January 1991), pp. 101–114.

207. Huddleston, G. K., and H. L. Bassett. "Radomes." *Antenna Engineering Handbook,* 3rd ed., R. C. Johnson (Ed.), New York: McGraw-Hill, 1993, Chap. 44.

208. Walton, J. D., Jr. *Radome Engineering Handbook.* New York: Marcel Dekker, 1970.

209. Schrank, H. E., G. E. Evans, and D. Davis. "Reflector Antennas." *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 6. Sec. 6.9.

210. Electronic Space Systems Corporation (Essco) Web site, 1998.

211. Hughes, D. "New FAA Radomes to Have 98% Transmission Efficiency." *Aviation Week & Space Technology* (January 10, 1994), pp. 66–67.

212. Vitale, J. A. "Large Radomes." *Microwave Scanning Antennas,* Vol. 1." R. C. Hansen (Ed.). New York: Academic, 1964, Chap. 5.

213. Effenberger, J. A., R. R. Strickland, and E. B. Joy. "The Effect of Rain on a Radome's Performance." *Microwave J.* 29 (May 1986), pp. 261–274.

214. Punnett, M. S. "Developments in Ground Mounted Air Supported Radomes." *IEEE 1977 Mechanical Engineering in Radar Symposium,* Nov. 8–10, 1977, Arlington, VA, pp. 40–45, IEEE Publication 77CH1250-0 AES.

215. Advertising brochure "Engineered Fabric Structures," from Birdair Structures Division, North Bennington, Vermont (no date).

216. Conti, D. A. "Special Problems Associated with Aircraft radomes." *IEE Proc.* 128, Pt. F, no. 7 (December 1981), pp. 412–418.

217. Pelton, E. L., and B. A. Munk. "A Streamlined Metallic Radome." *IEEE Trans.* AP-22 (November 1974), pp. 799–803.

218. Knott, E. F., J. F. Shaeffer, and M. T. Tuley. *Radar Cross Section,* 2nd ed. Boston: Artech House, 1993, Sec. 10.4.

219. Mittra, R., C. H. Chan, and T. Cwik. "Techniques for Analyzing Frequency Selective Surfaces—A Review." *Proc. IEEE* 76 (December 1988), pp. 1593–1615.

220. Farina, A. *Antenna-Based Signal Processing Techniques for Radar Systems.* Boston: Artech House, 1992.

221. Nitzberg, R. *Adaptive Signal Processing for Radar.* Boston: Artech House, 1992.

222. Widrow, B., and S. D. Stearns. *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

223. Gabriel, W. F. "Adaptive Arrays—An Introduction." *Proc. IEEE* 64 (February 1976), pp. 239–272.

224. Gabrial, W.F. "Adaptive Processing Array Systems." *Proc. IEEE* 80 (January 1992) pp. 1521–162.

225. Howells, P. W. "Explorations in Fixed and Adaptive Resolution at GE and SURC." *IEEE Trans.* AP-24 (September 1976), pp. 575–584.

226. Ward, J. "Space-Time Processing for Airborne Radar." MIT Lincoln Laboratory Technical Report 1015, December 13, 1994.

227. Staudaher, F. M. "Airborne MTI." *Radar Handbook,* M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 16, Sec. 16.8.

228. Gabriel, W. F. "Spectral Analysis and Adaptive Array Superresolution Techniques." *Proc. IEEE* 68 (June 1980), pp. 654–666.

229. Rihaczek, A. W. "The Maximum Entropy of Radar Resolution." *IEEE Trans.* AES-17 (January 1981), p. 144.

230. Hansen, R. C. *Phased Array Antennas.* New York: John Wiley, 1998, Chap. 9.

231. *IEEE Standard Dictionary of Electrical and Electronics Terms.* New York: IEEE, 1988.

232. Pugh, M. L., et al. "Electromechanically Scanned Arrays Using Micro Electro Mechancical Swith (MEMS) Technology." *Proc. 5th International Conf. on Radar Systems,* Brest, France, May, 1999, Subassemblies Section.

233. Smith, J. K., F. W. Hopwood, and K. A. Leahy, "MEM Switch Technology in Radar." *Record of the IEEE 2000 International Radar Conference,* Alexandria, VA, pp. 193–198.

234. Norvell, B. R., et al. "Micro Electro Mechanical Switch (MEMS) Technology Applied to Electrically Scanned Arrays (ESA)." *Proc. International Radar Symp. (IRS 98),* Munich Germany, September 15–17, 1998, vol. II, German Institute of Navigation.

PROBLEMS

9.1 (a) Derive the expression for the field-intensity pattern for a uniformly illuminated line-source aperture of dimension D. (b) Make a rough sketch of its radiation (power) pattern (with the ordinate in dB). (c) If the antenna dimension is 60 wavelengths, what is the width between the first nulls that define the main beam? (d) What is its half-power beamwidth?

9.2 The pattern of problem 9.1 also is the pattern from either of the two principal planes of a uniformly illuminated square aperture. Derive the field-intensity pattern in the diagonal plane of the square aperture, where the aperture illumination is triangular. (The triangular illumination is also known as a *gabled illumination.* It can be expressed as $A(z) = 1 - \frac{2}{D_g}|z|$, where $|z| \leq D_g/2$, and D_g is the diagonal of the square aperture.)

9.3 Determine and roughly sketch the field-intensity pattern for an antenna with the following aperture illumination pattern:

$$A(z) = -1 \text{ for } -D/2 < z < 0$$

$$A(z) = 0 \text{ for } z = 0$$

$$A(z) = +1 \text{ for } 0 < z < +D/2$$

where D = aperture dimension. What is the peak sidelobe level? (This is the difference pattern that would be obtained from an antenna whose sum pattern is of the form $(\sin x)/x$ obtained with a uniform aperture as in problem 9.1.)

9.4 Calculate and sketch the field-intensity pattern produced by an aperture illumination of a line source of dimension D given by $A(z) = \cos^2(\pi z/D)$. What is its null width (width between the two nulls defining the main beam), and what is the level of the first sidelobe?

9.5 In what manner is the field-intensity pattern $E(\phi)$ of an antenna, Eq. (9.10), and its aperture illumination $A(z)$ related to the time waveform $s(t)$ and the spectrum $S(f)$? Identify the analogous pairs of parameters between these two relationships?

9.6 Efficiency is generally defined as the ratio of the output power to the input power. The "aperture efficiency" of a reflector antenna, Eq. (9.9), is not defined in this manner and the term aperture "efficiency" can therefore be misleading. Discuss the effect of the aperture efficiency on radar performance and why it should not be interpreted as an indicator of a power loss.

9.7 Why does a parabolic surface make a good reflector antenna?

9.8 When might each of the following parabolic reflector antennas be used: (a) paraboloid, (b) section of a paraboloid, (c) parabolic cylinder, (d) parabolic torus, (e) offset-fed reflector, (f) Cassegrain, and (g) mirror-scan antenna? When might (h) a spherical reflector or (i) a lens antenna be used?

9.9 *Background.* In a Cassegrain antenna, Fig. 9.11, the feed is located at (or near) the apex of the primary paraboloid reflector. The radiation from the feed is reflected by a secondary reflector in front of the primary reflector. The radiation is returned to the primary reflector where it is reradiated in the forward direction. Blockage of the radiated energy occurs because of the obstruction by the secondary reflector and by the interception of energy by the feed. If the secondary reflector is made smaller so as to reduce its blockage, the feed has to be made larger so as to illuminate without spillover the smaller secondary reflector. Similarly, if the feed is made smaller to avoid blockage, the secondary reflector has to be made larger in order to intercept the wide-angle energy from the feed. If the secondary reflector and the feed are both circular with a diameter of S and d, respectively, the total blockage is due to their combined area $(\pi/4)(S^2 + d^2)$. *Problem:* What should the relationship be between the diameter S of the feed and the diameter d of the secondary reflector in order to minimize the total blockage in a Cassegrain antenna? (You may assume that the beamwidth of the feed is λ/d radians.)

9.10 List the five basic methods available for obtaining a phase shift, and give an example of a phase shifter based on each.

9.11 When might ferrite phase-shifters be used in an electronically steered phased array antenna, and when might diode phase-shifters be used?

9.12 (a) If the minimum range of radar is to be no greater than 1.2 nmi, what should be the maximum switching time of a phase shifter so that the array is ready to receive after it has transmitted? (Assume that the pulse duration is small and can be neglected.) (b) Select a type of ferrite phase shifter that can probably meet this requirement. (c) What limitation might occur with your selection? (d) If the minimum allowable range were increased to

12 nmi, how might your answer for (b) and (c) change? (e) What type of ferrite phase shifter can be used when the radar operates with dual orthogonal linear polarizations? (f) What type of ferrite phase shifter might be used in a reflectarray?

9.13 (a) Derive the array factor for a uniformly spaced linear array of N isotropic elements. [Suggest you start with Eq. (9.22); but instead of a summation of sinewaves, it will be easier to use the exponential relation such as $1 + e^{j\phi} + e^{j2\phi} + e^{j3\phi} \cdots$ where $\phi = 2\pi(d/\lambda) \sin \theta$, as in Eq. (9.22). You also will have to recall or rederive the expression for the sum of a geometric series that you learned in high school.] (b) At what angles will grating lobes appear (over a range of $\pm 90°$) when the element spacing is four wavelengths? (c) Compare, in words, the pattern produced by the linear array of equal amplitude elements to the pattern produced by a continuous line source with uniform illumination (as in problem 9.1) and the same aperture size.

9.14 When the beam of a phased array antenna is electronically steered to an angle θ_0 from broadside, show that its beamwidth varies inversely as $\cos \theta_0$.

9.15 (a) Show that grating lobes will not appear in a steered phased array if the element spacing is less than one-half wavelength. (b) What should the element spacing be if there can be grating lobes at $\pm 30°$, but not at smaller angles?

9.16 (a) What wrap-up factor is required in a frequency-scan array to scan the beam over an angle of $\pm 50°$ using no more than a total tunable (relative) bandwidth of 5 percent? (Assume a TEM transmission line where the velocity of propagation is the velocity of light.) (b) If there are 80 elements in the frequency scan array with an element spacing of 10 cm and a wrap-up factor the same as computed for part (a), what is the time required for a signal to fill the array aperture? (c) What bandwidth does this correspond to?

9.17 Consider the series-fed linear array of Fig. 9.16b. How far will the beam deviate from the broadside direction when the frequency is changed by 20 percent? (Assume the transmission feed-line propagates in the TEM mode so that the velocity in the line is c, the velocity of light.)

9.18 A frequency-scan array has an element spacing $d = 5$ cm, aperture dimension $D = 3$ m, and a feed system with a wrap-up factor = 16. As the beam is frequency scanned past the target, the echo will be frequency modulated with a bandwidth Δf_B [Eq. (9.42)]. (a) If a frequency $f = 1.05 f_0$ points the beam to 30°, what is the spectral width of the echo signal due to the linear FM modulation induced on the echo? (b) If pulse compression processing is used on receive to take advantage of the frequency modulation of the echo signal, what will be the compressed pulse width?

9.19 Why do you think problems might occur when attempting to apply the theory of the infinite array to a finite size array?

9.20 Compare the corporate-fed (passive) phased array, the active-aperture phased array, and the space-fed phased array with respect to loss, transmitter efficiency (Sec. 10.3), relative prime power required for the transmitter, and any other factors of concern.

9.21 The Dolph-Chebyshev antenna illumination produces an antenna pattern with a narrow beamwidth and all sidelobes equal. It would appear to be a good antenna pattern except

that it is not practical. What is it about the Dolph-Chebyshev illumination that makes it unrealizable, especially with large gain?

9.22 What approximate reduction in gain [Eq. (9.51)] results when an antenna pattern is rectangular from 0 to 3° and is shaped to have a cosecant-squared variation of gain over the angular region from $\theta_0 = 3°$ to $\theta_m = 25°$?

9.23 For a fixed-size reflector antenna with an rms surface error ϵ: (a) Show that the wavelength that results in maximum antenna directivity is $\lambda_m = 4\pi\epsilon$. [This is Eq. (9.53).] (b) How does this relate to the usual "rule of thumb" for reflector antenna tolerances which states that the rms error ϵ should be less than $\lambda/32$, where λ = wavelength? (c) Determine the loss of gain at the wavelength λ_m compared to the gain of a perfect (no error) antenna. (d) What is the maximum gain that is achievable for $D/\epsilon = 1000$ and $D/\epsilon = 10{,}000$? (Assume the aperture efficiency is 1.) (e) In your opinion, how high an antenna gain might be achieved with a practical radar antenna (and explain your answer)?

9.24 If a one-dB reduction in antenna gain is allowed due to errors in a 100-element linear array, what is the phase error (in degrees), the amplitude error (in dB), and the fraction of missing elements that can be tolerated when each one of these three factors is the only one contributing to the gain reduction? (That is, determine the phase error with the amplitude error zero and no missing elements, and so on.)

9.25 *Background.* A periodic error in an antenna will produce multiple equally spaced beams (or sidelobes) in sin θ space similar to the formation of grating lobes in an array antenna, except that these periodic error sidelobes are much smaller in gain. *Problem.* An antenna is suspected of having a periodic error since there are prominent sidelobes in its radiation pattern at angles of ± 37.3 and $\pm 53.9°$. If there is a periodic error, there will be other closer-in sidelobes; but in this problem we assume that these closer-in lobes are masked by the normal antenna sidelobe radiation. Based on the observed sidelobes whose directions are given above, what is the period (spacing) of this periodic error?

9.26 (a) With a uniformly illuminated planar array of 1000 isotropic elements, what should be the rms value of the phase error (in degrees) and the rms value of the amplitude error (in dB) to make the average error-sidelobes equal to -50 dB? (Assume that the contribution from the phase error equals that from the amplitude error.) (b) Repeat for an array antenna with 10,000 elements.

9.27 What determines the number of bits to be used in a digitally switched phased shifter?

9.28 According to G. Evans [Ref. 169, p. 115] the accuracy with which the antenna gain can be measured is commonly $\pm$ 0.5 dB. What is the maximum rms phase error in a reflector antenna that results in a loss of antenna gain of 0.5 dB?

9.29 (a) In a phased array with a square aperture of 100 by 100 elements, with half-wave spacing between elements, what is the rms angle error (in degrees) when the rms value of the normalized error current is $\sigma = 0.4$? (b) What fraction of a beamwidth would this be if the aperture illumination were uniform?

9.30 What are the characteristics of an aperture illumination that can achieve ultralow sidelobes? (b) When might ultralow sidelobes be needed? (c) What antenna characteristics

have to be sacrificed for ultralow sidelobes? (As a start, see Table 9.1) (d) What factors ultimately limit the sidelobe levels that can be achieved in practice?

9.31 When might a radar systems engineer decide not to use an ultralow sidelobe antenna?

9.32 Why is a phased array antenna more suitable than a reflector antenna as an ultralow sidelobe antenna?

9.33 Compare the advantages and disadvantages for shipboard air defense of a traditional four-faced phased array radar and a system consisting of two trainable phased array radars and one 2D air-surveillance radar with 360° azimuth coverage.

9.34 (a) What are the advantages and limitations of operating a ground-based air-surveillance radar under a radome? (b) Compare the air-inflated radome and the rigid geodesic radome for this application.

9.35 What types of antennas might be used for the detection and tracking of hostile ballistic missiles? (This is a question not just about antennas, but also about how radar systems might be applied for ballistic missile defense. There is no simple unique answer to this question, and there has not been general agreement as to the best approach.)

9.36 Why are phased array radars for air surveillance generally cheaper at the lower radar frequencies than at the higher frequencies?

9.37 For the application of air defense, compare the advantages and disadvantages of a single multifunction phased array that operates in one frequency band with a system having a separate air-surveillance radar and a separate phased array weapon control (tracking) radar, each operating in different bands.

9.38 A phased array has a beamwidth of 2° when pointed to broadside. If it is required to scan to an angle of 60° from broadside, what is the maximum signal bandwidth in MHz that a radar can have that operates at a center frequency of 3.3 GHz?

9.39 What does one have to do to obtain an electronically steered phased array with a large instantaneous signal bandwidth?

9.40 *Background.* It has sometimes been suggested (usually by non-radar administrators) that increasing the frequency of an air-defense phased array radar will result in a smaller size antenna for the same beamwidth (which is true), and thus provide a smaller radar system (which might not be true). If a smaller system were to result it can be made more mobile if a ground based system or, if a ship-based system, it can be employed on smaller ships. A smaller radar, however, does not result when the frequency is increased and all other requirements remain the same. *Problem.* What is wrong with reasoning that concludes that a smaller radar system will result if the array antenna is reduced in size by operating at a higher frequency? [To illustrate your answer, you can, if you wish, assume an S-band multifunction phased array radar (3.5 GHz) having a 12 by 12 foot aperture and 10 kW of average power. If the radar is increased to 35 GHz, the aperture is reduced in size to 1.2 by 1.2 foot.]

9.41 The typical *L*-band 2D air-surveillance radar usually has its maximum elevation coverage extending to about 20 to 40°, depending on the particular radar. If it is required to extend coverage of the radar to higher elevation angles, there are reasons why it might be better

to employ a separate antenna at a different frequency to fill the surveillance hole above the radar. (a) What are some reasons why a separate radar might be used rather than attempt to increase the elevation coverage of the 2D antenna? (b) If the elevation hole extends from 30° elevation angle to the zenith at 90°, what type of scanning patterns might be used? (There can be more than one choice of scanning pattern.) (c) What frequency band might be used for this hole-filler (and explain the reason for your selection)? (d) What type of antenna might be used for the hole-filler radar?

chapter

10

Radar Transmitters

10.1 INTRODUCTION

The radar systems engineer would like a transmitter to provide sufficient energy to detect a target, be easily modulated to faithfully produce the desired waveforms, generate a stable signal so that doppler signal processing can be performed without transmitter noise masking the doppler-shifted received signal, provide the needed signal bandwidth and tunable bandwidth, be of high efficiency, be of high reliability, be easy to maintain, have long life, be able to operate with a minimum of personnel, be of a size and weight suitable for the intended application, and be of affordable price. All of the above can be obtained, but seldom all together in one transmitter. Compromise is necessary.

Some radar transmitters have to generate large peak power as well as large average power; but it is the average power (which relates to energy) that is the measure of radar performance rather than peak power. It was seen in Chap. 2 that the range of a radar is proportional to the fourth root of the radar transmitter's average power. To increase the range of a radar by one order of magnitude (a factor of ten) the transmitter power has to be increased by ten thousand. Although there have been radars with average powers greater than a million watts, power cannot be increased without limit since high-power transmitters are heavy, take up space, and can consume much prime power (the power taken from the local power company) or fuel for motor-driven electrical generators.

An indicator of the performance of a radar is the product of the antenna area times the transmitter's average power. Tom Weil[1] described quite well the problem of choosing between high transmitter power and a large antenna as follows:

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.

Radar transmitters have been based on either a power amplifier, such as a klystron, or a power oscillator, such as a magnetron. In the early days of microwave radar in the 1940s and 1950s, the magnetron power oscillator was used almost exclusively since it was the only high-power microwave tube available at the time. It did an outstanding job in making microwave radar a reality in World War II, but it had many serious limitations. Magnetrons are noisy devices that limit the MTI improvement factor that can be obtained. Although they can produce high peak power (megawatts), they are not capable of large average power, and their signal output cannot be readily modulated to produce pulse-compression waveforms. All of these disadvantages are overcome with amplifiers such as the klystron, traveling wave tube, and the transistor. Modern high-performance radars almost always employ some sort of power amplifier as the transmitter. The magnetron appears to be limited to those applications where its relatively small size and lower cost are more important than its limitations.

Most of the discussion in this chapter is about the RF power source. A transmitter, however, is more than just the active RF power source. It includes the exciter and driver amplifiers that provide the signal to be amplified if the power source is an amplifier. If the transmitter generates a pulse waveform, a pulse modulator of some type is needed (except for RF power sources that are self-pulsed by the input waveform, as are transistors). There must be a d-c power supply for generating the necessary voltages and currents to operate the RF power device; means to remove the heat dissipated, including a heat exchanger when liquid cooling is used; protection devices for dissipating high-voltage arc discharges; safety interlocks; monitoring devices; isolators; high-voltage cable; insulating-oil tanks (to immerse high-voltage cathode bushings to prevent corona and high-voltage breakdown); and lead shielding of X-rays when high voltages are used. Not all high-power radar transmitters need all of the above, but an RF power source is useless without the ancillary devices required to make it function.

The efficiencies of RF power sources typically might range from about 10 percent to about 60 percent. This is the *RF conversion efficiency*, defined as the ratio of RF power output available from the device to the d-c power input to the electron stream. It is the efficiency of interest to the tube or RF power source designer, but it is not the best measure for the radar systems engineer. A better measure is the *transmitter system efficiency*, which is the ratio of the RF power available from the transmitter to the total power needed to operate the transmitter. The total power includes the power to generate the electrons at the cathode, the power to generate any electromagnetic fields required for containing the electron beam, the power to cool the device, and any other power needed for the proper operation of the transmitter. If, for example, the RF conversion efficiency were 40 to 50 percent, the transmitter system efficiency might be 20 to 25 percent, or less. Thus one usually doesn't want to start with a power source whose RF conversion efficiency is only 10 to 15 percent, unless the power is so low that efficiency is not an important consideration.

For maximum efficiency, most high-power RF power sources operate saturated, meaning they are either completely on or completely off, with no intermediate power levels. This is all right for a radar that generates a rectangular-like pulse waveform. There are times, however, when it might be desired to have an amplitude-tapered, or shaped, pulse (for example, to reduce the time sidelobes in pulse compression waveforms or to minimize the effects of RF interference to other users of the electromagnetic spectrum). Highly shaped transmission waveforms are seldom found in high-power microwave radar systems because of their lower efficiency. Transistor amplifiers can be operated in what is called class-A operation so that there is a linear relationship between the output and input signal amplitudes. The efficiency of a class-A amplifier, however, is much less than that which would be achieved with the same device operated class-C. (Class-C amplifiers are nonlinear and are self-pulsing in that they generate pulses when the RF drive is turned on and off.) Thus, a microwave radar transmitter is almost always operated in saturation and not as a linear device.

High reliability and long life are important for a transmitter. The life of most RF power sources can be many thousands or many tens of thousands of hours, as will be discussed when describing the individual devices later in this section. If a transmitter's mean time between failures (MTBF) is *not* as long as expected, factors other than the RF power source are often at fault. Likely candidates are fans and blowers, the wrong type of coolant, RF connectors and coolant fittings damaged, coolant lines clogged, and leads that are broken, mishandled, or abused. Conservative mechanical and electrical design and procurement practices that guarantee reliability from suppliers are needed for a trouble-free transmitter (or anything else). The user of RF power sources also needs to help in avoiding less than the achievable reliability. For example, A. S. Gilmour[2] states "If the truth were known, it could well be that over 50 percent of all failures are the fault of the users, rather than the tube manufacturers."

Summary of Radar RF Power Sources The different RF power sources available for high-power radar application include the klystron, traveling wave tube, solid-state transistor amplifier, Twystron, magnetron, crossed-field amplifier, amplitron, grid-controlled vacuum tube, extended interaction amplifier, gyrotron, and others. None provide all the desirable features that might be wanted. Some are no longer as popular as they once were. The choice depends in large part on the application and its constraints. Each of these devices will be briefly summarized below and described in more detail later in the chapter. All except the magnetron are power amplifiers. The gyrotron can be either an amplifier or an oscillator.

Klystron This is an excellent radar tube when it can be employed. It has high gain and good efficiency, and is capable of higher average and peak power than most other RF power sources. It can have wide bandwidth (in the vicinity of 8 to 10 percent relative bandwidth*) when its power is large, long life (tens of thousands of hours), low interpulse noise, and good stability for doppler processing. When large peak powers are generated it requires very high voltage and X-ray shielding.

*Relative, or fractional, bandwidth in percent = $(\Delta f/f_0) \times 100$, where Δf = absolute bandwidth and f_0 = center frequency.

Traveling Wave Tube (TWT) TWTs have slightly less power, slightly less gain, and slightly less efficiency than a klystron; but they are capable of wide bandwidth, especially at modest power levels. At high power levels the TWT bandwidth is lower than what can be achieved at lower powers, but it is still relatively large.

Hybrid Klystrons These are similar to klystrons, but with one or more of the resonant cavities replaced by multiple cavities similar to what are used in a TWT. There are at least three versions: the *Twystron*, the *extended interaction klystron*, and the *clustered cavity klystron*. Bandwidths can be 15 to 20 percent, or more. Extended interaction klystrons have also been used for modest power millimeter-wave transmitters.

Solid-State Transistor Amplifiers These are capable of wider bandwidth than most other RF power sources. They operate with low voltages, ease of maintenance, and have promise of long life. They are inherently of low power so that a large number of individual devices must be combined to generate sufficient power for most radar applications. For adequate efficiency, they should be operated at high duty cycles, which require that they generate long pulses and employ pulse compression.

Magnetron The magnetron is generally smaller in size and utilizes lower voltage than the klystron; but its average power is limited and it has poor noise and stability characteristics, which restrict its ability to cancel clutter when used in an MTI radar.

Crossed-Field Amplifiers These are capable of high power, good efficiency, and wide bandwidth, but are of relatively low gain (about 10 dB). Lower voltage, just as with the magnetron, means that X-rays are not usually a problem. Crossed-field amplifiers are generally noisier and less stable than most other RF power sources.

Grid-Control Tubes These are modern versions of the classical triode and tetrode vacuum tubes originally introduced early in the twentieth century. They are a good power source for UHF radars, but they have been largely replaced by solid-state devices.

Microwave Power Module This is a combination of a modest power TWT driven by a solid-state device that competes favorably with high-power solid-state modules in some applications.

Gyrotrons These can produce very high power in the millimeter wave region, but they require large magnetic fields. They have had only modest application in operational radar systems.

RF power sources may be grouped into four general classes: (1) linear-beam tubes, (2) solid state, (3) crossed-field tubes, and (4) others not included in the first three. The klystron, traveling wave tube, magnetron, and crossed-field amplifier are *slow-wave devices* in that the phase velocity of the electromagnetic wave in the RF structure is slowed so as to be approximately equal to the velocity of the electron stream in order for (d-c) energy from the electron beam to transfer to electromagnetic energy in the (RF) signal.[3] The gyrotron, on the other hand, is a *fast-wave device* in that the phase velocity of the electromagnetic wave exceeds the speed of light in the interaction region.

10.2 LINEAR-BEAM POWER TUBES

In the linear-beam tube the electrons emitted from the cathode are formed into a long cylindrical beam that receives the full potential energy of the electric field before the beam enters the RF interaction region. The klystron, traveling wave tube, Twystron, and extended interaction amplifier are examples of linear-beam tubes. The last two are basically hybrid devices that combine the technology of the klystron with the RF structure of the TWT. An axial magnetic field is used in linear-beam tubes to confine the electron beam and keep electrons from hitting the RF structure. Transit-time effects, which can limit conventional vacuum tube performance at high frequencies, are used to good advantage in linear-beam tubes to *density modulate* the uniform d-c electron beam to create bunches of electrons from which RF energy can be extracted.

