CHAPTER 1

Introduction to Radar Systems and Signal Processing

1.1 History and Applications of Radar

The word "radar" was originally an acronym, RADAR, for "*ra*dio *d*etection *a*nd *r*anging." Today, the technology is so common that the word has become a standard English noun. Many people have direct personal experience with radar in such applications as measuring fastball speeds or, often to their regret, traffic control.

The history of radar extends to the early days of modern electromagnetic theory (Swords, 1986; Skolnik, 2001). In 1886, Hertz demonstrated reflection of radio waves, and in 1900 Tesla described a concept for electromagnetic detection and velocity measurement in an interview. In 1903 and 1904, the German engineer Hülsmeyer experimented with ship detection by radio wave reflection, an idea advocated again by Marconi in 1922. In that same year, Taylor and Young of the U.S. Naval Research Laboratory (NRL) demonstrated ship detection by radar and in 1930 Hyland, also of NRL, first detected aircraft (albeit accidentally) by radar, setting off a more substantial investigation that led to a U.S. patent for what would now be called a *continuous wave* (CW) radar in 1934.

The development of radar accelerated and spread in the middle and late 1930s, with largely independent developments in the United States, Britain, France, Germany, Russia, Italy, and Japan. In the United States, R. M. Page of NRL began an effort to develop pulsed radar in 1934, with the first successful demonstrations in 1936. The year 1936 also saw the U.S. Army Signal Corps begin active radar work, leading in 1938 to its first operational system, the SCR-268 antiaircraft fire control system, and in 1939 to the SCR-270 early warning system, the detections of which were tragically ignored at Pearl Harbor. British development, spurred by the threat of war, began in earnest with work by Watson-Watt in 1935. The British demonstrated pulsed radar that year, and by 1938 established the famous Chain Home surveillance radar network that remained active until the end of World War II. They also built the first airborne interceptor radar in 1939. In 1940, the United States and Britain began to exchange information on radar development. Up to this time, most radar work was conducted at *high frequency* (HF) and *very high frequency* (VHF) wavelengths; but with the British disclosure of the critical cavity magnetron microwave power tube and the United States formation of the Radiation Laboratory at the Massachusetts Institute of Technology, the groundwork was laid for the successful development of radar at the microwave frequencies that have predominated ever since.

Each of the other countries mentioned also carried out CW radar experiments, and each fielded operational radars at some time during the course of World War II. Efforts in France and Russia were interrupted by German occupation. On the other hand, Japanese efforts were aided by the capture of U.S. radars in the Philippines and by the disclosure of German technology. The Germans themselves deployed a variety of ground-based, shipboard, and airborne systems. By the end of the war, the value of radar and the advantages of microwave frequencies and pulsed waveforms were widely recognized.

Early radar development was driven by military necessity, and the military is still a major user and developer of radar technology. Military applications include surveillance, navigation, and weapons guidance for ground, sea, air, and space vehicles. Military radars span the range from huge ballistic missile defense systems to fist-sized tactical missile seekers.

Radar now enjoys an increasing range of applications. One of the most common is the police traffic radar used for enforcing speed limits (and measuring the speed of baseballs and tennis serves). Another is the "color weather radar" familiar to every viewer of local television news. The latter is one type of meteorological radar; more sophisticated systems are used for large-scale weather monitoring and prediction and atmospheric research. Another radar application that affects many people is found in the air traffic control systems used to guide commercial aircraft both en route and in the vicinity of airports. Aviation also uses radar for determining altitude and avoiding severe weather, and may soon use it for imaging runway approaches in poor weather. Radar is commonly used for collision avoidance and buoy detection by ships, and is now beginning to serve the same role for the automobile and trucking industries. Finally, spaceborne (both satellite and space shuttle) and airborne radar is an important tool in mapping earth topology and environmental characteristics such as water and ice conditions, forestry conditions, land usage, and pollution. While this sketch of radar applications is far from exhaustive, it does indicate the breadth of applications of this remarkable technology.

This text tries to present a thorough, straightforward, and consistent description of the signal processing aspects of radar technology, focusing primarily on the more fundamental functions common to most radar systems. Pulsed radars are emphasized over CW radars, though many of the ideas are applicable to both. Similarly, *monostatic* radars, where the transmitter and receiver antennas are collocated (and in fact are usually the same antenna), are emphasized over *bistatic* radars, where they are significantly separated, though again many of the results apply to both. The reason for this focus is that the majority of radar systems are monostatic, pulsed designs. Finally, the subject is approached from a digital signal processing (DSP) viewpoint as much as practicable, both because most new radar designs rely heavily on digital processing and because this approach can unify concepts and results often treated separately.

1.2 Basic Radar Functions

Most uses of radar can be classified as *detection, tracking*, or *imaging*. This text addresses all three, as well as the techniques of signal acquisition and interference reduction necessary to perform these tasks.

The most fundamental problem in radar is detection of an object or physical phenomenon. This requires determining whether the receiver output at a given time represents the echo from a reflecting object or only noise. Detection decisions are usually made by comparing the amplitude $A(t)$ of the receiver output (where t represents time) to a threshold $T(t)$, which may be set a priori in the radar design or may be computed adaptively from the radar data; in Chap. 6 it will be seen why this detection technique is appropriate. The time required for a pulse to propagate a distance R and return, thus traveling a total distance $2R$, is just

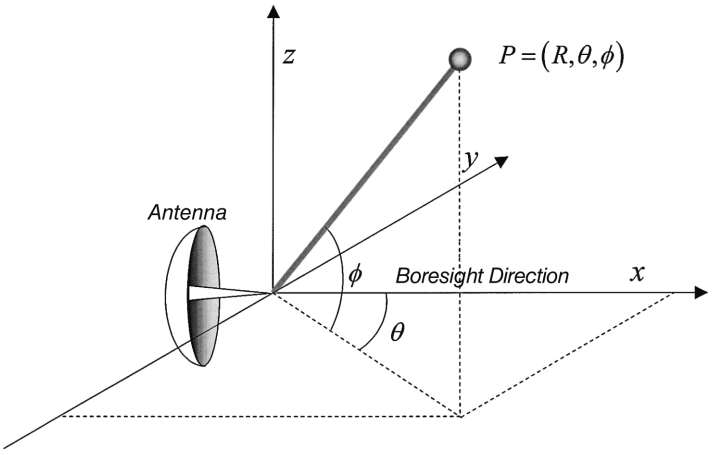


Figure 1.1 Spherical coordinate system for radar measurements.

$2R/c$; thus, if $A(t) > T(t)$ at some time delay t_0 after a pulse is transmitted, it is assumed that a target is present at range

$$R = \frac{ct_0}{2} \tag{1.1}$$

where c is the speed of light.[1]

Once an object has been detected, it may be desirable to track its location or velocity. A monostatic radar naturally measures position in a spherical coordinate system with its origin at the radar antenna's phase center, as shown in Fig. 1.1. In this coordinate system, the antenna look direction, sometimes called the *boresight* direction, is along the $+x$ axis. The angle θ is called *azimuth* angle, while ϕ is called *elevation* angle. Range R to the object follows directly from the elapsed time from transmission to detection as just described. Elevation and azimuth angle ϕ and θ are determined from the antenna orientation, since the target must normally be in the antenna main beam to be detected. Velocity is estimated by measuring the Doppler shift of the target echoes. Doppler shift provides only the radial velocity component, but a series of measurements of position and radial velocity can be used to infer target dynamics in all three dimensions.

Because most people are familiar with the idea of following the movement of a "blip" on the radar screen, detection and tracking are the functions most commonly associated with radar. Increasingly, however, radars are being used to generate two-dimensional images of an area. Such images can be analyzed for intelligence and surveillance purposes, for topology mapping, or for analysis of earth resources issues such as mapping, land use, ice cover analysis, deforestation monitoring, and so forth. They can also be used for "terrain following" navigation by correlating measured imagery with stored maps. While radar images have not achieved the resolution of optical images, the very low attenuation of electromagnetic waves at microwave frequencies gives radar the important advantage of "seeing" through clouds, fog, and precipitation very well. Consequently, imaging radars generate useful imagery when optical instruments cannot be used at all.

[1] $c = 2.99792458 \times 10^8$ m/s in a vacuum. A value of $c = 3 \times 10^8$ m/s is normally used except where very high accuracy is required.

The quality of a radar system is quantified with a variety of figures of merit, depending on the function being considered. In analyzing detection performance, the fundamental parameters are the *probability of detection* P_D and the *probability of false alarm* P_{FA}. If other system parameters are fixed, increasing P_D always requires accepting a higher P_{FA} as well. The achievable combinations are determined by the signal and interference statistics, especially the *signal-to-interference ratio* (SIR). When multiple targets are present in the radar field of view, additional considerations of resolution and sidelobes arise in evaluating detection performance. For example, if two targets cannot be resolved by a radar, they will be registered as a single object. If sidelobes are high, the echo from one strongly reflecting target may mask the echo from a nearby but weaker target, so that again only one target is registered when two are present. Resolution and sidelobes in range are determined by the radar waveform, while those in angle are determined by the antenna pattern.

In radar tracking, the basic figure of merit is *accuracy* of range, angle, and velocity estimation. While resolution presents a crude limit on accuracy, with appropriate signal processing the achievable accuracy is ultimately limited in each case by the SIR.

In imaging, the principal figures of merit are spatial resolution and dynamic range. Spatial resolution determines what size objects can be identified in the final image, and therefore to what uses the image can be put. For example, a radar map with 1 km by 1 km resolution would be useful for land use studies, but useless for military surveillance of airfields or missile sites. Dynamic range determines image contrast, which also contributes to the amount of information that can be extracted from an image.

The purpose of signal processing in radar is to improve these figures of merit. SIR can be improved by pulse integration. Resolution and SIR can be jointly improved by pulse compression and other waveform design techniques, such as frequency agility. Accuracy benefits from increased SIR and interpolation methods. Sidelobe behavior can be improved with the same windowing techniques used in virtually every application of signal processing. Each of these topics are discussed in the chapters that follow.

Radar signal processing draws on many of the same techniques and concepts used in other signal processing areas, from such closely related fields as communications and sonar to very different applications such as speech and image processing. Linear filtering and statistical detection theory are central to radar's most fundamental task of target detection. Fourier transforms, implemented using *fast Fourier transform* (FFT) techniques, are ubiquitous, being used for everything from fast convolution implementations of matched filters, to Doppler spectrum estimation, to radar imaging. Modern model-based spectral estimation and adaptive filtering techniques are used for beamforming and jammer cancellation. Pattern recognition techniques are used for target/clutter discrimination and target identification.

At the same time, radar signal processing has several unique qualities that differentiate it from most other signal processing fields. Most modern radars are coherent, meaning that the received signal, once demodulated to baseband, is complex-valued rather than real-valued. Radar signals have very high dynamic ranges of several tens of decibels, in some extreme cases approaching 100 dB. Thus, gain control schemes are common, and sidelobe control is often critical to avoid having weak signals masked by stronger ones. SIR ratios are often relatively low. For example, the SIR at the point of detection may be only 10 to 20 dB, while the SIR for a single received pulse prior to signal processing is frequently less than 0 dB.

Especially important is the fact that, compared to most other DSP applications, radar signal bandwidths are large. Instantaneous bandwidths for an individual pulse are frequently on the order of a few megahertz, and in some fine-resolution[2] radars may reach

[2]Systems exhibiting good or poor resolution are commonly referred to as high- or low-resolution systems, respectively. Since better resolution means a smaller numerical value, in this text the terms "fine" and "coarse" are used instead.

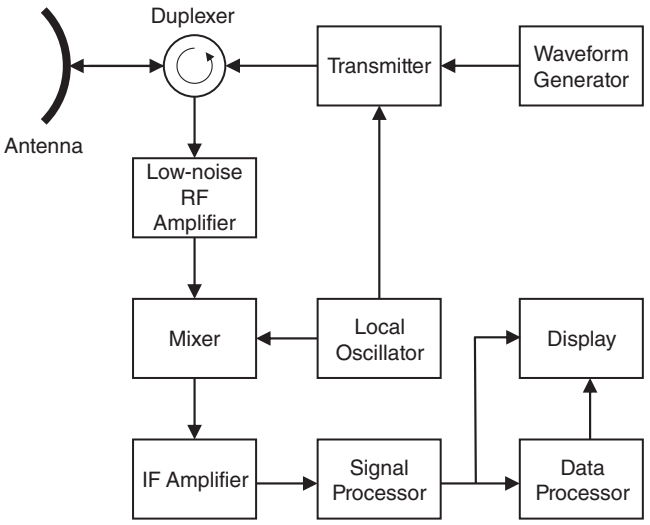


Figure 1.2 Block diagram of a pulsed monostatic radar.

several hundred megahertz and even as high as 1 GHz. This fact has several implications for digital signal processing. For example, very fast *analog-to-digital* (A/D) converters are required. The difficulty of designing good converters at multi-megahertz sample rates has historically slowed the introduction of digital techniques into radar signal processing. Even now, when digital techniques are common in new designs, radar word lengths in high-bandwidth systems are usually a relatively short 8 to 12 bits, rather than the 16 bits common in many other areas. The high data rates have also historically meant that it has often been necessary to design custom hardware for the digital processor in order to obtain adequate throughput, that is, to "keep up with" the onslaught of data. This same problem of providing adequate throughput has resulted in radar signal processing algorithms being relatively simple compared to, say, sonar processing techniques. Only in the late 1990s has Moore's Law[3] provided enough computing power to host radar algorithms for a wide range of systems on commercial hardware. Equally important, this same technological progress has allowed the application of new, more complex algorithms to radar signals, enabling major improvements in detection, tracking, and imaging capability.

1.3 Elements of a Pulsed Radar

Figure 1.2 is one possible block diagram of a simple pulsed monostatic radar. The waveform generator output is the desired pulse waveform. The transmitter modulates this waveform to the desired *radio frequency* (RF) and amplifies it to a useful power level. The transmitter output is routed to the antenna through a *duplexer*, also called a *circulator* or *T/R switch* (for transmit/receive). The returning echoes are routed, again by the duplexer, into the radar receiver. The receiver is usually a superheterodyne design, and often the first stage is a low-noise RF amplifier. This is followed by one or more stages of modulation of the received

[3]Gordon Moore's famous 1965 prediction was that the number of transistors on an integrated circuit would double every 18 to 24 months. This prediction has held remarkably true for nearly 40 years, enabling the computing and networking revolutions that began in earnest in the 1980s.

signal to successively lower *intermediate frequencies* (IFs) and ultimately to *baseband*, where the signal is not modulated onto any carrier frequency. Each modulation is carried out with a *mixer* and a *local oscillator* (LO). The baseband signal is next sent to the signal processor, which performs some or all of a variety of functions such as pulse compression, matched filtering, Doppler filtering, integration, and motion compensation. The output of the signal processor takes various forms, depending on the radar purpose. A tracking radar would output a stream of detections with measured range and angle coordinates, while an imaging radar would output a two- or three-dimensional image. The processor output is sent to the system display, the data processor, or both as appropriate.

The configuration of Fig. 1.2 is not unique. For example, many systems perform some of the signal processing functions at IF rather than baseband; matched filtering, pulse compression, and some forms of Doppler filtering are very common examples. The list of signal processing functions is redundant as well. For example, pulse compression and Doppler filtering can both be considered part of the matched filtering process. Another characteristic which differs among radars is at what point in the system the analog signal is digitized. Older systems are, of course, all analog, and many currently operational systems do not digitize the signal until it is converted to baseband. Thus, any signal processing performed at IF must be done with analog techniques. Increasingly, new designs digitize the signal at an IF stage, thus moving the A/D converter closer to the radar front end and enabling digital processing at IF. Finally, the distinction between signal processing and data processing is sometimes unclear or artificial.

