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 by radar (albeit accidentally), 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, Italy, Japan, and Russia. 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 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 or numerous online sources. More sophisticated meteorological radar 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 as one means for determining altitude and avoiding severe weather, and may soon use it to assist in imaging runway approaches in poor weather. Radar is commonly used by the shipping, heavy equipment, and automotive industries for collision avoidance, obstacle detection, and related safety functions. Indeed, one of the most important drivers in radar technology in the 2010s is the automotive industry, which now places millions of small radars on the road each year in driver assistance systems. Radar is also an essential sensor for emerging autonomous driving systems. 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 on fundamental range, Doppler, and angle processing techniques common to most radar systems. Previous editions emphasized pulsed over continuous wave radars. However, recent years have seen extensive proliferation of continuous wave radars in the automotive and other industries. Particularly common are linear frequency-modulated CW (linear FMCW) radars, which use many of the same signal processing and data organization techniques as pulsed systems. Consequently, this edition addresses both pulsed and linear FMCW systems, using the common data acquisition and organization construct of a *datacube* to unify their descriptions.

Similarly, because most radars are *monostatic*, meaning the transmitter and receiver antennas are collocated (and in fact are usually the same antenna), they are emphasized over *bistatic* radars where the antennas are significantly separated, though again many of the results apply to both. 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*. Higher-level capabilities are built on top of these basic functions. This text addresses all three and the techniques of signal acquisition and interference reduction necessary to perform them.

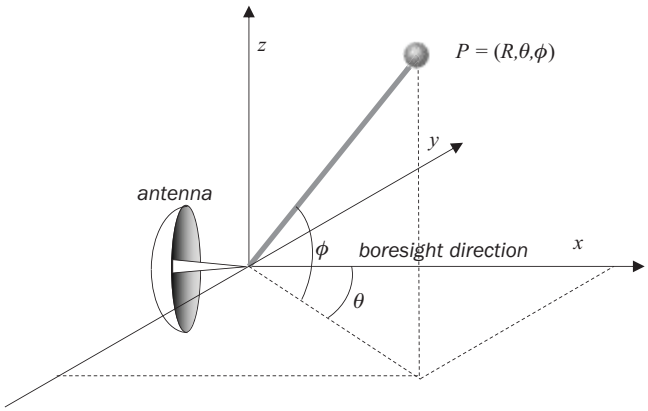


Figure 1.1 Spherical coordinate system for radar measurements.

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 includes the echo from a reflecting object or only noise or other interference. 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. The time required for a signal to propagate a distance R and return, thus traveling a total distance $2R$, is $2R/c$, where c is the speed of electromagnetic (EM) wave propagation ("speed of light").[1] Therefore, if $a(t) > T(t)$ at some time delay t_0 after a signal is transmitted, it is assumed that a target is present at range

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

Once an object has been detected, it may be desirable to track its location and velocity. A monostatic radar naturally measures position in a spherical coordinate system with its origin at the radar antenna's phase center (defined in Sec. 1.3.4), 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 the *azimuth* angle, while ϕ is called the *elevation* angle.[2]

The range R to the object is obtained 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 field of view 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

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

[2] In mathematics, the spherical coordinate system is often defined in terms of range, azimuth angle, and a *polar angle* ϕ_p (also called the *zenith* or *inclination* angle) measured from the z axis. The polar and elevation angles are related as $\phi = \pi/2 - \phi_p$ radians.

associated with radar. However, radars are increasingly used to generate two- and three-dimensional images of an area. Such images can be analyzed for intelligence and surveillance purposes, for topology mapping, or for "earth resources" applications such as analysis of land use, ice cover, deforestation, pollution spills, 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. In addition, radar imagery works day or night because the transmitter provides the "illumination." Consequently, imaging radars generate useful imagery when optical instruments cannot be used at all.

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 figures of merit are *accuracy* (bias) and *precision* (standard deviation) of the range, angle, and velocity estimates. With appropriate signal processing the achievable accuracy is typically limited by a combination of resolution and SIR. For example, if noise is the primary interference source, the limiting precision often is proportional to $\sqrt{\Delta/SNR}$, where Δ is the resolution in the coordinate of interest and SNR is the value of the *signal-to-noise ratio* (SNR).

In imaging, the principal figures of merit are spatial resolution and dynamic range. Spatial resolution determines what size objects can be distinguished in the final image and therefore to what uses the image can be put. For example, a radar map with 1 by 1 km resolution would be useful for large-scale land use studies but useless for detailed 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 extract end products such as detections or images from the raw radar data and to maximize the quality of those products by maximizing resolvability, SIR, and other relevant figures of merit. SIR can be improved by integration of multiple measurements. Resolution and SIR can be jointly improved by matched filters and other waveform design and processing 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 is explored 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 image formation. Modern model-based spectral estimation and adaptive filtering techniques are used for beamforming and jammer cancellation.

Pattern recognition and, more recently, machine learning techniques are used for target/clutter[3] discrimination and target identification.