Linear-beam tubes can produce much higher power than other power sources. Klystrons are capable of more than a megawatt of average power. High power is a result, in part, of their larger size and high voltages. Thomson-CSF in France produced a UHF klystron that delivered more than one MW CW power. It was 5 m (16.4 ft) long and weighed 1400 kg (3000 pounds).[4] On the other hand, X-band klystron and TWT transmitters producing many kilowatts of average power can be made light enough and small enough to fly in the nose of military fighter/attack aircraft.

Klystron A sketch of the principal parts of a klystron is shown in Fig. 10.1. At the left is the cathode which emits a stream of electrons that is formed into a narrow cylindrical beam by the *electron gun*. The electron gun consists of the cathode that is the source of electrons, a modulating anode or other beam-control electrode to provide a means for turning the beam on and off to generate pulses, and the anode. The electron emission density at the surface of the cathode is less than that required for the electron beam, so a large-area cathode surface is used and the emitted electrons are caused to converge to a narrow beam of high electron density. The multiple RF cavities, which correspond to the *LC* resonant circuits of conventional lower-frequency amplifiers, are at anode potential. Electrons are not intentionally collected by the anode, as in some other tubes; instead they are

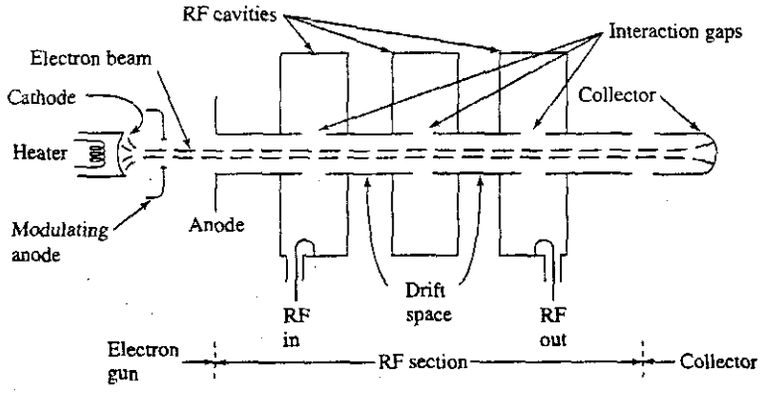


Figure 10.1 Representation of the principal parts of a three-cavity klystron amplifier.

removed by the collector electrode (shown on the right) after the beam has given up its RF energy to the output RF cavity.

The RF input signal is applied across the *interaction gap* of the first cavity. Those electrons which arrive at the gap when the input signal voltage is a maximum (peak of the sinewave) experience a voltage greater than those electrons which arrive at the gap when the input is at a minimum (trough of the sinewave). Thus the electrons that see the peak of the sinewave are speeded up and those that see the trough are slowed down. The process whereby some electrons are speeded up and others slowed down is called *velocity modulation* of the electron beam. In the *drift space,* electrons that are speeded up during the peak of one cycle catch up with those slowed down during the previous cycle. The result is that the electrons of the velocity-modulated beam become "bunched," or density modulated, after traveling through the drift space. A klystron usually has one or more appropriately placed intermediate cavities to enhance the bunching of the electron beam, which increases the gain. If the interaction gap of the output cavity is placed at the point of maximum bunching, power can be extracted from the density-modulated beam. The gain of a klystron might be 15 to 20 dB per stage when synchronously tuned (all cavities tuned to the same frequency), so that a four-cavity (three stage) klystron might provide over 50 dB gain.

After the bunched electron beam delivers its RF power to the output cavity, the energy of the electron beam that remains is dissipated when the spent electrons are removed by the collector. The energy dissipated by the collector is energy lost and reduces the efficiency of the tube. If the collector is insulated from the body of the tube and a negative voltage is applied to the collector, the electrons in the spent beam will have lower kinetic energy so that less heat is produced when they impact upon the collector.[5] This results in an increase in the efficiency of the tube. There is a spread, however, in the velocities of the electrons in the beam; so if the potential is too negative, some of the slower velocity electrons will be returned to the walls of the RF section of the tube and be collected as body current, with a decrease in efficiency. This problem is overcome by employing a collector with several segments insulated from one another and with different negative potentials so that electrons with different velocities can be separated and collected at their optimum potential. Figure 10.1 shows a single-stage collector, but both the klystron and TWT usually employ multiple-stage depressed collectors for greater efficiency. The multiple stages (three or four might be typical in a radar tube) are at intermediate voltages, which allow catching the spent electrons at a voltage near optimum.

According to Weil,[1] a klystron with a peak power of 1 MW requires a voltage of about 90 kV. Gains might vary from 30 to 70 dB, bandwidths from 1 to 8 percent, and efficiencies from 40 to 60 percent (with depressed collectors).

A long solenoid (not shown in Fig. 10.1) with iron shielding around its outside diameter surrounds the high-power klystron to provide an axial magnetic field that confines the electrons to a relatively long, thin beam and prevents the beam from dispersing. Cooling might have to be provided for the electromagnets. In a high-power klystron, from 2 to 5 percent of the beam power might normally be intercepted by the interaction structure, or body of the tube. If the beam were not properly confined by the external magnetic field, the stray electrons impinging on the structure of the tube could cause it to overheat and possibly be destroyed. Protective circuitry is normally employed to remove

the electron beam voltage in the event the magnetic field fails to keep the beam properly focused.

The electron beams of klystrons and TWTs also can be confined with permanent magnets. They do not require power or cooling, and the various protective circuits needed with solenoids are eliminated. Permanent magnets have been used with high power tubes, but they are quite heavy.[6,7] A significant reduction in weight, however, can be obtained with a periodic permanent-magnet (PPM) focusing system that consists of a series of magnetic lenses. These lenses employ washer-shaped disk magnets separated by iron pole pieces. The PPM replaces the uniform field of a solenoid with a periodic, essentially sinusoidal, field having the same rms value as the uniform field. Samarium cobalt is an example of a magnet material that has been widely used for tubes requiring permanent magnets. In the past, PPM focusing was usually not suited for large average-power tubes, but it has been successfully applied to high peak-power klystrons, as described next.

High-power klystrons have been used ever since the 1950s for linear accelerators to generate energetic beams of charged particles for research on the physics of high-energy particles. Many advances in klystron capability have been obtained from the development of klystrons for this purpose. These advances have, of course, been of benefit to radar as well. Although the invention of the klystron amplifier was reported in 1939, before the invention of the magnetron, it was not used or further developed significantly during World War II. It did not find its way into radar application until the development of a 20-MW peak power klystron, used in one of the first linear accelerators, was reported in 1953 by Stanford University. Thus the high-power klystron, which is a very important power source for radar, was a by-product of basic research in science. Work continued at the Stanford Linear Accelerator Center (SLAC) to develop high peak-power klystrons for electron-positron colliders. The klystron for the Next Linear Collider (NLC)[8] is at a frequency of 11.424 GHz, with a peak power of 60 to 75 MW using periodic permanent magnets made of neodymium-iron boron. The PPM with 10 pounds of permanent magnets replaces a 1/2-ton, 10-kW focusing solenoid. The NLC uses 6528 klystrons, which means that a total of 65 MW of solenoid power is avoided. The use of PPM, as well as having a tube with 60 percent efficiency, reduces "the NLC electric power bill by tens of millions of dollars per year." The tube requires an electron beam voltage of 490 kV. The average life of the *S*-band klystrons used for the previous Stanford Linear Collider (SLC) was 50,000 hours, and it is expected that a similar life will be obtained with the *X*-band NLC klystron. The manufacturing cost was said to be $30,000 per tube. If the tube delivers 75 MW with a pulse width of 1.5 μs and a prf of 120 Hz, its average power is 13.5 kW.

Bandwidth of a Klystron The frequency of a klystron is determined by its resonant cavities. When all the cavities are tuned to the same frequency, the gain of the tube is high but the bandwidth is narrow, usually a fraction of one percent for a tube of modest power output. This is called *synchronous tuning*. To maximize the klystron's efficiency the next to last (penultimate) cavity is tuned upward in frequency and is outside the passband. Although the gain is reduced by about 10 dB, the improved electron bunching results in greater efficiency and in 15 to 25 percent more output power.[9] Broadbanding of a multicavity klystron may be accomplished by *stagger tuning* the cavities, similar to the method for broadbanding a conventional multistage IF amplifier. Stagger tuning a klystron is not

precisely analogous to stagger tuning an IF amplifier because of interactions among the cavities that can cause the tuning of one cavity to affect the tuning of the others. The VA-87 four-cavity S-band synchronously tuned klystron amplifier with a 20 MHz bandwidth and a 61 dB gain can have a 27 MHz bandwidth and a 57.6 dB gain when tuned for maximum power.[10] When stagger tuned it has a 77-MHz bandwidth (2.8 percent) and a gain of 44 dB.

Theory shows that the bandwidth of a klystron can be significantly increased by increasing its power and its beam perveance (which is defined as the beam current I divided by beam voltage V to the 3/2 power, or $p = I/V^{3/2}$). A 10 MW peak power klystron, for example, can have an 8 percent bandwidth, as compared to a 200 kW tube which might have a 2 percent bandwidth, and a 1 kW tube having only a 0.5 percent bandwidth.[10] High-power multicavity klystrons can be designed with bandwidths as large as 10 to 12 percent.

Frequency Changing, or Tuning[6,9] Conventional narrowband klystrons may have their frequency changed mechanically over a relatively wide frequency range. The individual cavities of a klystron can be changed in frequency (or tuned) by having a flexible wall in the resonant cavity (tuning range of about 2 or 3 percent), by a movable capacity element in the cavity (10 to 20 percent tuning range) or by a sliding contact movable cavity wall.

It can be tedious to tune a multiple-cavity klystron because of the interactions among the several cavities. Conventional gang-tuning is complicated since the resonant cavities do not have the same tuning rates. The *channel tuner mechanism*[6] avoids the problem of frequency tracking of the resonant cavities by pretuning the cavities, generally at the factory. The tuning information is stored mechanically within the tuner mechanism. When a particular frequency is selected, the tuner mechanism provides the correct tuner position for each cavity to furnish the desired klystron frequency response. The channel tuner mechanism is in a box attached to the klystron with gears to simultaneously set the tuning plungers at each cavity to their predetermined positions for a given frequency. The tuning plungers can be actuated manually or remotely by push buttons and a servomotor. The frequency can be changed in seconds.

Power Some of the highest power radar transmitters have used klystrons. The ability of a klystron to produce higher power than other microwave power sources is, in part, due to its geometry. The regions of beam formation, RF interaction, and beam collection are separate. Each can be designed to best perform its own particular function independently of the others. The cathode, for example, is outside the RF field and need not be restricted to sizes that are small compared to a wavelength. Large cathode area and large interelectrode spacings may be used to keep the emission current densities and voltage gradients to reasonable values. The only function of the collector is to dissipate heat. It can be a shape and size best suited for satisfying the average or peak power requirements without regard for conducting RF currents, since none are present.

Efficiency It is said in Ref. 8 that Robert Symons found that the best values of efficiency reported for klystrons worldwide followed the relation

$$\text{RF efficiency (percent)} = 90 - 20 \times \text{microperveance} \qquad [10.1]$$

where the microperveance is the perveance $I/V^{3/2}$ times 10^6, where I = beam current and V = beam voltage. Thus the lower the microperveance (or perveance) the higher will be the klystron efficiency. The perveance affects other properties of the klystron, including its bandwidth and power. Higher efficiency often requires, therefore, a reduction in bandwidth and lower power.

Reliability and Life High-power transmitters employing power vacuum tubes have sometimes had the unwarranted reputation for poor reliability and short life. There is much evidence to the contrary for the klystron tube. Gilmour[11] reports the mean time between failures (MTBF) of eleven different applications of klystrons in radar systems (not identified by type or power). The MTBF for these examples varied from 75,000 hours to 5000 hours, with an average value of about 37,000 hours for all eleven applications. (There are 8760 hours in a year.) The VA-842 high-power klystron tube used by the U.S. Air Force in the original Ballistic Missile Early Warning System (BMEWS) had a demonstrated life in excess of 50,000 hours. Symons[12] reports that one of the BMEWS tubes he designed in 1958 was still operating after 240,000 hours when the radar in Greenland was replaced by the solid-state Pave Paws radar.

An S-Band Klystron The venerable VA-87 klystron built by Varian (now Communications & Power Industries) was widely used in the FAA's S-band Air Surveillance Radars commonly found at major airports. It was a six-cavity tube tunable from 2.7 to 2.9 GHz, the frequency band reserved for air-traffic control radars. It had a peak power of from 0.5 to 2 MW, average power of from 0.5 to 3.5 kW, 50 dB gain, 45 percent efficiency, and a one-dB bandwidth of 39 MHz. It demonstrated a mean-time-between-failure (MTBF) rate of 72,000 hours. A similar tube was used in the Nexrad Doppler Weather Radar, but with bandwidth from 2.7 to 3.0 GHz.

Traveling Wave Tubes (TWT) Like the klystron, the traveling wave tube is also a linear-beam tube with the cathode, RF circuit, and collector separated from one another. The klystron and the TWT were invented at different times in different parts of the world, but they are similar to one another. There is continuous interaction of the electron beam and the RF field over the entire length of the propagating structure of the traveling wave tube. In the klystron, on the other hand, the interaction occurs only at the gaps of a relatively few resonant cavities. The chief characteristic of a TWT is that it has wide bandwidth. Low power TWTs with a helix slow-wave RF structure are capable of octave bandwidths. With the high peak powers required of most radar applications, the bandwidths available with high-power TWTs are, however, much less than an octave.

The major parts of a TWT are indicated in Fig. 10.2. A helix is shown for the slow-wave RF structure even though the helix is seldom used in TWTs found in radar applications. The electron beam is similar to that of the klystron. Both the TWT and the klystron employ the principle of velocity modulation to cause the electron-beam current to be periodically bunched (density modulation). The electron beam passes thorough the RF interaction circuit known as the *slow-wave structure*, or periodic delay line. The velocity of propagation of the RF signal is slowed down by the periodic delay line so that it is nearly equal to the velocity of the electron beam. This is the reason that the helix and

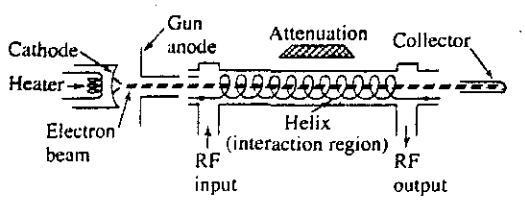


Figure 10.2 Representation of the principal parts of a traveling-wave tube.

other microwave circuits used in TWTs are called slow-wave structures. The synchronism between the electromagnetic wave propagating along the slow-wave structure and the d-c electron beam propagating inside the helix results in a cumulative interaction which transfers d-c energy from the electron beam to increase the energy of the RF wave, causing the wave to be amplified. Just as in the klystron, an axial magnetic field keeps the electron beam from dispersing as it travels down the tube.

After delivering their d-c energy to the RF field on the slow-wave structure, the electrons are removed by the collector, which is usually a multistage depressed collector, as was described for the klystron. It is easier to design a depressed collector for a TWT than for a klystron since the spent electron beam of a TWT might have a 20 percent spread in velocity, but the klystron might have a velocity spread of almost 100 percent.[5] Because the efficiency of a conventional TWT is usually lower than that of a klystron, the increase in efficiency in the TWT provided by the depressed collector has a greater relative effect than with a klystron.

Although a helix is shown in Fig. 10.2 as the slow-wave structure, it is seldom found in TWTs used for radar. The helix TWT is limited to voltages of about 10 kV and a peak power output of a few kilowatts,[1] which is generally too low for most radar applications. Other types of RF slow-wave structures have to be employed instead, and these do not have as wide a bandwidth as the helix. A modification of the helix known as the ring-bar circuit can be used if the peak power is less than about 100 to 200 kW. One example is the Raytheon QKW-1671A, a tube suitable for air-surveillance radars. It has a peak power of 160 kW, duty cycle of 0.036, 70-μs pulse width, 45-dB gain, and a 100-MHz bandwidth. The Air Force Cobra Dane phased array radar uses 96 QKW-1723 ring-bar TWTs, each with a peak power of 175 kW and average power of 10.5 kW. The Cobra Dane operates from 1175 to 1375 MHz.

Powers greater than 200 kW are obtained with the coupled-cavity circuit, of which the so-called "cloverleaf" is an example. The bandwidth, however, is less than that of lower power TWTs. The individual unit cells of the coupled-cavity circuit resemble klystron resonant cavities. Several tens of these klystron-like cavities are used for the slow-wave structure of a high-power TWT.[13] There is no direct coupling between the cavities of a klystron, but in the traveling wave tube, coupling is provided by a long slot in the wall of each cavity. There are two slots in each cavity (input and output) that are 180° apart in rotational position so they act similar to a folded waveguide.

An example of a TWT using a cloverleaf coupled-cavity slow-wave structure is the *S*-band VA-125A. It is liquid cooled and is capable of 3 MW peak power over a 300-MHz bandwidth, 0.002 duty cycle, 2-μs pulse width, and a gain of 33 dB. It was originally designed to be used interchangeably with the VA-87 klystron, except that the VA-125 TWT has a wider bandwidth and requires a larger power input signal because of its lower gain.

The bandwidth of a coupled-cavity TWT can be from 10 to 15 percent. When the power of a TWT is increased, its wide bandwidth decreases. On the other hand, when the power of a klystron is increased, its narrow bandwidth widens. As the power of these two tubes increase, their bandwidths become comparable. With high power, the klystron tends to be the preferred tube since it doesn't experience the stability problems of the TWT.

Although the TWT and the klystron are similar in many respects, a major difference is that there is feedback along the slow-wave structure of the TWT, but the back coupling of the RF energy in the klystron is negligible. If there is a mismatch at the input of the TWT and if sufficient energy is fed back to the input, undesired oscillations can result. To reduce the amount of feedback energy to an insignificant level, attenuation has to be inserted in the slow-wave structure. The attenuation may be distributed or lumped, but it is usually found within the middle third of the tube. The loss introduced to attenuate the feedback also reduces the power of the forward-traveling wave, and is therefore undesirable. The loss in the forward wave can be avoided by the use of *severs*, which are short internal terminations designed to dissipate the reverse-directed power without seriously affecting the power in the forward direction. The number of severs depends on the gain of the tube; one sever is used for each 15 to 30 dB of tube gain.[1]

The efficiency of TWTs is less than that of the klystron because of the necessity for including attenuation or severs. Efficiency is also reduced by the presence of relatively high RF power over an appreciable fraction of the entire structure.

In some traveling-wave tubes with coupled-cavity circuits, oscillations appear for an instant during the turn-on and turn-off portions of the pulse.[1] These are called *rabbit-ear oscillations* because of their characteristic appearance when the RF envelope of the pulse waveform is displayed visually on a CRT. They are undesirable in some military applications since they might provide a distinctive feature for recognizing a particular radar. Weil[1] describes some of the ways rabbit-ear oscillations can be avoided.

TWT MTBF The mean time between failures (MTBF) is given by Gilmour[11] for nine different types of coupled-cavity TWTs. (The type of TWT, frequency, and power are not mentioned.) The MTBFs of these nine classes vary from a high of 17,800 hours to 2200 hours, with an average of 7000 hours for all nine classes of tubes. TWTs for space applications, which are of lower power than radar tubes, are said by Gilmour to have MTBFs of the order of one million hours.

Hybrid Variants of the Klystron By judiciously combining the best features of the klystron and the traveling wave tube, an RF power source can be obtained which has bandwidth, efficiency, and gain flatness better than either the conventional klystron or TWT. This is achieved by replacing one or more of the klystron resonant cavities with broader bandwidth cavities that are more like the coupled-cavity circuits used in traveling wave tubes. There have been three variants of the klystron in which this is done: the *Twystron*, the *extended interaction klystron*, and the *clustered-cavity klystron*. Such combinations of klystrons and TWTs are sometimes called *hybrid tubes*.[14]

Twystron The bandwidth of a klystron is limited by the output resonant cavity. It cannot be made broadband without a decrease in efficiency. Since coupled-cavity slow-wave

circuits have broader bandwidth than klystron resonant cavities, replacing the output cavity of a klystron with a TWT coupled-cavity circuit can significantly increase the bandwidth as well as achieve a slight increase in efficiency. Although the output is a TWT slow-wave circuit, the driver portion of the tube (the intermediate cavities and the input cavity) consists of resonant cavities that are stagger tuned. Such a tube is called a Twystron, a trademark name of Varian (now Communications & Power Industries, or CPI). The VA-145 Twystron has demonstrated a 14 percent 3-dB bandwidth (12 percent 1-dB bandwidth), 48 percent efficiency, and 41 dB gain at midband.[1]

Extended Interaction Klystron (EIK) In this device, the single-gap resonant cavity of a klystron is replaced by a resonated slow-wave TWT-like circuit. The use of slow-wave coupled resonators can be applied to the prior cavities as well as the output cavity. The extended interaction amplifier (EIA) klystron can have a high average power; for example, 1 MW CW at *X* band using a five-cavity resonator.[15] It has broader bandwidth than a klystron, but less than that of a TWT. EIKs also have been used for low-power millimeter wave tubes. A 150-W average power 95-GHz klystron, advertised by CPI Canada, is claimed to have a 1.5-kW peak power, 500-MHz bandwidth, 25 percent efficiency, 45-dB gain, and to weigh 4.5 kg. There is also an extended interaction oscillator, or EIO, which has been used at millimeter wavelengths.

Clustered-Cavity Klystron The technique of grouping cavities was extended in what has been called a *clustered-cavity klystron* (CCK). In this tube, Fig. 10.3, the individual intermediate cavities of a multicavity klystron are replaced by pairs or triplets of artificially loaded low-Q cavities with Qs of one half or one third that of the single cavity they replace.[16] Similar groups of cavities are used in both the EIA and the CCK, but there is

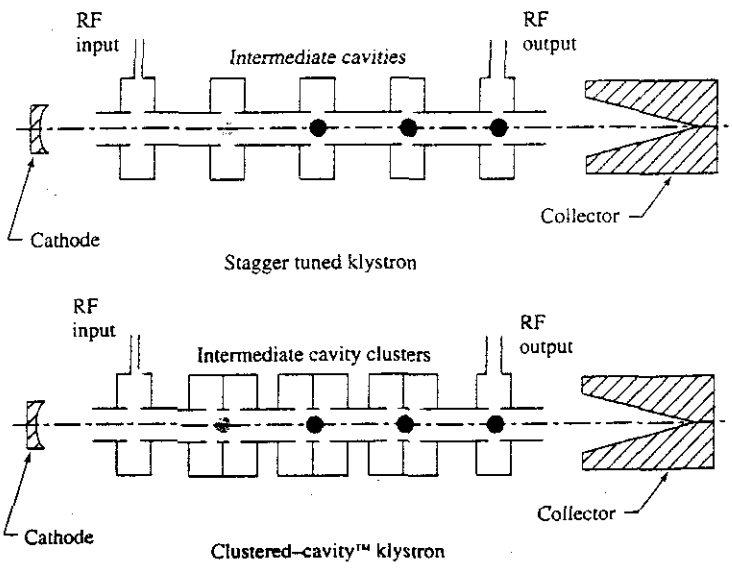


Figure 10.3 Comparison of a conventional staggered-tuned klystron (top) and a clustered-cavity klystron (bottom). The shaded circular regions represent bunching of the electrons. (From Symons and Vaughan,[16] Copyright 1994 IEEE.)

no inductive or other coupling between cavities in the CCK as there is in the EIA. Theory indicates that bandwidths of 30 percent should be obtained in megawatt klystrons using fifteen intermediate cavities in triplets. In practice, 20 percent bandwidths have been observed. This form of structure also provides the greatest bandwidth in the shortest length. Clustered-cavity klystrons might be more complex and costly than a klystron, but they are less complex and of less cost than a comparable TWT or Twystron.[1] Symons,[12] the inventor of the clustered-cavity klystron, states that two of these tubes can be used instead of the two narrower-band klystrons in the AWACS radars. Redundant operation is provided, without a large weight penalty, since either of the clustered-cavity klystrons provide full operational capability similar to the redundancy commonly employed in FAA radars.

10.3 SOLID-STATE RF POWER SOURCES

Use of Solid-State Amplifiers for Radar Transmitters The solid-state RF power generation device usually of interest for radar application has been the transistor amplifier, both silicon bipolar and gallium arsenide FET. An individual transistor amplifier device is inherently of low power and low gain, but it operates with low voltages and has high reliability.

A single microwave transistor might have an average power capability from a few watts to over a hundred watts, depending on the frequency and the duty cycle. The lower the frequency the greater can be the power. To increase the power, transistors may be operated in parallel, and with more than one stage to increase the gain. A single power *module* might, for example, consist of eight transistors, with four operating in parallel as the final stage, two in parallel as the next to last stage, and two in series as the driver stages. To achieve the high powers required for most radar applications, the outputs of many solid-state devices have to be combined in some manner. Combining of many devices can be achieved with microwave circuitry or by combining in space (radiating from many individual antenna elements of an array antenna).

Solid-state power devices of a given average power cannot be operated at high peak powers as can vacuum tubes. According to Borkowski,[17] "a microwave transistor capable of perhaps 50-W average power cannot handle much more than 100 to 200 W of peak power without overheating during the pulse." For this reason solid-state amplifiers for radar generally operate in the vicinity of 0.1 duty factor, instead of the 0.001 to 0.01 duty factors common with high-power vacuum tube RF power sources. Thus when solid-state devices are employed in radar transmitters they have long pulse widths and require pulse compression to obtain useful range resolution. Long pulses are not always desired by the radar engineer, but they have been accepted as one of the prices to be paid for the use of solid state.

There are at least four ways that solid-state devices can be employed in radar: (1) as a transmitter for a low-power application, (2) as a high-power transmitter where a large number of individual transistors are combined with microwave circuitry, (3) with many modules distributed on a mechanically steered planar array (such as a 3D radar), and (4) with a module at each of the many elements of an electronically scanned phased array

(also called an active aperture). In the last two, the power from the many solid-state transmitter modules is "combined in space."

Low-Power Transmitter The solid-state device is used as a direct replacement for a vacuum tube when the radar waveform is of low power and of high duty cycle or CW. Examples are the FM-CW radar altimeter, doppler (police) speedmeter, and the airborne doppler navigator. The solid-state transmitter has been highly successful in such applications. It has been difficult, however, for solid-state to replace the small magnetron in the civil marine radars found on many ships and pleasure boats because this radar market is highly competitive and low price is important for success. The same appears to be true for the absence of solid-state transmitters for the nonradar application of microwave ovens for the household market.

High-Power Transmitter The solid-state transmitter has replaced the high-power vacuum tube in some air-surveillance radars. A large number of transistors are combined to produce a single output that feeds a conventional antenna. (This was at one time called a "solid-state bottle," but such transmitters are housed in cabinets which do not resemble "bottles.") Two such transmitters will be described.