In the next few subsections, the major characteristics of these principal radar subsystems are briefly discussed.

1.3.1 Transmitter and Waveform Generator

The transmitter and waveform generator play a major role in determining the sensitivity and range resolution of radar. Radar systems have been operated at frequencies as low as 2 MHz and as high as 220 GHz (Skolnik, 2001); laser radars operate at frequencies on the order of 10^{12} to 10^{15} Hz, corresponding to wavelengths on the order of 0.3 to 30 μm (Jelalian, 1992). However, most radars operate in the microwave frequency region of about 200 MHz to about 95 GHz, with corresponding wavelengths of 0.67 m to 3.16 mm. Table 1.1 summarizes the letter nomenclature used for the common nominal radar bands (IEEE, 1976). The millimeter wave band is sometimes further decomposed into approximate subbands of 36 to 46 GHz (Q band), 46 to 56 GHz (V band), and 56 to 100 GHz (W band) (Richards et al., 2010).

Within the HF to K_a bands, specific frequencies are allocated by international agreement to radar operation. In addition, at frequencies above X band, atmospheric attenuation of electromagnetic waves becomes significant. Consequently, radar in these bands usually operates at an "atmospheric window" frequency where attenuation is relatively low. Figure 1.3 illustrates the atmospheric attenuation for one-way propagation over the most common radar frequency ranges under one set of atmospheric conditions. Most K_a band radars operate near 35 GHz and most W band systems operate near 95 GHz because of the relatively low atmospheric attenuation at these wavelengths.

Lower radar frequencies tend to be preferred for longer range surveillance applications because of the low atmospheric attenuation and high available powers. Higher frequencies tend to be preferred for finer resolution, shorter range applications due to the smaller achievable antenna beamwidths for a given antenna size, higher attenuation, and lower available powers.

Weather conditions can also have a significant effect on radar signal propagation. Figure 1.4 illustrates the additional one-way loss as a function of RF frequency for rain rates ranging from a drizzle to a tropical downpour. X-band frequencies (typically 10 GHz) and

Band	Frequencies	Wavelengths
HF	3–30 MHz	100–10 m
VHF	30–300 MHz	10–1 m
UHF	300 MHz–1 GHz	1–30 cm
L	1–2 GHz	30–15 cm
S	2–4 GHz	15–7.5 cm
C	4–8 GHz	7.5–3.75 cm
X	8–12 GHz	3.75–2.5 cm
K_u	12–18 GHz	2.5–1.67 cm
K	18–27 GHz	1.67–1.11 cm
K_a	27–40 GHz	1.11 cm–7.5 mm
mm	40–300 GHz	7.5–1 mm

Table 1.1 Letter Nomenclature for Nominal Radar Frequency Bands

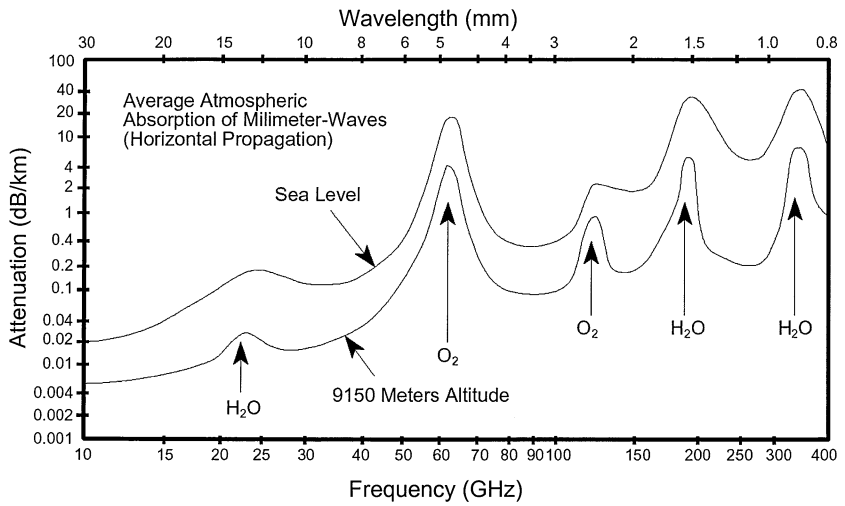


Figure 1.3 One-way atmospheric attenuation of electromagnetic waves. (*Source:* EW and Radar Systems Engineering Handbook, Naval Air Warfare Center, Weapons Division, http://ewhdbks.mugu.navy.mil/)

below are affected significantly only by very severe rainfall, while millimeter wave frequencies suffer severe losses for even light-to-medium rain rates.

Radar transmitters operate at peak powers ranging from milliwatts to in excess of 10 MW. One of the more powerful existing transmitters is found in the AN/FPS-108 COBRA DANE radar, which has a peak power of 15.4 MW (Brookner, 1988). The interval between pulses is called the *pulse repetition interval* (PRI), and its inverse is the *pulse repetition frequency* (PRF). PRF varies widely but is typically between several hundred pulses per second (pps) and several tens of thousands of pulses per second. The duty cycle of pulsed systems is

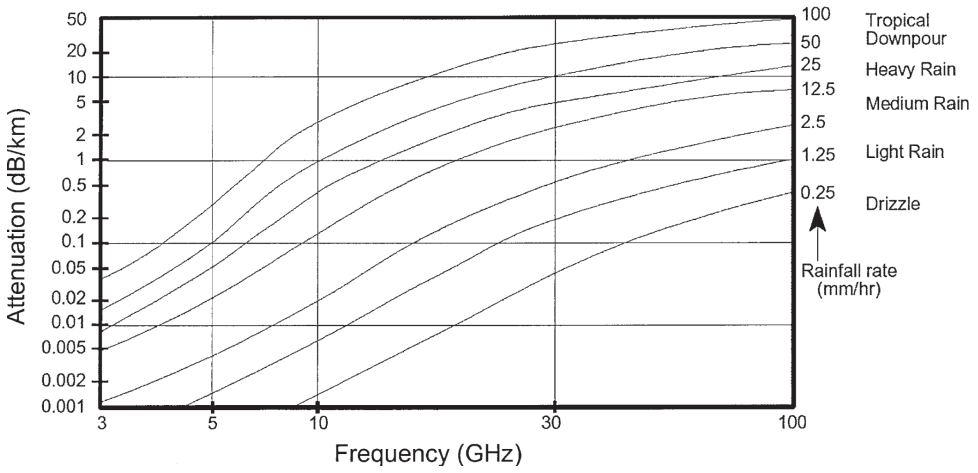


Figure 1.4 Effect of different rates of precipitation on one-way atmospheric attenuation of electromagnetic waves. (*Source:* EW and Radar Systems Engineering Handbook, Naval Air Warfare Center, Weapons Division, http://ewhdbks.mugu.navy.mil/)

usually relatively low and often well below 1 percent, so that average powers rarely exceed 10 to 20 kW. COBRA DANE again offers an extreme example with its average power of 0.92 MW. Pulse lengths are most often between about 100 ns and 100 μs, though some systems use pulses as short as a few nanoseconds while others have extremely long pulses, on the order of 1 ms.

It will be seen (Chap. 6) that the detection performance achievable by a radar improves with the amount of energy in the transmitted waveform. To maximize detection range, most radar systems try to maximize the transmitted power. One way to do this is to always operate the transmitter at full power during a pulse. Thus, radars generally do not use amplitude modulation of the transmitted pulse. On the other hand, the nominal range resolution ΔR is determined by the waveform bandwidth β according to Chap. 4.

$$\Delta R = \frac{c}{2\beta} \tag{1.2}$$

For an unmodulated pulse, the bandwidth is inversely proportional to its duration. To increase waveform bandwidth for a given pulse length without sacrificing energy, many radars routinely use phase or frequency modulation of the pulse.

Desirable values of range resolution vary from a few kilometers in long-range surveillance systems, which tend to operate at lower RFs, to a meter or less in very fine-resolution imaging systems, which tend to operate at high RFs. Corresponding waveform bandwidths are on the order of 100 kHz to 1 GHz, and are typically 1 percent or less of the RF. Few radars achieve 10 percent bandwidth. Thus, most radar waveforms can be considered narrowband, bandpass functions.

1.3.2 Antennas

The antenna plays a major role in determining the sensitivity and angular resolution of the radar. A wide variety of antenna types are used in radar systems. Some of the more common

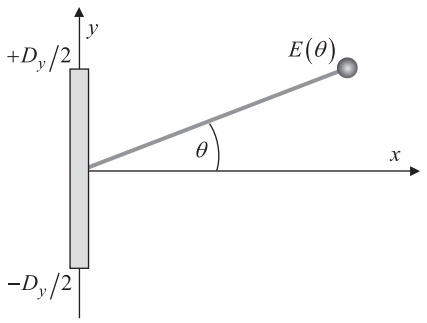


Figure 1.5 Geometry for one-dimensional electric field calculation on a rectangular aperture.

types are parabolic reflector antennas, scanning feed antennas, lens antennas, and phased array antennas.

From a signal processing perspective, the most important properties of an antenna are its gain, beamwidth, and sidelobe levels. Each of these follows from consideration of the antenna *power pattern*. The power pattern $P(\theta, \phi)$ describes the radiation intensity during transmission in the direction (θ, ϕ) relative to the antenna boresight. Aside from scale factors, which are unimportant for normalized patterns, it is related to the radiated electric field intensity $E(\theta, \phi)$, known as the antenna *voltage pattern*, according to

$$P(\theta, \phi) = |E(\theta, \phi)|^2 \tag{1.3}$$

For a rectangular aperture with an illumination function that is separable in the two aperture dimensions, $P(\theta, \phi)$ can be factored as the product of separate one-dimensional patterns (Stutzman and Thiele, 1998):

$$P(\theta, \phi) = P_\theta(\theta)\, P_\phi(\phi) \tag{1.4}$$

For most radar scenarios, only the *far-field* (also called *Fraunhofer*) power pattern is of interest. The far-field is conventionally defined to begin at a range of D^2/λ or $2D^2/\lambda$ for an antenna of aperture size D. Consider the azimuth (θ) pattern of the one-dimensional linear aperture geometry shown in Fig. 1.5. From a signal processing viewpoint, an important property of aperture antennas (such as flat plate arrays and parabolic reflectors) is that the electric field intensity as a function of azimuth $E(\theta)$ in the far field is just the inverse Fourier transform of the distribution $A(y)$ of current across the aperture in the azimuth plane (Bracewell, 1999; Skolnik, 2001):

$$E(\theta) = \int_{-D_y/2}^{D_y/2} A(y) e^{j(2\pi y/\lambda)\sin\theta} dy \tag{1.5}$$

where the "frequency" variable is $(2\pi/\lambda)\sin\theta$ and is in radians per meter. The idea of spatial frequency is introduced in App. B.

To be more explicit about this point, define $s = \sin\theta$ and $\zeta = y/\lambda$. Substituting these definitions in Eq. (1.5) gives

$$\frac{1}{\lambda}\int_{-D_y/2\lambda}^{D_y/2\lambda} A(\lambda\zeta) e^{j2\pi\zeta s} d\zeta = \hat{E}(s) \tag{1.6}$$

which is clearly of the form of an inverse Fourier transform. (The finite integral limits are due to the finite support of the aperture.) Because of the definitions of ζ and s, this transform relates the current distribution as a function of aperture position normalized by the wavelength to a spatial frequency variable that is related to the azimuth angle through a nonlinear mapping. It of course follows that

$$A(\lambda\zeta) = \int_{-\infty}^{+\infty} \hat{E}(s) e^{-j2\pi\zeta s} ds \tag{1.7}$$

The infinite limits in Eq. (1.7) are misleading, since the variable of integration $s = \sin\theta$ can only range from –1 to +1. Because of this, $\hat{E}(s)$ is zero outside of this range on s.

Equation (1.5) is a somewhat simplified expression that neglects a range-dependent overall phase factor and a slight amplitude dependence on range (Balanis, 2005). This Fourier transform property of antenna patterns will, in Chap. 2, allow the use of linear system concepts to understand the effects of the antenna on cross-range resolution and the pulse repetition frequencies needed to avoid spatial aliasing.

An important special case of Eq. (1.5) occurs when the aperture current illumination is a constant, $A(y) = A_0$. The normalized far-field voltage pattern is then the familiar sinc function,

$$E(\theta) = \frac{\sin[\pi(D_y/\lambda)\sin\theta]}{\pi(D_y/\lambda)\sin\theta} \tag{1.8}$$

If the aperture current illumination is separable, then the far-field is the product of two Fourier transforms, one in azimuth (θ) and one in elevation (ϕ).

The magnitude of $E(\theta)$ is illustrated in Fig. 1.6, along with the definitions for two important figures of merit of an antenna pattern. The angular resolution of the antenna is determined by the width of its mainlobe, and is conventionally expressed in terms of the *3-dB beamwidth*.

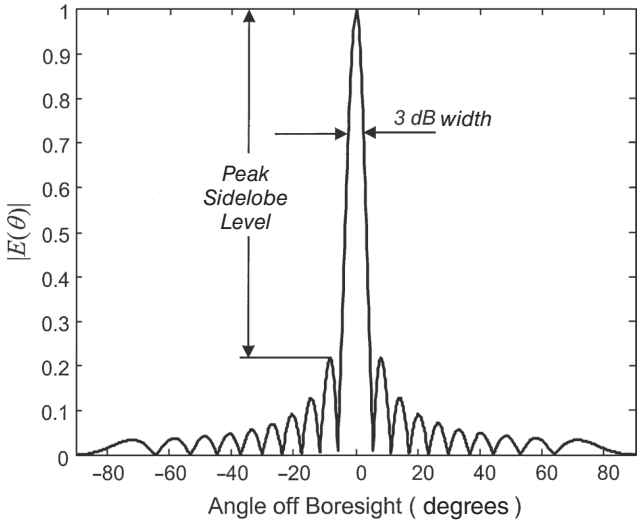


Figure 1.6 One-way radiation pattern of a uniformly illuminated aperture. The 3-dB beamwidth and peak sidelobe definitions are illustrated.

This can be found by setting $E(\theta) = 1/\sqrt{2}$ and solving for the argument $\alpha = \pi(D_y/\lambda)\sin\theta$. The answer can be found numerically to be $\alpha = 1.4$, which gives the value of θ at the –3-dB point as $\theta_0 = 0.445\lambda/D_y$. The 3-dB beamwidth extends from $-\theta_0$ to $+\theta_0$ and is therefore

$$\text{3-dB beamwidth} = \theta_3 = 2\sin^{-1}\left(\frac{1.4\lambda}{\pi D_y}\right) \approx 0.89\frac{\lambda}{D_y} \text{ radians} \tag{1.9}$$

Thus, the 3-dB beamwidth is 0.89 divided by the aperture size in wavelengths. Note that a smaller beamwidth requires a larger aperture or a shorter wavelength. Typical beamwidths range from as little as a few tenths of a degree to several degrees for a *pencil beam antenna* where the beam is made as narrow as possible in both azimuth and elevation. Some antennas are deliberately designed to have broad vertical beamwidths of several tens of degrees for convenience in wide area search; these designs are called *fan beam antennas*.

The *peak sidelobe* of the pattern affects how echoes from one scatterer affect the detection of neighboring scatterers. For the uniform illumination pattern, the peak sidelobe is 13.2 dB below the mainlobe peak. This is often considered too high in radar systems. Antenna sidelobes can be reduced by use of a nonuniform aperture distribution (Skolnik, 2001), sometimes referred to as *tapering* or *shading* the antenna. In fact, this is no different from the window or weighting functions used for sidelobe control in other areas of signal processing such as digital filter design, and peak sidelobes can easily be reduced to around 25 to 40 dB at the expense of an increase in mainlobe width. Lower sidelobes are possible, but are difficult to achieve due to manufacturing imperfections and inherent design limitations.