At the same time, radar signal processing has several unique qualities that differentiate it from many other signal processing fields. Most modern radars are coherent, resulting in a received signal that, 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. In addition, received signals are many decibels weaker than transmitted signals due to propagation and other losses, so that SIR ratios at the receiver are often relatively low. For example, successful detection typically requires an SIR at the point of detection of 10 to 20 dB, but the SIR of the received signal at the antenna will be much less than 0 dB. Large signal processing gains are needed to overcome this deficit.

Another very important distinguishing feature of radar signal processing is the large signal bandwidths compared to most other DSP applications. Instantaneous bandwidths for an individual pulse or CW transmission are frequently on the order of a few megahertz. In some fine resolution[4] radars, they may reach several hundred megahertz and even low gigahertz levels. The difficulty of designing good converters at multi-megahertz or gigahertz sample rates has historically slowed the introduction of digital techniques into radar signal processing for two primary reasons. First, very fast *analog-to-digital* (A/D) converters are required to digitize the high-bandwidth data. Even now that digital techniques are standard in new designs, A/D converter effective word lengths in high-bandwidth systems (high tens of megahertz to ones of gigahertz) are usually a relatively short 8 to 12 bits, rather than the 16 bits common in many other areas.

Second, the high data rates require high-speed computing capability to implement the algorithms. Historically, this meant that it was often necessary to design custom hardware for the digital processor in order to obtain adequate throughput, that is, to "keep up with" the torrent of data. It also meant that radar signal processing algorithms had to be relatively simple compared to lower-bandwidth applications such as sonar in order to minimize the processing load. Only in the late 1990s and later have improved analog semiconductor technology and Moore's law[5] provided enough computing power to host radar algorithms for a wide range of high-performance systems on commercial hardware. This technological progress has enabled rapid development and introduction of new, more complex algorithms to radar signal processing, enabling major improvements in detection, tracking, and imaging capability.

1.3 Elements of a Radar

Figure 1.2 is one possible block diagram of a basic monostatic radar. The waveform generator output is the desired pulse or CW waveform to be transmitted. The transmitter comprises a series of *mixers* and *local oscillators* (LOs) to modulate this waveform to a desired *intermediate*

[3]"Clutter" is an interference signal consisting of unwanted echoes of the radar's transmitted signal from objects not of interest.

[4]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 to reduce confusion.

[5]Gordon Moore's famous 1965 prediction was that the number of transistors on an integrated circuit would double every 18 to 24 months. That prediction held remarkably true for at least 40 years, enabling the computing and networking revolutions that began in earnest in the 1980s. Whether it is being maintained, or can be, in the 2020s and later is a perennial subject of debate.

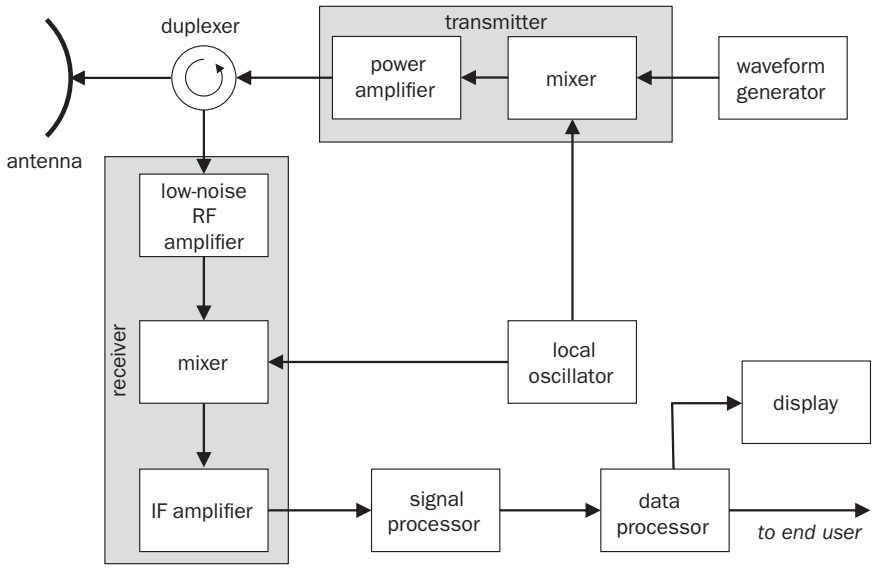


Figure 1.2 Block diagram of a monostatic radar.

frequency (IF) and then to the desired *radio frequency* or *radar frequency* (RF), followed by amplifiers to boost the signal power a useful level. The transmitter output is routed to the antenna through a *duplexer*, *circulator,* or *T/R switch* (for transmit/receive).

The antenna focuses and concentrates the transmitted radiation into a narrow region of space in a particular direction, providing a power gain to the transmitted signal in that direction. Similarly, it selectively provides greater sensitivity on receive to echoes from that same direction. In doing so, the antenna both amplifies weak target echo signals and localizes them in azimuth and elevation.

The returning echoes are routed by the duplexer into the radar receiver. The receiver is usually a superheterodyne design (Bruder, 2010), and often the first stage is a low-noise RF amplifier. This is followed by one or more stages of demodulation of the received signal to successively lower *IFs* and ultimately to *baseband*, where the signal is not modulated onto any carrier frequency.