One of the first solid-state radars to have its tube transmitter replaced by solid state was the AN/SPS-40, a modest UHF 2D shipboard air-surveillance radar used mainly by United States ASW (antisubmarine warfare) destroyers to provide conventional air-surveillance for keeping track of ASW aircraft.[18] It was developed by Westinghouse, Baltimore (now known as Northrop Grumman). This was a good example of a direct replacement since the tube transmitter operated with a long pulse (60 μs), pulse compression, and a moderate duty cycle (0.018), so that the solid-state transmitter could utilize the same waveform as did the original radar system. The basic transistor building block operated from 400 to 450 MHz, with 400- to 500-W peak power, 8-dB gain, and 55 percent efficiency. A module consisted of two stages with a total of 10 silicon bipolar transistors that produced 2500-W peak power out when the input was 120 W. There were 112 of these modules combined in two groups of 56 each to produce 250-kW peak power and 4 kW of average power. Each of the two groups of 56 modules was housed in its own cabinet. There was a third cabinet with the driver, power supplies, and some other devices. The transmitter was designed so that no damage occurred when a full short circuit was applied across the load. Both liquid and forced air cooling were used; and if the liquid cooling was lost, the transmitter could operate with 80 percent power (200-kW peak) with only air cooling. The loss of one module reduced the transmitter power output by 0.08 dB. The transmitter had good reliability due in part to its built-in spare modules. The solid-state transmitter for this radar cost more and was larger than the vacuum tube transmitter it replaced; but it was considered a success.

The Ramp (Radar Modernization Project) radar system was an L-band (1250 to 1350 MHz) air-traffic control Primary Surveillance Radar (PSR) located at major airports across Canada.[19] It was developed by Raytheon Canada and had a range of 80 nmi and an altitude coverage of 23,000 ft against a 2 m^2 aircraft target with 80 percent probability of detection. It used a solid-state transmitter with a peak power of 28 kW and an average power of 1.2 kW, which corresponded to a duty cycle of 0.068.

There were a total of 14 modules used in the RAMP PSR. Each module consisted of 42 identical class-C 100-W peak-power silicon bipolar transistors arranged in a 2-8-32 configuration to produce 2350 W of peak power. As described by Merrill,[20] "The transistors were arranged in a one-driving-four 'unit amplifier' format, with eight unit amplifiers in parallel so that 10 transistors were drivers while 32 were output devices." The 50-pound air-cooled module had a measured efficiency greater than 25 percent and a power gain greater than 16 dB.[17]

The 14 modules were combined as pairs to form seven transmitting channels. Only six of the seven channels were needed to meet the system requirement for a minimum peak power of 21 kW. The extra seventh channel permitted maintenance and repair to be performed on a failed channel while the remaining six channels were continuously available. The extra channel, therefore, allowed the radar to maintain a high availability. [The theoretical power output when N out of 7 channels are operating is $P_{out} = P_7 (N/7)^2$, where P_7 = the power delivered by all seven channels.] The antenna for this radar was a 33 ft wide by 22 ft high reflector with 33.5-dB gain. It rotated at 12 rpm.

Modules Arranged on a Mechanically Scanned Planar Array, AN/TPS-59 Individual transmitter modules can be arranged on a mechanically scanning array antenna by placing one module at each element, but it has been more usual in such a radar to employ one module at each row of the antenna. This is the arrangement in the AN/TPS-59,[17] a transportable *L*-band 3D air-surveillance radar developed by GE, Syracuse, N.Y. (now Lockheed Martin) for the U.S. Marine Corps for air defense and ground-control of intercept (GCI). It was designed to detect a 1-m^2 fluctuating target within a 200-nmi range with 90 percent probability of detection, but it was also required to cover a volume out to 300 nmi in range and an altitude up to 100,000 ft. The rotating planar array antenna is 30 ft high by 15 ft wide and consisted of 54 rows, each with 24 dipole elements. The radar can operate within 1200 to 1400 MHz with 54-kW peak power and 9.7-kW average power, which results in the relatively high duty cycle of 0.18. A single pencil beam is electronically scanned over an elevation angle of 20° as the antenna rotates 360° in azimuth.

At each row is a transceiver, which is a miniature radar containing transmitter, receiver preamplifier, duplexer, phase shifter for steering in elevation, logic control, cooling, and power supply. Each transmitter module has ten 100-W amplifier units consisting of two 55-W silicon bipolar transistors (with 7-dB gain) driven by a smaller 25-W device. There are a total of 540 modules on the antenna.

Fixed-site variants of this radar are known as the GE-592 and the AN/FPS-117. The latter uses a 24 ft by 24 ft antenna with 44 rows. It is a minimally attended radar whose antenna operates within a radome for use in northern regions. The tactical mobile version of this radar is the TPS-117, which was shown in Fig. 1.8.

Active Aperture, Electronically Steered Phased Array An example of the composition of a T/R module as might be used for an active-aperture phased array antenna is illustrated in Fig. 10.4. The T/R switches select between the transmitter and the receiving paths. The circulator, which might be the largest and heaviest component on the module, performs the function of the duplexer. The receiver protector (which is a diode limiter) provides further protection of the low-noise amplifier (LNA). The same phase shifter is

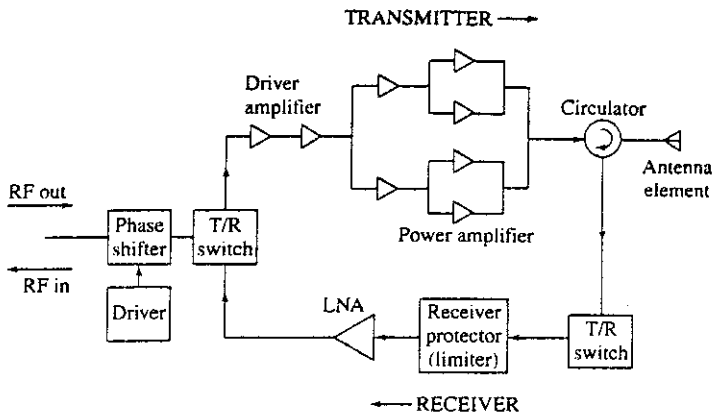


Figure 10.4 Example of the composition of a T/R module that might be used for an active-aperture phased array radar.

used for both transmitting and receiving. Many other parts of a T/R module are not shown in this illustration, such as the module controller and the power conditioner.[21] The module controller obtains a beamsteering command from a central computer and calculates the correct settings for the phase shifter. To minimize power consumption, the power amplifiers and the LNA might be gated off when the controller is on. The module controller might also perform self-testing and reporting of the status of the module. The power conditioner is important in keeping the efficiency of the module as high as practical.

The Pave Paws electronically steered array radar, also known as the AN/FPS-115, is a UHF radar developed by Raytheon that was the first all solid-state active aperture electronically steered phased array radar. Its function was to detect and warn of sea-launched ballistic missiles fired at the United States. This radar was discussed in Sec. 9.9. It operated from 420 to 450 MHz with a peak power of 600 kW and an average power of 150 kW, which corresponds to a duty cycle of 0.25. Its diameter was 72.5 ft, with room to grow to 102 ft. An individual module delivered 340 W of peak power with 39 percent efficiency.[17,22] The pulse width was 16 ms. There were two phased array faces per site to cover approximately 240° in azimuth. Each antenna face had the capability of operating with 5354 elements, but only 1792 active transceiver modules were used for transmission. Extra elements and a narrow beam were used on receive. (The remaining elements were for future growth.) A transmitting module was made up of seven Class-C silicon bipolar transistors in a 1-2-4 amplifier configuration. It has been said[22] that "system performance is maintained with as many as 200 modules per face inoperative." An enlarged version of the Pave Paws radar replaced the parabolic torus antennas used in the original Ballistic Missile early Warning System (BMEWS). This version had 850-kW peak power with a 0.30 duty cycle.

THAAD Ground Based Radar This is an example of an active-aperture phased array radar with 25,344 elements, a very large number. It was discussed in Sec. 9.9 and its picture shown in Fig. 9.36. With such a large number of elements, the cost of an individual T/R module must be kept small. If, for example, the cost of an individual T/R module were $1000 each, the total cost of just the modules for this radar would be twenty-five million dollars.

Solid-State Devices Used in Radar[17] As has been mentioned, the transistor amplifier has been the device usually used for radars with high-power solid-state transmitters. At the lower microwave frequencies, the silicon bipolar transistor is usually used; and at the higher microwave frequencies, it has been the gallium arsenide (GaAs) FET transistor. At the higher frequencies, the solid-state power device can also be incorporated as part of a microwave monolithic integrated circuit (MMIC).

The silicon bipolar transistor has been used at microwave frequencies below about 3 GHz (S band). According to Olson,[23] a typical internally matched silicon bipolar transistor operating as a class-C amplifier over the frequency range from 2.7 to 2.9 GHz with a pulse width of 50 μs, 10 percent duty cycle and a supply voltage of 40 V, can have a minimum power output greater than 100 W, a minimum gain of 6.5 dB, and a minimum efficiency of 35 percent.

The power output of the silicon bipolar transistor decreases with increasing frequency. At the higher microwave frequencies, the gallium arsenide FET, often in the form of a MESFET (metal semiconductor field-effect transistor) is capable of greater power than the silicon bipolar transistor. Transistors should be operated at a high duty cycle since the peak power output for pulsed operation cannot be significantly increased over that of CW operation. At *X* band, the power output of such devices might be 10 W. Other devices that have been considered for solid-state power sources include[23] GaAs HEMT (high electron mobility transistor); GaAs-based pseudomorphic HEMT, or PHEMT; GaAs heterojunction bipolar transistor (HBT); and devices employing unconventional materials such as silicon carbide and semiconductor diamond for high temperature, high power operation.[24]

At the higher microwave frequencies where compactness in size is desired, the microwave monolithic integrated circuit (MMIC) has been of interest for T/R modules. Active and passive circuit elements are formed on a semi-insulating semiconductor substrate, usually GaAs to create system architectures that are difficult to realize with less integrated technologies. The benefits of MMIC are due in large part to the batch processing of both the active and passive components on the same substrate. Borkowski[17] lists the advantages of MMIC for radar as low cost, increased reliability, increased reproducibility, and small size and weight. The nonrecurring costs of engineering design of MMICs, however, can be high and design might require a relatively long time. Since MMICs are not well suited for tweaking of the circuits once manufactured, the designs must be tolerant to variations in the processing. Olson[23] states that the power available from MMICs is about 10 W in the frequency range from 3 to 10 GHz and then decreases with increasing frequency at a rate of 6 dB per octave.

Advantages of Solid State The solid-state transistor amplifier has been of interest for radar transmitter applications because of the following:[17]

- Individual solid-state devices have long MTBF (mean time between failures).
- Maintenance is relatively easy with the modular construction of solid state. (A defective module is pulled out and replaced by another.)
- Very wide bandwidths can be obtained (up to 50 percent or more).
- No cathode heater is required (no warm-up time and no heater power to reduce the overall transmitter efficiency).

- Solid-state devices operate at much lower voltages (order of tens of volts) than RF power tubes (order of tens of kilovolts).
- No pulse modulator is required. (When operated as a class-C amplifier, the transistor is self-pulsing in that it automatically turns on when the RF drive signal is applied and it automatically turns off when the drive signal is turned off.)
- Solid-state transistor amplifiers have low noise and good stability (important for detection of small targets in the presence of large clutter echoes).

Another attribute often claimed for a solid-state radar is that many individual devices can fail without significant effect on the overall transmitter power (graceful degradation). The power output in dB varies as 20 log r, where r = ratio of the number of operating devices to the total number of devices.[17] This is correct in principle, but in practice there can be catastrophic failure modes for a solid-state transmitter and, eventually, the modules that fail must be replaced, even if they fail "gracefully." Except for its long pulses and high duty cycle the solid-state transistor is well suited for use in an active-aperture phased array where each element contains its own transmitter/receiver module.

When solid-state transmitters were first proposed for radar, it was said they would be lighter in weight and lower in cost than vacuum tube transmitters. It is not obvious that this has occurred. In some cases in the past when a solid-state transmitter replaced a high-power vacuum-tube transmitter in an existing radar, the solid-state transmitter was heavier and cost more.

Systems Implications of Solid-State Devices Pulse radars typically have been characterized in the past by low duty-cycle waveforms with typical values of duty cycle ranging from approximately 0.001 to 0.01. Power vacuum tubes are well suited for low duty cycles. For a given average power their peak power can be increased by a factor of 1000 or more with little penalty other than the practical problem of making the insulators able to stand off the higher voltages. Semiconductor power devices, on the other hand, cannot be efficiently operated at low duty cycles. For a given average power, the peak power might be less than ten times the average power. Thus replacing a vacuum-tube transmitter with a solid-state device usually means the radar must use high duty-cycle waveforms. High duty cycles mean long pulse widths which have the disadvantage of long minimum ranges. When a short minimum range is important, more than one pulse width might have to be transmitted. The long pulses of solid-state transmitters require pulse compression to achieve good range resolution. The technology of pulse compression has been widely used in radar, but it does have some limitations that short-pulse waveforms do not have. Seldom, however, is the cost of pulse compression and the cost of multiple waveforms considered as a solid-state transmitter cost, even though they increase the total cost of the radar system and are not needed with many vacuum-tube transmitters.

There are at least two reasons why the cost of radars with solid-state transmitters is often higher than those with comparable vacuum-tube transmitters.[12] One is that the efficiency of solid-state transmitters is generally less than that of a high-power vacuum-tube transmitter. The other is that the cost of obtaining power is greater when the total power is obtained with more than one power-source unit. An advantage of low-power solid-state devices for computers and low-power transmitters is that they can be made very compact.

A lot can be placed on a small chip. When the solid-state device has to handle high power, however, as it would for a radar transmitter, there can be a problem in dissipating the heat generated by the power sources. Solid-state devices that have to handle high power have to be spread out over a greater extent of circuit board area to avoid exceeding the heat transfer limits. The size and weight of the solid-state amplifier are therefore determined by the power densities that the amplifier can handle rather than by the size of the individual components. Thus the advantages of photolithographic fabrication of high-density low-power solid-state integrated circuits are not available when the power is high. Spreading the solid-state devices over a wide area in order to provide for heat dissipation can result in lower efficiency due to higher power-combining losses in transmission lines and in combiners. Dissipation of the higher heat levels requires heavier heat sinks and results in a heavier transmitter.[12]

It has been known for a long time that the cost of a single high-power vacuum tube varies as the square root of the output average power. Thus, the lowest cost tube transmitter for a given total power output is the transmitter that uses a single high-power tube rather than multiple tubes. Symons[12] indicates that the cost of a transmitter made up of multiple devices, such as is necessary when using high-power solid-state devices, increases almost linearly with the number of devices. At high powers, therefore, it should be expected that the single vacuum-tube transmitter will be of lower cost than a solid-state transmitter made up of many modules generating the same total power.

Individual solid-state devices can have a much longer life and lower failure rate than an individual vacuum-tube power source. The life of a solid-state transmitter, however, is determined not by the life of a single transistor or a single module but by the life of all the modules and the many other components that make up a transmitter. Vacuum-tube transmitters, when well designed and properly operated, have been known to achieve very long lifetimes, mentioned previously in this chapter. Maintenance of a solid-state transmitter should be easier than maintenance of a high-power vacuum tube transmitter, but the life of a properly designed and operated vacuum tube transmitter should be quite long and not be a serious system problem.

Compared to the high-power vacuum tube, the solid-state transmitter has advantages, but it also has some serious limitations. It is not obvious that current solid-state transmitter technology will cause the high power RF vacuum tube to disappear. Solid-state will be used when its particular advantages are more important than its limitations and higher cost. As happens often in engineering, the radar systems engineer has a number of choices when it comes to selecting the type of RF power source to use for a particular radar system application. The solid-state transmitter is just one of many possibilities that have to be considered for any particular application—unless the customer insists otherwise.

10.4 MAGNETRON

The magnetron has been the only high-power RF power source used for radar that is a power oscillator rather than a power amplifier. It is a crossed-field device in that its electric field and its magnetic field are perpendicular to one another. The compact size and efficient operation of the magnetron at microwave frequencies allowed radars to be

small enough to fly in military aircraft, be mobile for ground warfare, and even be used on submarines.

Coaxial Magnetron A major improvement in the power, efficiency, stability, and life of the original magnetron architecture came about with the coaxial magnetron introduced in the 1960s. The key difference was the incorporation of a built-in stabilizing cavity surrounding the conventional magnetron. Figure 10.5 is a sketch of the cross section of the circular geometry of a coaxial magnetron. At the center is the "fat" oxide-coated cathode. Surrounding the cathode are a number of RF resonant cavities defined by the radial vanes. Between the cathode and the resonant cavities is the interaction space where the electrons interact with the d-c electric field and the static magnetic field in such a manner that the electrons give up their d-c energy to the RF field. The crossed electric and magnetic fields cause the electrons to be "bunched" almost as soon as they are emitted from the cathode. After bunching, the electrons move along in a traveling-wave field that is almost the same speed as that of the electrons.

The frequency of a coaxial magnetron can be changed by mechanically moving one of the end plates, called a tuning piston, of the stabilizing cavity. (The end plate is located in the plane of the paper of Fig. 10.5 and is not shown.) The tuning piston can be positioned mechanically from outside the vacuum by means of a vacuum bellows.

There is also an inverted form of the coaxial magnetron (an inside-out) version with the anode and resonant cavities in the center and the cathode around the outer perimeter of the tube. It is supposed to provide better performance at higher frequencies when the cavity becomes small and the regular type of coaxial magnetron would result in a small cathode.

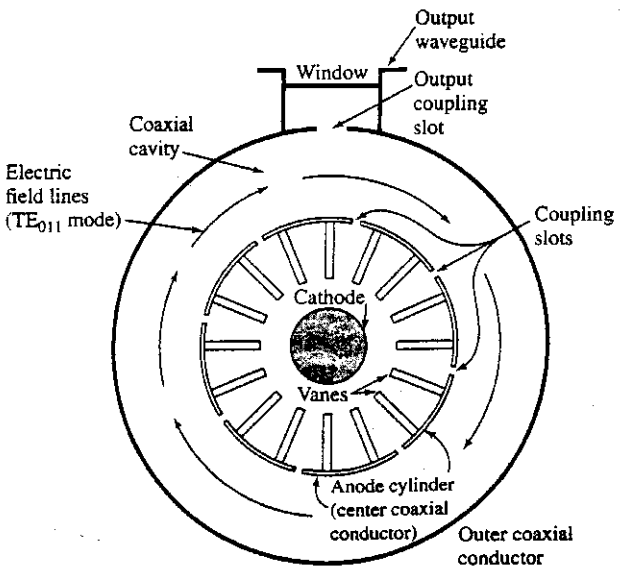


Figure 10.5 Cross-sectional sketch of a coaxial cavity magnetron.

π Mode It is not easy to describe the theory of operation of a magnetron in a simple manner, so no attempt is made to do so here. A magnetron can oscillate at a number of different, closely spaced frequencies due to various possible configurations of the RF field that can exist between the cathode and the resonant cavities. These different RF field configurations, along with coupling among the many cavity resonators of the magnetron, result in different modes of oscillation. The magnetron can shift, almost unpredictably, from one mode to another (which means from one frequency to another) as the voltage changes or as the input impedance that the magnetron sees changes. The shift from one mode to another (often called *moding*) is especially bad since it can occur when the radar antenna scans and views different environments. It is important to avoid moding.

The preferred magnetron mode of operation is the so-called *π mode* that occurs when the RF field configuration is such that the RF phase alternates 180° (π radians) between adjacent cavities. The advantage of the π mode is that its frequency can be more readily separated from the frequencies of the other possible modes. (An N-cavity magnetron has $N/2$ possible modes of oscillation. The π mode oscillates at only a single frequency, but the other modes can oscillate at two different frequencies, so that the magnetron can oscillate at a total of $N - 1$ different frequencies.)

In the coaxial magnetron, the output of every other resonant cavity is coupled to a stabilizing cavity that surrounds the anode structure, as indicated in Fig. 10.5. The output power is then coupled from the stabilizing cavity. The cavity operates in the TE_{011} mode with the electric lines closed on themselves and concentric with the circular cavity. The RF current at every point on the circumference of the cavity has the same phase, so that the alternate slots which couple to the stabilizing cavity are of the same phase as required for π-mode operation.

Coaxial Magnetron Life The power that can be produced by a magnetron depends on its size. A larger size means more resonators, which then makes it more difficult to separate the various modes of oscillation in a conventional magnetron. The coaxial magnetron, however, with stabilization controlled by the TE_{011} outer cavity permits stable operation with a larger number of resonant cavities, and thus with greater power. The anode and cathode structures of a coaxial magnetron can also be bigger than those of a conventional magnetron, which further allows operation at larger powers. The larger structures permit more conservative design, with the result that the coaxial magnetron exhibits longer life and better reliability than conventional magnetrons, in addition to having a more stable operation. The operating life of a coaxial magnetron tube can be between 5000 and 10,000 hours, which represents a five- to twenty-fold increase compared to conventional magnetrons.[25] It has been said[26] that a VMS-1143 S-band coaxial magnetron operating at 3 MW of peak power in an AN/FPS-6 height-finder radar exceeded a life of 50,000 hours. This tube was likely one-of-a-kind, but it indicates that a magnetron does not necessarily have to have a short life, as once was the case.

Just after World War II and during the 1950s, the life of the early magnetrons was as low as 200 hours mean time between failures, which probably explains why some have had the impression that power vacuum tubes were inherently unreliable and of short life. The demonstrated long lifetimes of the coaxial magnetron and even longer lifetimes of the modern klystron and TWT linear-beam tubes offer ample evidence that the use of

power vacuum tubes need not result in unreliable radars. (The reader might try to recall how often, when watching TV or listening to the radio, the transmission shut down and went off the air because of a failure in the tube transmitter.)

Systems Aspects of the Magnetron The magnetron transmitter was used in a large number of different types of radar transmitters. At one time it was the most popular radar transmitter. Its use, however, has diminished greatly because of the more demanding requirements of modern radars that it cannot readily meet, but which can be satisfied by other RF power sources. Weil[27] describes the problems encountered with the use of the magnetron in his chapter "Transmitters" in the *Radar Handbook*. The major limitations of the magnetron are its limited average power and poor ability to see moving targets in heavy clutter.

Although the magnetron can produce a peak power of several megawatts, its average power is limited to about one or two kilowatts. This may be sufficient for some medium-range radars and for civilian air-traffic control radars that use large antennas, but it is not large enough for many military radar applications. The magnetron is usually smaller in size than many other types of RF power sources, but this is due in part to its low average power.

During World War II, air-surveillance radars did not have an MTI (moving target indication) capability, except for the S-band AN/CPS-1 MEW (Microwave Early Warning) radar which only became available in small numbers near the end of the war. In order for radar to detect aircraft in World War II, targets had to be in the clear outside of the clutter. The bomber aircraft of that time unknowingly cooperated since they were not designed to fly at low altitude below the radar coverage. The early analog MTI radars could use only a single or a double delay-line canceler, so they were limited in MTI improvement factor (or clutter attenuation) to about 20 dB. The magnetron itself limits the improvement factor to perhaps 30 or 40 dB. Therefore, the magnetron was not the limiting factor in early MTI radar performance. This changed when digital signal-processing of MTI signals became available that allowed much better values of MTI improvement factor. Now power amplifiers such as the klystron, TWT, or the transistor have to be used—and not the magnetron—in order to detect small moving targets in heavy clutter, consistent with the full capabilities of digital doppler-signal-processing.

There are several other factors that are not favorable with the magnetron. Its pulse widths are limited from just under 0.1 μs to about 100 μs, but this is usually not a problem. However, modulating the pulse with frequency or phase to achieve pulse compression is quite difficult with a magnetron and has not been done operationally. The signal is not coherent from pulse to pulse so that in MTI radar the phase of the coho (reference oscillator) in the receiver has to be reset every time a new pulse is transmitted. The magnetron frequency drifts with time, which requires that the frequency of the local oscillator in the receiver be continually tuned to the transmitter's frequency (whatever it might be). Magnetrons are noisy and can produce electromagnetic interference at frequencies outside their design frequency range.

Marine Radar Magnetrons The magnetron has proven to be a tube well suited for civil marine radars. These are small devices that generate peak powers between 3 and 75 kW

with low average powers of a few watts to a few tens of watts. The marine radar customer demands reliability. When a commercial vessel goes to sea, its captain wants the radar to still be operating when the ship returns to its home port. An example of a magnetron for a civil marine radar is the third generation MG5241 manufactured by EEV of Chelmsford, England. It is an 18-cavity *X*-band magnetron, which produces a peak power of 12.5 kW with an efficiency of 43 percent. The manufacturer claims an expected typical life of over 10,000 h and guarantees a minimum life of 3000 h. Its weight is 625 g (1.4 lb) and it has a volume of 315 cm^3. It operates at a fixed frequency within the band 9.38 to 9.44 GHz.

10.5 CROSSED-FIELD AMPLIFIERS

The crossed-field amplifier (CFA) resembles the magnetron in that it employs a magnetic and electric field that are perpendicular to one another.[1,28] It is also similar in appearance to the magnetron, except that the RF circuit is interrupted to provide the input and output connections, Fig. 10.6. CFAs have high efficiency (40 to 60 percent), use lower voltage than linear-beam tubes, and are lighter in weight and smaller in size. They are of wide band (10 to 20 percent), with high peak and average power, and have good phase stability; but their gain is relatively low. They are sometimes used with a high-gain, but lower power, TWT that serves as the driver for the CFA.

CFA Operation There are several different types of crossed-field amplifiers, and they all employ a slow-wave circuit, cathode, and input and output ports. For radar, CFAs usually have the form diagrammed in Fig. 10.6, which is reentrant with distributed emission. Distributed emission means that, like the magnetron, the cathode is adjacent to the full length of the RF structure. Electrons are emitted from the cylindrical cathode, which is coaxial to the RF slow-wave circuit that acts as the anode. The electrons, under the action of the crossed electric and magnetic fields, form into rotating electron (space-charge) bunches, or spokes. These bunches drift along the slow-wave circuit in phase with the RF signal

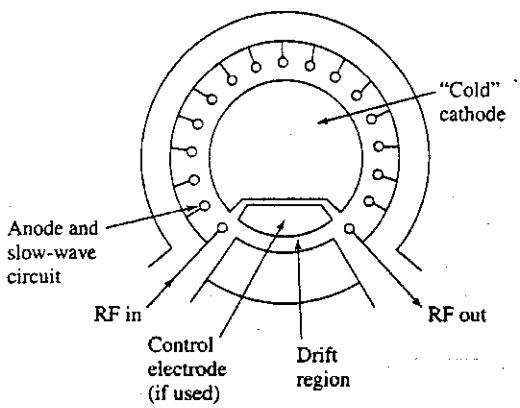


Figure 10.6 Simple representation of a crossed-field reentrant CFA showing the drift region and the control electrode.

and transfer their d-c energy to the RF wave to produce amplification. The spent electrons that remain after their energy is extracted are collected by the slow-wave anode structure. The electrons that are not collected after their energy is extracted at the output are permitted to reenter the RF interaction area at the input, which is the reason such a tube is called reentrant. Some of the reentering electrons contain modulation (bunched electrons) that will be amplified in the next pass around the RF circuit. To prevent this, a drift space is included between the output and input ports. In the drift space, space charge forces cause the electron bunches to disperse, removing any modulation that accompanies the reentering electrons.