The factor of 0.89 in Eq. (1.9) is often dropped, thus roughly estimating the 3-dB beamwidth of the uniformly illuminated aperture as just λ/D_y radians. In fact, this is the 4-dB beamwidth, but since aperture weighting spreads the mainlobe it is a good rule of thumb.

The antenna *power gain* G is the ratio of peak radiation intensity from the antenna to the radiation that would be observed from a lossless, isotropic (omnidirectional) antenna if both have the same input power. Power gain is determined by both the antenna pattern and by losses in the antenna. A useful rule of thumb for a typical antenna is (Stutzman, 1998)

$$\begin{aligned} G &\approx \frac{26{,}000}{\theta_3\phi_3} \qquad (\theta_3, \phi_3 \text{ in degrees}) \\ &= \frac{7.9}{\theta_3, \phi_3} \qquad (\theta_3, \phi_3 \text{ in radians}) \end{aligned} \tag{1.10}$$

Though both higher and lower values are possible, typical radar antennas have gains from about 10 dB for a broad fan-beam search antenna to approximately 40 dB for a pencil beam that might be used for both search and track.

Effective aperture A_e is an important characteristic in describing the behavior of an antenna being used for reception. If a wave with power density W W/m^2 is incident on the antenna, and the power delivered to the antenna load is P, the effective aperture is defined as the ratio (Balanis, 2005)

$$A_e = \frac{P}{W} \qquad m^2 \tag{1.11}$$

Thus, the effective aperture is the area A_e such that, if all of the power incident on the area was collected and delivered to the load with no loss, it would account for all of the observed

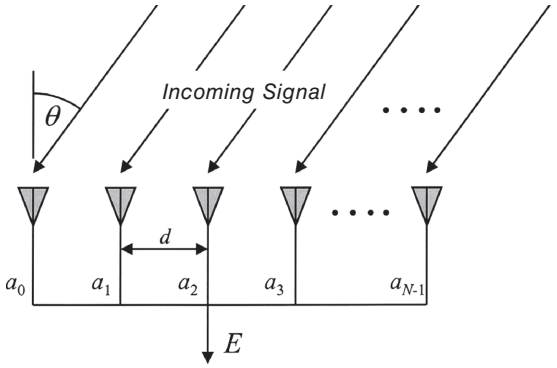


Figure 1.7 Geometry of the uniform linear array antenna.

power output of the actual antenna. (Note, however, that A_e is *not* the actual physical area of the antenna. It is a fictional area that accounts for the amount of incident power density captured by the receiving antenna.) Effective aperture is directly related to antenna directivity, which in turn is related to antenna gain and efficiency. For most antennas, the efficiency is near unity and the effective aperture and gain are related by (Balanis, 2005)

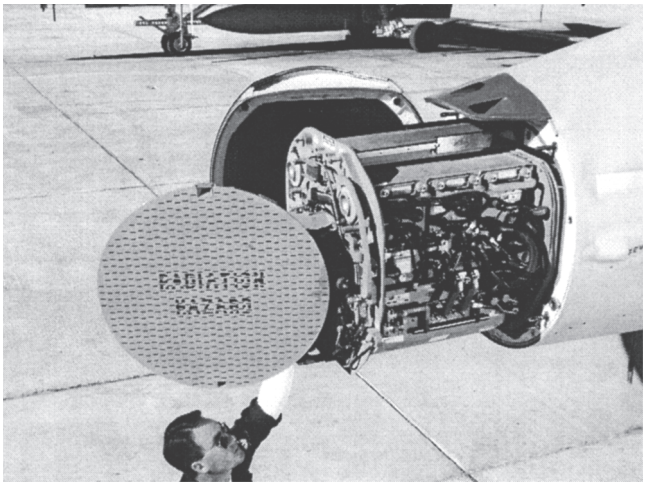
$$G = \frac{4\pi}{\lambda^2} A_e \tag{1.12}$$

Two more useful antenna concepts are the antenna *phase front* (or *wave front*) and *phase center* (Balanis, 2005; Sherman, 1984). A phase front of a radiating antenna is any surface on which the phase of the field is a constant. In the far-field, the phase fronts are usually approximately spherical, at least over localized regions. The phase center of the antenna is the center of curvature of the phase fronts. Put another way, the phase center is the point at which an isotropic radiator should be located so that the resulting phase fronts best match those of the actual antenna. The phase center concept is useful because it defines an effective location of the antenna, which can in turn be used for analyzing effective path lengths, Doppler shifts, and so forth. For symmetrically illuminated aperture antennas, the phase center will be centered in the aperture plane, but may be displaced forward or backward from the actual aperture. Referring to Fig. 1.5, the phase center would occur at $y = 0$, but possibly $x \neq 0$, depending on the detailed antenna shape.

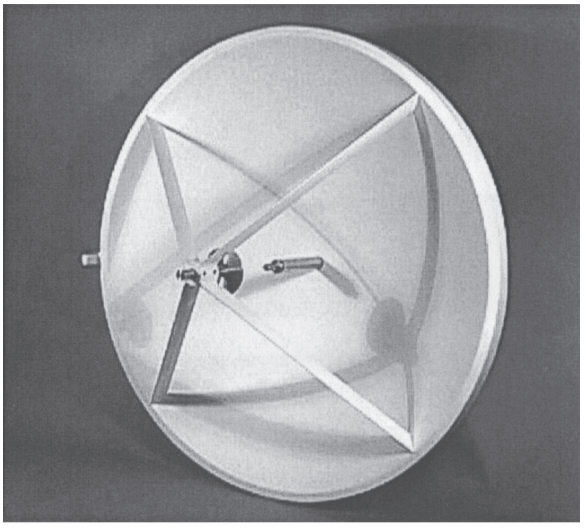
Another important type of antenna is the *array antenna*. An array antenna is one composed of a collection of individual antennas called array *elements*. The elements are typically identical dipoles or other simple antennas with very broad patterns. Usually, the elements are evenly spaced to form a *uniform linear array* as shown in Fig. 1.7. Figure 1.8 illustrates examples of real array and aperture antennas.

The voltage pattern for the linear array is most easily arrived at by considering the antenna in its receive, rather than transmit mode. Suppose the leftmost element is taken as a reference point, there are N elements in the array, and the elements are isotropic (unity gain for all θ). The signal in branch n is weighted with the complex weight a_n. For an incoming electric field $E_0 \exp(j\Omega t)$ at the reference element, the total output voltage E can easily be shown to be (Stutzman and Thiele, 1998; Skolnik, 2001)

$$E(\theta) = E_0 \sum_{n=0}^{N-1} a_n e^{j(2\pi/\lambda)nd \sin\theta} \tag{1.13}$$



(*a*)



(*b*)

Figure 1.8 Examples of typical array and aperture antennas. (*a*) Slotted phased array in the nose of an F/A-18 aircraft. This antenna is part of the AN/APG-73 radar system. (*b*) A Cassegrain reflector antenna. (*Image (a) courtesy of Raytheon Corp. Image (b) courtesy of Quinstar Corp. Used with permission*.)

This is similar in form to the *discrete Fourier transform* (DFT) of the weight sequence $\{a_n\}$. Like the aperture antenna, the antenna pattern of the linear array thus involves a Fourier transform, this time of the weight sequence (which determines the current distribution in the antenna). For the case where all the $a_n = 1$, the pattern is the familiar "aliased sinc" function, whose magnitude is

$$|E(\theta)| = E_0 \left| \frac{\sin[N(\pi d/\lambda)\sin\theta]}{\sin[(\pi d/\lambda)\sin\theta]} \right| \tag{1.14}$$

This function is very similar to that of Eq. (1.8) and Fig. 1.6. If the number of elements N is reasonably large (nine or more) and the product Nd is considered to be the total aperture size D, the 3-dB beamwidth is $0.89\lambda/D$, and the first sidelobe is 13.2 dB below the mainlobe peak; both numbers are the same as those of the uniformly illuminated aperture antenna. Of course, by varying the amplitudes of the weights a_n, it is possible to reduce the sidelobes at the expense of a broader mainlobe. The phase center is at the center of the array.

Actual array elements are not isotropic radiators. A simple model often used as a first-order approximation to a typical element pattern $E_{el}(\theta)$ is

$$E_{el}(\theta) \approx \cos\theta \tag{1.15}$$

The right-hand side of Eq. (1.13) is then called the *array factor* $AF(\theta)$, and the composite radiation pattern becomes

$$E(\theta) = AF(\theta)E_{el}(\theta) \tag{1.16}$$

Because the cosine function is slowly varying in θ, the beamwidth and first sidelobe level are not greatly changed by including the element pattern for signals arriving at angles near broadside (near $\theta = 0°$). The element pattern does reduce distant sidelobes, thereby reducing sensitivity to waves impinging on the array from off broadside.

The discussion so far has been phrased in terms of the transmit antenna pattern (for aperture antennas) or the receive pattern (for arrays), but not both. The patterns described have been *one-way antenna patterns*. The reciprocity theorem guarantees that the receive antenna pattern is identical to the transmit antenna pattern (Balanis, 2005). Consequently, for a monostatic radar, the *two-way antenna pattern* (power or voltage) is just the square of the corresponding one-way pattern. It also follows that the antenna phase center is the same in both transmit and receive modes.

1.3.3 Receivers

It was shown in Sec. 1.3.1 that radar signals are usually narrowband, bandpass, phase- or frequency-modulated functions. This means that the echo waveform $r(t)$ received from a single scatterer is of the form

$$r(t) = A(t)\sin[\Omega t + \theta(t)] \tag{1.17}$$

where the amplitude modulation $A(t)$ represents only the pulse envelope. The major function of the receiver processing is demodulation of the information bearing part of the radar signal to baseband, with the goal of measuring $\theta(t)$. Figure 1.9 illustrates the conventional approach to receiver design used in most classical radars.

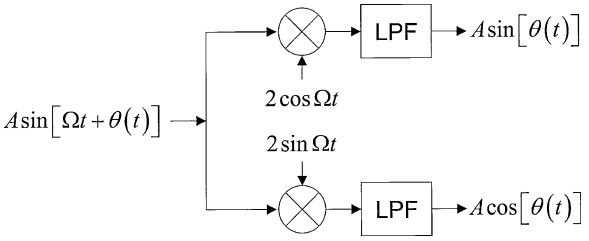


Figure 1.9 Conventional quadrature channel receiver model. In this illustration, the lower channel is the in-phase ("I") channel, and the upper is the quadrature phase ("Q") channel.

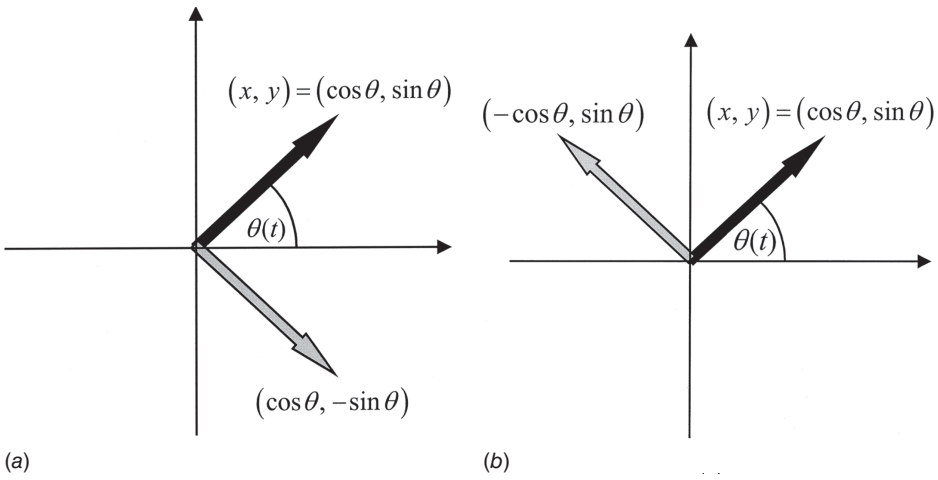


FIGURE 1.10 (*a*) The I channel of the receiver in Fig. 1.9 measures only the cosine of the phasor $\theta(t)$. (*b*) The Q channel measures only the sine of the phasor.

The received signal is split into two channels. One channel, called the *in-phase* or "I" channel of the receiver (the lower branch in Fig. 1.9) mixes the received signal with an oscillator, called the *local oscillator* (LO), at the radar frequency. This generates both sum and difference frequency components:

$$2\sin(\Omega t)\,A(t)\sin[\Omega t + \theta(t)] = A(t)\cos[\theta(t)] - A(t)\cos[2\Omega t + \theta(t)] \tag{1.18}$$

The sum term is then removed by the lowpass filter, leaving only the modulation term $A(t)\cos[\theta(t)]$. The other channel, called the *quadrature* phase or "Q" channel, mixes the signal with an oscillator having the same frequency but a 90° phase shift from the I channel oscillator. The Q channel mixer output is

$$2\cos(\Omega t)\,A(t)\sin[\Omega t + \theta(t)] = A(t)\sin[\theta(t)] + A(t)\sin[2\Omega t + \theta(t)] \tag{1.19}$$

which, after filtering, leaves the modulation term $A(t)\sin[\theta(t)]$. If the input $r(t)$ is written as $A(t)\cos[\Omega t + \theta(t)]$ instead, the upper channel of Fig. 1.9 becomes the I channel and the lower the Q channel, with outputs $A(t)\cos[\theta(t)]$ and $-A(t)\sin[\theta(t)]$, respectively. In general, the I channel is the one where the oscillator function (sine or cosine) is the same as that used in modeling the signal.

The reason that both the I and Q channels are needed is that either one alone does not provide sufficient information to determine the phase modulation $\theta(t)$ unambiguously. Figure 1.10 illustrates the problem. Consider the case shown in Fig. 1.10*a*. The signal phase $\theta(t)$ is represented as a solid black phasor in the complex plane. If only the I channel is implemented in the receiver, only the cosine of $\theta(t)$ will be measured. In this case, the true phasor will be indistinguishable from the gray phasor $-\theta(t)$. Similarly, if only the Q channel is implemented so that only the sine of $\theta(t)$ is measured, then the true phasor will be indistinguishable from the gray phasor of Fig. 1.10*b*, which corresponds to $\pi - \theta(t)$. When both the I and Q channels are implemented, the phasor quadrant is determined unambiguously.[4] In fact, the signal processor will

[4]This is analogous to the use of the two-argument atan2() function instead of the single-argument atan() function in many programming languages such as FORTRAN or C.

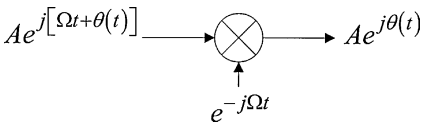


Figure 1.11 Simplified transmission and receiver model using complex exponential signals.

normally assign the I signal to be the real part of a complex signal and the Q signal to be the imaginary part, forming a single complex signal

$$x(t) = \mathrm{I}(t) + j\mathrm{Q}(t) = \mathrm{e}^{j\theta(t)} \tag{1.20}$$

Equation (1.20) implies a more convenient way of representing the effect of an ideal coherent receiver on a transmitted signal. Instead of representing the transmitted signal by a sine function, an equivalent complex exponential function is used instead.[5] The echo signal of (1.17) is thus replaced by

$$r(t) = A(t)e^{j[\Omega t+\theta(t)]} \tag{1.21}$$

The receiver structure of Fig. 1.9 is then replaced with the simplified model of Fig. 1.11, where the echo is demodulated by multiplication with a complex reference oscillator $\exp(-j\Omega t)$. This technique of assuming a complex transmitted signal and corresponding complex demodulator produces exactly the same result obtained in Eq. (1.20) by explicitly modeling the real-valued signals and the I and Q channels, but is much simpler and more compact. This complex exponential analysis approach is used throughout the remainder of the book. It is important to remember that this is an analysis technique; actual analog hardware must still operate with real-valued signals only. However, once signals are digitized, they may be treated explicitly as complex signals in the digital processor.

Figure 1.9 implies several requirements on a high-quality receiver design. For example, the local oscillator and the transmitter frequencies must be identical. This is usually ensured by having a single *stable local oscillator* (STALO) in the radar system that provides a frequency reference for both the transmitter and the receiver. Furthermore, many types of radar processing require *coherent* operation. The IEEE *Standard Radar Definitions* defines "coherent signal processing" as "echo integration, filtering, or detection using amplitude *and phase* of the signal referred to a coherent oscillator" (emphasis added) (IEEE, 1982). Coherency is a stronger requirement than frequency stability. In practice, it means that the transmitted carrier signal must have a fixed phase reference for several, perhaps many, consecutive pulses. Consider a pulse transmitted at time t_1 of the form $a(t - t_1)\sin[\Omega(t - t_1) + \phi]$, where $a(t)$ is the pulse shape. In a coherent system, a pulse transmitted at time t_2 will be of the form $a(t - t_2)\sin[\Omega(t - t_1) + \phi]$. Note that both pulses have the same argument $(t - t_1) + \phi$ for their sine term; only the envelope term changes location on the time axis. Thus, both sinusoids are referenced to the same absolute starting time and phase. This is as opposed to the second pulse being of the form $a(t - t_2)\sin[\Omega(t - t_2) + \phi]$, which is nonzero over the same time interval

[5]Although these formalizations are not needed for the discussions in this text and are therefore avoided for simplicity, it is worthwhile to note that the complex signal in Eq. (1.21) is the *analytic signal* associated with the real-valued signal of Eq. (1.17). The imaginary part of Eq. (1.21) is the *Hilbert transform* of the real part.

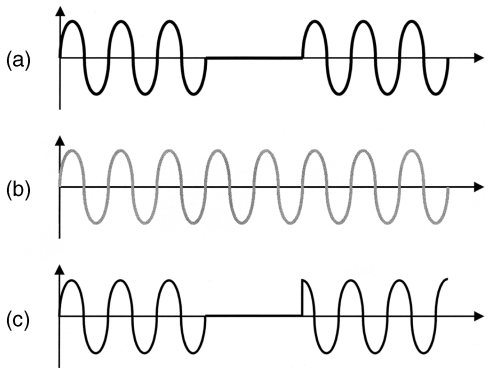


Figure 1.12 Illustration of the concept of a fixed phase reference in coherent signals. (*a*) Coherent pulse pair generated from the reference sinusoid. (*b*) Reference sinusoid. (*c*) Noncoherent pulse pair.

as the coherent pulse $a(t-t_2)\sin[\Omega(t-t_1)+\phi]$ and has the same frequency, but has a different phase at any instant in time. Figure 1.12 illustrates the difference visually. In the coherent case, the two pulses appear as if they were excised from the same continuous, stable sinusoid; in the noncoherent case, the second pulse is not in phase with the extension of the first pulse. Because of the phase ambiguity discussed earlier, coherency also implies a system having both I and Q channels.

Another requirement is that the I and Q channels have perfectly matched transfer functions over the signal bandwidth. Thus, the gain through each of the two signal paths must be identical, as must be the phase delay (electrical length) of the two channels. Of course, real receivers do not have perfectly matched channels. The effect of gain and phase imbalances will be considered in Chap. 3. Finally, a related requirement is that the oscillators used to demodulate the I and Q channels must be exactly in quadrature, that is, 90° out of phase with one another.

In the receiver structure shown in Fig. 1.9, the information-bearing portion of the signal is demodulated from the carrier frequency to baseband in a single mixing operation. While convenient for analysis, pulsed radar receivers are virtually never implemented this way in practice. One reason is that active electronic devices introduce various types of noise into their output signal, such as *shot noise* and *thermal noise*. One noise component, known as *flicker noise* or $1/F$ *noise*, has a power spectrum that behaves approximately as F^{-1} and is therefore strongest near zero frequency. Since received radar signals are very weak, they can be corrupted by $1/F$ noise if they are translated to baseband before being amplified.

Figure 1.13 shows a more representative *superheterodyne* receiver structure. The received signal, which is very weak, is amplified immediately upon reception using a *low-noise amplifier* (LNA). The LNA, more than any other component, determines the *noise figure* of the

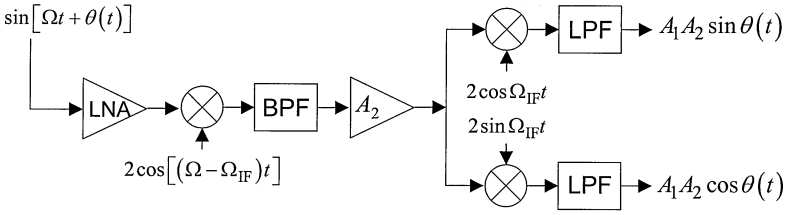


Figure 1.13 Structure of a superheterodyne radar receiver.

overall receiver. It will be seen in Sec. 2.3 that this is an important factor in determining the radar's *signal-to-noise* ratio (SNR), so good design of the LNA is important. The key feature of the superheterodyne receiver is that the demodulation to baseband occurs in two or more stages. First, the signal is modulated to an IF, where it receives additional amplification. Amplification at IF is easier because of the greater percentage bandwidth of the signal and the lower cost of IF components compared to microwave components. In addition, modulation to IF rather than to baseband incurs a lower conversion loss, improving the receiver sensitivity, and the extra IF amplification also reduces the effect of flicker noise. Finally, the amplified signal is demodulated to baseband. Some receivers may use more than two demodulation stages (so that there are two or more IF frequencies), but two stages is the most common choice. One final advantage of the superheterodyne configuration is its adaptability. The same IF stages can be used with variable RFs simply by tuning the LO so as to track changes in the transmitted frequency.

1.4 Common Threads in Radar Signal Processing

A radar system's success or failure in detecting, tracking, and imaging objects or features of interest in the environment is affected by various characteristics of those objects, the environment, and the radar itself, and how they are reflected in the received signals available for processing. Two of the most basic and important signal quality metrics are the signal-to-interference ratio and the resolution. Because of their importance, improving SIR and resolution is the major goal of most of the basic radar signal processing discussed in this text.

While subsequent chapters discuss a wide variety of signal processing techniques, there are a few basic ideas that underlie most of them. These include *coherent* and *noncoherent integration*, target *phase history modeling*, *bandwidth expansion*, and *maximum likelihood estimation*. The remainder of this section gives a heuristic definition of SIR and resolution, and then illustrates the simplest forms of integration, phase history modeling, and bandwidth expansion and how they affect SIR and resolution. Maximum likelihood estimation is deferred to Chap. 9 and App. A.

1.4.1 Signal-to-Interference Ratio and Integration

Consider a discrete-time signal $x[n]$ consisting of the sum of a "desired signal" $s[n]$ and an interfering signal $w[n]$:

$$x[n] = s[n] + w[n] \tag{1.22}$$

The discussion is identical for continuous time signals. The SIR χ of this signal is the ratio of the power of the desired signal to that of the interference. If $s[n]$ is deterministic, the signal power is usually taken as the peak signal value, and may therefore occur at a specific time t_0. In some deterministic cases, the average signal power may be used instead. The interference is almost invariably modeled as a random process, so that its power is the mean-square $\mathbf{E}\{|w[n]|^2\}$. If the interference is zero mean, as is very often the case, then the power also equals the variance of the interference, σ_w^2. If the desired signal is also modeled as a random process, then its power is also taken to be its mean-square or variance.

As an example, let $s[n]$ be a complex sinusoid $A\exp[j\omega n]$ and let $w[n]$ be complex zero mean white Gaussian noise of variance σ_w^2. The SIR of their sum $x[n]$ is

$$\chi_x = \frac{A^2}{\sigma_w^2} \tag{1.23}$$

In this case, the peak and average signal power are the same. If $s[n]$ is a real-valued sinusoid $A\cos[\omega n]$ and $w[n]$ is real-valued zero mean white Gaussian noise of variance σ_w^2, the peak

SIR would be the same but the average SIR would be $A^2/2\sigma_w^2$ because the average power of a real cosine or sine function of amplitude A is $A^2/2$.

A variation is the "energy SIR," defined as the ratio of the total energy $E_s = \sum |s[n]|^2$ in the signal $s[n]$ divided by the average noise power:

$$\chi_x = \frac{E_s}{\sigma_w^2} \tag{1.24}$$

The proportionality between E_s and A depends on the signal shape. For a rectangular pulse or a complex exponential of amplitude A and duration N samples, it is just $E_s = N \cdot A^2$. It can be seen in Chap. 6 that when matched filters are used, the peak SIR at the filter output is the energy SIR of the original signal.

SIR affects detection, tracking, and imaging performance in different ways. In general, detection performance improves with SIR in the sense that P_D increases for a given P_{FA} as SIR increases. For instance, it will be seen in Chap. 6 that for one particular model of the target behavior and detection algorithm, P_D is related to P_{FA} according to

$$P_D = (P_{FA})^{-1/(1+\chi)} \tag{1.25}$$

which shows that $P_D \to 1$ as $\chi \to \infty$ for fixed P_{FA}. As another example, the limit on precision (standard deviation of repeated measurements) due to additive noise of typical estimators of range, angle, frequency, or phase tends to decrease as $1/\sqrt{\chi}$; this behavior will be demonstrated in Chap. 9. In radar imaging (Chap. 8), SIR directly affects the contrast or dynamic range (ratio of reflectivity of brightest to dimmest visible features) of the image. These considerations make it essential to maximize the SIR of radar data, and many radar signal processing operations discussed in this text have as their primary goal increasing the SIR. The ways in which this is done will be discussed along with each technique.

1.4.2 Resolution

The closely related concepts of *resolution* and a *resolution cell* will arise frequently. Two equal-strength scatterers are considered to be *resolved* if they produce two separately identifiable signals at the system output, as opposed to combining into a single undifferentiated output.[6] The idea of resolution is applied in range, cross-range, Doppler shift or velocity, and angle of arrival. Two scatterers can simultaneously be resolved in one dimension, say range, and be unresolved in another, perhaps velocity.

Figure 1.14 illustrates the concept of resolution, in this case in frequency. Part (*a*) of the figure shows a portion of the positive frequency spectrum of the sum of two unit amplitude cosine functions with zero initial phase, one at 1000 Hz and one at 1500 Hz. This signal could represent the Doppler spectrum of two moving targets with the same echo strength but different radial velocities. The observation time is such that the mainlobe of the sinc function contributed by each has a Rayleigh width (peak to first null width) of 100 Hz. The two vertical dotted lines mark the two cosine frequencies. There are two distinct, well-separated peaks in the spectrum. The actual frequency of each peak is perturbed very slightly from the expected value by the sidelobes of the other sinusoid. Nonetheless, these two signal components are considered well resolved. Parts (*b*) through (*d*) of the figure repeat this measurement with the frequency spacing reduced to 100, 75, and 50 Hz. At 100 Hz spacing the two spectral peaks are still well resolved, though with more perturbation of the apparent frequencies, but as the separation drops below the Rayleigh width to 75 and then to 50 Hz, they

[6]The effects of unequal signal strength and noise on resolution are considered in (Mir and Wilkinson, 2008).

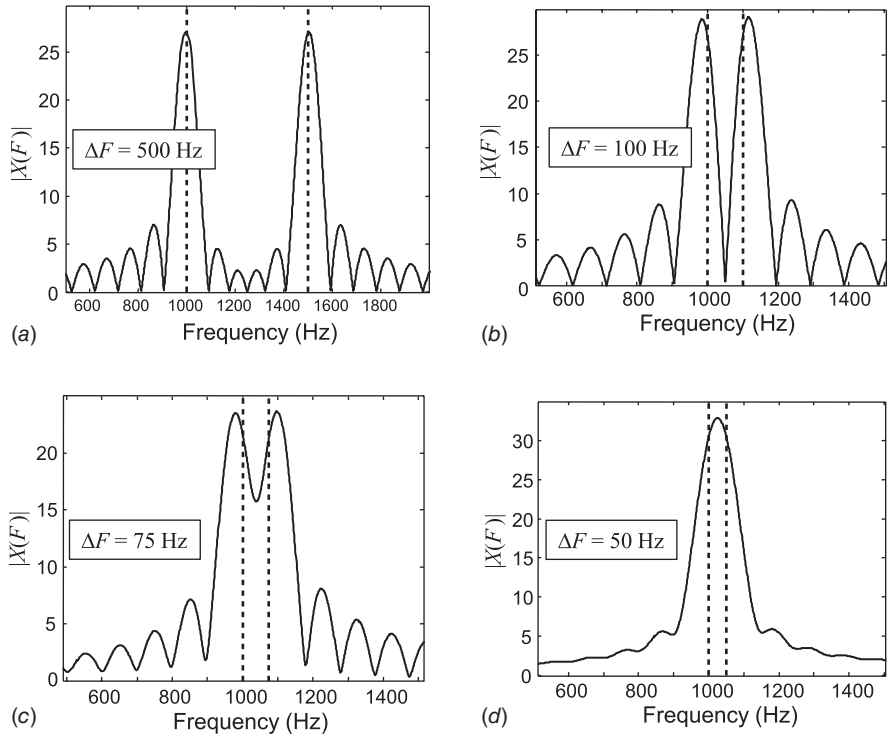


Figure 1.14 Resolution of two sinusoids in frequency, each having a Rayleigh width of 100 Hz. (*a*) Well resolved at 500 Hz spacing. (*b*) Well resolved at 100 Hz spacing. (*c*) Marginally resolved at 75 Hz spacing. (*d*) Unresolved at 50 Hz spacing.

blur into a single spectral peak. At 50 Hz, they are no longer resolved; the spectrum measurement does not show two separate signals. At 75 Hz they are marginally resolved, although a little noise added to the data would make that a precarious claim. It appears that a separation of about the Rayleigh width or greater is needed for clear resolution of the two frequencies. This demonstration also suggests that the width of the signature of a single isolated target is the major determinant of the system's resolution.

The resolution of a radar in turn determines the size of a resolution cell. A resolution cell in range, velocity, or angle is the interval in that dimension that contributes to the echo received by the radar at any one instant. Figure 1.15 illustrates resolution and the resolution interval in the range dimension for a simple constant-frequency pulse. If a pulse whose leading edge is transmitted at time $t = 0$ has duration τ seconds, then at time t_0 the echo of the leading edge of the pulse will be received from a scatterer at range $ct_0/2$. At the same time, echoes of the trailing edge of the pulse from a scatterer at range $c(t_0 - \tau)/2$ are also received. Any scatterers at intermediate ranges would also contribute to the voltage at time t_0. Thus, scatterers distributed over $c\tau/2$ in range contribute simultaneously to the received voltage. In order to resolve the contributions from two scatterers into different time samples, they must be spaced by more than $c\tau/2$ meters so that their individual echoes do not overlap in time. The quantity $c\tau/2$ is called the *range resolution* ΔR. Similarly, two- and three-dimensional resolution cells can be defined by considering the simultaneous resolution in, say, range, azimuth angle, and elevation angle.