Cold-Cathode Emission In high-power CFAs, the electrons can be generated by cold-cathode field emission rather than thermionic emission with a heated cathode. Cold-cathode emission requires the presence of both the d-c voltage between cathode and anode as well as the RF drive signal applied to the tube. The initiation of electron-emission build-up results from the relatively small number of free electrons present near the cathode surface. Emission is sustained by those electrons not collected by the anode and which are returned to the cathode by the action of the RF field and the crossed electric and magnetic fields. When these electrons strike the cathode they produce secondary electrons that maintain the electron emission process. There is little pulse-to-pulse time jitter in the starting process, and the buildup time is quite rapid (<10 ns).[28]

Insertion Loss A CFA has low insertion loss, perhaps less than 0.5 dB. Sometimes this can be an advantage in a multistage transmitter. By omitting the application of d-c voltage to the final stage of a multistage transmitter, the lower level RF drive can be fed directly through the final stage with little attenuation. This allows a radar with such a transmitter to have two power levels, which might be of interest for some system applications.

The low insertion loss of a CFA means that the RF drive power will appear at the output tube with little attenuation. In a low-gain amplifier, such as the CFA, the input power that appears at the output can be a sizable fraction of the total, perhaps one-tenth or more. The *conversion efficiency* of a CFA is defined as

$$\text{Efficiency} = \frac{\text{RF power output} - \text{RF drive power}}{\text{d-c power input}} \qquad [10.2]$$

When the RF drive power is included in the output power rather than omitted as it is in Eq. (10.2), it is sometimes called the *power-added efficiency*. That is, power-added efficiency is the total RF power out divided by the d-c power in. Tube engineers like to quote the power-added efficiency instead of Eq. (10.2), since it results in a higher value.

Forward- and Backward-Wave CFAs The interaction of the electron bunches with the RF signal may be with either a forward traveling wave or a backward traveling wave. The type of interaction is determined by the slow-wave circuit. A forward-wave interaction takes place when the phase velocity and the group velocity of the propagating signal along the slow-wave circuit are in the same direction. (The *group velocity* is the velocity with which energy is propagated along the slow-wave circuit, and the *phase velocity* is the velocity of the RF signal on the slow-wave circuit as it appears to the electrons.) To achieve

amplification, the phase velocity must be near the velocity of the electron stream. A backward-wave interaction, as in the CFA device known as the *amplitron,* takes place when the phase velocity and the group velocity are in opposite directions. The forward-wave CFA can operate over a broad range of frequency with a constant anode voltage, and with only a small variation in the output power. On the other hand, the power output of a backward-wave CFA, with a constant anode voltage, varies with frequency. It is like a voltage-tuned amplifier. The power output can vary 100 percent for a 10 percent change in frequency.[29] It is possible, however, with conventional modulator techniques to operate a backward-wave CFA over a wide band with little change in output power. The line-type modulator can be compensated for the power variation with frequency and can hold the variation of output power within acceptable levels. When the anode voltage is properly adjusted, the bandwidth of a backward-wave CFA might be 10 percent. Operation of a forward-wave CFA is more like that of a TWT, and it can obtain bandwidths up to about 20 percent.[28]

High-Gain CFA[1,30] The gain of conventional pulsed crossed-field amplifiers is typically 8 to 15 dB. By designing the cold cathode as a slow-wave circuit, and introducing the RF drive at the cathode emitting surface, it is possible to achieve 15 to 30 dB of gain in a high-power pulse CFA with power, bandwidth, and efficiency comparable to conventional designs. The RF output is taken from the anode slow-wave circuit. This is known as a *high-gain CFA* or a *cathode-driven CFA.* Its higher gain means that less drive power is required. The cathode-driven CFA can also be made to have lower noise than a conventional CFA by 10 to 20 dB if a slightly different configuration is used. It has been said[31] that signal-to-noise ratios greater than 70 dB/MHz were obtained, which are claimed to be comparable to linear beam tubes and 20 to 30 dB better than conventional CFAs. However, both high gain and low noise cannot be obtained simultaneously with the same configuration.

A 1.2-MW peak power *S*-band cathode-driven CFA operating from 3.1 to 3.5 GHz (12 percent bandwidth) achieved 23-dB gain at saturation and 30 dB at reduced peak power, with an efficiency of 53 percent when employing a drive power of 6 kW.[29,32]

Modulating a CFA The CFA can be pulsed-modulated by turning on and off the high voltage between the anode (at ground potential) and the cathode (at a large negative potential). This is called *cathode pulsing.* A forward-wave CFA with cold-cathode electron emission, however, can be pulsed without the need for a high-power modulator as required for cathode pulsing. The d-c operating voltage is applied continuously between cathode and anode. The tube remains inactive until the application of the RF input pulse starts the electron emission process, and amplification then takes place. At the end of the RF drive-pulse the electrons remaining in the tube must be cleared from the interaction area to avoid causing feedback that generates oscillation or noise. In reentrant CFAs, the electron stream can be collected after removal of the RF drive-pulse by mounting an electrode in the cathode, but insulated from it, in the region of the drift space between the RF input and output ports. This is called a *cutoff,* or *control electrode,* and was indicated in Fig. 10.6. A short positive pulse is applied to the cutoff electrode at the termination of the pulse to quench the remaining electron current. This method of modulation is called *d-c*

operation Weil[1] has said that d-c operation has been seldom used because it requires a large capacitor bank to limit droop of the pulse and because an arc in a d-c operated tube requires a crowbar protective device to quench the arc, which interrupts the operation for a few seconds instead of for only a single pulse.

It is also possible to turn the CFA on and off with just the RF drive pulse, without the need for a positive pulse applied to the cutoff electrode at the end of the drive pulse. The is called *RF keying*, and is a simple method for modulating a CFA. It has not been widely used, however, since there are factors other than modulator size that determine the best method of modulation.[1]

System Implications CFAs in the past have been employed for high-power air-surveillance radars, for power applied at the subarrays of a high-power phased array radar, and as a power booster following a magnetron oscillator. The low gain of CFA, however, requires that there be multiple stages. When used in an amplifier chain, the CFA is generally found in only one or two of the highest power stages. It can be preceded by a medium-power, high-gain traveling-wave tube, a combination that takes advantage of the best qualities of both tube types. The TWT provides high gain and the CFA allows high power to be obtained with high efficiency and lower voltage.

The electrons in the rotating space-charge bunches do not have identical velocities so that random currents are induced in the slow-wave structure, which generates broadband noise. The noise levels in a CFA, therefore, will be higher than those of a linear-beam tube by about 20 to 30 dB.[32]

The possibility of pulsing the CFA in a simpler manner than with a cathode pulser has been an attractive feature; but d-c operation and RF keying have limitations that have made cathode-pulsing preferred even though it is heavier. Since high voltage remains on the tube between pulses with d-c operated CFAs, serious levels of noise may be generated even though there is only a small amount of beam current flowing through the tube. With cathode pulsing, on the other hand, high voltage is removed between pulses so that noise is not normally encountered. The increased interpulse noise of the CFA without a cathode pulser, as well as its high level of in-band spurious noise can prevent attaining good MTI performance (large MTI improvement factors) and low time-sidelobes in pulse-compression systems.

The CFA does not seem to generate as much interest as it once did, and it appears to have been overtaken by linear-beam tubes when high performance is required.

10.6 OTHER RF POWER SOURCES

In addition to the RF power sources already mentioned, there have been several other RF devices that been used or proposed for radar application.

Microwave Power Modules (MPM)[33] The microwave power module, outlined in Fig. 10.7, combines in a single unit a solid-state MMIC (monolithic microwave integrated circuit) amplifier driving a moderate-power helix traveling-wave tube, along with an integrated

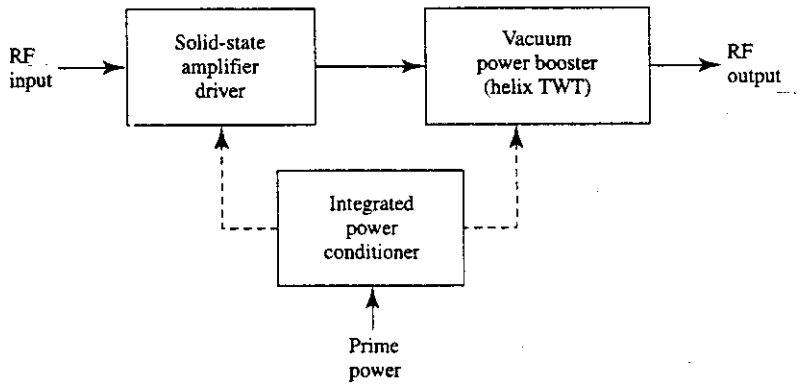


Figure 10.7 Microwave power module.

power conditioner in a compact lightweight package. It provides an RF power source with high efficiency, wide instantaneous bandwidth, low noise, and average power levels from several tens to several hundreds of watts. It is smaller and lighter than comparable TWT and solid-state RF power sources and is also capable of operating at high ambient temperatures. The gain of an MPM (nominally 50 dB) is divided between the solid-state driver and the TWT power booster in the ratios from 20/30 to 30/20. The MPM is claimed to be competitive to a TWT or a solid-state power amplifier for radar applications as well as for electronic warfare. It leverages the advantages of both solid-state and vacuum-tube technologies while minimizing their disadvantages.

The MPM seems best suited for the higher microwave frequencies, from S band to K_u band, approximately 2 GHz to 40 GHz. It is expected that an MPM operating over a frequency range from 6 to 18 GHz would produce a peak power of 100 W (up to a 100 percent duty cycle) and be able to be packaged in a space 5/16 in × 4 in × 6 in. The MPM is claimed to have an efficiency three times better than that of a comparable solid-state power amplifier and an improvement in noise 30 dB better than a TWT alone. It has been suggested that the MPM can be employed in *C*- and *X*-band phased arrays, synthetic aperture radars for unmanned airborne vehicles (UAV), missile seekers, and airborne turbulence-warning radars.

Although the MPM has been considered mainly for wideband (nonradar) applications operating over one or more octaves, it can also be used for radar with more usual bandwidths. An example is a C-band module described by Smith, Armstrong, and Duthie[34] which can operate over the band from 4 to 6 GHz with an efficiency greater than 35 percent and CW power greater than 125 W. If its operation is restricted to frequencies from 4.6 to 4.9 GHz, the output CW power is 200 W and its efficiency is 50 percent. It is of low weight and small size, producing more than 70 W per pound and 4 W per cubic inch. A limitation for radar application is that it requires a high duty factor (preferably unity or perhaps 50 percent).

Grid-Control Tubes This is the microwave version of the classical triode or tetrode vacuum tube. These tubes employ a cathode to generate electrons, an anode to collect them, and one (if a triode) or two (if a tetrode) control grids in between. A voltage applied to

the control grid acts as a gate, or valve, to control the number of electrons traveling from the cathode to the anode. By varying the voltage on the control grid, the number of electrons that reach the plate also varies. The process by which the electron density of the electron stream is modulated by the signal on the control grid to produce amplification is called density modulation. The grid-controlled tube is capable of high power, wide bandwidth, good efficiency, and inherent long life; but it is of low or moderate gain. It can be used only at the lower radar frequencies.

The performance of density modulated grid-control vacuum tubes is limited by the time it takes for the electrons to transit from the cathode to the anode. The transit time should be small compared to the period of the RF signal to be amplified. For this reason, grid-control tubes have been limited to frequencies below 1 GHz. To minimize transit-time effects, the complete RF input and output circuits and the electrical interaction system are within the vacuum envelope.

Grid-control tubes have been used with success in HF over-the-horizon radar, and in VHF and UHF radars, including the U.S. Navy's E2C airborne early warning radar,[35] and the U.S. Air Force's AN/FPS-85 satellite surveillance radar.[36] Outside of radar, the grid-control vacuum tube has been widely used in commercial VHF and UHF TV transmitters. It is not likely, however, that the grid-control tube will be used much for radar in the future. Solid-state transmitters, even though they may cost more than vacuum tubes, seem to be preferred by those who buy radars at the frequencies where grid-control vacuum tubes have been used previously.

Inductive Output Tube (IOT), or Klystrode[3] This device dates back to work in 1939 by Andrew Haeff,[37] who tried to extend density modulated vacuum tubes to microwave frequencies. Haeff called his device an *Inductive Output Tube* (IOT). He described an IOT that produced 100-W CW power at 450 MHz with 35 percent efficiency and 10-dB gain. This was quite good for its time. Nothing further happened since interest then was mainly on velocity modulated linear-beam tubes.

The IOT was independently reinvented about 30 years later by D. H. Preist and M. B. Shrader under the name *Klystrode*. The name was chosen to signify that the device resembled the klystron in the region between the anode and the collector and it resembled a tetrode in the region between cathode and anode. According to Priest and Shrader,[38] Haeff realized that the conventional triodes and tetrodes were limited by their use of intercepting grids, so in his IOT he replaced the wire grids with an aperture that did not intercept the electrons. A coaxial magnetic field confined the electron stream, as in a klystron or TWT. The action of the grid was to density modulate the electrons, as in a conventional triode, to form bunches of electrons. RF energy was removed from the bunched electron beam by passing it through a resonant cavity that extracted the kinetic energy of the high-velocity electrons. The spent electrons were not collected by the resonators, but by a separate collector. The IOT was like the klystron, except that the density modulation of the electron beam was performed by a grid rather than by an input resonant cavity and drift space that induced velocity modulation on the electrons as in the klystron.

The Klystrode was developed mainly for UHF TV. It can produce many tens of kilowatts average power at high efficiencies (50 to 60 percent) with power gains of 18 to

25 dB. Although it has not been used in radar, it has the potential to provide better performance for UHF radars than the tetrodes used previously.

Constant-Efficiency Amplifier (CEA) One reason for mentioning the IOT and the Klystrode in the above is that they can be modified to produce an amplifier whose efficiency is approximately independent of the power output. Such a device would be of interest for radar when it is desired to shape the radar pulses in order to reduce the time sidelobes of pulse-compression waveforms or to reduce the out-of-band interference generated by a rectangular waveform. Conventional radar RF power sources, such as discussed in this chapter, cannot operate with shaped pulses without a loss in efficiency. A so-called *Constant Efficiency Amplifier* (CEA), however, can be obtained by combining the Inductive Output Tube (IOT) with a multistage depressed collector similar to that used in klystrons and TWTs.[12] The CEA was developed for the television industry. It is claimed that with a CEA the prime power requirements of a TV transmitter can be reduced to one-half compared to a conventional tube transmitter[39] and to one third the prime power of a silicon-carbide solid-state transmitter.[40] The CEA, however, does not operate at frequencies higher than UHF.

Gyrotrons[41–43] The power available from the microwave-radar power sources discussed thus far in this chapter decreases as the frequency is increased. This is due to the resonant structures of the devices becoming smaller with increasing frequency (decreasing wavelength) and the difficulty in removing heat dissipated in small structures. Consequently the power output of a particular type of generator varies approximately inversely as the square of the frequency. The *gyrotron*, on the other hand, does not have this type of frequency dependence since it does not employ a resonant slow-wave structure. Instead, it is based on a fast-wave structure such as a smooth circular tube (one where the phase velocity of the electromagnetic wave is greater than the speed of light). The diameter of the gyrotron circuit can be several wavelengths and the electron beam need not be placed close to the RF structure. Since the size limitations of conventional microwave power sources with resonant circuits are not present in gyrotrons, their power handling capability can be considerably greater. The gyrotron is of interest as a potential source of high peak and average power at millimeter wavelengths. It has also been considered for operation at microwave frequencies, but it has not been able to compete with more conventional microwave power sources.

The gyrotron, also known as a *cyclotron resonance maser*, employs a strong externally applied axial magnetic field to cause electrons within the circular fast-wave structure to rotate at the electron cyclotron frequency, which is $\omega_c = eB_0/m\gamma$, where e = electron charge, m = electron rest mass, B_0 = axial magnetic field, γ is the relativistic factor which is $[1 + (e/mc^2)V_0]$, c = velocity of light, and V_0 = beam voltage. The beam voltage and the corresponding electron velocity are high enough to cause relativistic effects. The electrons follow helical paths around the lines of the magnetic field in the presence of an electromagnetic wave with a transverse component of electric field. The electrons become phase-bunched in their cyclotron orbits as a result of the relativistic mass change of the electrons. Electrons that lose energy to the electromagnetic wave become lighter and accumulate phase lead and catch up with the electrons that gain energy and become heavier and accumulate phase lag.

The frequency of the gyrotron is determined by the magnetic field and not by the characteristic size of the interaction region as it is in microwave power tubes. The magnetic field must be quite large in order to generate cyclotron oscillations at the higher frequencies. For this reason, the magnets used for millimeter wavelength gyrotrons are usually superconducting, which can be a burden for some applications, especially if the device has to be operated in a cryostat at liquid helium temperatures. However, the development of magnets based on high-temperature superconducting materials and efficient closed-cycle refrigerators, or cryocoolers, offers the possibility of using supercooled gyrotrons in mobile platforms such as ships and aircraft.

Since the diameter of the gyrotron RF circuit is normally large compared to the wavelength of the electromagnetic wave it generates, it is possible to have a higher order mode or multiple modes of the electromagnetic field. Operation in more than one mode can result in operation at more than one frequency. The design of a gyrotron requires that care be given to insure stable, single-frequency operation in only one mode.

The gyrotron oscillator at 94 GHz can produce CW power greater than 100 kW and peak pulse power of 1 MW, with efficiencies of about 30 percent. A quasi-optical gyrotron was tunable from 80 to 130 GHz (a half octave) by varying the magnetic field.[44] The power was relatively constant over this frequency range, averaging about 60 kW.

The gyrotron can be operated as an amplifier as well as an oscillator. Generally, more power can be obtained from a gryrotron oscillator than the gyrotron amplifier, but the amplifier might have some advantages in radar when doppler processing is important—just as at microwave frequencies. The electron beam of a *gyroklystron* passes through two or more resonant cavities with standing-wave fields separated by drift spaces. A *gyroTWT* is one which operates with traveling wave fields, similar to a microwave TWT. The gyroklystron has a smaller relative bandwidth than the gyroTWT, but it is capable of higher gain, efficiency, and output power. A *gyrotwystron* operates similarly to a microwave twystron in that it uses standing-wave fields to bunch the electrons and a traveling-wave field to extract the energy. The relative bandwidths of millimeter-wave gyrotron amplifiers are generally less than the relative bandwidths that can be obtained with microwave power amplifiers.

The specifications for a particular experimental gyroklystron designed for radar operation at a center frequency of 94 GHz required that it have an average power of 10 kW, peak power of 80 kW, efficiency of 20 percent, and bandwidth of 600 MHz.[45]

Multiple-Beam Klystrons[46] In a conventional klystron with a single electron beam, the power can be increased by increasing the already high beam voltage. Instead of increasing the beam voltage to obtain greater power, it is possible to employ many electron beams that pass through individual channels located in a single multichannel drift tube. The total power is the sum of the power extracted from each of the lower-current electron beams. The number of beams has been from 6 to 61. Such a power generator is known as a *multiple-beam klystron* (MBK).

The significant reduction in beam voltage results in reduced size and weight compared to a conventional klystron of comparable power. Its magnet and power supply are smaller and lighter. The geometry of the multiple beams of the MBK allows an increase in bandwidth because of an increase in perveance. The lower voltage also can eliminate the need for lead shielding to screen against X-ray radiation.

The high-power multiple-beam klystron was first seriously examined in the U.S. in the 1950s, but interest was not sustained because the needs for high power were satisfied by other more conventional types of klystrons. Tube engineers in the Soviet Union began seriously investigating the MBK in the 1970s, and were successful in producing RF power sources that have been widely used in Russian radar systems. An example of an MBK produced and marketed by the Russian Company ISTOK is their IKS-9007, a six-cavity, 36-beam klystron. It operates at 3.3 GHz with a 200-MHz bandwidth (6 percent), peak power from 500 to 800 kW, duty cycle of 0.02, gain of 40 dB, and an efficiency of 40 to 50 percent. The beam voltage is 28 to 32 kW. The klystron tube weight is 25 kg and the solenoid magnet weight is 95 kg, which is said to be 2 to 3 times less than single-beam klystrons of similar performance.

ISTOK has also applied the multiple-beam technology to the traveling wave tube and the inductive output tube.

10.7 OTHER ASPECTS OF RADAR TRANSMITTERS

Pulse Modulators[47,48] The function of the modulator is to turn the transmitter on and off to generate the desired waveform. When the waveform is a pulse, the modulator is sometimes called a *pulser.* Each RF power source has its own particular characteristics that determine the type of modulator to be used. The magnetron modulator, for instance, has to handle the full pulse power. The transistor amplifier, on the other hand, requires no modulator at all since the transistor is turned on and off by the presence or absence of the input pulse. The full power of the klystron and the traveling-wave tube can be switched by a modulator handling only a small fraction of the total electron-beam power if the tubes are designed with a modulating anode or a shadow grid. The crossed-field amplifier (CFA) is often cathode-pulsed, requiring a full-power modulator. Some CFAs, however, can be d-c operated, which means they can be turned on by the start of the RF input pulse and turned off by a short, low-energy pulse applied to a cut-off electrode. Similar to the transistor amplifier, some lower power CFAs require no modulator since they can be turned on and off by the start and stop of the input RF pulse.

The basic elements of one type of radar modulator is shown in Fig. 10.8. Energy from an external power source is accumulated in the energy-storage element at a slow rate during the interpulse time. The charging impedance limits the rate at which energy is delivered to the storage element. When the pulse is ready to be formed, the switch is closed and the stored energy is quickly discharged through the load to form the d-c pulse that is applied to the RF power device. During the discharge part of the cycle, the charging impedance prevents energy from the storage element from being returned to the energy source.

Line-Type Modulator In this device a delay line, or pulse forming network (PFN), is used as the storage element. The switch can be a hydrogen thyratron, mercury ignitron, a silicon controlled rectifier (SCR), saturable reactor, or other device that can initiate the discharge of the PFN to form a rectangular pulse. The shape and duration of the pulse are

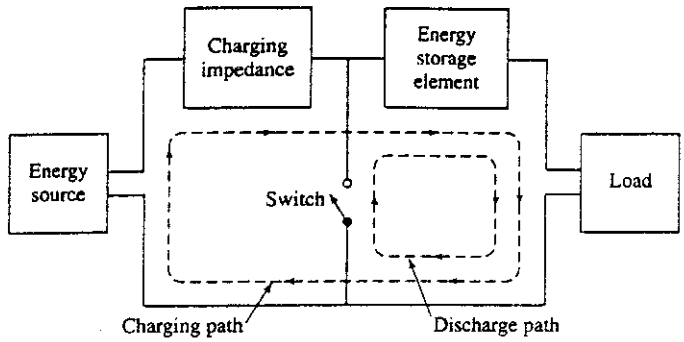


Figure 10.8 Basic elements of one type of radar pulse modulator.

determined by the passive elements of the PFN. The switch has no control over the pulse shape other than to initiate it. The pulse ends when the PFN has discharged sufficiently to allow the switch to close and recover its voltage hold-off capability. The trailing edge of the pulse is usually not sharp since it depends on the discharge characteristics of the PFN. The line-type modulator is simple, compact in size, and can tolerate abnormal load conditions. It has been widely used in the past for magnetron pulsing.

Active-Switch Modulator The switching element in this type of modulator has to be able to be turned off as well as on. Originally the switch was a vacuum tube and the modulator was called a hard-tube modulator to distinguish it from the gas-tube switch often used in a line-type modulator. In addition to the vacuum tube, the switch can be a semiconductor device such as a silicon controlled rectifier.

There are three types of active-switch modulators: (1) cathode pulsers that control the full beam power of the RF tube, (2) mod-anode pulsers that are required to switch at the full beam voltage of the RF tube but with little current, and (3) grid pulsers that operate at a far smaller voltage than that of the beam.

Unlike the line-type modulator, the switch in the active-switch modulator controls both the beginning and the end of the pulse. The energy storage element is a capacitor. To prevent droop in the pulse shape due to the exponential nature of a capacitor discharge, only a small fraction of the stored energy is extracted from the capacitor for the pulse to be delivered to the tube. In high-power transmitters with long pulses the capacitance must be very large. A large capacitance might be obtained with a collection of capacitors known as a *capacitor bank.*

The active-switch modulator permits more flexibility and precision than the line-type modulator. It can provide excellent pulse shape, varying pulse durations and prfs, including mixed pulse lengths and bursts of pulses with close spacings. It is, however, of greater complexity and weight than a line-type modulator.

Crowbar Protective Device Power tubes can develop internal arc discharges with little warning. When an arc occurs in an unprotected device, the capacitor bank discharges large currents through the arc and the tube can be damaged. The tube can be protected with a device called an *electronic crowbar,* which places a short circuit across the capacitor bank

to transfer its stored energy. When a sudden surge of current due to a fault in a protected power tube is sensed, the crowbar switch is activated within a few microseconds. The current surge also causes the circuit breaker to open and deenergize the primary source of power. The name "crowbar" is derived from the analogous action of placing a heavy conductor, like a crowbar, directly across the capacitor bank. A crowbar is required for a high-power active-switch modulator because of the large amount of energy that is stored in its capacitor bank. Crowbars are not usually needed with line-type modulators which store less energy in their pulse-forming network. Line-type modulators are designed to discharge safely all the stored energy each time a pulse is generated.

Transmitter Noise and Spectrum[49] An RF power source can produce spurious, unwanted outputs as harmonics of the fundamental frequency, adjacent-band (out-of-band) noise, and in-band noise. Harmonics and adjacent-band noise can be reduced 30 to 60 dB by using high-power filters. Shaping of the pulse to make it more rounded and less rectangular reduces out-of-band signal energy. In-band spurious signals and noise cannot be readily filtered since these unwanted signals are within the frequency range of the desired signal spectrum. The in-band noise is greater in some RF power sources than in others. For example, Weil states[49] that the noise level in a 1-MHz bandwidth of a conventional CFA is typically 50 to 60 dB down, but is down 90 dB or better in linear-beam tubes.

Section 3.7 discussed the effects of equipment instabilities on the amount of clutter cancellation, or improvement factor, that can be achieved in MTI radars. The typical noise level in conventional CFAs can set a limit on the achievable MIT improvement factor to perhaps the vicinity of 45 dB or so. Linear beam tubes, on the other hand, are capable of very high MTI improvement factors except for limitations introduced by their modulators and power supplies. The ripple on the modulator voltage and the variation of the high-voltage power supply (HVPS) must be sufficiently small to obtain the large improvement factors needed in high-performance radars.