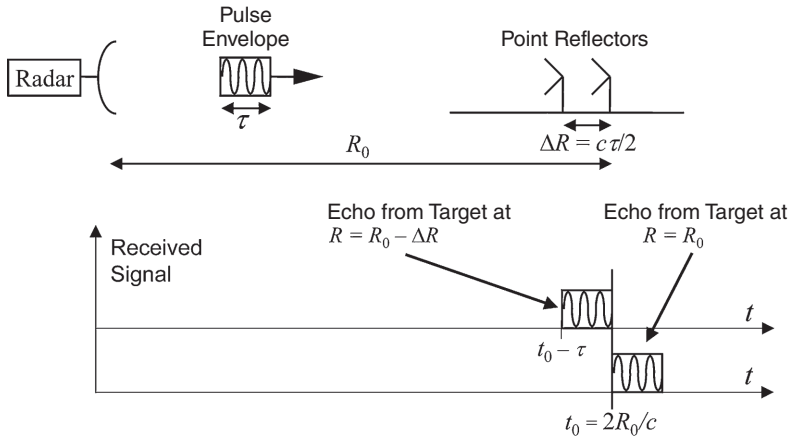


Figure 1.15 Geometry for describing conventional pulse range resolution. See text for explanation.

This description of range resolution applies only to unmodulated, constant frequency pulses. As will be seen in Chap. 4, pulse modulation combined with matched filtering can be used to obtain range resolution finer than $c\tau/2$.

Angular resolution in the azimuth and elevation dimensions is determined by the antenna beamwidths in the same planes. Two scatterers at the same range but different azimuth (or elevation) angles will contribute simultaneously to the received signal if they are within the antenna mainlobe and thus are both illuminated at the same time. For the purpose of estimating angular resolution, the mainlobe width is typically taken to be the 3-dB beamwidth θ_3 of the antenna. Thus, the two point scatterers in Fig. 1.16 located at the 3-dB edges of the beam define the angular resolution of the radar. The figure illustrates the relation between the angular resolution in radians and the equivalent resolution in units of distance, which will be called the *cross-range resolution* to denote resolution in a dimension orthogonal to range. The arc length at a radius R for an angle subtending θ_3 radians is exactly $R\theta_3$. The cross-range resolution ΔCR is the distance between two scatterers located at the 3-dB edges of the beam, corresponding to the dashed line in Fig. 1.16, and is given by

$$\Delta CR = 2R\sin\left(\frac{\theta_3}{2}\right) \approx R\theta_3 \tag{1.26}$$

where the approximation holds when the 3-dB beamwidth is small, which is usually the case for pencil beam antennas. This result is applicable in either the azimuth or elevation dimension.

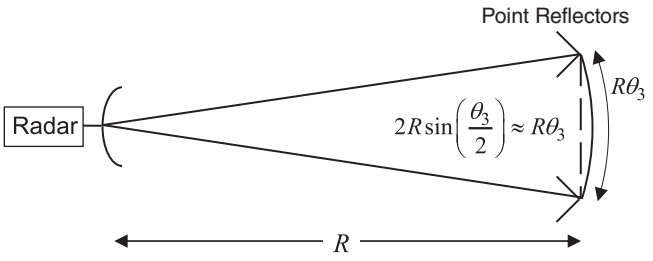


Figure 1.16 The angular resolution is determined by the 3-dB antenna beamwidth θ_3.

Three details bear mentioning. First, the literature frequently fails to specify whether one- or two-way 3-dB beamwidth is required or given. The two-way beamwidth should be used for monostatic radar. Second, note that cross-range resolution increases linearly with range, whereas range resolution was a constant. Finally, as with range resolution, it will be seen later (Chap. 8) that signal processing techniques can be used to improve resolution far beyond the conventional $R\theta$ limit and to make it independent of range as well.

The radar resolution cell volume V is approximately the product of the total solid angle subtended by the 3-dB antenna mainlobe, converted to units of area, and the range resolution. For an antenna having an elliptical beam with azimuth and elevation beamwidths θ_3 and ϕ_3, this is

$$\begin{aligned} \Delta V &= \pi\left(\frac{R\theta_3}{2}\right)\left(\frac{R\phi_3}{2}\right)\Delta R = \frac{\pi}{4}R^2\theta_3\phi_3\Delta R \\ &\approx R^2\theta_3\phi_3\Delta R \end{aligned} \tag{1.27}$$

The approximation in the second line of Eq. (1.27) is 27 percent larger than the expression in the first line, but is widely used. Note that resolution cell volume increases with the square of range because of the two-dimensional spreading of the beam at longer ranges.

1.4.3 Data Integration and Phase History Modeling

A fundamental operation in radar signal processing is *integration* of samples to improve the SIR. Both *coherent integration* and *noncoherent integration* are of interest. The former refers to integration of complex (i.e., magnitude and phase) data, while the latter refers to integration based only on the magnitude (or possibly the squared or log magnitude) of the data.

Suppose a pulse is transmitted, reflects off a target, and at the appropriate time the receiver output signal is measured, consisting of a complex echo amplitude $Ae^{j\phi}$ corrupted by additive noise w. The noise is assumed to be a sample of a random process with power σ_w^2. The single-pulse SNR is

$$\chi_1 = \frac{\text{signal power}}{\text{noise power}} = \frac{A^2}{\sigma_w^2} \tag{1.28}$$

Now suppose the measurement is repeated $N - 1$ more times. One expects to measure the same deterministic echo response, but with an independent noise sample each time. Form a single measurement z by integrating (summing) the individual measurements; this sum of complex samples, retaining the phase information, is a coherent integration:

$$\begin{aligned} z &= \sum_{n=0}^{N-1}\{Ae^{j\phi} + w[n]\} \\ &= NAe^{j\phi} + \sum_{n=0}^{N-1} w[n] \end{aligned} \tag{1.29}$$

The power in the integrated signal component is N^2A^2. Provided the noise samples $w[n]$ are independent of one another and zero mean, the power in the noise component is the sum of the power in the individual noise samples. Further assuming each has the same power σ_w^2, the total noise power is now $N\sigma_w^2$. The integrated SNR becomes

$$\chi_N = \frac{N^2A^2}{N\sigma_w^2} = N\left(\frac{A^2}{\sigma_w^2}\right) = N\chi_1 \tag{1.30}$$

Coherently integrating N measurements has improved the SNR by a factor of N; this increase is called the *integration gain*. Later chapters show that, as one would expect,

increasing the SNR improves detection and parameter estimation performance. The cost is the extra time, energy, and computation required to collect and combine the N pulses of data.

In coherent integration, the signal components added in phase, i.e., coherently. This is often described as adding on a *voltage* basis, since the amplitude of the integrated signal component increased by a factor of N, with the result that signal power increased by N^2. The noise samples, whose phases varied randomly, added on a *power* basis. It is the alignment of the signal component phases that allowed the signal power to grow faster than the noise power.

Sometimes the data must be preprocessed to ensure that the signal component phases align so that a coherent integration gain can be achieved. If the target had been moving in the previous example, the signal component of the measurements would have exhibited a Doppler shift, and Eq. (1.29) would instead become

$$z = \sum_{n=0}^{N-1} \{Ae^{j(2\pi f_D n+\phi)} + w[n]\} \tag{1.31}$$

for some value of normalized Doppler frequency f_D. The signal power in this case will depend on the particular Doppler shift, but except in very fortunate cases will be less than $A^2 N^2$. However, if the Doppler shift is known in advance, the phase progression of the signal component can be compensated before summing:

$$\begin{aligned} z' &= \sum_{n=0}^{N-1} e^{-j2\pi f_D n} \{Ae^{j(2\pi f_D n+\phi)} + w[n]\} \\ &= NAe^{j\phi} + \sum_{n=0}^{N-1} e^{-j2\pi f_D n} w[n] \end{aligned} \tag{1.32}$$

The phase correction aligns the signal component phases so that they add coherently. The noise phases are still random with respect to one another. Thus, the integrated signal power is again N^2A^2, while the integrated noise power is again $N\sigma_w^2$, and therefore an integration gain of N is again achieved. Compensation for the phase progression so that the compensated samples add in phase is an example of phase *history modeling*: if the sample-to-sample pattern of target echo phases can be predicted or estimated (at least to within a constant overall phase), the data can be modified with a countervailing phase so that the full coherent integration gain is achieved. Phase history modeling is central to many radar signal processing functions and is essential for achieving adequate gains in SNR.

In noncoherent integration, the phases are discarded and some function of the magnitudes of the measured data samples are added, such as the magnitude, magnitude-squared, or log-magnitude. If the magnitude-squared is chosen, then z is formed as

$$\begin{aligned} z &= \sum_{n=0}^{N-1} \left| Ae^{j\phi} + w[n] \right|^2 \\ &= \sum_{n=0}^{N-1} \left| Ae^{j\phi} \right|^2 + \sum_{n=0}^{N-1} \left| w[n] \right|^2 + \sum_{n=0}^{N-1} 2\,\mathrm{Re}\{Ae^{j\phi} \cdot w^*[n]\} \\ &= NA^2 + \sum_{n=0}^{N-1} \left| w[n] \right|^2 + \sum_{n=0}^{N-1} 2\,\mathrm{Re}\{Ae^{j\phi} \cdot w^*[n]\} \end{aligned} \tag{1.33}$$

The important fact is that phase information in the received signal samples is discarded.

The first line of Eq. (1.33) defines noncoherent square-law integration. The next two lines show that, because of the nonlinear magnitude-squared operation, z cannot be expressed as the sum of a signal-only part and a noise-only part due to the presence of the third term involving cross-products between signal and noise components. A similar situation exists if the magnitude or log-magnitude is chosen for the noncoherent integration. Consequently, a noncoherent integration gain cannot be simply defined as it was for the coherent case.

It is possible to define a noncoherent gain implicitly. For example, in Chap. 6 it will be seen that detection of a constant-amplitude target signal in complex Gaussian noise with a probability of detection of 0.9 and a probability of false alarm of 10^{-8} requires a single-sample SNR of 14.2 dB (about 26.3 on a linear scale). The same probabilities can be obtained by integrating the magnitude of 10 samples each having an individual SNR of only 5.8 dB (3.8 on a linear scale). The reduction of 8.4 dB (a factor of 26.3/3.8 = 6.9) in the required single-sample SNR when 10 samples are noncoherently integrated is the implied noncoherent integration gain.

Noncoherent integration is much more difficult to analyze than coherent integration, typically requiring derivation of the probability density functions of the noise-only and signal-plus-noise cases in order to determine the effect on detection and parameter estimation.

Chapter 6 will show that in many useful cases, the noncoherent integration gain is approximately N^{α}, where α ranges from about 0.7 or 0.8 for small N to about 0.5 ($\sqrt{N}$) for large N, rather than in direct proportion to N. Thus, noncoherent integration is less efficient than coherent integration. This should not be surprising, since not all of the signal information is used.

1.4.4 Bandwidth Expansion

The scaling property of Fourier transforms states that if $x(t)$ has Fourier transform $X(\Omega) = \mathbf{F}\{x(t)\}$, then

$$\mathbf{F}\{x(\alpha t)\} = \frac{1}{|\alpha|} X\left(\frac{\Omega}{\alpha}\right) \tag{1.34}$$

Equation (1.34) states that if the signal x is compressed in the time domain by the factor $\alpha > 1$, its Fourier transform is stretched (and scaled) in the frequency domain by the same factor (Papoulis, 1987). When $\alpha < 1$, Eq. (1.34) shows that stretching in the time domain results in compression in the frequency domain. This *reciprocal spreading* behavior is illustrated in Fig. 1.17. Part (*a*) shows a sinusoidal pulse with a frequency of 10 MHz and a duration of 1 μs and its Fourier transform, which is a sinc function centered on 10 MHz and with a Rayleigh mainlobe width of 1 MHz, the reciprocal of the 1 μs pulse duration. In part (*b*) the pulse has the same frequency but only one-quarter the duration. Its spectrum is still a sinc centered at 10 MHz, but the Rayleigh width is now four times larger at 4 MHz. The spectrum amplitude is also reduced by a factor of four. This effect can also be viewed in the opposite direction: if the signal gets wider in the frequency domain, it must get narrower in the time domain.

Combining the reciprocal spreading property of Fourier transforms with the observation that resolution depends on signal width shows that improving resolution requires increasing "bandwidth" in the opposite Fourier domain. For example, improving range resolution for simple pulses requires using shorter pulses, as was seen in Sec. 1.4.2; but Fig. 1.17 shows that a shorter pulse implies a wider spectrum, i.e., more bandwidth. Conversely, it was also shown in Sec. 1.4.2 that improving resolution in the frequency domain requires a narrower spectrum mainlobe and thus according to Fig. 1.17, a longer observation (more "bandwidth") in the time domain. This behavior holds for any two functions related by a Fourier transform: finer resolution in one domain requires wider support in the opposite domain.

Radar designers have developed techniques for increasing the appropriate bandwidth to obtain improved resolution in various dimensions. For example, improving resolution in range requires increasing waveform bandwidth, which has led to the use of wideband

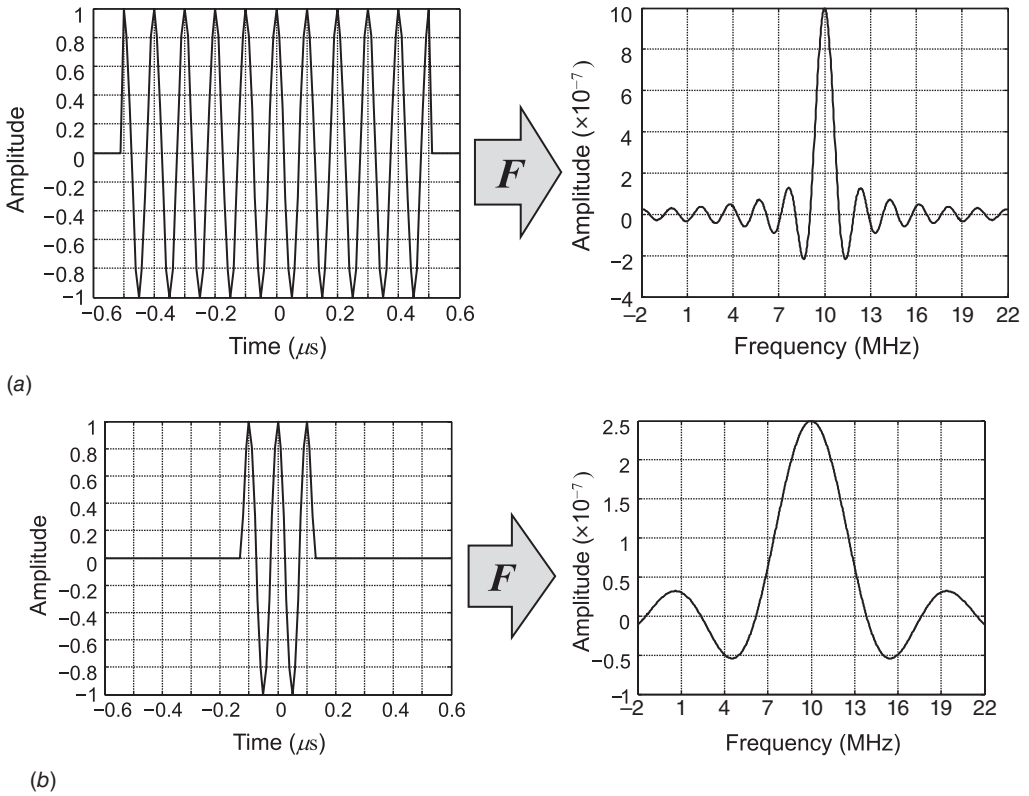


Figure 1.17 Illustration of reciprocal spreading property of Fourier transforms. (*a*) A sinusoidal pulse and the main portion of its Fourier transform. (*b*) A narrower pulse has a wider transform. See text for details.

phase- and frequency-modulated waveforms in place of the simple pulse (Chap. 4). Improving cross-range resolution requires viewing a scene over a wide angular interval to increase cross-range spatial frequency bandwidth, and leads to the synthetic aperture techniques of Chap. 8. Improving velocity (equivalently, Doppler) resolution requires a long time observation and is accomplished with multipulse waveforms. Because the antenna far-field pattern is the Fourier transform of the aperture current distribution, improved angular resolution can be obtained with larger apertures, i.e., bigger antennas.

1.5 A Preview of Basic Radar Signal Processing

There are a number of instances where the design of a component early in the radar signal processing chain is driven by properties of some later component. For example, in Chap. 4 it will be seen that the matched filter maximizes SNR; but it is not until the performance curves for the detectors that follow the matched filter are derived that it will be seen that maximizing SNR also optimizes detection performance. Until the detector is considered, it is hard to see exactly how performance depends on SNR. Having seen the major components of a typical pulsed coherent radar system, the most common signal processing operations in the radar signal processing chain are now described heuristically. By sketching out this preview of the "big picture" from beginning to end, it may be easier to understand the motivation for and interrelation of many of the processing operations to be described in later chapters.

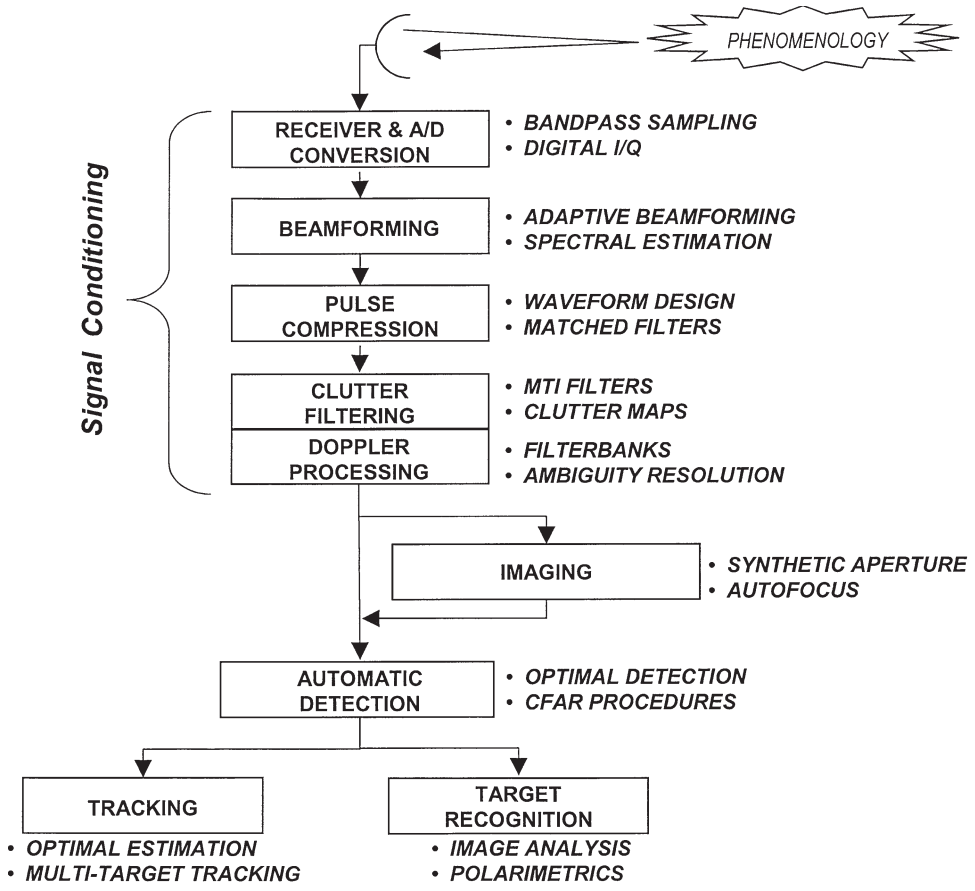


Figure 1.18 One example of a generic radar signal processor flow of operations.

Figure 1.18 illustrates one possible sequence of operations in a generic radar signal processor. The sequence shown is not unique, nor is the set of operations exhaustive. In addition, the point in the chain at which the signal is digitized varies in different systems; it might occur as late as the output of the clutter filtering step. The operations can be generally grouped into *signal conditioning and interference suppression; imaging; detection*; and *postprocessing*. Radar signal *phenomenology* must also be considered. In the next few subsections the basic purpose and operation of each block in this signal processing chain is described.

1.5.1 Radar Time Scales

Radar signal processing operations take place on time scales ranging from less than a nanosecond to tens of seconds or longer, a range of 10 to 12 orders of magnitude. Different classes or levels of operations tend to operate on significantly different time scales. Figure 1.19 illustrates one possible association of operations and time scale.

Operations that are applied to data from a single pulse occur on the shortest time scale, often referred to as *fast time* because the sample rate, determined by the instantaneous pulse bandwidth (see Chap. 2), is on the order of hundreds of kilohertz (kHz) to as much as a few gigahertz in some cases. Corresponding sampling intervals range from a few microseconds

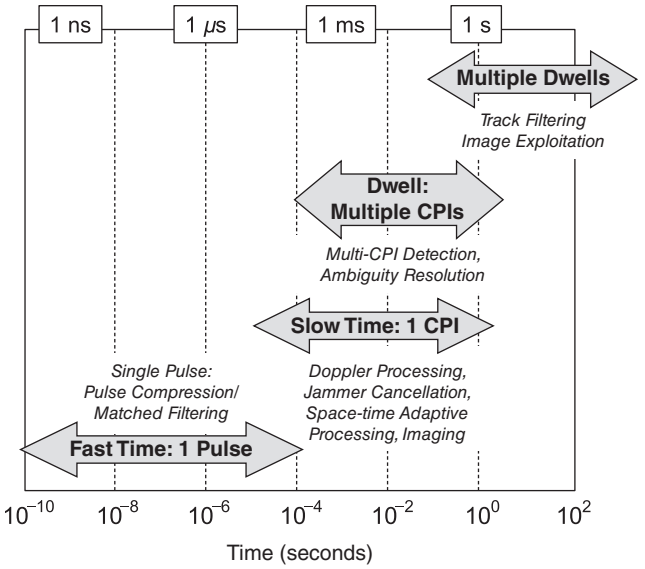


Figure 1.19 Illustration of the range of time scales over which radar signal processing is performed.

down to a fraction of a nanosecond, and signal processing operations on these samples therefore tend to act over similar time intervals. Typical fast time operations are digital I/Q signal formation, beamforming, pulse compression or matched filtering, and sensitivity time control.

The next level up in signal processing operations operates on data from multiple pulses. The sampling interval between pulses (the PRI) is typically on the order of tens of microseconds to hundreds of milliseconds, so again operations that involve multiple pulses occupy similar time scales. Due to the much slower sampling rate compared to single-pulse operations, such operations are said to act in *slow time*. Typical operations include coherent and noncoherent integration, Doppler processing of all types, synthetic aperture imaging, and space-time adaptive processing. The idea of slow and fast time will be revisited in the discussion of the data organizational concept of the datacube in Chap. 3.

A group of pulses that are to be somehow combined coherently, for example via Doppler processing or *synthetic aperture radar* (SAR) imaging, are said to form a *coherent processing interval* (CPI). A still higher level of radar processing acts on data from multiple CPIs and therefore operates on an even longer time scale often called a *dwell* and typically lasting milliseconds to ones or tens of seconds. Operations on this scale include multiple-CPI detection and ambiguity resolution techniques, multilook SAR imaging, and track filtering. Some radars may track detected targets for many seconds or minutes using data from multiple dwells. Track filtering operates in this regime. Finally, some imaging radars may monitor an area over days, months, or even years.

1.5.2 Phenomenology

To design a successful signal processor, the nature of the signals to be processed must be understood. *Phenomenology* refers to the characteristics of the signals received by the radar. Relevant characteristics include signal power, frequency, phase, polarization, or angle of arrival; variation in time and spatial location; and randomness. The received

signal phenomenology is determined by both intrinsic features of the physical object(s) giving rise to the radar echo, such as their physical size or their orientation and velocity relative to the radar; and the characteristics of the radar itself such as its transmitted waveform, polarization, or antenna gain. For example, if more power is transmitted a more powerful received echo is expected, all other things being equal.

In Chap. 2, models of the behavior of typical measured signals that are relevant to the design of signal processors are developed. The radar range equation will give a means of predicting signal power. The Doppler phenomenon will predict received frequency. It will be seen that the complexity of the real world gives rise to very complex variations in radar signals; this will lead to the use of random processes to model the signals, and to particular probability density functions that match measured behavior well. A (very) brief overview of the behavior of the variation of ground and sea echo with sensing geometry and radar characteristics will be given. It will also be shown that measured signals can be represented as the convolution of the "true" signal representing the ideal measurement with the radar waveform (in the range dimension) or its antenna pattern (in the azimuth or elevation dimension, both also called cross-range dimension). Thus, a combination of random process and linear systems theory will be used to describe radar signals and to design and analyze radar signal processors.

1.5.3 Signal Conditioning and Interference Suppression

The first several blocks after the antenna in Fig. 1.18 can be considered as signal conditioning operations whose purpose is to improve the SIR of the data prior to detection, parameter measurement, or imaging operations. That is, the intent of these blocks is to "clean up" the radar data as much as possible. This is done in general with a combination of fixed and adaptive *beamforming*, *pulse compression*, *clutter filtering*, and *Doppler processing*.

Beamforming is applicable when the radar antenna is an array, i.e., when there are multiple phase center signals, or *channels*, available to the signal processor. Fixed beamforming is the process of combining the outputs of the various available phase centers to form a directive gain pattern, similar to that shown in Fig. 1.6. The high-gain mainlobe and low sidelobes selectively enhance the echo strength from scatterers in the antenna look direction while suppressing the echoes from scatterers in other directions, typically clutter. The sidelobes also provide a measure of suppression of jamming signals so long as the angle of arrival of the jammer is not within the mainlobe of the antenna. By proper choice of the weights used to combine the channels, the mainlobe of the beam can be steered to various look directions, and the tradeoff between the sidelobe level and the mainlobe width (angular resolution) can be varied.

Adaptive beamforming takes this idea a step further. By examining the correlation properties of the received data across channels, it is possible to recognize the presence of jamming and clutter entering the antenna pattern sidelobes and design a set of weights for combining the channels such that the antenna not only has a high-gain mainlobe and generally low sidelobes, but also has a null in the antenna pattern at the angle of arrival of the jammer. Much greater jammer suppression can be obtained in this way. Similarly, it is also possible to increase clutter suppression by this technique. *Space-time adaptive filtering* (STAP) combines adaptive beamforming in both angle and Doppler for simultaneous suppression of clutter and jammer interference. Figure 1.20 illustrates interference suppression using STAP, allowing a previously invisible target signal to be seen and perhaps detected. The two vertical bands in Fig. 1.20*a* represent jammer energy, which comes from a fixed angle of arrival but is usually in the form of relatively wideband noise; thus it is present at all Doppler frequencies observed by the radar. The diagonal band in Fig. 1.20*a* is due to ground clutter, for which the Doppler shift depends on the angle from the radar to the ground patch contributing energy. Figure 1.20*b* shows that the adaptive filtering has created nulls along the loci of the jammer and clutter energy, making the target at 0° angle of arrival and

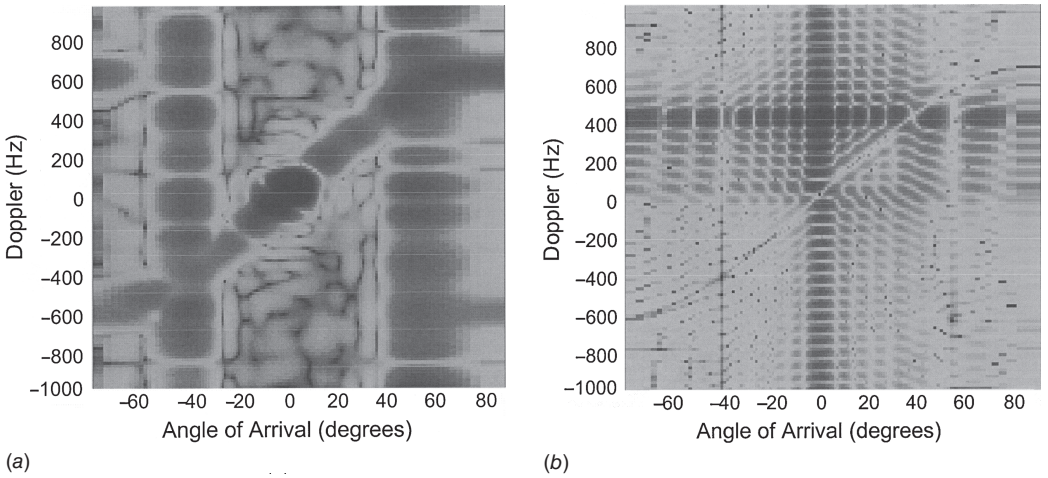


Figure 1.20 Example of effect of adaptive beamforming. (*a*) Map of received signal power as a function of angle of arrival and signal Doppler shift. (*b*) Angle-Doppler map after adaptive processing. (*Images courtesy of Dr. W. L. Melvin. Used with permission*.)

400 Hz Doppler shift apparent. Adaptive interference suppression will be introduced in Chap. 9.

Pulse compression is a special case of *matched filtering*. Many radar system designs strive for both high sensitivity in detecting targets and fine range resolution (the ability to distinguish closely spaced targets). Upcoming chapters show that target detectability improves as the transmitted energy increases, and that range resolution improves as the transmitted waveform's instantaneous bandwidth increases. If the radar employs a simple, constant-frequency rectangular envelope pulse as its transmitted waveform the pulse must be lengthened to increase the transmitted energy for a given power level. However, lengthening the pulse also decreases its instantaneous bandwidth, degrading the range resolution. Thus sensitivity and range resolution appear to be in conflict with one another.

Pulse compression provides a way out of this dilemma by decoupling the waveform bandwidth from its duration, thus allowing both to be independently specified. This is done by abandoning the constant-frequency pulse and instead designing a modulated waveform. A very common choice is the linear frequency modulated (linear FM, LFM, or "chirp") waveform, shown in Fig. 1.21*a*. The instantaneous frequency of an LFM pulse is swept over the desired bandwidth during the pulse duration; the frequency may be swept either up or down, but the rate of frequency change is constant.

The matched filter is by definition a filter in the radar receiver designed to maximize the SNR at its output. The impulse response of the filter having this property turns out to be a replica of the transmitted waveform's modulation function that has been reversed in time and conjugated; thus the impulse response is "matched" to the particular transmitted waveform modulation. Pulse compression is the process of designing a waveform and its corresponding matched filter so that the matched filter output in response to the echo from a single point scatterer concentrates most of its energy in a very short duration, thus providing good range resolution while still allowing the high transmitted energy of a long pulse. Figure 1.21*b* shows the output of the matched filter corresponding to the LFM pulse of Fig. 1.21*a*; note that the mainlobe of the response is much narrower than the duration of the

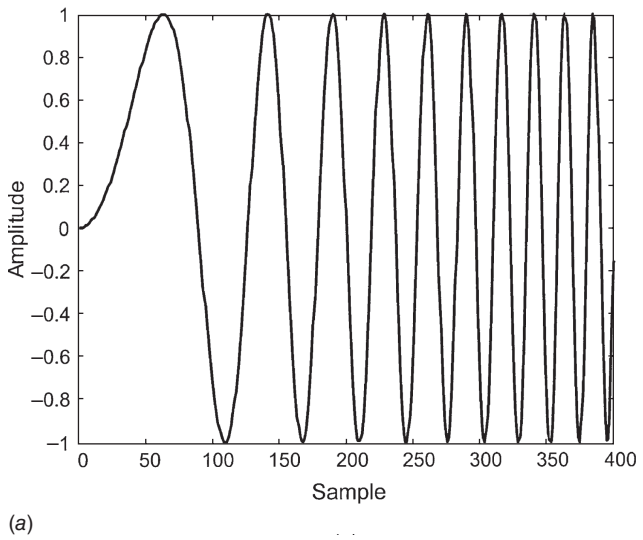


Figure 1.21 (*a*) Linear FM waveform modulation function, showing an increasing instantaneous frequency. (*b*) Output of the matched filter for the LFM waveform of (*a*).

original pulse. The concepts of matched filtering, pulse compression, and waveform design, as well as the properties of linear FM and other common waveforms, are described in Chap. 4. There it is seen that the 3-dB width of the mainlobe in time is approximately $1/\beta$ seconds, where β is the instantaneous bandwidth of the waveform used. This width determines the ability of the waveform to resolve targets in range. Converted to equivalent range units, the range resolution is given by

$$\Delta R = \frac{c}{2\beta} \tag{1.35}$$

[This is the same as Eq. (1.2) presented earlier.]