In a staggered prf MTI system (Sec. 3.3) the variation of the interpulse period causes a variation in the HVPS voltage, which can be a significant source of transmitter instability. The consequent reduction in improvement factor that would result needs to be corrected, as indicated by Weil.[49]

The transmitter and its modulator can also distort pulse compression waveforms and introduce spurious time-sidelobes. Active-switch modulators are more likely to allow low pulse-compression time-sidelobe levels as compared with the time sidelobes produced by line-type modulators.

REFERENCES

1. Weil, T. A. "Transmitters." In *Radar Handbook,* 2nd ed. M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap 4.
2. Gilmour, A. S., Jr. *Microwave Tubes.* Norwood, MA: Artech House, 1986, Chap. 16.
3. Granatstein, V. L., R. K. Parker, and C. M. Armstrong. "Vacuum Electronics at the Dawn of the Twenty-First Century." *Proc. IEEE* 87 (May 1999), pp. 702–718.

4. News article from *Microwaves & RF,* (November 1984), p. 31.
5. Smith, M. J., and G. Phillips. *Power Klystrons Today.* New York: John Wiley, 1995, Sec. 7.2.3.
6. Staprans, A. "Linear Beam Tubes." In *Radar Technology,* E. Brookner (Ed.). Boston: Artech House, 1977, Chap. 22.
7. Gilmour, A. S., Jr. Ref. 2, Chap. 4.
8. Phillips, R. M., and D. W. Sprehn. "High-Power Kystrons for the Next Linear Collider." *Proc. IEEE* 87 (May 1999), pp. 738–751.
9. Staprans, A., E. W. McCune, and J. A. Ruetz. "High-Power Linear-Beam Tubes." *Proc. IEEE* 61 (March 1973), pp. 299–330.
10. Dodds, W. J., T. Moreno, and W. J. McBride, Jr. "Methods for Increasing the Bandwidth of High Power Microwave Amplifiers." *IRE WESCON Conv. Rec.* 1, pt. 3 (1957), pp. 101–110.
11. Gilmour, A. S., Jr. Principles of Traveling Wave Tubes. Boston: Artech House, 1994, Sec. 18.4.
12. Symons. R. S. "Tubes: Still Vital After All These Years." *IEEE Spectrum* 35 (April 1998), pp. 52–63.
13. Gilmour, A. S., Jr. Ref. 11, Chap. 13.
14. Gilmour, A. S., Jr. Ref. 2, Chap. 11.
15. Luebke, W. and G. Caryotakis. "Development of a One Megawatt CW Klystron." *Microwave J.* 9, no. 8 (August 1966), pp. 43–47.
16. Symons, R. S., and J. R. M. Vaughan. "The Linear Theory of the Clustered-Cavity™ Klystron." *IEEE Trans.* PS-22 (October 1994), pp. 713–718.
17. Borkowski, M. T. "Solid-State Transmitters." *Radar Handbook,* 2nd ed., M. Skolnik (Ed.). New York: McGraw-Hill, 1990, Chap. 5.
18. Lee, K., C. Corson, and G. Mols. "A 250 kW Solid State AN/SPS-40 Radar Transmitter." *Microwave J.* 26 (July 1983), pp. 93–105.
19. Dyck, J. D., and H. R. Ward. "RAMP's New Primary Surveillance Radar." *Microwave J.* 27 (December 1984), pp. 105–113.
20. Merrill, P. R. "A 20 kW Solid-State L-Band Transmitter for the RAMP PSR Radar." *Microwave J.* 31 (March 1988), pp. 165–173.
21. Chilton, R. H. "MMIC T/R Modules and Applications." *Microwave J.* 30 (September 1987), pp. 131–146.
22. Hoft, D. J. "Solid State Transmit/Receive Module for the Pave Paws Phased Array Radar." *Microwave J.* 21 (October 1978), pp. 33–35.
23. Olson, F. A. "Microwave Solid-State Power Amplifier Performance: Present and Future." *Microwave J.* 38 (February 1995), pp. 24–46.
24. Trew, R. J., J-B Yan, and P. M. Mock. "The Potential of Diamond and SiC Electronic Devices for Microwave and Millimeter-Wave Power Applications." *Proc. IEEE* 79 (May 1991), pp. 598–620.

25. Butler, N. "The Microwave Tube Reliability Problem." *Microwave J.* 16 (March 1973), pp. 41–42.

26. Advertisement of the Electron Device Group, Varian Beverly Division, Beverly, MA. (Varian is now known as CPI, Inc.)

27. Weil, T. A. Ref. 1, Sec. 4.2.

28. Gilmour, A. S., Jr. Ref. 2, Sec. 13.3.

29. Weil, T. A. "Comparison of CFA's for Pulsed-Radar Transmitters." *Microwave J.* 16 (June 1973), pp. 51–54, 72.

30. Kaisel, S. F. "Microwave Tube Technology Review." *Microwave J.* 20 (July 1977), pp. 23–42.

31. Anonymous. "Cathode-Driven Crossed-Field Amplifer." *Microwave J.* 31 (February 1988), pp. 208–209.

32. Sivan, L. *Microwave Tube Transmitters.* London: Chapman & Hall, 1994, Sec. 7.4.

33. Abrams, R. H., Jr. "The Microwave Power Module: A 'Supercomponent' for Radar Transmitters." *Record of the 1994 IEEE National Radar Conf.*, Atlanta, GA, pp. 1–6, IEEE No. 94CH3359–7.

34. Smith, C. R., C. M. Armstrong, and J. Duthie. "The Microwave Power Module: A Versatile RF Building Block for High-Power Transmitters." *Proc. IEEE* 87 (May 1999), pp. 717–737.

35. Yingst, T. E., et al. "High-Power Gridded Tubes—1972." *Proc. IEEE* 61 (March 1973), pp. 357–381.

36. Reed, J. E. "The AN/FPS-85 Radar System." *Proc. IEEE* 57 (March 1969), pp. 324–335.

37. Haeff, A. V. "An Ultra-High-Frequency Power Amplifier of Novel Design." *Electronics* 10 (February 1939), pp. 30–32.

38. Preist, D. H., and M. B. Shrader. "The Klystrode—An Unusual Transmitting Tube with Potential for UHF-TV." *Proc. IEEE* 70 (November 1982), pp. 1318–1325.

39. Symons, R. S. "The Constant Efficiency Amplifier." *NAB Broadcast Engr. Conf. Proc.* (1997), pp. 523–530.

40. Symons, R. S., et al. "The Constant Efficiency Amplifier—A Progress Report." *NAB Broadcast Engr. Conf. Proc.*, 1998.

41. Granatstein, V. L., and I. Alexoff. *High-Power Microwave Sources.* Boston: Artech House, 1987.

42. Gilmour, A. S., Jr. Ref. 2, Chap. 14.

43. Felch, K. L., et al. "Characteristics and Applications of Fast-Wave Gyrodevices." *Proc. IEEE* 87 (May 1999), pp. 752–781.

44. Manheimer, W. M. "On the Possibility of High Power Gyrotrons for Super Range Resolution Radar and Atmospheric Sensing." *Int. J. Electronics* 72, nos. 5 and 6 (1992), pp. 1165–1189.

45. Blank, M., B. G. Danly, and B. Levush. "Circuit Design of a Wideband W-Band Gyroklystron Amplifier for Radar Applications." *IEEE Trans.* PS-26 (June 1998), pp. 426–432.

46. Gelvich, E. A., et al. "The New Generation of High-Power Multiple-Beam Klystrons." *IEEE Trans.* MTT-41 (January 1993), pp. 15–19. See also the ISTOK Web Site at www.istok.com.

47. Weil, T. A. Ref. 1, Sec. 4.8.

48. Ewell, G. W. *Radar Transmitters.* New York: McGraw-Hill, 1981.

49. Weil, T. A. Ref. 1, Secs. 4.6 and 4.7.

PROBLEMS

10.1 One way to define the efficiency of a transmitter is RF power out, P_{out}, divided by the prime power in, P_{in}. (a) Plot the power dissipated, $P_{dis} = P_{in} - P_{out}$, as a function of the transmitter efficiency, ϵ, for a fixed power out. [Make the ordinate the ratio (power dissipated)/(power out).] (b) If the output power has to be 30 kW, what power will be dissipated with a transmitter having a 15 percent efficiency? (c) If the transmitter efficiency can be increased to 50 percent, what is the amount of power to be dissipated? (d) What disadvantages occur with low efficiency?

10.2 (a) In a solid-state transmitter (a solid-state "bottle") with 300 modules, what would be the reduction in output power if 20 modules were to fail? (b) What fractional reduction in radar range would this cause?

10.3 (a) If 10 percent of the modules in an active-aperture phased array radar fail, what would be the reduction in transmitter power? (b) What would be the reduction in the maximum radiation power density? (c) What would be the reduction in radar range?

10.4 For an air-traffic control radar application, compare the advantages and disadvantages of a solid-state transmitter, klystron transmitter, and magnetron transmitter.

10.5 (a) If one wanted a radar transmitter with a 10 percent bandwidth, what options are available to the radar system designer and which RF power source appears the most desirable? (You may have to make some assumptions about the application.) (b) If one wanted a radar transmitter with 40 percent bandwidth, what options are available and which of these options might you choose? (Include the reason for your choice.) (c) Are there any undesirable consequences with your choice for (b) that you might have to live with?

10.6 What are the advantages and disadvantages of the gyroklystron (amplifier) for radar application at 94 GHz (millimeter waves)?

10.7 How can a tube designer achieve a large bandwidth with a klystron type of power tube?

10.8 When might the systems engineer choose to use a traveling wave tube over a klystron for a high-power radar application?

10.9 For a high-power UHF radar transmitter application, compare the advantages and disadvantages of solid state, the grid-control vacuum tube, the constant-efficiency amplifier, and the klystron.

10.10 When might a magnetron be desirable for use in radar applications?

10.11 What factors might be involved when a radar systems engineer tries to choose among the crossed-field amplifier, TWT, and Twystron for some radar application?

10.12 If the R&D genie were to grant a radar systems engineer many wishes, what improvements might the radar systems engineer want to have for a radar transmitter?

chapter

11

Radar Receiver

11.1 THE RADAR RECEIVER

The function of the receiver in early radar systems was to extract the weak echo signals that appeared at the antenna terminals and amplify them to a level where they could be displayed to a radar operator who then made the decision as to whether or not a target echo signal was present. The modern radar receiver still has to extract the weak echo signals and amplify them, but it does much more. It employs a matched filter (Sec. 5.2) whose purpose is to maximize the peak-signal-to-mean-noise-ratio and discriminate against unwanted signals whose waveforms are different from those transmitted by the radar. When the clutter echoes are large enough to mask desired target echoes, the receiver also has to incorporate means for separating the moving targets from stationary clutter echoes by recognizing the doppler frequency shift of the moving targets (Chap. 3).

In modern radars the decision whether a target is present or absent is seldom made by an operator viewing on a display the unprocessed output of a receiver. Instead, the detection decision is made automatically based on threshold detection (Sec. 5.5). Information about a target's location in range and angle can also be extracted automatically instead of manually by an operator. In an operational air-surveillance radar, tracking of targets is no longer performed by an operator marking with a grease pencil on a radar display the location of blips (targets) from scan to scan and calculating the target speed and

estimating its direction. Targets are acquired and tracked automatically (Sec. 4.9) and only processed tracks are displayed to the operator or sent to some automatic device, such as an air-traffic control system or weapons control computer, for further use. When the radar cannot remove all the clutter echoes, constant false alarm rate (CFAR) circuitry is employed to prevent the tracking computer from becoming overloaded when trying to establish tracks using clutter echoes. The receiver is also the place where external interference and hostile electronic-countermeasures are kept from interfering with the detection of targets.

Thus, in addition to detection and amplification of signals, a radar receiver performs many other functions either directly as part of the receiver or in conjunction with it. These other functions include signal processing, information extraction, data processing, electromagnetic compatibility, and electronic counter-countermeasures. (The modern receiver might be thought of as the *receiver/processor*.) Sometimes the display is considered part of the receiver system. In this and other radar books, these other functions are often considered separately from the discussion of the receiver. The interested reader will find a more thorough review of the radar receiver by John W. Taylor, Jr. in the *Radar Handbook*.[1]

The radar receiver is almost always a *superheterodyne*, or superhet. It was shown in the block diagram of Fig. 1.4 and briefly described in Sec. 1.3. The essential characteristic of a superheterodyne is that it converts the RF input signal to an intermediate frequency (IF) where it is easier than at RF to achieve the necessary filter shape, bandwidth, gain, and stability. An advantage of the superheterodyne receiver is that its frequency can be readily changed by changing the frequency of the local oscillator (LO). The first stage, or *front-end*, of a radar superheterodyne receiver can be an RF low-noise amplifier (LNA) such as a transistor.

Before the availability of low-noise transistors, the receiver front-end was the mixer stage without an RF amplifier preceding it. In some applications the mixer stage might still be desired as the receiver front-end instead of a low-noise amplifier. A receiver with a mixer as the first stage has a greater dynamic range than one with a low-noise amplifier, which might be important when large MTI improvement factors are needed to remove clutter echoes. The extra dynamic range available with a mixer as the front-end can also be of value in reducing the likelihood of receiver saturation when large signals or jamming are present. The larger receiver noise figure of a mixer might be compensated with greater transmitter power and/or a bigger antenna, both of which are beneficial when a military radar is faced with hostile noise jamming. Although the mixer front-end might have some advantage over a low-noise transistor amplifier front-end, a low-noise amplifier as the first stage of a superheterodyne receiver generally seems to be preferred by those who buy radars.

A high-performance air-surveillance radar sometimes employs more than one type of receiver. Each would share the front-end, mixer, and IF stages. One receiver might be a linear amplifier and envelope detector for detection of targets in the clear (no competing clutter). A second receiver might be for doppler processing to remove clutter, as in an MTI radar. It would use I and Q channels and digital signal processing to filter the moving targets. A third receiver might be a log-FTC (Sec. 7.8), or something similar, to aid in detecting targets located within moving weather clutter beyond the range of surface clutter.

It was said in Chap. 2 that if the radar designer wishes to increase the detection range of a radar, the chief factors available are the average power of the transmitter and the area of the antenna. The classical radar equation also indicates that the range can be increased by reducing the receiver noise figure; but, in practice, the noise figures of radar receivers are already quite low and any further decrease can produce marginal results and, sometimes, unwanted effects. Further lowering of the noise figure might not be justified if it significantly increases receiver cost, lowers the dynamic range, subjects the device to a greater risk of burn-out, and results in less reliability. A very sensitive receiver also allows more interference to enter the receiver. Sometimes an increase in interference is a price that may be worth the benefits of improved sensitivity, but there can be limits.

A radar receiver has to have sufficient gain to increase the level of the weak echo signal to where it is large enough to be processed or displayed. In the superheterodyne the total receiver gain is divided between the IF and the video amplifiers. The receiver should have adequate dynamic range (where the receiver is linear) so that large clutter echoes do not cause the receiver to saturate and reduce the MTI improvement factor. It must not introduce unwanted phase or amplitude changes that could distort the echo signals. It must be protected from overload, saturation, and damage (burnout) by strong unwanted signals. Timing and reference signals are needed to properly extract target information and take advantage of the doppler frequency shift of echo signals.

A limitation with early radar receivers when vacuum tubes were the only available technology was that they were relatively large in size. The size of radar receivers is no longer a problem with modern technology. The trend is to make the receiver as digital as is practical, with analog components confined to the RF and perhaps the IF.

There can be many demands on the radar receiver, but the receiver designer has responded well to the challenges and there exists a highly developed state of receiver technology. Radar receiver design and implementation may not always be an easy task, but receiver designers have usually been able to provide the radar systems engineer the means to accomplish the desired objectives.

11.2 RECEIVER NOISE FIGURE[2,3]

Definition The receiver noise figure was described in Sec. 2.3 as a measure of the noise produced by a practical receiver compared to the noise of an ideal receiver. The noise figure of a linear network may be defined as either

$$F_n = \frac{N_{out}}{kT_0B_nG} \quad \text{or} \quad \frac{S_{in}/N_{in}}{S_{out}/N_{out}} \qquad [11.1]$$

where N_{out} = available output noise power; $kT_0B_n = N_{in}$ = available input noise power; k = Boltzmann's constant = 1.38×10^{-23} J/deg; T_0 = standard temperature of 290 K (approximately room temperature); B_n = noise bandwidth defined by Eq. (2.3); $G = S_{out}/S_{in}$ = available gain; S_{out} = available output signal power; and S_{in} = available input signal power. The term "available power" refers to the power that would be delivered to a matched load. (The term "available" will be understood in the following discussion of noise figure and is not mentioned further.) The product $kT_0 = 4 \times 10^{-21}$ W/Hz.

The reason for a standard temperature T_0 in the definition of noise figure is to refer measurements made under differing temperature conditions to a common basis of comparison.

Equation (11.1) permits two different, but equivalent, interpretations of the noise figure. It may be considered (right-hand side) as the degradation of the signal-to-noise ratio as the signal passes through the network, or (left-hand side) it may be interpreted as the ratio of the noise-power out of the actual network to the noise-power out of an ideal network that amplifies the input thermal noise and introduces no additional noise of its own. The noise figure of Eq. (11.1) can be expanded as

$$F_n = \frac{kT_0B_nG + \Delta N}{kT_0B_nG} = 1 + \frac{\Delta N}{kT_0B_nG} \qquad [11.2]$$

where ΔN is the additional noise introduced by the practical (nonideal) network.

The noise figure is commonly expressed in decibels; that is, 10 log F_n. The term *noise factor* has also been used at times instead of noise figure. The definition of noise figure assumes that the input and output of the network are matched. In some devices, less noise is obtained under mismatched, rather than matched, conditions. In spite of definitions, such networks would be operated so as to achieve the maximum output signal-to-noise ratio.

Noise Figure of Networks in Cascade Consider two networks in cascade, each with the same noise bandwidth B_n but with different noise figures and gain, Fig. 11.1. Let F_1, G_1 be the noise figure and gain, respectively, of the first network and F_2, G_2 be similar parameters for the second network. The problem is to find F_0, the overall noise-figure of the two networks in cascade. From the definition of noise figure given by Eqs. (11.1) and (11.2), the output noise N_{out} of the two networks in cascade is

N_{out} = noise from network 1 at output of network 2 + noise ΔN_2 introduced by network 2

$$= F_0kT_0B_nG_1G_2 = F_1kT_0B_nG_1G_2 + \Delta N_2 = F_1kT_0B_nG_1G_2 + (F_2 - 1)\,kT_0B_nG_2$$

which results in

$$F_0 = F_1 + \frac{F_2 - 1}{G_1} \qquad [11.3]$$

It is not sufficient that the first stage of a low-noise receiver have a low noise figure. The second stage must also have a low noise figure or, if not, the gain of the first stage needs to be large. Too large a first-stage gain, however, is not always desirable since the dynamic range of the receiver is reduced by the gain G_1 of the low-noise amplifier. If the first network is not an amplifier, but is a diode mixer, the gain G_1 should be interpreted as a number less than unity (a loss).

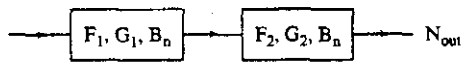


Figure 11.1 Two networks in cascade with different noise figures and gains, but the same noise bandwidths.

The noise figure of N networks in cascade may be shown to be

$$F_0 = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_N - 1}{G_1 G_2 \cdots G_{N-1}} \qquad [11.4]$$

Similar expressions may be derived when the bandwidth and/or temperature of the individual networks are not the same.[4]

Noise Figure Due to Loss in the Transmission Line Any losses in the RF portion ahead of the receiver front-end result in an increase in the apparent overall noise figure. Such losses can be due to the transmission line between antenna and receiver, the duplexer, receiver protector, rotary joint, preselector filter, STC if applied at RF, monitoring devices, and the radome. The noise figure due to these RF losses, obtained from the second part of the definition of Eq. (11.1), is equal to the RF loss L_{RF}. (This can be seen since L_{RF} is the loss in signal-to-noise ratio as the signal travels from the antenna to the receiver.) It can also be obtained from the first part of Eq. (11.1) since the noise out of a lossy transmission line is kT_0B_n, and its gain $G = 1/L_{RF}$.

If the loss in the transmission line and its associated devices is incorporated in the receiver noise figure, it should not also be a part of the system losses. Most radar analyses treat these losses as system losses rather than as part of the receiver noise figure. There are some RF devices that are likely to be closely associated with the low-noise receiver. These include the circulator that provides isolation, the receiver protector, waveguide to coax transition, and receiver performance monitoring (which might produce a loss). When a noise figure for a receiver is quoted in a publication or given in a manufacturer's catalog, it might not always be obvious what is included. It might be the receiver noise figure without the associated receiver losses or it might include the loss of those devices that are a close part of the receiver, as mentioned above. With a mixer front-end receiver the noise figure quoted has sometimes been that of the mixer alone and not of the entire receiver. There seems to be no standard method for reporting receiver noise figure.

Noise Temperature The noise introduced by a network may also be expressed as the *effective noise temperature*, T_e, defined as the (fictional) temperature at the input of the network, that accounts for the additional noise ΔN at the output. Therefore $\Delta N = kT_eB_nG$ and from Eq. (11.2) we have

$$F_n = 1 + \frac{T_e}{T_0} \qquad [11.5]$$

$$T_e = (F_n - 1)T_0 \qquad [11.6]$$

The *system noise temperature* T_s is defined as the effective noise temperature of the receiver including the effects of antenna temperature T_a. If the receiver effective noise temperature is T_e, then

$$T_s = T_a + T_e = (F_s - 1)T_0 \qquad [11.7]$$

where F_s is the *system noise figure*. This equation also defines the system noise figure when it includes the effects of the antenna temperature T_a and receiver effective noise temperature T_e.

The effective noise temperature of a receiver consisting of a number of networks in cascade is

$$T_e = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \cdots \qquad [11.8]$$

where T_i and G_i are the effective noise temperature and gain of the ith network.

The effective noise temperature and the noise figure both describe the same characteristic of a network. The effective noise temperature generally is used to describe the noise performance of very low-noise receivers, lower than might be of interest for radar. It is also preferred by some radar engineers and many receiver designers as being more useful than noise figure for analysis purposes. The noise figure, however, seems to be the more widely used term to describe radar receiver performance, and is used in this text for that purpose.

11.3 SUPERHETERODYNE RECEIVER

The discussion of the superheterodyne receiver in this section does not include all aspects of the receiver, but only with those component parts that have an effect on the radar system design. This includes the low-noise RF amplifier, the mixer, receiver dynamic range, the $1/f$ noise at IF, oscillator noise, and the detector.

Low-Noise Front-End The first stage of a superheterodyne receiver for radar application can be a transistor amplifier. At the lower radar frequencies the silicon bipolar transistor has been used. Gallium-arsenide field-effect transistors (FET) are found at the higher frequencies. Other types of transistors also can be used, depending on the trade-off between the desired noise figure and the ability of the transistor to withstand burnout. An X-band transistor can provide a noise figure of about one dB and can withstand a leakage peak power of 0.2 W.[5] With a diode limiter ahead of the transistor, the peak power can be as great as 50 W before burnout. The diode limiter increases the noise figure about 0.5 dB at X band and 0.2 dB at C band. The lower the frequency the lower can be the transistor noise figure. At C band the noise figure might be around 0.6 dB. These values are more than adequate for radar. (Early microwave radars had noise figures of 12 to 15 dB and radars in the 1960s had noise figures of 7 to 8 dB.) It is not necessary for the radar systems engineer to have extremely low noise figures in most radar applications, especially when the unavoidable losses in the transmission line between receiver and antenna are considered. If improved radar system performance is of concern, it is probably more fruitful to try to reduce some of the many system losses that occur elsewhere in a radar rather than try to reduce further the noise figure of the low-noise amplifier (LNA). It is usually good enough.

Prior to the low-noise transistor amplifier, the parametric amplifier and the maser were available as low-noise receiver front-ends. Although their noise figures were low (lower than those of transistors, which came later), they were seldom used operationally for radar. They were expensive, of large size, and often did not have sufficient dynamic

range. Until low noise transistor amplifiers were developed, the radar receiver seldom employed an RF amplifier stage except perhaps at UHF or lower frequencies. Before the low-noise transistor, the mixer was the receiver front-end. As already mentioned, a mixer as the front-end without a low-noise amplifier preceding it is a valid option for some radar applications in spite of its higher receiver noise figure.

Achieving low receiver noise is no longer the problem it once was, and designers of high-performance radar receivers usually are more concerned with obtaining large dynamic range and low oscillator noise.

Mixers[6] Whether or not it is used as the front-end, the mixer is a key element in a superheterodyne receiver since it is the means by which the incoming RF signal is converted to IF (intermediate frequency). When the down conversion from RF to IF is performed in one step, it is called single conversion. Sometimes the down conversion is done in two steps with two mixers and IF amplifiers. This is known as dual conversion. Dual conversion superheterodyne receivers are used to avoid some forms of interference and (spoofing) electronic countermeasures. The mixer should have a low conversion loss, introduce little additional noise of its own, minimize spurious responses, and not be susceptible to burnout, especially when it is used as the front-end without a low noise amplifier ahead of it. An integral part of the mixer is the local oscillator.

Noise Figure of a Mixer Used as a Front End The noise figure of a mixer is determined by its conversion loss and noise-temperature ratio. The conversion loss of a mixer is defined as

$$L_c = \frac{\text{available RF power}}{\text{available IF power}} \qquad [11.9]$$

It is a measure of the efficiency of the mixer in converting RF signal power into IF. The conversion loss of typical microwave diodes in a conventional single-ended mixer configuration varies from about 5 to 6.5 dB. Schottky diodes in an image recovery mixer have been reported to have a minimum conversion loss of 3.5 dB over a narrow band and a 4-dB conversion loss over a 10 percent bandwidth at S band.[7] The noise-temperature ratio of a mixer (not to be confused with the effective noise temperature) is defined as

$$t_r = \frac{\text{actual available IF noise power}}{\text{available noise power from an equivalent resistance}} \qquad [11.10]$$

or

$$t_r = \frac{F_m k T_0 B G_c}{k T_0 B} = F_m G_c = \frac{F_m}{L_c}$$

where F_m = mixer noise figure and L_c = conversion loss = $1/G_c$. The noise temperature ratio of a mixer varies inversely with the IF frequency from about 100 kHz down to a small fraction of a hertz. At a frequency of 30 MHz, the noise temperature ratio might range from 1.2 to 2.0. Generally, the lower the conversion loss, the larger is the noise temperature ratio.

The noise figure of a mixer based on Eq. (11.10) is $F_m = L_c t_r$. It is, however, not a complete measure of the sensitivity of a receiver with a mixer front-end. The overall noise figure depends not only on the mixer stage, but also on the noise figure of the IF amplifier. The latter becomes a significant factor in the overall noise figure since the mixer has a loss rather than a gain. Using Eq. (11.3), the noise figure of the first network (the mixer) is $F_1 = L_c t_r$ and its gain is $G_1 = 1/L_c$. The noise figure of the second network is that of the IF amplifier, so that $F_2 = F_{IF}$. The receiver noise figure with a mixer front-end is then

$$F_R = F_1 + \frac{F_2 - 1}{G_1} = L_c t_r + (F_{IF} - 1)L_c = L_c(t_r + F_{IF} - 1) \qquad [11.11]$$

(This does not include any losses in the RF transmission line.) If, for example, the conversion loss of the mixer were 5.5 dB, the IF noise figure 0.5 dB, and the noise-temperature ratio 1.2, the receiver noise figure would be 6.7 dB. For low noise-temperature diodes, the receiver noise figure is approximately equal to the conversion loss times the IF noise figure.