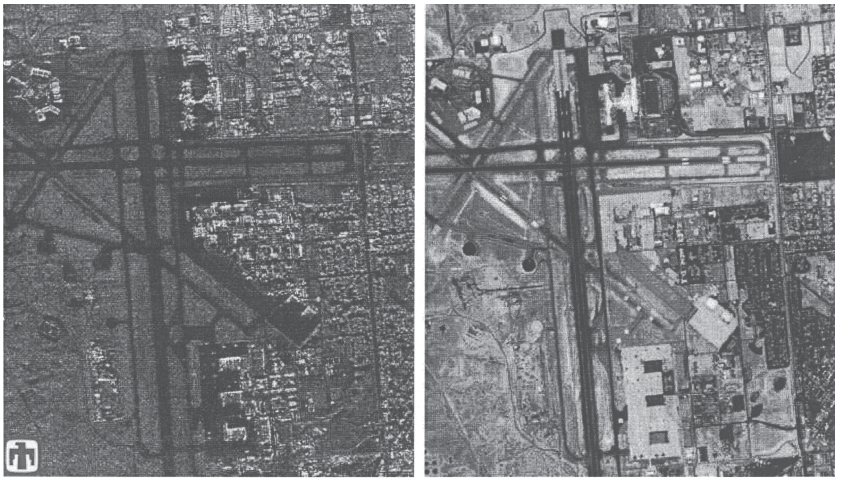
Clutter filtering and Doppler processing are closely related. Both are techniques for improving the detectability of moving targets by suppressing interference from clutter echoes, usually from the terrain in the antenna field of view, based on differences in the Doppler shift of the echoes from the clutter and from the targets of interest. The techniques differ primarily in whether they are implemented in the time or frequency domain and in historical usage of the terminology.

Clutter filtering usually takes the form of *moving target indication*, or MTI, which is simply pulse-to-pulse highpass filtering of the radar echoes at a given range to suppress constant components, which are assumed to be due to nonmoving clutter. Extremely simple, very low-order (most commonly first- or second-order) digital filters are applied in the time domain to samples taken at a fixed range but on successive transmitted pulses.

The term "Doppler processing" generally implies the use of the fast Fourier transform algorithm, or occasionally some other spectral estimation technique, to explicitly compute the spectrum of the echo data for a fixed range across multiple pulses. Due to their different Doppler shifts, energy from moving targets is concentrated in different parts of the spectrum from the clutter energy, allowing detection and separation of the targets. Doppler processing obtains more information from the radar signals, such as number and approximate velocity of moving targets, than does MTI filtering. The cost is more required radar pulses, thus consuming energy and timeline, and greater processing complexity. Many systems use both techniques in series. Clutter filtering and Doppler processing are the subjects of Chap. 5.

1.5.4 Imaging

Most people are familiar with the idea of a radar producing "blips" on a screen to represent targets, and in fact systems designed to detect and track moving targets may do exactly that. However, radars can also be designed to compute fine-resolution images of a scene. Figure 1.22 compares the quality routinely obtainable in SAR imagery in the mid-1990s to that of an aerial photograph of the same scene; close examination reveals many similarities and many significant



(a) (b)

Figure 1.22 Comparison of optical and SAR images of the Albuquerque airport. (*a*) K_u band (15 GHz) SAR image, 3-m resolution. (*b*) Aerial photograph. (Images courtesy of Sandia National Laboratories. Used with permission.)

differences in the appearance of the scene at radar and visible wavelengths. Not surprisingly, the photograph is easier for a human to interpret and analyze, since the imaging wavelengths (visible light) and phenomenology are the same as the human visual system. In contrast, the radar image, while remarkable, is monochromatic, offers less detail, and exhibits a "speckled" texture, some seemingly unnatural contrast reversals, and some missing features such as the runway stripes. Given these drawbacks, why is radar imaging of interest?

While radars do not obtain the resolution or image quality of photographic systems, they have two powerful advantages. First, they can image a scene through clouds and inclement weather due to the superior propagation of RF wavelengths. Second, they can image equally well 24 hours a day since they do not rely on the sun for illumination; they provide their own "light" via the transmitted pulse. If the example of Fig. 1.21 were repeated in the middle of a rainy night, the SAR image on the left would not be affected in any noticeable way, but the optical image on the right would disappear entirely.

To obtain fine-resolution imagery, radars use a combination of high-bandwidth waveforms to obtain good resolution in the range dimension and the synthetic aperture radar technique to obtain good resolution in the cross-range dimension. The desired range resolution is obtained while maintaining adequate signal energy by using pulse compression waveforms, usually linear FM. A long pulse that is swept over a large enough bandwidth β and processed using a matched filter can provide very good range resolution according to Eq. (1.35). For example, range resolution of 1 m can be obtained with a waveform swept over 150 MHz. Depending on their applications, modern imaging radars usually have range resolution of 30 m or better; many systems have 10 m or better resolution, and some advanced systems have resolution under 1 m.

For a conventional nonimaging radar, referred to as a *real aperture* radar, the resolution in cross-range is determined by the width of the antenna beam at the range of interest and is given by $R\theta_3$ as shown in Eq. (1.26). Realistic antenna beamwidths for narrow-beam antennas are typically 1° to 3°, or about 17 to 52 mrad. Even at a relatively short imaging range of 10 km, the cross-range resolution that results would be 170 to 520 m, much worse than typical range resolutions and too coarse to produce useful imagery. This poor cross-range resolution is overcome by using SAR techniques.

The synthetic aperture technique refers to the concept of synthesizing the effect of a very large antenna by having the actual physical radar antenna move in relation to the area being imaged. Thus, SAR is most commonly associated with moving airborne or space-based radars, rather than with fixed ground-based radars. Figure 1.23 illustrates the concept. By transmitting pulses at each indicated location, collecting the range data for each pulse, and properly processing it together, a SAR system creates the effect of a much larger phased array antenna being flown along the aircraft flight path. As suggested by Eq. (1.9) (though some details differ in the SAR case), a very large aperture size produces a very narrowly focused effective antenna beam, thus making possible very fine cross-range resolution. The SAR concept is explained more fully in Chap. 8. A more modern and robust viewpoint based on integrating over a range of angles is also given there.

1.5.5 Detection

The most basic function of a radar signal processor is detection of the presence of one or more targets of interest. Information about the presence of targets is contained in the echoes of the radar pulses. These echoes compete with receiver noise, undesired echoes from clutter signals, and possibly intentional or unintentional jamming. The signal processor must somehow analyze the total received signal and determine whether it contains a desirable target echo and, if so, at what range, angle, and velocity.

Because the complexity of radar signals leads to the use of statistical models, detection of target echoes in the presence of competing interference signals is a problem in statistical

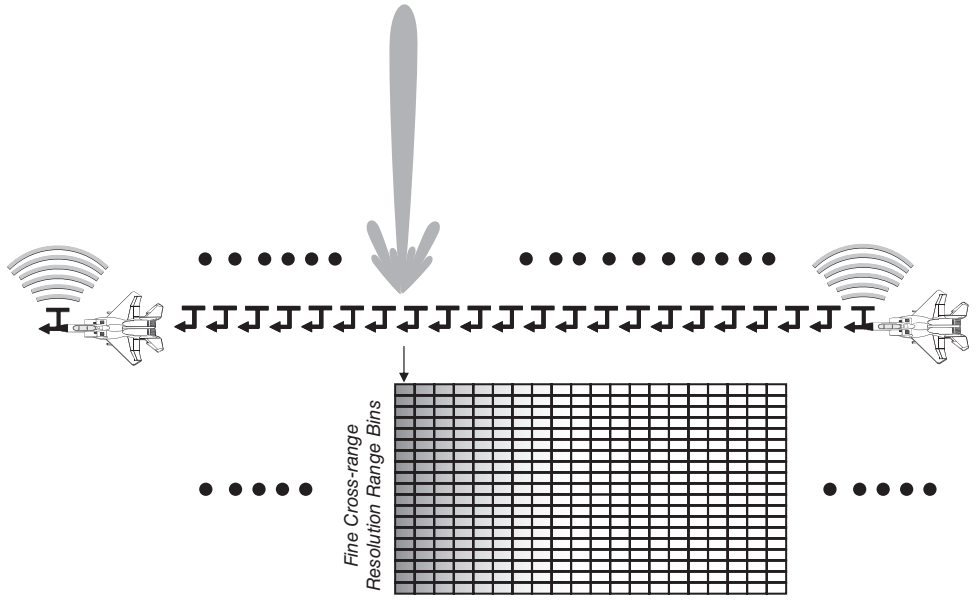


Figure 1.23 The concept of synthetic aperture radar.

decision theory. The theory as applied to radar detection will be developed in Chap. 6. There it will be seen that in most cases optimal performance can be obtained using *threshold detection*. In this method, the magnitude of each complex sample of the radar echo signal, possibly after signal conditioning and interference suppression, is compared to a precomputed threshold. If the signal amplitude is below the threshold, it is assumed to be due to interference signals only. If it is above the threshold, it is assumed that the stronger signal is due to the presence of a target echo in addition to the interference, and a detection or "hit" is declared. In essence, the detector makes a decision as to whether the energy in each received signal sample is too large to likely have resulted from interference alone; if so, it is assumed a target echo contributed to that sample. Figure 1.24 illustrates the concept. The "clutter + target" signal might represent the variation in received signal strength versus range (fast time) for a single transmitted pulse. It crosses the threshold at three different times, suggesting the presence of three targets at different ranges.

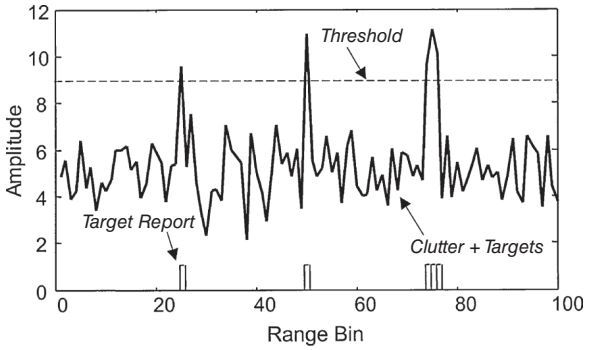


Figure 1.24 Illustration of threshold detection.

Because they are the result of a statistical process, threshold detection decisions have a finite probability of being wrong. For example, a noise spike could cross the threshold, leading to a false target declaration, commonly called a *false alarm*. These errors are minimized if the target spikes stand out strongly from the background interference, i.e., if the SIR is as large as possible. If this is the case the threshold can be set relatively high, resulting in few false alarms while still detecting most targets. This fact also accounts for the importance of matched filtering in radar systems. The matched filter maximizes the SIR, thus providing the best threshold detection performance. Furthermore, the achievable SIR increases monotonically with the transmitted pulse energy E, thus encouraging the use of longer pulses to get more energy on the target. Since longer simple pulses reduce range resolution, the technique of pulse compression is also important so that fine resolution can be obtained while maintaining good detection performance.

The concept of threshold detection can be applied to many different radar signal processing systems. Figure 1.24 illustrates its application to a fast-time (range) signal trace, but it can be equally well applied to a signal composed of measurements at different Doppler frequencies for a fixed range, or in a two-dimensional form to combined range-Doppler data or to SAR imagery.

There are numerous significant details in implementing threshold detection. Various detector designs work on the magnitude, squared-magnitude, or even log-magnitude of the complex signal samples. The threshold is computed from knowledge of the interference statistics so as to limit false alarms to an acceptable rate. However, in real systems the interference statistics are rarely known accurately enough to allow for precomputing a fixed threshold. Instead, the required threshold is set using interference statistics estimated from the data itself, a process called *constant-false-alarm rate* (CFAR) detection. Detection processing is described in detail in Chap. 6.

1.5.6 Measurements and Track Filtering

Radar systems employ a wide variety of processing operations after the point of target detection. One of the most common postdetection processing steps, and one of the three major functions of interest in this text, is *tracking* of targets, an essential component of many radar systems. Tracking is comprised of (usually multiple) *measurements* of the position of detected targets followed by *track filtering*.

The radar signal processor detects the presence of targets using threshold detection methods. The range, angle, and Doppler resolution cell in which a target is detected provide a coarse estimate of its location in those coordinates. Once detected, the radar will seek to refine the estimated range by using signal processing methods to more precisely estimate the time delay after pulse transmission at which the threshold crossing occurred, the angle of the target relative to the antenna mainbeam direction, or its radial velocity. Individual measurements will have some error due to interference, and so provide a noisy snapshot of the target location and motion at one instant in time.

The term track filtering describes a higher-level, longer time scale process of integrating a series of such measurements to estimate a complete trajectory of the target over time. It is often described as *data processing* rather than signal processing. Because there may be multiple targets with crossing or closely spaced trajectories, track filtering must deal with the problems of determining which measurements to associate with which targets being tracked, and with correctly resolving nearby and crossing trajectories. A variety of optimal estimation techniques have been developed to perform track filtering. An excellent reference in this area is Bar-Shalom (1988).

Figure 1.25 illustrates a series of noisy measurements in one dimension of the position of two targets and the filtering of that noisy trajectory using an extremely simple alpha-beta filter, to be discussed in Chap. 9. The position in the x dimension versus time for each target is shown by the gray lines, so the two targets are moving at different velocities along the x axis and one passes the other at around time step 32. The circle and diamond markers

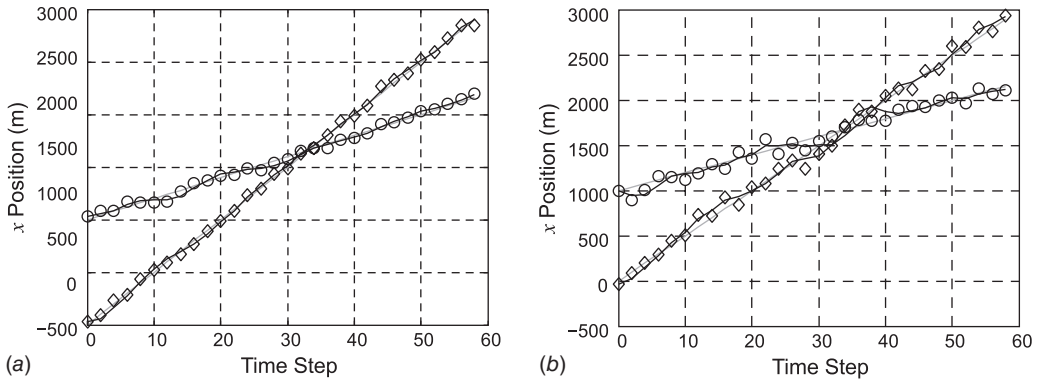


Figure 1.25 Track filtering of noisy measurements for two targets in one dimension using an alpha-beta filter. Markers show individual measurements. Gray lines are actual position, black lines are filtered position estimates. (*a*) Low measurement noise. (*b*) Tracks incorrectly switch targets in higher measurement noise.

indicate the noisy radar measurements of position for each. The solid black lines are the smoothed estimates of position produced by the alpha-beta filter. In part (*a*) of the figure, the filter correctly associates the measurements with each target when they cross, so that each smoothed estimate follows the same target over the observation time. In Fig. 1.25*b*, the noise variance is higher, causing the filter to incorrectly swap the tracks around time step 40. This represents an error in measurement-to-track *data association*. A variety of techniques are available to attempt to address this problem; a few are discussed in Chap. 9.