Some manufacturers have used Eq. (11.11) to determine the noise figures of the mixers listed in their catalogs. Other manufacturers have used the expression $L_c t_r$, which is lower than that of Eq. (11.11). When using mixer noise figures quoted in advertising brochures or in the literature, one should be aware of how it was determined.

Types of Mixers[8,9] An ideal mixer is one whose output is proportional to the product of the RF echo signal and the local oscillator (LO) signal. The mixer provides two output frequencies that are the sum and difference of the two input frequencies, or $f_{RF} \pm f_{LO}$, assuming $f_{RF} > f_{LO}$. The difference frequency $f_{RF} - f_{LO}$ is the desired IF frequency. The sum frequency $f_{RF} + f_{LO}$ is rejected by filtering. There are however, two possible difference frequency signals at the IF when a signal appears at the RF. One is $f_{IF} = f_{RF} - f_{LO}$, assuming the input RF signal is of greater frequency than the LO frequency. The other possible difference frequency occurs when the RF signal is at a lower frequency than the LO frequency such that $f_{IF} = f_{LO} - f_{RF}$. If one of these is at the desired signal frequency, the other is the *image frequency*. Signals and receiver noise that appear at the image frequency need to be rejected using either an RF filter or an image-reject mixer described later in this subsection.

A relatively simple mixer is the *single-ended mixer*, which uses a single diode, as in Fig. 11.2a. The diode terminates a transmission line and the LO is inserted via a directional coupler. A low-pass filter, not shown, following the diode allows the IF to pass while rejecting the RF and LO signals. In a single-ended mixer the image frequency is short-circuited or open-circuited so as to avoid having the noise from the image frequency affect the mixer output.

The diode of a mixer is a nonlinear device and, in theory, can produce intermodulation products at other frequencies, called *spurious responses*. These occur for any RF signal that satisfies the relation[10]

$$mf_{RF} + nf_{LO} = f_{IF} \qquad [11.12]$$

where m and n are integers such that $m,n = \ldots, -2, -1, 0, 1, 2, \ldots$. These are unwanted since they appear within the radar receiver bandwidth. Spurious responses that are

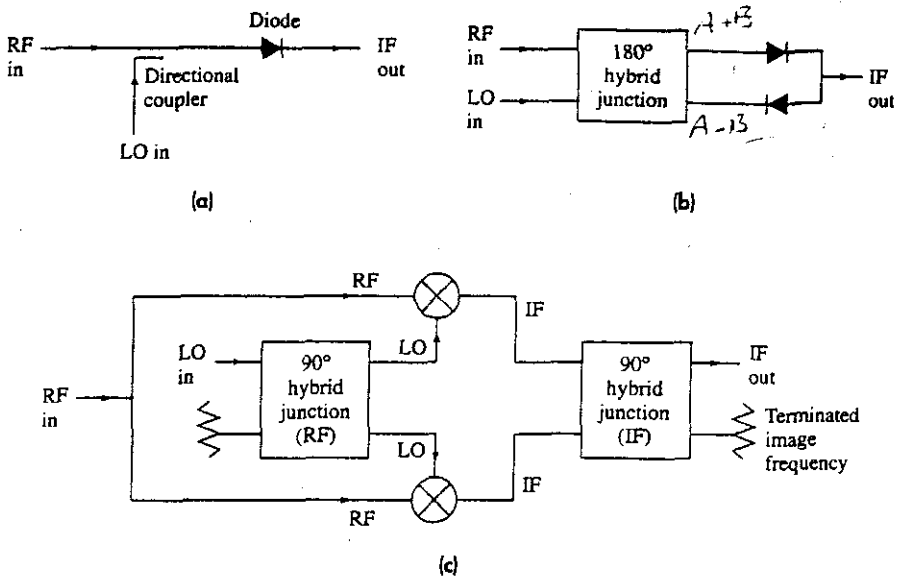


Figure 11.2 Types of mixers: (a) single-ended mixer, (b) balanced mixer, (c) image-rejection mixer.

IF outputs due to the action of the mixer should not be confused with spurious signals, or spurs, that are due to the LO or the receiver power supply and can occur even in the absence of an RF signal. Taylor[7] describes the so-called *mixer chart*, which allows one to determine the combinations of the RF and LO frequencies that are free of strong spurious components. Such a chart indicates the bandwidth available for the mixer as a function of the ratio of the RF and LO frequencies. Taylor points out that the nature of the spurious responses are such that single-conversion receivers generally provide better suppression of spurious responses than double-conversion receivers. The third-order intermodulation product generally affects the dynamic range of the receiver, and is mentioned later under the discussion of dynamic range. There also can be other spurious, or intermodulation, responses from a mixer when two or more RF signals are present at the mixer's input and produce responses within the IF bandwidth.

Noise that accompanies the local oscillator (LO) signal in a single-ended mixer can appear at the IF frequency because of the nonlinear action of the mixer. This noise can be eliminated by inserting a narrowband RF filter between the LO and the mixer. It also has to be a tunable filter if the LO frequency is also tunable. A method to eliminate LO noise that doesn't have these disadvantages is a *balanced mixer*. The balanced mixer also can remove much of the mixer intermodulation products.

A diagram of a balanced mixer is shown in Fig. 11.2b. It can be thought of as two single-ended mixers in parallel and 180° out of phase. At the left of the figure is a four-port junction such as a magic-T, hybrid junction, 3-dB coupler, or equivalent. (Either a 90° or a 180° hybrid can be employed; here it is 180°.) In Fig. 11.2b the LO is applied to one port and the RF is applied to a second port. The signals inserted at these two ports appear in the third port as their sum and in the four port as their difference. A diode mixer is at the output of each of the other two ports. The hybrid junction has the property that

the sum of the RF and LO signals appears at a port containing one of the diode mixers, and at the other port the difference of the RF and LO appears at the diode. The two diode mixers should have identical characteristics and be well matched. The IF signal is obtained by subtracting the outputs of the two diode mixers. In Fig. 11.2b, the balanced diodes are shown reversed so that the IF outputs can be added to obtain the required difference between the two channels. Local-oscillator AM noise at the two diode mixers will be in phase and will be canceled at the output. This mixer configuration also suppresses the even harmonics of both the RF and LO signals.

A *double-balanced mixer* (not shown) utilizes four diodes in a ring, or bridge, network to reduce the LO reflections and noise at the RF and IF ports, achieve better isolation between the RF and LO ports, reject spurious responses and certain intermodulation products, provide good suppression of the even harmonics of both the RF and LO signals, and permit wide bandwidth.[11]

In an *image-rejection mixer*, Fig. 11.2c, the RF signal is split and fed to the two mixers. The LO is fed into one port of a 90° hybrid junction that produces a 90° phase difference between the LO inputs to the two mixers. On the right is an IF hybrid junction that imparts another 90° phase difference in such a manner that the signal frequency and the image frequency are separated. The port with the image signal can be terminated in a matched load. According to Maas,[12] to reduce the image frequency by 20 dB requires that the phase error of the image-rejection mixer be less than 10° and the gain imbalance to be less than 1 dB. Dixon states[13] that the image-rejection mixer provides only about 30 dB of image rejection, which might not be sufficient for some applications. The image-rejection mixer is capable of wide bandwidth, and is restricted only by the frequency sensitivity of the structure of the microwave circuit. It is attractive because of its high dynamic range, good VSWR, low intermodulation products, and less susceptibility to burnout. The noise figure of the image-rejection mixer as well as the balanced mixer will be higher than that of a single-ended mixer because of the loss associated with the hybrid junctions.

The *image-recovery mixer* is an image-rejection mixer designed to reduce the mixer conversion loss by properly terminating the diode in a reactance at the image frequency. Sometimes the lower conversion loss is offset by an increase in noise temperature, a mismatch at the IF, and higher intermodulation products. The improvement using image enhancement is about 1 or 2 dB; hence, the mixer needs to be of low loss so as not to negate the benefit.[14]

Dynamic Range There seems to be no unique definition for the *dynamic range* of a radar receiver. It can generally be described as the ratio of the maximum input signal power to the minimum input signal power the receiver can handle without degradation in performance. "Degradation in performance," however, is not easy to define since it depends on the application. The minimum signal is sometimes taken to be the receiver rms noise level, which depends on the receiver bandwidth. The *minimum detectable signal* S_{min} might be selected as the minimum in the definition of dynamic range; but it also depends on bandwidth and apparently it is seldom used for this purpose.

The maximum signal might be the signal that causes the receiver to saturate (the output no longer increases with an increase in input). Saturation, however, is gradual as the

signal increases in power, so the signal level that results in saturation is not precise. The maximum signal is more usually defined by the acceptable amount of gain compression, which is the deviation of the gain curve (output vs. input) from a straight line. The signal that causes a gain compression of one dB is commonly used for defining the maximum signal. Another criterion for the maximum signal power is based on the onset of *intermodulation distortion*. Intermodulation distortion generally occurs with large signals in the later stages of the receiver. The mixer generates a unique form of intermodulation product called *spurious responses*. It occurs when a harmonic of the local oscillator frequency mixes with a harmonic of the RF signal frequency to create difference frequencies that appear within the IF bandwidth. *Third-order intermodulation* occurs when two equal-amplitude signals within the receiver pass-band at two different frequencies f_1 and f_2 are input to the receiver and produce at the output the frequencies $2f_1 - f_2$ and $2f_2 - f_1$. The maximum signal power is then specified by the third-order intermodulation that can be tolerated when two signals are present within the pass band. The *third-order intermodulation* product is difficult to eliminate by filtering when the two frequencies are close to one another. Another possible indication of the maximum signal is when the echo signal at the mixer approaches the power level of the local oscillator. (The local oscillator power should be at least 7 dB greater than the largest received signal.[15]) No matter what definition is used, the dynamic range is almost always expressed in dB.

A large dynamic range is important if receiver saturation is to be avoided. Once the receiver saturates it can take a finite time to recover before targets again can be detected. Furthermore, when clutter is large enough to saturate the receiver, the MTI improvement factor will be reduced (Sec. 3.7). Saturation of the receiver by clutter echoes causes weak target echoes to be suppressed and not detected even though there might have been adequate improvement factor otherwise. In high-performance radars that must detect small moving targets in the presence of large clutter echoes by doppler processing, the receiver dynamic range must be at least equal to the required improvement factor.

As an example of the variation of the target echo signal power that might be experienced by a radar receiver, assume that an air-surveillance radar has to detect aircraft at ranges from 4 to 200 nmi. This corresponds to a variation in signal power of $(200/4)^4$, which is 68 dB. The average cross section of aircraft might vary from 2 to 100 m^2 (a variation of 17 dB), and the fluctuations in cross section might range over 30 dB. Adding all three factors, the variation of the total target echo signal might be 115 dB, more or less. This might be an extreme value, but radars that have to detect low cross-section targets could require even greater dynamic range.

Clutter echoes might vary over a range from 60 to 70 dB, or more. The use of STC (sensitivity time control), where the receiver gain is made to vary with time (Sec. 7.8), can reduce the variation in the target echo signal as well as the clutter echo signal. Not all radars, however, can employ STC. Pulse doppler radars, for example, cannot.

The mixer stage is often the limiting factor in dynamic range. A radar receiver that uses a doppler filter bank will have a higher dynamic range because of the narrower bandwidth of each filter. Pulse compression can also increase the dynamic range in proportion to the pulse compression ratio, if the clutter seen by the time sidelobes is not too large. The wider the bandwidth of the receiver (the IF stage) the less will be the dynamic range because of the greater likelihood that mixer intermodulation products (spurious responses)

will be within the frequency band to limit the maximum signal that can be received. A wide bandwidth, as mentioned, also increases the noise level, which reduces dynamic range.

Large dynamic range may be obtained in some radar applications by inserting variable attenuation into the receiver as needed to keep the receiver from being overloaded, but this solution is limited to situations where rapid changes in the input signal are not expected.

Flicker Noise, or 1/f Noise There exists in semiconductors a noise mechanism whose spectral density is inversely proportional to the frequency. It is called *flicker noise* or *1/f noise*,[16] and can be of importance at the lower frequencies. It is quite different from thermal or shot noise, which are independent of frequency. Flicker noise occurs in semiconductor devices such as diodes or transistors, and also in vacuum tubes with oxide-coated cathodes. The frequency relationship of flicker noise is more like $1/f^{\alpha}$, where α varies between 0.8 and 1.3, but it is more common to characterize it as $\alpha \approx 1$.[17] This relationship holds for very low frequencies, lower than might be of practical interest for radar.

The $1/f$ noise is not important for radar receivers whose IF frequencies are greater than a few hundred kilohertz. This is the case for most radar IF frequencies. It can be a factor limiting sensitivity, however, in radars that employ a *homodyne receiver*, also known as a superheterodyne receiver with zero IF. Homodyne receivers are sometimes used in CW radars because of their simplicity. The decrease in sensitivity due to $1/f$ noise at low frequencies might be tolerated for very short-range systems; but when maximum performance is necessary, the effect of the $1/f$ noise can be avoided by use of a superheterodyne receiver with an IF frequency where $1/f$ noise is low.

Oscillator Stability In conventional pulse radars that do not perform doppler processing, stability of the local oscillator, or LO, cannot be ignored but it is usually not a major concern. However, when doppler processing is used to detect moving targets in clutter, as in the MTI radar, the LO has to be quite stable in order to reliably detect the doppler shift. This is why the LO in an MTI radar is called a *stalo*, or stable local oscillator. The MTI improvement factor that can be achieved with a magnetron oscillator transmitter is limited to modest values, so that the demands on oscillator stability can readily be met when a magnetron is used. Power amplifiers such as the klystron, TWT, and the transistor, however, allow much larger improvement factors than a magnetron. Thus greater demands are placed on the stability of the stalos used in such radars. Some high-prf pulse doppler radars that have to detect small targets in the midst of large clutter might encounter clutter echoes that could be 100 dB, or greater, than the target echoes, and thus require highly stable RF sources.

MTI and pulse doppler radars that employ a power amplifier use the sum of the receiver stalo and the coho as the input signal to the power amplifier. (This was indicated in Fig. 3.7 for the MTI radar.) Since the stalo is at a much higher frequency than the coho, it is the stalo that usually sets the limits on what can be achieved. The stalo can have a greater effect on performance than the power amplifier of the transmitter.[18] Thus we only consider the stalo here.

Phase Noise Instability or phase noise in a stalo can be caused by power supply ripple; mechanical and acoustic vibrations from fans, motors, and cooling systems; or by vibrations of the platform (such as an aircraft or ship); as well as spurious responses and

noise from the stalo itself. Phase noise is usually considered in the frequency domain, but in the time domain it can be thought of as being due to the deviation of the oscillator signal from a perfect sinewave. There is also amplitude-modulation noise associated with oscillators, but AM noise is usually small compared to phase noise. If not, it can be reduced by balanced mixers or other means.

In Sec. 3.7 the effect of equipment instabilities on MTI performance was mentioned. There it was shown that a pulse-to-pulse change in phase, $\Delta\phi$, limits the improvement factor of a two-pulse MTI to $I_f = (\Delta\phi)^{-2}$. When an MTI or pulse doppler radar uses many pulses to perform doppler filtering, this simple expression for improvement factor no longer applies. A different model has to be considered.

The reader is referred to Fig. 3.36 for an example of the spectrum of an oscillator as might be used in a mixer. (There is also further discussion of oscillator stability for doppler radars where this figure appears in Sec. 3.7.) In addition to the narrow spike at d-c due to the carrier (not shown in the figure), there is a noise spectrum that decreases monotonically with increasing frequency. There are also spikes, or *spurs*, that are usually caused by the power supply or vibrations. At the higher frequencies the phase noise levels off and is characterized by a uniform noise floor. The ordinate in Fig. 3.36 is the noise power within a one-hertz bandwidth relative to that of the carrier. It should be multiplied by the receiver bandwidth to obtain the actual power at the receiver.

Although Fig. 3.36 might be the noise from a stalo, it also represents the noise radiated by the transmitter since the stalo is a major part of the signal that excites the power amplifier transmitter. This noise may seem far down from the peak of the carrier, but the spectrum of the echo from stationary clutter is the same as the spectrum of the transmitter. (Internal motion of the clutter can further increase the spectrum of the received echo signal.) As mentioned, clutter can be quite large compared to the weak moving target echo. The MTI or pulse doppler radar may be able to attenuate the main clutter line at d-c, but the clutter spectrum often has components at frequencies where doppler-shifted echoes from moving targets are expected. These components can mask the desired target echoes. Good performance of an MTI or a pulse doppler radar requires that the transmitter spectrum, and the clutter-echo spectrum it produces, be low enough to detect the slowly moving weak targets that are of interest. It would not be unusual to find that oscillator noise can be the limiting factor in some high-performance radars that must detect low-speed, low cross-section moving targets in heavy clutter. Good oscillator design is therefore important for achieving good radar performance.

The effect of phase noise can be determined by measuring the phase-modulation spectrum of the stalo and using it to obtain the MTI improvement factor. The procedure will not be given here. It is outlined by Taylor[18] and given in more detail with examples by Goldman.[19] Since the stalo is part of the transmitter as well as part of the receiver, the effect of phase noise on MTI performance will be range dependent. At the shorter ranges, or time delays, greater stalo noise can be tolerated at frequencies closer to the carrier (lower target doppler frequency shifts) than at longer ranges. For this reason, the effect of stalo stability needs to be computed for several ranges.

Oscillator phase noise can be a serious limitation to the performance of a modern high-performance MTI or a pulse doppler radar. Its effect has to be found with measurements and analysis more complicated than was indicated in Sec. 3.7.

Types of Stable Oscillators[20,21] Almost all of the oscillators used for stable sources can be thought of as consisting of an amplifier, a resonant circuit that determines the frequency and the phase noise, and feedback to generate oscillation. The amplifier is often a transistor. The following is a brief listing of the various oscillators that have been considered for use as stable sources.

Crystal Oscillator The mechanically vibrating piezoelectric quartz crystal has been an important device for producing stable oscillators ever since the early days of commercial radio.[22] A piezoelectric material is one which mechanically deforms along one crystal axis when an electric potential is applied along another axis. Conversely an electrical potential is obtained when a mechanical deformation occurs. The piezoelectric crystal is used as the resonator in the feedback circuit of a transistor oscillator. It is often mounted in a small-size temperature-controlled oven and isolated from vibrations. It is a very stable source at low frequencies (10 to 180 MHz), but its output can be multiplied in frequency to provide stable signals in the microwave region.

Frequency Multiplier A low-frequency stable oscillator can be multiplied to a higher frequency by applying its signal to a nonlinear device such as a diode or varactor to generate harmonics of the fundamental frequency. A filter is used to select the desired harmonic. The phase noise power, however, increases as the square of the frequency-multiplication ratio. For example, when a 10-MHz stable source is multiplied to 10 GHz, its noise is increased by 60 dB. In addition, there can also be additive phase noise produced in frequency multipliers. In spite of the increase in noise with multiplication in frequency, multiplication is a good method for taking advantage of the excellent stability of a low-frequency sources to obtain a stable oscillator at radar frequencies.

Dielectric Resonator Oscillator (DRO) The resonant circuit in this type of oscillator is a dielectric material such as a sapphire crystal, ceramic,[23] or titanate in a regular geometric form that acts as a microwave resonant cavity. The high dielectric constant of the resonator allows it to be much smaller in size compared to a metallic cavity resonator. It is among the most stable of room-temperature oscillators. Because of its small size it has a relatively high Q and may be quite rigid so as to reduce its sensitivity to shock and vibration. When the dielectric resonator is made larger to obtain even higher values of Q and enhanced frequency stability, it might be more sensitive to temperature changes and vibration. The DRO has been a popular device for application as a low-noise, stable oscillator at microwave frequencies.

SAW Oscillator The surface acoustic wave (SAW) device can also be used as the resonator in a feedback oscillator. SAW oscillators can be quite small and can be obtained from about 100 MHz to 3 GHz. Ewell[20] states that the phase noise of a SAW oscillator can be worse than that of a frequency-multiplied crystal oscillator at low offset frequencies (1 kHz for example) from the carrier, but it can be better at high frequency offsets (greater than 10 kHz).

YIG Oscillator A small sphere of yttrium iron garnet (YIG) suspended within a resonant cavity with an applied magnetic field can act as the resonant device of an oscillator. The resonant frequency of a spherical YIG crystal depends only on the applied magnetic field and not on its dimensions. It has a relatively high level of phase noise, but it has the advantage of being tunable by changing the applied magnetic field.

Klystron Oscillator and Gunn Oscillator The reflex klystron oscillator (originally used as the local oscillator of many a World War II radar receiver) and the Gunn diode oscillator are two very different type of devices. Both have relatively high phase noise, but when coupled to a high-Q external resonant cavity they can be of high stability. The use of superconducting high-Q cavities can produce "extremely low phase noise levels."[20]

High-Temperature Superconducting Oscillators As was mentioned in Sec. 3.7, the phase noise of oscillators can be improved by employing very low loss superconductive resonators, especially those that are superconductive at the temperature of liquid nitrogen, 77 degrees.[24]

Direct Digital Synthesis[25] A *frequency synthesizer* produces one or more frequencies over a wide spectrum by translating the stable frequency of a precision frequency source, such as a crystal-controlled oscillator. In *direct synthesis* a single precision oscillator is multiplied and/or divided to obtain a desired frequency. When this process is performed digitally, it is called *direct digital synthesis* (DDS).[26] A DDS can generate the multiple frequencies needed for the stalo, coho, a second LO if dual conversion is used, and timing frequencies. It can also provide linear or nonlinear FM for pulse compression systems. A DDS generally uses a phase accumulator (to establish the time sequence), sine lookup table (to establish the amplitude of the signal waveform), a digital-to-analog converter, low pass filter, and frequency multiplier or heterodyne to translate to a higher frequency.[27] It has the advantage of extremely fast frequency switching, small size steps in frequency, excellent phase noise, reasonably good spurious performance, transient-free (phase continuous) changes in frequency, flexibility in applying modulation, and it achieves its good performance in a small volume.

A/D Converters The A/D converter, which changes analog signals to digital signals, is an important component of digital processing. There are many different ways it has been implemented.[28] Its performance for radar is judged by the number of bits into which it can quantize a signal and the sampling rate at which it can operate. As was mentioned in Sec. 3.5 (where the effect of the A/D converter on MTI performance was discussed and some examples of performance were given), the number of bits into which the A/D converter can quantize a signal decreases as the sampling rate, or bandwidth, increases. Thus the larger the bandwidth of the signal the more difficult it is to maintain good performance. The A/D converter sometimes can be a limitation in wideband radar or when large clutter attenuation is required.

Bandpass Sampling at IF Digital signal processing that is conducted at baseband (video) requires two baseband A/D converters and an in-phase and a quadrature channel. Although

baseband digital processing has been widely used, there are limitations. The two baseband converters have to be well balanced over a wide dynamic range and there cannot be significant phase errors between the two channels (the phase difference between the two channels cannot differ significantly from 90°). Waters and Jarrett[29] indicate that these problems do not appear if the A/D conversion is performed in the bandpass portion of the receiver at IF. The in-phase and quadrature components are obtained by a single A/D converter from the samples taken directly from the original IF signal. The phase errors between the two channels are considerably smaller with IF sampling than with baseband sampling. Although only one channel is needed in bandpass sampling, its sampling rate has to be greater than that of the A/D converters used in baseband sampling. Further discussion can be found in Sec. 3.5.

Digital Radar Receiver There does not seem to be a unique, well-accepted definition of a digital receiver. A digital radar receiver, ideally, could be thought of as one that is completely digital with a wide dynamic range A/D converter that operates directly on the signal received at the antenna terminals. This would be followed by a highly capable computer to perform the functions found in a radar receiver. It is difficult, however, to achieve such a receiver with the bandwidth and large dynamic range required for high-performance microwave MTI and pulse doppler radars. More realistically, a digital radar receiver might be one that uses an analog RF amplifier and mixer, and even analog IF circuitry, followed by an IF A/D converter and digital video processing.

A different and more practical definition of a digital radar receiver was proposed by Wu and Li,[30] who stipulate that such a receiver have two significant differences compared to analog radar receivers. It should utilize (1) a direct digital synthesizer (DDS) as the local oscillator and (2) direct bandpass sampling at IF before detection, with all subsequent processing being done digitally.

In addition to a high-speed A/D converter with many bits of quantization, a digital receiver requires the digital processing to have sufficient speed to operate in real time and to have a large enough information storage memory. There is little doubt that the major advances in radar and its increased applicability since the 1970s have been due to the phenomenal advances in digital processing technology. It is likely that "digits" will continue to be a major driver of future advances in radar performance.

Russian Cyclotron Wave Electrostatic Amplifier[31–33] Solid-state amplifiers have been a popular choice for the front-end of a radar receiver, but they are not the only choice. A Russian receiver development, called the cyclotron wave electrostatic amplifier (CWESA), has been a popular receiver for certain types of radars because of characteristics not available with other devices. It is also more usually known as an *electrostatic amplifier* (ESA). The ESA is said to have a low noise figure, bandwidths of 5 to 10 percent, linear phase variation with frequency, and other attributes suitable for a receiver front end; but its uniqueness is that it can sustain a high level of input power without additional protection and it can recover quickly from overload. Duplexers or receiver protectors are not needed.

In the ESA an electrostatic cyclotron wave is launched on a thin electron beam in an input structure; it is amplified in an intermediate structure, and then coupled to an output

structure. The thin electron beam at the cathode might have dimensions of 0.03 by 0.7 mm and a current of 250 to 280 μA. The theory[31] of this device will not be summarized here except to say that a longitudinal magnetic field is required so that the input signal, when coupled to the electron beam, results in cyclotron motion of the electrons. Permanent magnets are used to reduce weight. At *S* or *C* bands, these units are said to weigh approximately 2 kg, have an approximate volume of one liter, and a power consumption of 1 to 1.5 W. A 1.0 dB noise figure was achieved at frequencies up to 3 GHz and a 2.4 dB noise figure at 10 GHz.

When a large signal appears at the input to the ESA, it causes a large reflection coefficient (a large VSWR) so that the signal is entirely reflected and is not absorbed, which is unlike diode receiver protectors that absorb the input energy. Thus these receivers have been used in radars without additional duplexers or diode receiver protectors. When the overload is removed the device quickly returns to service, typically in about 20 ns at frequencies above *S* band.[31,32] Longer values of recovery time are experienced at lower frequencies. It has been claimed that in radar applications the ESA can withstand peak powers of 10 kW and average powers of 300 W at frequencies above *S* band, and higher powers at lower frequencies.

Sometimes a transistor amplifier is added as a second stage to obtain higher gain. Such a combined ESA and transistor, operating from 7 to 7.4 GHz with a single high-voltage supply of 400 V in addition to small filament and transistor amplifier supply voltages, produced a noise figure of 3.4 dB and a gain of approximately 23 dB. It could withstand in excess of 5 kW of peak and 150 W of average power at the input and recover in less than 50 ns. When a transistor second stage is used, such a device sometimes is called an *electrostatic combined amplifier* (ESCA).

A tunable version of the ESA was said to demonstrate very rapid tuning over a 50 percent bandwidth, with an instantaneous bandwidth of 1 percent.

The rapid recovery time of this amplifier makes it attractive for use with pulse doppler radars which require a high prf. Pulse doppler radars operate with high duty cycles so there is little range-space available. Long recovery times reduce the available range-space. If the duty cycle were 10 percent and the pulse width were 1 μs, a diode protector recovery time of 1 μs would significantly increase the receiver dead time and increase the minimum range. The 20 ns recovery time of the ESA would hardly be noticed. The duty cycles of high-prf pulse doppler radars can be as high as 0.3 to 0.5, which makes receiver recovery time even more important.

The testing of a 200-MHz bandwidth *X*-band ESA combined with a transistor second stage as an ESCA receiver for a high-prf pulse doppler radar was described by Ewell.[31] Pulse repetition rates were from 1 kHz to several hundred kHz and pulse durations less than 1 μs. For this application, the ESCA was considered superior to conventional gas TR tubes whose recovery times were too long and too unpredictable. They were also superior to multipactor discharges which had good recovery time but had high spike leakage that required varactor diode receiver protectors. They also were costly and required additional components such as an oxygen generator, ion pump, and cooling system. Ewell's measurements appear to confirm the consistency of this device to meet pulse doppler radar system requirements. It provided protection from overload, fast recovery time, linearity, and electronic control of dynamic range.

Because of its size, the ESA is not suitable for most applications of active-aperture phased array radar; but there are many important radar applications where an active aperture is not necessary. The ESA is attractive in those radar systems where a single or only a few receiver channels are used. At the time of publication of the cited references for this subsection, there were roughly ten thousand of these devices manufactured and in use in a number of systems around the world, mainly in Russia and China. An example is its use in the Russian S300 PMU air defense and anti-tactical ballistic missile (ATBM) system (NATO designation SA-10). This employs an *X*-band space-fed phased array radar with pulse doppler waveform designed to operate in high clutter and electromagnetic countermeasure environments.[31] The ESA is also found in the Russian S300V (NATO designation SA-12) air defense system. Barton points out that the electrostatic amplifier tube helps make the total RF loss in these Russian receiving systems significantly lower than the loss found in comparable Western systems.[34]

Phase Detector, Phase-Sensitive Detector In Fig. 3.7 of Sec. 3.1, the *phase detector* was introduced as the device in an MTI receiver that extracted the doppler frequency shift of an echo signal. It compared the echo signal to a reference signal (the coho) which was coherent with the MTI transmitter signal. In the MTI phase detector, it is the rate of change of phase of the echo signal with time that is of interest since it determines the doppler frequency shift of the echo from a moving target. In Fig. 4.4 of Sec. 4.2, the *phase-sensitive detector* was shown in the amplitude-comparison monopulse tracking radar to allow the extraction of the sign of the angle error along with its magnitude. The input to this detector was the angle-error signal and the signal from the sum channel which acted as the reference. In both the MTI radar and the monopulse tracker two sinusoidal voltage inputs were available to a nonlinear device. The two were coherent with respect to one another in that they could be thought of as being from the same source. In both these detectors, one of the two voltages is the reference and the other is the received echo signal.

Taylor[35] points out that the distinction between a phase detector and a phase-sensitive detector is not always clear because of the similarity of the analog circuits that perform these two functions. He states that it is generally agreed that a phase detector is one in which only phase information is present in the output; a phase-sensitive detector is one in which both phase and amplitude information are in the output; and a mixer when phase, amplitude, and frequency information are present in the output. He also points out that "doppler frequency shifts are excepted in this convention."

Krishnam[36] indicates that the difference between these two detectors is in the actual operating conditions and not the hardware. He states that it had been usual to assume that the reference and the signal are of the same amplitude for the phase detector. For the phase-sensitive detector it was common to assume that the reference is much larger than the signal. He then shows that other assumptions can be made. He denotes $V_1 = E_1 \sin \omega t$ as the reference and $V_2 = E_2 \sin (\omega t + \phi)$ as the signal. For his particular detector model, he then shows that when E_1 is exactly equal to E_2, the output is $E_0/E_1 = 2(|\cos \phi/2| - |\sin \phi/2|)$, which is approximately linear with respect to ϕ over the range $0 < \phi < \pi/2$. Under these conditions, the device can be considered as a phase detector. When $E_2 >> E_1$ (signal is large compared to the reference), the output is $E_0 = 2E_1 \cos \phi$, which also is a phase detector. When $E_2 < E_1$ so that the reference is larger than the largest

signal E_2 , the device is shown to be a "perfectly linear" phase-sensitive detector with an output $E_0 = +2E_2$ when $\phi = 0$, and $E_0 = -2E_2$ when $\phi = \pi$.

Example of a Receiver One seldom finds in the radar literature a paper on the design of a radar receiver. For some reason, receiver designers do not prepare such papers, or the journal editors and referees do not accept them. There is, however, at least one paper of which I am aware that describes the receiver for the original Aegis AN/SPY-1A shipboard air-defense system.[37] The receiver is in two parts. One part is the on-array portion which contains the low-noise amplifiers and related components. It is mounted to the rear of each of the four antenna faces of Aegis to minimize pre-RF amplifier losses. The other part is located both in the fore and aft deckhouses and contains components with minimal impact on the noise figure. There are eleven receiver channels: three monopulse tracking channels, one sidelobe interference blanker channel, six auxiliary ECCM sidelobe canceler channels, and one auxiliary channel that acts as a spare. Each channel has two inputs so they can be time-shared between two antenna arrays to reduce cost.

There is too much in the paper to adequately summarize here, but it is recommended as being one of the few examples available that provides an overview of radar receiver engineering not usually found in radar texts.

11.4 DUPLEXERS AND RECEIVER PROTECTORS

A pulse radar can time share a single antenna between the transmitter and receiver by employing a fast-acting switching device called a *duplexer*. On transmission the duplexer must protect the receiver from damage or burnout, and on reception it must channel the echo signal to the receiver and not to the transmitter. Furthermore it must accomplish the switching rapidly, in microseconds or nanoseconds, and it should be of low loss. For high-power applications, the duplexer is a gas-discharge device called a TR (transmit-receive) switch. The high-power pulse from the transmitter causes the gas-discharge device to break down and short circuit the receiver to protect it from damage. On receive, the RF circuitry of the "cold" duplexer directs the echo signal to the receiver rather than the transmitter. Solid-state devices have also been used in duplexers. In a typical duplexer application, the transmitter peak power might be a megawatt or more, and the maximum safe power that can be tolerated by the receiver might be less than a watt. The duplexer, therefore, must provide more than 60 to 70 dB of isolation between the transmitter and recovery with negligible loss on transmit and receive.

The duplexer cannot always do the entire job of protecting the receiver. In addition to the gaseous TR switch, a receiver might require diode or ferrite limiters to limit the amount of leakage that gets by the TR switch. These limiters, which have been called *receiver protectors*, also provide protection from the high-power radiation of other radars that might enter the radar antenna with less power than necessary to activate the duplexer, but with greater power than can be safely handled by the receiver. There might also be a mechanically actuated shutter to short-circuit and protect the receiver whenever the radar is not operating. Sometime the entire package of devices has been known as a *receiver*

protector.[38] The term is ambiguous, since receiver protector is also the name for the diode limiter or similar device that follows the duplexer for the purpose of reducing the leakage power passed by the duplexer. In this text the term receiver protector is used to denote a limiter that follows the duplexer. The duplexer, receiver protector, and other devices for preventing receiver damage are better known as the *duplexer system*, so as to prevent confusion by the same term (receiver protector) being used to describe the entire receiver protection system as well as one part of it.

Balanced Duplexer The balanced duplexer, shown in Fig. 11.3, is based on the short-slot hybrid junction which consists of two sections of waveguides joined along one of their narrow walls with a slot cut in the common wall to provide coupling between the two.[39] (The short-slot hybrid junction may be thought of as a broadband directional coupler with a coupling ratio of 3 dB.) Two TR tubes are used, one in each section of waveguide. In the transmit condition, Fig. 11.3a, power is divided equally into each waveguide by the first hybrid junction (on the left). Both gas-discharge TR tubes break down and reflect the incident power out the antenna arm as shown. The short-slot hybrid junction has the property that each time power passes through the slot in either direction, its phase is advanced by 90°. The power travels as indicated by the solid lines. Any power that leaks through the TR tubes (shown by the dashed lines) is directed to the arm with the matched dummy load and not to the receiver. In addition to the attenuation provided by the TR tubes, the hybrid junctions provide an additional 20 to 30 dB of isolation.

On reception the TR tubes do not fire and the echo signals pass through the duplexer and into the receiver as shown in Fig. 11.3b. The power splits equally at the first junction and because of the 90° phase advance on passing through the slot, the signal recombines in the receiving arm and not in the arm with the dummy load.

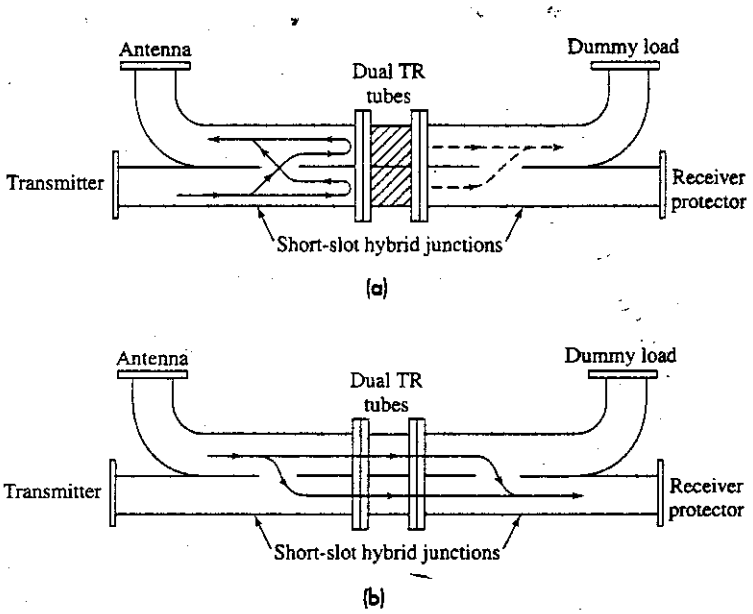


Figure 11.3 Balanced duplexer using dual TR tubes and two short-slot hybrid junctions. (a) Transmit condition and (b) receive condition.

The balanced duplexer is a popular form of duplexer with good power handling capability and wide bandwidth.

TR Tube The TR tube is a gas-discharge device designed to break down and ionize quickly at the onset of high RF power, and to deionize quickly once the power is removed. One construction of a TR consists of a section of waveguide containing one or more resonant filters and two glass-to-metal windows to seal in the gas at low pressure. A noble gas like argon in the TR tube has a low breakdown voltage, and offers good receiver protection and relatively long life. TR tubes filled only with pure argon, however, have relatively long deionization times (long recovery times) and are not suitable for short-range applications. Adding water vapor or a halogen gas to the tube speeds up the deionization time, but such tubes have shorter lifetimes than tubes filled only with a noble gas. Thus a compromise must usually be accepted between fast recovery time and long life.

To insure reliable and rapid breakdown of the TR tube on application of high power, an auxiliary source of electrons is supplied to the tube to help initiate the discharge. This may be accomplished with a "keep-alive," which produces a weak d-c discharge that generates electrons that diffuse into the TR where they assist in triggering the breakdown once RF power is applied by the transmitter. An alternative is to include a small source of radioactivity, such as tritium (a radioactive isotope of hydrogen), which produces low-energy-level beta rays to generate a supply of electrons.[40] The tritium is in compounded form as a tritide film. The radioactive source, sometimes called a *tritiated ignitor*, has the advantage of not increasing the wideband noise level as does a keep-alive discharge (by about 50 K) and has longer life (by an order of magnitude), but it passes more leakage energy so that it requires one or more cascaded PIN diode limiter stages to further attenuate the leakage.[41] The tritium ignitor needs no active voltages, so it allows the receiver protector to function with the radar off without the need for a mechanical shutter to protect the radar from nearby transmissions. Being a radioactive device, however, does cause concern about its handling and disposal. The combination of the tritium-activated TR followed by a diode limiter has been called a *passive TR-limiter*.

The TR is not a perfect switch; some transmitter power always leaks through to the receiver. The envelope of the RF leakage might be similar to that shown in Fig. 11.4. The short-duration, large-amplitude *spike* at the leading edge of the leakage pulse is the result of the finite time required for the TR to ionize and break down. Typically, this time is of the order of 10 nanoseconds. After the gas in the TR tube is ionized, the power leaking through the tube is considerably reduced from the peak value of the spike. This portion

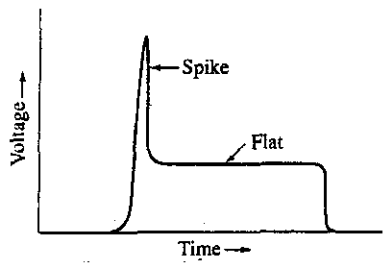


Figure 11.4 Leakage pulse through a TR tube.

of the leakage pulse is called the *flat*. Damage to the receiver front-end may result when either the energy contained within the spike or the power in the flat portion of the pulse is too large. The spike leakage of TR tubes varies with frequency and power and whether or not the tube is primed with electrons, but might be "typically" about one erg. The attenuation of the incident transmitter power might be of the order of 70 to 90 dB.

A fraction of the transmitter power incident on the TR tube is absorbed by the discharge. This is called *arc loss*. It might be 0.5 to 1 dB in tubes with water vapor and 0.1 dB or less with argon filling. On reception, the TR tube introduces an insertion loss of about 0.5 to 1 dB. The life of a TR tube is determined more by the amount of leakage power it allows to pass or when its recovery time becomes excessive rather than by its physical destruction or wear.

Solid-State Receiver Protectors, Diode Limiters Improvements in receiver sensitivity sometimes are obtained with front-ends and mixers that are more sensitive to damage from RF leakage. Such sensitive devices require better protection from the RF leakage of conventional duplexers. A PIN diode limiter placed in front of the receiver helps reduce the leakage and act as a *receiver protector*. A diode limiter passes low power with negligible attenuation, but above some threshold it attenuates the signal so as to maintain the output power constant. This property can be used for the protection of radar receivers in two different implementations depending whether the diodes are operated unbiased (self-actuated) or with a d-c forward-bias current. Unbiased operation without the use of an external current supply is also known as *passive*. It has the advantage of almost unlimited operating life, fast recovery time, no radioactive priming, and versatility to perform multiple roles.[38] Its chief limitation is its low power handling. A passive solid-state limiter for *X*-band WR90 waveguide with 7 percent bandwidth and a 1-μs pulse width had a peak power capability of 10 kW and a CW power capability of 10 W.[42] Its insertion loss was 0.6 dB, leakage power was 10 mW, and a 1.0-μs recovery time. When used with a 40-μs pulse width, this limiter could withstand 2 kW of peak power and 300 W CW, with a loss of 0.8 dB.

Biasing of the diodes during the high-power pulse, also known as *active*, is capable of handling a great deal more power than when operated passively. The diode is biased into its low impedance mode prior to the onset of the transmitter pulse. Although the active diode-limiter offers many advantages for use with duplexers, it does not protect the receiver when the bias is off. It thus offers poor protection against nearby asynchronous transmissions that arrive at the radar during the interpulse period or when the radar is shut down.

Incorporation of Sensitivity Time Control[43,44] In Sec. 7.8 the use of sensitivity time control (STC) was described as a method to reduce the effects of nearby large clutter echoes without seriously degrading the detection of desired targets at short range. STC is the programmed change of receiver gain with time, or range. At short ranges the receiver gain is lowered to reduce large nearby clutter echoes. As the pulse travels out in range the gain is increased until there are no more clutter echoes.

There are advantages for having the STC in the RF portion of the radar just ahead of the receiver. STC can be applied by biasing the diodes of a receiver protector to provide

a time-varying attenuation without adding to the receiver noise figure. There is no increase in insertion loss to obtain the STC action, above that inherent in the design of the receiver protector. The PIN diode stages provide the self-limiting action during transmit and the STC function during receive. The nonlinear nature of the diode requires a linearizing circuit to achieve the desired variation of attenuation with time. The STC variation depends on the nature of the terrain seen by the radar. A digitally controlled STC drive with random access memory allows the radar designer to employ different STC response profiles according to the various types of terrain that might be seen by the radar.

Varactor Receiver Protectors With fast-rise-time, high-power RF sources, the receiver protector may be required to self-limit in less than one nanosecond. This can be achieved with fast-acting PN (varactor) diodes. A number of diode stages, preceded by plasma limiters, might be employed. In one design, an *X*-band passive receiver protector was capable of limiting 1-ns rise time, multikilowatt RF pulses to 1-W spike levels.[45]

Ferrite Limiters The ferrite limiter has very fast recovery time (can be as low as several tens of nanoseconds), and if the power rating is not exceeded, it should have long life. The spike and flat leakage are low and it has been able to support a peak power of 100 kW;[46] but the insertion loss is usually higher (1.5 dB) and the package is generally longer, heavier, and more expensive than other receiver protectors. Except for the initial spike, the ferrite limiter is an absorptive device rather than a reflective device (as is a gas-tube TR) so that the average power capability of these devices can be a problem. Air or liquid cooling might be required. A diode limiter usually follows the ferrite limiter to reduce the leakage at high peak power.

Pre-TR Limiter[38,47] A pre-TR is a gaseous tube placed in front of a solid-state limiter. The function of the pre-TR is to reduce the power that has to be handled by the diode limiter. (It is similar to what was called a passive TR-limiter earlier in this section.) The pre-TR gas tube has high power handling capability, can operate with long pulses, has very fast recovery time, contains a radioactive priming source, but has limited operating life. Very high average power levels may require liquid cooling of the pre-TR mount. End of life for a pre-TR tube usually is caused by the increased recovery times that result from the cleanup of the gas within the tube.

The pre-TR tube can be a quartz cylinder filled with chlorine or a mixture of chlorine and an inert gas. Chlorine, a halogen gas, has a very rapid recovery time; typically a fraction of a microsecond for pulse widths up to 10 μs. The tube is mounted in a waveguide iris. In some cases, the quartz pre-TR tube can be designed to be field replaceable once it reaches end of life.

Mutipactor[46,48] The recovery times of high-power duplexers discussed thus far are from a fraction of a microsecond to several tens of microseconds. By employing the principle of multipacting, a recovery time as short as 5 or 10 ns is possible. Fast recovery time is important for high prf and high duty cycle radars. The multipactor is a vacuum tube and does not have the long recovery characteristics of a gas-filled tube. It contains surfaces capable of large secondary electron emission upon impact by electrons. The secondary

emission surfaces are biased with a d-c potential. The presence of RF energy causes electrons to make multiple impacts that generates by secondary emission a large electron cloud. The electron cloud moves in phase with the oscillations of the applied RF electric field to absorb energy from the RF field. RF power is dissipated thermally at the secondary emission surfaces, and the device requires liquid cooling to remove the absorbed power. Since it is a vacuum device, the recovery time of the multipactor is extremely fast. The flat-leakage power passed by the multipactor is often high enough to require a passive diode limiter to follow it. The multipactor offers no protection when the power is turned off. It has the disadvantage of being complex in that it requires liquid cooling, an ignitor electrode to ensure that multipacting starts quickly, an oxygen source to maintain the magnesium oxide surface that provides the secondary emission electrons, and a pump to maintain a good vacuum.

Solid-State Duplexers There has always been a desire to replace the gas-discharge duplexer with an all-solid-state duplexer because of the potential for long life, fast recovery time, no radioactive priming, and versatility. Although passive operation is desired, it is limited in power. The lowest loss and highest power handling are obtained with active circuits in which the PIN diodes are switched in synchronism with the transmitter pulses. Generally, diodes that can handle high power will have longer recovery times and tend to have higher leakage power—so that they might require additional stages of lower level limiters with increased loss and increased cost. A failure of the active drive circuit, however, could cause destruction of the diode switches as well as the receiver.

Several examples of all solid-state duplexers have been described in the literature. An *L*-band self-switching duplexer design used four PIN diodes that were biased by four fast-acting decoupled varactor detector diodes.[49] These detector diodes bias the PIN diodes into conduction in a time considerably shorter than the rise time of the RF power pulse. The device could handle 100-kW peak power with 100-W average power and a 3-μs pulse width. Its insertion loss was 0.5 dB. The duplexer was followed by a low-power multiple stage varactor limiter that reduced the spike and flat leakage of 2.8 kW and 32 W peak respectively to levels low enough that low-noise amplifiers were adequately protected. The recovery time was about 15 μs. A UHF solid-state duplexer also using four diodes was reported to have 300-kW peak power, 5-kW average power, 60-μs pulse width, and an insertion loss of 0.75 dB. A *C*-band solid-state duplexer with 16 PIN diodes was capable of 1-MW peak power with a 14-μs pulse width, 0.01 duty cycle, and insertion loss of less than 1 dB.[50] This device was followed by an additional low-power diode switch with an insertion loss of about 0.6 dB. It provided an isolation of 60 dB, making the total isolation of the duplexer system over 100 dB.

Circulators as Duplexers The ferrite circulator is a three- or four-port device that can, in principle, offer isolation of the transmitter and receiver. In the three-port circulator, the transmitter may be connected to port 1. It radiates out of the antenna connected to port 2. The received echo signal from the antenna is directed to port 3 which connects to the receiver. The isolation between the various ports might be from 20 to 30 dB, but the limitation in isolation is determined by the reflection (due to impedance mismatch) of the transmitter signal from the antenna that is then returned directly to the receiver. For

Table 11.1 Comparison of various types of duplexing devices

Device	Recovery Time	Average Power	Peak Power
TR tube	<1 μs to 100 μs		1 MW
Pre-TR	50 ns to 1 μs	50 kW	5 MW
Diode limiter	50 ns to 10 μs	1 kW	100 kW
Ferrite limiter	20 ns to 120 ns	10 W	100 kW
Multipactor	1 ns to 20 ns	500 W	80 kW
Electrostatic amplifier	20 ns	300 W, or higher	10 kW, perhaps as high as 500 kW

example, if the VSWR (voltage standing wave ratio) of the antenna were 1.5 (a pretty good value), about 4 percent of the transmitter power will be reflected by the antenna and return to the receiver. This corresponds to an isolation of 14 dB. If the VSWR were 2.0, the effective isolation is only 10 dB. To limit damage, a good receiver protector needs to be included. Circulators can be made to withstand high peak and average power; but large power capability generally comes with large size and weight. For example, an *S*-band differential phase-shift waveguide circulator that weighs 80 pounds has essentially the same insertion loss, isolation, and bandwidth of an *S*-band miniature coaxial Y-junction circulator that weighs 1.5 oz.[51] The larger circulator, however, can handle 50 kW of average power while the smaller circulator is rated at 50 W. (The ratio of powers exceeds the ratio of weights.)

Small-size circulators, usually in conjunction with a receiver protector, often are used as the duplexer in solid-state TR modules for active aperture phased arrays. (Note that, unfortunately, the term "TR" has been used for both T/R modules and TR duplexer gas-discharge tubes.)

Summary of Performance Table 11.1 summarizes the performance of various duplexer devices in term of recovery time and power handling. This table is adapted from the paper by Bilotta,[38] but modified from information in other references cited previously. The values in this table depend on frequency and other factors so they should not be considered as absolute limits, but only as an approximate guide.

11.5 RADAR DISPLAYS

Originally the radar display had the important purpose of visually presenting the output of the radar receiver in a form such that an operator could readily and accurately detect the presence of a target and extract information about its location. The display had to be designed so as not to degrade the radar information and to make it easy for the operator to perform with effectiveness the detection and information extraction function. It was not uncommon for an operator to employ a grease pencil to mark on the face of a cathode-ray-tube display the location of a target from scan to scan and manually extract the

target speed and direction. As digital signal processing and digital data processing improved, more and more of the detection and information extraction process was performed automatically by electronic means so that the role of the operator was less. Processed detection and target information now are displayed to the operator who has little responsibility for making the actual detection decision. Instead of displaying only detections, many surveillance radars display target track vectors along with auxiliary alphanumeric information to an operator.

When the display is connected directly to the output of the radar receiver without further processing, the output is called *raw video*. When the receiver output is first processed by an automatic detector or an automatic detector and tracker before display, it is called *synthetic video* or *processed video*. The requirements for the display differ somewhat depending whether raw or processed video is displayed. Some radar operators prefer to see on a display the raw video lightly superimposed on the processed video.

In many cases the operator does not see the unprocessed output of the radar. An example is the Nexrad doppler weather radar in which the radar measures three parameters in each resolution cell: the amplitude of the echo signal (proportional to its radar cross section), the mean radial velocity (from the doppler frequency shift) of the meteorological scatterers, and the variance of the radial velocity (a measure of the motion of the individual scatterers within the resolution cell). These three meteorological echo parameters in each resolution cell are passed to a computer that generates the many different types of weather products such as maps of precipitation, wind shear at various horizontal and vertical planes, mesocyclones, tornadoes, prediction of flooding, and many others.

The radar display is now more like the familiar television monitor or computer display that shows the entire scene continuously rather than just indicate the echoes from the region currently illuminated by the narrow antenna beam. Thus the role of the display has changed as the need for operator interpretation has decreased.

Types of Display Presentations The IEEE Standard on Radar Definitions includes 19 different types of display formats.[52] Most date to World War II and many are now seldom used. The standardized definitions do not cover all possible display formats. Given below are some of the more popular formats that have been employed. The IEEE uses the term "display" in its definitions but here we use either "scope" or "display" depending on what is perceived to be the more common usage. The following definitions are not precisely identical with the IEEE definitions, but they are consistent with them.

A-scope. *A deflection-modulated rectangular display in which the vertical deflection is proportional to the amplitude of the receiver output and the horizontal coordinate is proportional to range (or time delay).* This display is well suited to a staring or manually tracking radar, but it is not appropriate for a continually scanning surveillance radar since the ever-changing background scene makes it difficult to detect targets and interpret what the display is seeing.

B-scope. *An intensity-modulated rectangular display with azimuth angle indicated by one coordinate (usually horizontal) and range by the orthogonal coordinate (usually vertical).* It has been used in airborne military radar where the range and

angle to the target are more important than concern about distortion in the angle dimension.

C-scope. *A two-angle intensity-modulated rectangular display with azimuth angle indicated by the horizontal coordinate and elevation angle by the vertical coordinate.* One application is for airborne intercept radar since the display is similar to what a pilot might see when looking through the windshield. It is sometimes projected on the windshield as a heads-up display. The range coordinate is collapsed on this display so a collapsing loss might occur, depending how the radar information is processed.