1.6 Radar Literature

This text covers a middle ground in radar technology. It focuses on basic radar signal processing from a digital signal processing point of view. It does not address radar systems, components, or phenomenology in any great depth except where needed to explain the signal processing aspects; nor does it provide in-depth coverage of advanced radar signal processing specialties. Fortunately, there are many excellent radar reference books that address both needs. Good books appear every year; those listed in the paragraphs that follow are current as of the year 2013.

1.6.1 Radar Systems and Components

Probably the most classic introductory text to radar systems, now in its third edition, is by Skolnik (2001). The newest and one of the best "radar 101" introductions is the new text by Richards et al. (2010), the first of a three-volume series. The 1990s saw the introduction of several general radar system textbooks. The text by Edde (1995) also has an associated self-study course. Peebles (1998) provides a recent, comprehensive introduction, while Mahafza (2000) provides a number of useful MATLAB® files to aid in simulation and experimentation. Morris and Harkness (1996) provides a good introduction to airborne pulsed Doppler systems specifically. A newer discussion of pulsed Doppler systems is given by Alabaster (2012). An up-to-date survey of a broad range of traditional and modern radar applications is given in Scheer and Melvin (2014), showing how many of the techniques discussed in both these introductory texts and the more specialized ones discussed below are brought together into complete systems.

1.6.2 Basic Radar Signal Processing

It is this author's opinion that there are a number of excellent books about radar systems in general, including coverage of components and system designs, and several on advanced radar signal processing topics, especially in the area of synthetic aperture imaging. There have been few books that address the middle ground of basic radar signal processing, such as pulse compression, Doppler filtering, and CFAR detection. Such books are needed to provide greater quantitative depth than is available in the radar systems books without restricting themselves to in-depth coverage of a single advanced application area, and this text aims to fill that gap. Nonetheless, there are a few texts that fit somewhat into this middle area. Nathanson (1991) wrote a classic book, now in its second edition, that covers radar systems in general but in fact concentrates on signal processing issues, especially RCS and clutter modeling, waveforms, MTI, and detection. Probably the closest text in intent to this one is by Levanon (1988), which provides excellent analyses of many basic signal processing functions. The new text by Levanon and Mozeson (2004) addresses the widening variety of radar waveforms in detail. A recent text by Sullivan (2000) is interesting especially for its introductory coverage of both SAR and space-time adaptive processing (STAP), thus providing a bridge between basic signal processing and more advanced texts specializing in SAR and STAP.

1.6.3 Advanced Radar Signal Processing

Two very active areas of advanced radar signal processing research are SAR imaging and STAP. SAR research extends back to 1951, but only in the 1990s did open literature textbooks begin to appear in the market. There are now many good textbooks on SAR. The first comprehensive text was by Curlander and McDonough (1991). Based on experience gained at the NASA Jet Propulsion Laboratory, it emphasizes space-based SAR and includes a strong component of scattering theory as well. Cumming and Wong (2005) is a newer text that also emphasizes spaced-based SAR. The spotlight SAR mode received considerable development in the 1990s, and two major groups published competing texts in the mid-1990s. Carrara, Goodman, and Majewski (1995) represented the work of the group at the Environmental Research Institute of Michigan (ERIM, now a part of General Dynamics, Inc.); Jakowatz, Jr., et al. (1996) represented the work of a group at Sandia National Laboratories, a unit of the U.S. Department of Energy. Franceschetti and Lanari (1999) provide a compact, unified treatment of both major modes of SAR imaging, namely stripmap and spotlight. The book by Soumekh (1999) is the most complete academic reference on synthetic aperture imaging and includes a number of MATLAB® simulation resources.

STAP, one of the most active radar signal processing research areas, began in earnest in 1973 and is correspondingly less mature than SAR processing. Klemm (1998) wrote the first significant open literature text on the subject. Just as with the Curlander and McDonough book in the SAR community, this book was the first sign that a series of STAP texts can be expected as that research topic matures and reaches mainstream use. The book by Guerci (2003) is the newest primer on this subject at this writing, while Van Trees (2002) prepared a detailed text that continues his classic series on detection and estimation. Additionally, there are other texts on more limited forms of adaptive interference rejection. A good example is the one by Nitzberg (1999), which discusses several forms of sidelobe cancellers. An excellent new book covering a wide range of advanced radar signal processing techniques, including such new topics as multi-input, multi-output radar, and compressive sensing, is the companion volume to (Richards et al., 2010) by Melvin and Scheer (2013).

1.6.4 Radar Applications

The preceding sections have cited a number of books addressing general radar applications, such as imaging or pulse Doppler. There are a number of books in the literature devoted to

more specific application areas. The forthcoming text by Melvin and Scheer (2014) will provide an excellent survey of and introduction to a wide range of applications in a single text, and will complete the *Principles of Modern Radar* series.

1.6.5 Current Radar Research

Current radar research appears in a number of scientific and technical journals. The most important in the United States are the Institute of Electrical and Electronics Engineers (IEEE) *Transactions on Aerospace and Electronic Systems, Transactions on Geoscience and Remote Sensing*, *Transactions on Signal Processing,* and *Transactions on Image Processing*. Radar-related material in the latter is generally limited to papers related to SAR processing, especially interferometric three-dimensional SAR. In the United Kingdom, radar technology papers are often published in the Institution of Engineering and Technology (IET) [formerly the Institution of Electrical Engineers (IEE)] journal *IET Radar, Sonar, and Navigation.*

References

Alabaster, C. M., *Pulse Doppler Radar: Principles, Technology, Applications*. SciTech Publishing, Raleigh, NC, 2012.

Balanis, C. A., *Antenna Theory,* 3d ed., Harper & Row, New York, 2005.

Bar-Shalom, Y., and T. E. Fortmann, *Tracking and Data Association*. Academic Press, Boston, MA, 1988.

Bracewell, R. N., *The Fourier Transform and Its Applications*, 3d ed., McGraw-Hill, New York, 1999.

Brookner, E. (ed.), *Aspects of Modern Radar*. Artech House, Boston, MA, 1988.

Carrara, W. G., R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar*. Artech House, Norwood, MA, 1995.

Cumming, I. G., and F. N. Wong, *Digital Processing of Synthetic Aperture Radar Data*. Artech House, Norwood, MA, 2005.

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

Edde, B., *Radar: Principles, Technology, Applications*. Prentice Hall PTR, Upper Saddle River, NJ, 1995.

EW and Radar Systems Engineering Handbook, Naval Air Warfare Center, Weapons Division. Available at http://ewhdbks.mugu.navy.mil/.

Franceschetti, G., and R. Lanari, *Synthetic Aperture Radar Processing*. CRC Press, New York, 1999.

Guerci, J. R., *Space-Time Adaptive Processing for Radar*. Artech House, Norwood, MA, 2003.

Institute of Electrical and Electronics Engineers, "IEEE Standard Letter Designations for Radar-Frequency Bands," Standard 521-1976, Nov. 30, 1976.

Institute of Electrical and Electronics Engineers, "IEEE Standard Radar Definitions," Standard 686-1982, Nov. 30, 1982.

Jakowatz, C. V., Jr., et al., *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Kluwer, Boston, MA, 1996.

Jelalian, A. V., *Laser Radar Systems*. Artech House, Boston, MA, 1992.

Johnson, D. H., and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, Englewood Cliffs, NJ, 1993.

Klemm, R., *Space-Time Adaptive Processing: Principles and Applications*. INSPEC/IEEE, London, 1998.

Levanon, N., *Radar Principles*. J. Wiley, New York, 1988.

Levanon, N., and E. Mozeson, *Radar Signals*. J. Wiley, New York, 2004.

Mahafza, B. R., *Radar Systems Analysis and Design Using MATLAB®*. Chapman & Hall/CRC, New York, 2000.

Melvin, W. L., and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Raleigh, NC, 2013.

Mir, H. S., and J. D. Wilkinson, "Radar Target Resolution Probability in a Noise-Limited Environment," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44(3), pp. 1234–1239, July 2008.

Morris, G. V., and L. Harkness (eds.), *Airborne Pulsed Doppler Radar*, 2d ed. Artech House, Boston, MA, 1996.

Nathanson, F. E., (with J. P. Reilly and M. N. Cohen), *Radar Design Principles*, 2d edition. McGraw-Hill, New York, 1991.

Nitzberg, R., *Radar Signal Processing and Adaptive Systems*, 2d ed. Artech House, Boston, MA, 1999.

Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing*, 2d ed. Prentice Hall, Englewood Cliffs, NJ, 1999.

Papoulis, A., *The Fourier Integral and Its Applications*, 2d ed. McGraw-Hill, New York, 1987.

Peebles, Jr., P. Z., *Radar Principles*. Wiley, New York, 1998.

Richards, M. A., J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.

Scheer, J. A. and W. L. Melvin (eds.), *Principles of Modern Radar: Radar Applications*. SciTech Publishing, Edison, NJ, to appear 2014.

Sherman, S. M., *Monopulse Principles and Techniques*. Artech House, Boston, MA, 1984.

Skolnik, M. I., *Introduction to Radar Systems*, 3d ed. McGraw-Hill, New York, 2001.

Soumekh, M., *Synthetic Aperture Radar Signal Processing with MATLAB Algorithms*. J. Wiley, New York, 1999.

Stutzman, W. L., "Estimating Gain and Directivity of Antennas," IEEE *Transactions on Antennas and Propagation*, vol. 40(4), pp. 7–11, Aug. 1998.

Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*. J. Wiley, New York, 1998.

Sullivan, R. J., *Microwave Radar: Imaging and Advanced Concepts*. Artech House, Boston, MA, 2000.

Swords, S. S., *Technical History of the Beginnings of RADAR*. Peter Peregrinus Ltd., London, 1986.

Van Trees, H. L., *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. J. Wiley, New York, 2002.

Problems

1. Compute the range R corresponding to echo delays t_0 of 1 ns, 1 μs, 1 ms, and 1 second.
2. Compute the time delays for two-way propagation to targets at distances of 100 km, 100 statute miles, and 100 ft.
3. Radar is routinely used as one means of measuring the distance to objects in space. For example, it has been used to calculate the orbital parameters and rate of rotation of the planet Jupiter. The distance from Earth to Jupiter varies from 588.5×10^6 to 968.1×10^6 km. What are the minimum and maximum time delays *in minutes* from the time a pulse is transmitted in the direction of Jupiter until the time the echo is received? If pulses are transmitted at a rate of 100 pulses per second, how many pulses are in flight, either on their way to Jupiter or back again, at any given instant?
4. Table 1.1 defines the millimeter wave (MMW) band to extend from 40 to 300 GHz. Only certain frequencies in this band are widely used for radar. This is partly due to frequency allocation rules (which frequencies are allotted to which services), but also due to atmospheric propagation. Based on Fig. 1.3, list two frequencies in the

MMW band that might be preferable for radar use, and two that would not be suitable. Explain.

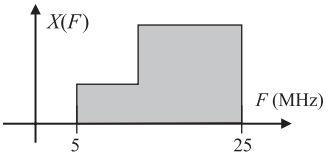
5. Compute the bandwidth β needed to achieve range resolutions of 1 m, 1 km, and 100 km. What is the length of a rectangular pulse having this Rayleigh bandwidth (peak-to-first null width of the Fourier transform) for each value of resolution?
6. In terms of D_y and λ, what is the peak-to-first null beamwidth (called *Rayleigh beamwidth*) in radians of the antenna pattern for an aperture antenna with constant illumination? Give both the general result, and a small-angle approximation.
7. How large must a uniformly illuminated aperture antenna be (value of D_y) in terms of wavelengths so that its 3-dB beamwidth is 1°? What is the estimated gain in decibels of an antenna having azimuth and elevation beamwidths $\theta_3 = \phi_3 = 1°$, based on the approximation in Eq. (1.10)?
8. Suppose a police "speed gun" radar has a rectangular antenna. It is desired to have a cross-range resolution ΔCR of 10 ft at a distance of one-quarter mile. What is the required antenna width in inches if the radar frequency is 9.4 GHz? Repeat for 34.4 GHz.
9. Continuing problem 8, what is the actual cross-range resolution in feet at each RF if the antenna width is 6 in.?
10. Starting from Eq. (1.13) and setting $a_n = 1$, derive Eq. (1.14).
11. What is the maximum 3-dB beamwidth θ_3 in degrees such that the approximation for the cross-range resolution, $R\theta_3$, in the last step of Eq. (1.26) has an error of no more than 1 percent?
12. Determine the cross-range resolution ΔCR in meters at ranges of 10, 100, and 1000 km for a 3-dB beamwidth $\theta_3 = 3°$.
13. Determine the approximate size of a volume resolution cell in cubic meters, ΔV, for $R = 20$ km, $\Delta R = 100$ m, and $\theta_3 = \phi_3 = 3°$.
14. Suppose Eq. (1.31) is modified to consider the magnitude-squared of the signal-plus-noise data:

$$z = \sum_{n=0}^{N-1} \left| Ae^{j(2\pi f_D n + \phi)} + w[n] \right|^2$$

Show explicitly that z cannot be expressed as the sum of a signal-only and a noise-only term.

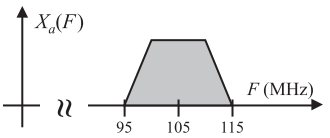
The remaining problems relate to topics covered in Appendix B.

15. What is the Nyquist sampling rate (minimum rate to avoid aliasing) for the signal $x(t)$ having the spectrum $X(F)$ shown in the figure below? Sketch a block diagram of a system for recovering a new signal $\hat{x}(t)$ from samples of $x(t)$ taken at the Nyquist rate, such that the spectrum $\hat{X}(F)$ has the same shape as $X(F)$, but is centered at $F = 0$.



16. In some cases, the spectrum replication property of sampling can be used as a substitute for demodulation. Given a signal $x_a(t)$ with the spectrum shown, what is

the lowest sampling rate that will ensure both no aliasing of the spectrum, and that one of the spectrum replicas is centered at $F = 0$?



17. How many bits are required in an A/D converter to provide a dynamic range of at least 40 dB? What is the expected SQNR with this number of bits, assuming $k = 3$?
18. Numerical values of spatial frequency differ greatly from the usual temporal frequency values. What is the spatial frequency in cycles per meter for a 1-GHz electromagnetic wave?
19. What is the instantaneous frequency in hertz of the waveform $x(t) = \exp[j \cdot \exp(-\alpha t)]$?
20. Determine a phase function $\psi(t)$ such that the instantaneous frequency in hertz of the waveform $x(t) = \cos[\psi(t)]$, $-\tau/2 \le t \le \tau/2$ seconds, sweeps linearly from $-\beta/2$ Hz to $+\beta/2$ Hz.
21. Suppose signal x_1 is 30 dB greater in power than signal x_2. What is the ratio of their power in linear units (i.e., not in dB)? What is the ratio of the corresponding voltages?