E-scope. *An intensity-modulated rectangular display with range indicated by the horizontal coordinate and elevation angle by the vertical coordinate.* The E-scope provides a vertical profile of the radar coverage at a particular azimuth. It is of interest with 3D radars and in military airborne terrain-following radar systems in which the radar antenna is scanned in elevation to obtain vertical profiles of the terrain ahead of the aircraft. The E-scope is related to the RHI display.

PPI-display, or plan-position indicator. *An intensity-modulated circular display in which echo signals from reflecting objects are shown in plan view with range and azimuth angle displayed in polar (rho-theta) coordinates to form a map-like display.* Usually the center of the display is the location of the radar. A *sector-scan PPI* might be used with a forward-looking airborne radar to provide surveillance or ground mapping over a limited azimuth sector. An *offset PPI* is one where the origin (or location of the radar) is at a location other than the center of the display. This provides a larger display area for a selected portion of the coverage. The location of the radar with an offset PPI may be outside the face of the display.

RHI-display, or range-height indicator. *An intensity-modulated rectangular display with height (target altitude) as the vertical axis and range as the horizontal axis.* The scale of the height coordinate is usually expanded relative to the range coordinate. It has been used with meteorological radars to observe the vertical profile of weather echoes.

In addition, imaging radars such as synthetic aperture radar (SAR) and side-looking airborne radar (SLAR) generally display their output as a strip map with range as one coordinate and cross-range as the other coordinate. With the expanding graphics technology available from the computer industry, there is much more flexibility available in displaying radar information than previously.

Cathode Ray Tube Display The cathode ray tube (CRT), the origin of which dates back to the end of the nineteenth century, has been widely used as a radar display. There are two basic types of CRT displays. One is the *deflection-modulated CRT*, such as the A-scope, in which a target is indicated by the deflection of the electron beam. The A-scope displays the receiver output amplitude as a function of range, or time. An example was shown in Fig. 7.21, which illustrated the effect of frequency agility on clutter and target echoes. The other type is the *intensity-modulated CRT*, in which an echo is indicated by intensifying the electron beam and presenting a luminous spot on the face of

the CRT. An example is the PPI shown in Fig. 1.5. The TV CRT is also an example of an intensity-modulated display.

In general, a deflection-modulated display has the advantage of simpler circuits, and a target may be more readily discerned in the presence of noise, clutter, or interference. On the other hand, an intensity-modulated display, such as the PPI, has the advantage of presenting data in a more convenient and easily interpreted form. The deflection of the beam or the appearance of an intensity-modulated spot of a radar display caused by the presence of a target is commonly referred to as a *blip*.

Even though the CRT display has been widely used in radar, as well as in TV and in computers, it is by no means ideal. It employs a relatively large vacuum tube, which is a disadvantage compared to other types of displays. The entire display with its necessary circuits and controls can be even larger. The amount of information that can be presented is limited by the spot size of the electron beam. The number of resolution cells (pixels) per diameter might be one or two thousand, or more. In some high-resolution radars the number of resolvable range cells from the radar might be greater than the number of resolution cells available on the PPI. The result can be a collapsing loss. Increasing the CRT diameter does not necessarily increase the number of resolvable pixels since the spot diameter varies linearly with screen diameter. Another limitation of the intensity-modulated CRT is that its inherent dynamic range, or contrast ratio, might be of the order of 10 dB. This can cause blooming of the display by large targets so as to mask the blips from nearby smaller targets.

The decay characteristics of CRT phosphors are important when an operator views the screen to detect targets and extract information. The decay time of the visual information displayed should be long enough to allow the operator to detect the echo, yet short enough that the information painted on one scan does not interfere with the information entered from the succeeding scan. When processed information rather than raw video is displayed, the display characters might be dots, vectors to indicate direction and velocity, alphanumerics, or other appropriate symbols.

The improvements in electronic circuitry that have allowed major advances in signal and data processing have also benefited the CRT display. Character generators can be on a chip rather than occupy a bulky box. A complete deflection system can be placed on a chip. High-voltage power supplies that used to occupy a cubic foot and weigh 50 pounds[53] have been decreased considerably. Digital memories are small enough to replace the bulky analog scan converter. The required decay times for a PPI display need not depend on the decay characteristics of the phosphor, but an artificial persistence can be achieved with electronic circuitry that controls the refresh rate of the display. In the past, CRTs often had to be viewed in a darkened room or with a hood, but the brightness of CRTs has been increased so that they can be used in ambient light or in sunlit aircraft cockpits. There have been significant advances in other types of displays due to the demands of TV and computers, but the CRT has been able to make significant advances as well. In spite of limitations, the CRT has been a competitive display because of its ruggedness, cost, color capability, ability to operate over wide temperature ranges, its wide viewing angle, and ability to conveniently display the type of information obtained by a radar.

In a conventional PPI display when raw, unprocessed video is displayed to an operator, some background noise should be present since it improves the operator's ability to

make a detection decision. A completely "black scope" has reduced sensitivity compared to one with some background noise. This applies to a radar with raw video and not to a display presenting processed data where the detection decision is made by automatic circuitry without the intervention of an operator.

Stroke and Raster Displays The conventional stroke PPI display is generated in synchronism with the rotating antenna rather than all at once as it might appear in a photo. The photo of a PPI display is usually made by opening the shutter of a camera and holding it open for one or more scans, as was done in Fig. 1.5. An operator viewing a normal PPI, on the other hand, sees a rotating radial line, or strobe, that rotates in synchronism with the scanning antenna. The trace of the rotating strobe with a raw-video display with no echoes present is normally dim, but is brightened to indicate the location of an echo signal when detected by the radar. The brightened blip fades with time depending on the persistence characteristics of the phosphor or the refresh characteristics of the electronic circuitry. The operator focuses his or her attention on the rotating radial strobe line to detect targets. This type of display is called a *stroke* display. In a stroke display the operator concentrates on that portion of the display in the vicinity of the strobe line since that is usually all that will be strong enough to be seen.

A TV-like display based on a raster scan* to provide a continuous picture of the radar output has some advantages over the stroke display. It can be made brighter than a stroke display. Information from other sensors such as other radars, the Air Traffic Control Radar Beacon System (ATCRBS), military identification friend or foe (IFF), low light level TV, forward-looking IR (FLIR), collision avoidance systems, or information from civil or military data links, can all be combined on one display. Maps of the region viewed by the radar as well as alphanumeric information and graphics can also be superimposed on a raster display, in addition to the processed radar video and raw video. A scan converter is required to change the format of the stroke display to that of a raster TV-like display.

Scan Converter[54] A scan converter changes the r,θ coordinates of a PPI into the x,y coordinates of a raster (TV-like) display. The r,θ coordinates are natural for the radar, but the x,y coordinates are more natural for viewing the output of the radar on a display. Early analog scan-converters were bulky, had low resolution, and poor presentation of gray scales. They were of limited utility compared to modern digitally generated scan converters in which the r,θ polar coordinates (range and azimuth) of the radar information are converted to rectangular x,y coordinates and stored in digital memory to generate a raster TV-like display. The raster display can be presented continuously to the operator since it can be refreshed at a rapid rate. If desired, an artificial decay can be inserted to imitate the decay characteristics of natural phosphors. Alternatively, there need be no decay and the image can be frozen for the time equal to the radar revisit time, and then updated. Displays with 2560×2048 pixels can be accommodated. The outputs of multiple radars can be shown together with appropriate symbology even though they may be air-surveillance radars with considerably different antenna rotation rates (revisit times) and weapon control radars with pie-shaped angular coverage of a limited sector. The use of a

*A raster is a scan pattern in which an area is scanned from side to side in lines from top to bottom.

scan converter usually does not seriously extend the latency of the display; that is, the time between echo detection and its display is held to a minimum. The format of a raster display can be that of a TV display or that of a computer monitor. The advantage of a TV-display format is that it can be recorded on tape with an inexpensive consumer video cassette recorder (VCR), viewed with a conventional TV monitor, and easily remoted using standard video cabling.

Flat Panel Displays (FPD)[55] Interest in the flat panel display for radar evolved from its successful development for commercial computer and TV applications. There have been several different types of FPDs produced or explored, but not all are suitable for radar.

The *liquid crystal display* (LCD) has been widely used for nonradar applications where low weight, volume, and power consumption are important as in laptop computers, watches, instruments, and calculators. The LCD does not emit light of its own, but operates by controlling the light that either passes through it or reflects off it. Usually the light is directed from behind, and the display is said to be *backlit*. There are two types of liquid crystal displays: the passive-matrix LCD and the active-matrix LCD. In the latter, a thin-film transistor is associated with each pixel of the display. The passive-matrix has seen much wider nonradar application than the active matrix because of its lower cost, but the active-matrix LCD (AMLCD) has much higher resolution, better image quality, it can display in color, and has faster response (greater video bandwidths). Thus the active-matrix LCD has more potential for radar applications than does the passive-matrix.

Other types of FPDs are the *plasma display* which can produce large flat full-color displays, electroluminescent displays, light-emitting diodes, and field emitter displays.

Flat panel displays such as the AMLCD and the plasma display have several important advantages over conventional CRTs. They are smaller, lighter, occupy less volume (reduced depth), and require less power than CRTs. In addition they are expected to have better reliability and reduced life-cycle cost. For most radar applications, however, they have to be more rugged than for commercial applications in that they usually have to withstand greater shock and vibration, as well as extremes in temperature.

The FPD is especially well suited for cockpit displays in military airborne applications and is replacing the CRT in many airborne systems.[56] In addition to presenting radar information, military cockpit displays must also handle data provided by electronic warfare sensors, command and control information for situation awareness, navigation information, alphanumeric data, graphics, and others.

Color in Radar Displays The availability of color in a radar display allows another "dimension" for the presentation of information. It can aid in providing a clear, easily understood picture of the situation as observed by radar. It is also an "attention getter" to alert the operator to something special or dangerous. Different colors can be used to indicate such things as the outputs of different radars presented on the same display; the outputs from multiple beams of a 3D radar; the areas of adverse weather with color coding of the weather by rain intensity; range rings; target tracks along with single detections; identification information from civil ATCRBS and/or military IFF; superimposed video maps of the area being observed by radar; and superimposed raw video. It can also be used to indicate the altitude or cross section of individual radar

Table 11.2 An Example of the Display Colors Used for an Airborne Weather Avoidance Radar[57]

Storm Intensity	Rainfall Rate (mm/h)	Rainfall Rate (dBZ)	Display Color
Drizzle	0.25 mm/h	13 dB	Black
Light	1.0 mm/h	23 dB	Green
Moderate	4.0 mm/h	33 dB	Yellow
Industry standard pilot alert	11.5 mm/h	40 dB	Red

The parameter dBZ was explained in Sec. 7.6. The "industry standard pilot alert" is the rainfall rate above which there might be hail that could damage an aircraft or cause sufficient turbulence to disturb passengers. Pilots are advised to stay clear of such areas.

echoes by color coding the target blips or by use of alphanumeric color symbols inserted on the display.

An example of the use of color is in the airborne weather-avoidance radar display, where the intensity of precipitation is designated by a distinctive color. A listing of the colors used by one radar manufacturer to indicate storm intensities is shown in Table 11.2.[57]

The original tricolor shadow-mask cathode-ray tube used for color TV did not have the resolution capability of monochrome displays or the penetration color tube which used a multilayer screen. This has changed with the increasing demands for high resolution computer color graphics as well as high-definition TV. Although a monochrome display with various shades of gray can be made to exhibit much of the same information that a color display can, color is capable of providing a greater number of distinguishable shades than can a monochrome display, is more pleasing, and has been widely accepted.

REFERENCES

1. Taylor, J. W., Jr. "Receivers." *Radar Handbook*, M. Skolnik (Ed). New York: McGraw-Hill, 1990, Chap. 3.
2. Mumford, W. W., and E. H. Scheibe. *Noise Performance Factors in Communication Systems*. Dedham, MA: Horizon House—Microwave, Inc., 1968.
3. Pettai, R. *Noise in Receiving Systems*. New York: John Wiley, 1984.
4. Goldberg, H. "Some Notes on Noise Figures." *Proc. IRE 36* (October 1948), pp. 1205–1214.
5. Heil, T., B. Roehrich, and J. Hakoupian. "Advances in Receiver Front-End and Processing Components." *Microwave J.* 40 (January 1997), pp. 174–180.
6. Maas, S. A. *Microwave Mixers*, 2nd ed. Boston: Artech House, 1993.
7. Neuf, D. "Extended Dynamic Range Mixers." *Applied Microwave & Wireless* (Winter 1996), pp. 24–39.

8. Taylor, J. W., Jr. Ref. 1, Sec. 3.4.
9. Eaves, J. L., and E. K. Reedy. *Principles of Modern Radar*. New York: Van Nostrand Reinhold, 1987, Chap. 7.
10. Maas, S. A. "Microwave Mixers in the 90s." *Microwave J. 1990 State of the Art Reference*, pp. 61–72.
11. Maas, S. A. Ref. 6, Sec. 7.3.
12. Maas, S. A. Ref. 6, Sec. 7.3.5.
13. Dixon, R. C. *Radio Receiver Design*. New York: Marcel Dekker, 1998, Sec. 6.4.
14. Maas, S. A. Ref. 6, Sec. 4.6.1.
15. Dixon, R. C. Ref. 13, Sec. 6.5.
16. van der Ziel, A. "Unified Presentation of $1/f$ Noise in Electronic Devices: Fundamental $1/f$ Noise Sources." *Proc. IEEE* 76 pp. (March 1988), pp. 233–258.
17. Halford, D. "A General Model for f^{α} Spectral Density Random Noise with Special Reference to Flicker Noise $1/f$." *Proc. IEEE*. 56 (March 1968), pp. 251–257.
18. Taylor, J. W., Jr. Ref. 1, Sec. 3.5.
19. Goldman, S. J. *Phase Noise Analysis in Radar Systems Using Personal Computers*. New York: John Wiley, 1989.
20. Ewell, G. W. "Stability and Stable Sources." In *Coherent Radar Performance Estimation*, Scheer, J. A., and J. L. Kurtz (Eds.). Boston: Artech House, 1993, Chap. 2.
21. Losee, Ferril. *RF Systems, Components, and Circuits Handbook*. Boston: Artech House, 1997, Chap. 16.
22. Terman, F. *Radio Engineering*. New York: McGraw-Hill, 1937, Sec. 70.
23. Elmi, N., and M. Radmanesh. "Design of Low-Noise, Highly Stable Dielectric Resonator Oscillators." *Microwave J.* 39 (November 1996), pp. 104–112.
24. Khanna, A. P. S., M. Schmidt, and R. B. Hammond. "A Superconducting Resonator Stablized Low Phase Noise Oscillator." *Microwave J.* 34 (February 1991), pp. 127–130.
25. Galani, Z., and R. A. Campbell. "An Overview of Frequency Synthesizers for Radars." *IEEE Trans.* MTT-39 (May 1991), pp. 782–790.
26. Kroupa, V. F. (Ed.). *Direct Digital Frequency Synthesizers*. Piscataway, NJ.: IEEE Press, 1999.
27. Crawford, J. A. *Frequency Synthesizer Design Handbook*. Boston: Artech House, 1994, Secs. 7.3 and 7.5.
28. Hoeschele, D. F. *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, 2nd ed., New York: John Wiley, 1994.
29. Waters, W. M., and B. R. Jarrett. "Bandpass Signal Sampling and Coherent Detection." *IEEE Trans.* AES-18 (November 1982), pp. 731–736.

30. Wu, Y., and J. Li. "The Design of Digital Radar Receivers." *Proc. 1997 IEEE National Radar Conf.* pp 207–210; also reprinted in the *IEEE AES Systems Magazine* 13 (January 1998) pp. 35–41.

31. Manheimer, W. M., and G. Ewell, "Cyclotron Wave Electrostatic and Parametric Amplifiers." Naval Research Laboratory, Washington, D.C., Memorandum Rep. MR/6707-97-7910, February 28, 1997.

32. Budzinsky, Yu. A., and S. P. Kantyuk. "A New Class of Self-Protecting Low-Noise Microwave Amplifiers." *1993 IEEE International Microwave Symp. Digest*, Atlanta GA., vol. 2, p. 1123, June, 1993. See also information on the ISTOK Web Site http://www.istok.com.

33. Manheimer, W. M., and G. W. Ewell. "Electrostatic and Parametric Cyclotron Wave Amplifiers." *IEEE Trans.* PS-26 (August 1998), pp. 1282–1296.

34. Barton, D. K. "The 1993 Moscow Air Show." *Microwave J.* 37 (May 1994), pp. 24–39.

35. Taylor, J.W., Jr. Ref. 1, Sec. 3.10.

36. Krishnam, S. "Diode Phase Detectors." *Electronic & Radio Engr.* 36 (February 1959), pp. 45–50.

37. Socci, R. J. "The Aegis Radar Receiver." *Microwave J.* 21 (October 1978), pp. 38–47.

38. Bilotta, R. F. "Receiver Protectors: A Technology Update." *Microwave J.* 40 (August 1997), pp. 90–96.

39. Riblet, H. J. "The Short-Slot Hybrid Junction." *Proc. IRE* 40 (February 1952), pp. 180–184.

40. Golde, H. "Radioactive (Tritium) Ignitors for Plasma Limiters." *IEEE Trans.* ED-19 (August 1972), pp. 917–928.

41. Golde, H. "What's New with Receiver Protectors?" *Microwaves* 15 (January 1976), pp. 44–52.

42. Roberts, N. "A Review of Solid-State Radar Receiver Protection Devices." *Microwave J.* 34 (February 1991), pp. 121–125.

43. Ratliff, P. C., W. Cherry, M. J. Gawronski, and H. Goldie. "L-Band Receiver Protection Using Sensitivity Time Control." *Microwave J.* 19 (January 1976), pp. 57–60.

44. Goldie, H. "Combined Receiver Protector, AGC Attenuator and Sensitivity Time Control Device." *United States Patent* 4,194,200, March 18, 1980.

45. Nelson, T. M., and H. Goldie. "Fast Acting X-band Receiver Protector Using Varactors." *IEEE MTT Symp. Digest* (1974), pp. 176–177.

46. Brown, N. J. "Modern Receiver Protection Capabilities with TR-Limiters." *Microwave J.* 17 (February 1974), pp. 61–64.

47. "Product and Engineering Data, Receiver Protectors," Communication and Power Industries, Beverly Microwave Division, Beverly, MA, Brochure no. EDB-2417/273 (no date, but circa 1995)

48. Ferguson, P., and R. D. Dokkem. "For High-Power Protection . . . Try Multipacting." *Microwaves* 13 (July 1974), pp. 52–53.

49. Patel, S. D., and H. Goldie "A 100 kW Solid-State Coaxial Limiter for L-band, Part I." *Microwave J.* 25 (December 1981), pp. 61–65; Part II, vol. 26, (January 1982), pp. 93–97.

50. Hamilton, C. H. "A 1 MW C-band PIN Diode Duplexer." *1978 Conf. Proc. Military Microwaves*, pp.103–107, Microwave Exhibitions and Publishers, Ltd., Sevenoaks, Kent, England.

51. Rodrigue, G. P. "Circulators from 1 to 100 GHz." *Microwave J. 1989 State of the Art Reference*, vol. 32, pp. 115–132.

52. IEEE Standard Radar Definitions, *IEEE Std 686–1997,* Piscataway, NJ.

53. Wurtz, J. E. "CRT Update." *IEEE EASCON-77*, paper 12-2, 1977.

54. Some of the information in this subsection was obtained from the advertising literature of Folsom Research, Rancho Cordovia, CA, and from Robert W. Cribs, the CEO of Folsom Research.

55. Werner, K. "U. S. Display Industry on the Edge." *IEEE Spectrum* 32 (May 1995), pp. 62–69.

56. Hopper, D. G. (Ed.). *Cockpit Displays V: Displays for Defense Applications.* Proc. SPIE (International Society for Optical Engineering) 3363, 1998, Bellingham, WA.

57. Aires, R. H., and G. A. Lucchi. "Color Displays for Airborne Weather Radar." *RCA Engineer* 23 (February/March 1978), pp. 54–60.

PROBLEMS

11.1 (a) Find the overall noise figure of a superheterodyne receiver consisting of a low-noise RF amplifier with noise figure of 1.4 dB and gain of 15 dB, a mixer with 6.0-dB conversion loss and noise-temperature ratio of 1.2, and an IF amplifier with a noise figure of 1.0 dB. (b) What would be the noise figure of the receiver in (a) if the RF low-noise amplifier had a gain of 30 dB instead of 15 dB? (c) What would be the overall receiver noise figure if the IF amplifier noise figure in part (a) were 3.0 dB instead of 1.0 dB, and do you think this change is significant?

11.2 (a) Derive the overall noise figure of a receiver with noise figure F_r that is preceded by an RF device with a loss L_{RF}. (b) What is the overall noise figure of a transmission line and duplexer, which have a loss of 1.2 dB, connected to a receiver whose noise figure is 2.3 dB?

11.3 The greater the gain of the RF low-noise amplifier, the lower will be the overall noise figure. What adverse effect, however, occurs with an increase in the gain of the RF low-noise amplifier?

11.4 Show that the noise figure of a mixer is approximately the product of its conversion loss and the IF amplifier noise figure, when the diode mixer has a low noise-temperature ratio.

11.5 (a) Show that when a receiver of noise figure F_{RF} is attached to an antenna with antenna temperature T_a, the system noise figure F_s [Eq. 11.7] is

$$f_s = \frac{T_a}{T_0} + F_r$$

where T_0 is the standard temperature 290 K. (b) What is the system noise figure if the antenna temperature is 300 K, transmission line loss is 1.5 dB, and the receiver noise figure is 2.6 dB?

11.6 Consider a radar system with a receiver noise figure of 1.0 dB preceded by a transmission line with a loss of 0.5 dB. If the antenna temperature is 300 K, how important is it, in general, to attempt to reduce the receiver noise figure from 1.0 dB to 0.5 dB?

11.7 (a) Show that a radome with a loss L at a physical temperature T_{rd}, used with an antenna whose noise temperature in the absence of the radome is T_a', has an antenna noise temperature given by

$$T_a = \frac{T_a'}{L} + T_{rd}\frac{(L-1)}{L}$$

(b) Starting with the above, derive the *change* in antenna noise temperature $\Delta T_a = T_a - T_a'$ due to the presence of the radome. (c) Using the result of (a), what is the system noise figure? (d) What is the system noise figure when the receiver has a noise figure of 2.6 dB, a transmission line loss of 1.5 dB, an antenna with a noise temperature of 110 K, radome at a physical temperature of 310 K, and a loss of 0.6 dB through the radome?

11.8 (a) Find the noise bandwidth [Eq. (2.3)] of a network whose frequency response function $H(f) = (1 + jf/B)$, where B is the half-power bandwidth. (b) The above network is a single-stage low-pass RC filter. What is the expression for the noise bandwidth of a single-stage RLC bandpass network? (Do by inspection.) (c) Find the noise bandwidth for a low-pass filter with a gaussian shaped response $\exp[-a^2(f - f_0)^2]$, with $f > 0$.

11.9 Why might a double-conversion superheterodyne receiver be used instead of a single-conversion receiver? What limitation might there be in using a double-conversion receiver?

11.10 What effect does the local oscillator have on the receiver's dynamic range?

11.11 A receiver with a mixer front end has a noise figure of 6.6 dB. A low-noise amplifier (LNA) with a noise figure of 1.2 dB and gain of 10 dB is inserted ahead of the mixer to reduce the overall receiver noise figure. (a) How much of the new receiver noise figure is due to the mixer noise, and by how much has the dynamic range of the receiver been reduced? (b) If the gain of the LNA were increased to 20 dB, what would be the receiver noise figure and the decrease in dynamic range?

11.12 Why is a diode-limiter following the duplexer sometimes used as a receiver protector?

11.13 What duplexer options are available for a pulse doppler radar with a 10 percent duty cycle and a 0.1-μs pulse duration?

11.14 What limitations might there be in using an all-solid-state duplexer?

11.15 What is the usual cause, or criterion, for the end-of-life of a gas-discharge TR tube?

11.16 Consider a high prf pulse doppler radar with a 1-μs pulse width and a 10 percent duty cycle. (a) If a receiver protector is used that has a 1.5-μs recovery time, what fraction of the range coverage is blanked out? (b) If an electrostatic amplifier with a 30-ns recovery time is used, what fraction of the range coverage would be blanked out?

INDEX

Introduction to Radar Systems 3d edition
Errata

Page 7. Eq. [1.10] should be $R_{max} = \left[\frac{P_t A_e^2 \sigma}{4\pi\lambda^2 S_{min}} \right]^{1/4}$

Page 30. Eq. [2.1] should be $R_{max} = \left[\frac{P_t G A_e \sigma}{(4\pi)^2 S_{min}} \right]^{1/4}$

Page 33. The missing denominator of Eq. [2.3] is $|H(f_o)|^2$

Page 36. The middle left-hand part of Eq. [2.15] should be $\left\langle (x - m_1)^2 \right\rangle_{av}$

Page 40. The right-hand side of Eq. [2.23] should be $\exp\left(-\frac{V_T^2}{2\Psi_o} \right)$

Page 41. Eq. [2.25] should be $P_{fa} = \frac{\sum_{k=1}^{N} t_k}{\sum_{k=1}^{N} T_k} = \frac{\langle t_k \rangle_{av}}{\langle T_k \rangle_{av}} = \frac{1}{T_{fa} B}$

Page 44. The missing denominator on the far right-hand side of the first equation on this page is "rms noise voltage."

Page 49. The square root in the middle of Eq. [2.35] should be $\sqrt{n + 0.44}$. (A plus sign rather than a minus sign.)

Page 49. The missing denominator on the far right-hand side of Eq. [2.36] is $|E_i|^2$

Page 65. The missing denominator on the far right-hand side of Eq. [2.39] is $\sum_i a_i \cos\phi_i$.

Page 109. The last line of the footnote is "during the time of a pulse repetition period."

Page 117. The denominator of Eq. [3.15] should be $\int_0^\infty W(f)|H(f)|^2 \, df$.

Page 118. Delete the extraneous equal sign at the end of the first line of Eq. [3.20].

Page 222. At the bottom of Fig. 4.9b the left-hand side of the single bracket indicating the extent of ($d \sin \theta$) should extend from the middle of the dot, rather than from beyond the left of the dot.

Page 247. There should be a right-hand parenthesis at the end of the denominator of Eq. [4.7].

Page 288. The right-hand side of Eq. [5.23] should be $\sqrt{(av)^2 + 4} - 2$.

Page 319. The denominator of Eq. [6.5] should be $2BE/N_o$.

Page 332. The exponent in Eq. [6.41] should be $j2\pi f_d t$.

Page 497. In the caption for Fig. 8.9, change Lemont to Lamont.

Page 559. In the caption of Fig. 9.13, change Northrup to Northrop.

Page 636. The right-hand part of Eq. [9.36b] should be $G_o\left(1 - \frac{\pi^2}{3 \times 2^{2B}}\right)$

Page 643. In the caption of Fig. 9.49, change Northrup to Northrop.