The baseband signal is next sent to the *signal processor*, which performs some or all of a variety of functions such as matched filtering, Doppler filtering, integration, and motion compensation. The output of the signal processor typically becomes the input to a *data processor*. Typical data processor functions might include target classification, target tracking, or image processing operations, depending on the radar purpose. The data processor output is sent to the system display, passed to other systems, or both as appropriate. The distinction between the signal processor and the data processor is somewhat arbitrary. Generally, the signal processor is associated with "lower" level, higher speed, streaming operations, while the data processor is associated with "higher" level operations that tend to be more data-dependent but require less computation.

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 and some forms of Doppler filtering are common examples. Also, radars differ in which portions of the signal flow are analog and which are digital. Older systems are 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 received signal at an IF or even RF stage, moving the A/D converter closer to the radar front end and performing more of the processing digitally. Waveform generation is also done digitally in many modern systems.

The next few subsections provide some additional detail on these major radar subsystems and also discuss some additional radar systems issues.

1.3.1 Radar Frequencies

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 of 0.3 to 30 μm (Jelalian, 1992). However, most radars operate in the microwave frequency region of about 200 MHz to about 30 GHz, with corresponding wavelengths of 0.67 m to 1 cm. There are also numerous systems in the *millimeter wave* (MMW) region of 30 to 300 GHz (1 cm to 1 mm), especially in the 35 and 95 GHz regions. Table 1.1 summarizes the letter nomenclature used for the common nominal radar bands (IEEE, 2019).

Band	Frequencies	Wavelengths	Common Uses
HF	3–30 MHz	100–10 m	Over-the-horizon surveillance
VHF	30–300 MHz	10–1 m	Long range surveillance, foliage penetration, ground penetrating radar, counter-stealth
UHF	300 MHz–1 GHz	1–30 cm	Long range surveillance, foliage penetration
L	1–2 GHz	30–15 cm	Long range surveillance, long range air traffic control
S	2–4 GHz	15–7.5 cm	Moderate range surveillance, terminal air traffic control, airborne early warning, long range weather observation
C	4–8 GHz	7.5–3.75 cm	Long range tracking, weather observation, weapon location
X	8–12 GHz	3.75–2.5 cm	Short range tracking, missile guidance, marine radar, ground imaging, airborne intercept, weapon location
K_u	12–18 GHz	2.5–1.67 cm	High resolution mapping, satellite altimetry, UAV radar
K	18–27 GHz	1.67–1.11 cm	Police radar, automotive radar
K_a	27–40 GHz	1.11 cm–7.5 mm	Short-range fine resolution imaging, airport surveillance
V	40–75 GHz	7.5–4 mm	Scientific remote sensing
W	75–110 GHz	4–2.73 mm	Automotive radar, missile seekers, very fine resolution imaging
Millimeter wave (includes V and W)	30–300 GHz	1 cm–1 mm	Experimental

Table 1.1 Letter Nomenclature and Common Uses for Nominal Radar Frequency Bands

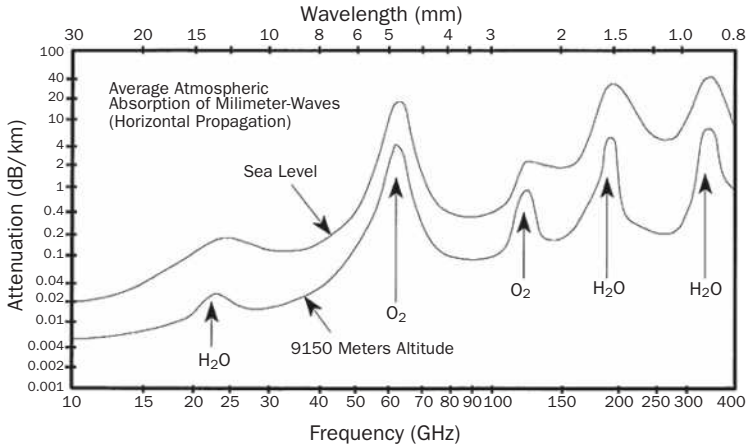


Figure 1.3 One-way atmospheric attenuation of electromagnetic waves. (*Source: EW and Radar Systems Engineering Handbook*, Naval Air Warfare Center, Weapons Division, http://ewhdbks.mugu.navy.mil/.)

Not all frequencies in these bands are suitable for or available to radar operation. Within the HF to K_a bands, specific frequencies are allocated by international agreement to radar operation. At frequencies above X band, atmospheric attenuation of electromagnetic waves becomes significant. Consequently, radars in these bands usually operate at one of several "atmospheric window" frequencies where attenuation is relatively low. Figure 1.3 illustrates the atmospheric attenuation for one-way propagation over the most common radar frequency ranges under approximately "clear air" 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 power available in transmitters at these frequencies. 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 transmitter 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 radar frequency for rain rates ranging from a drizzle to a tropical downpour. X-band frequencies (typically about 10 GHz) and below are affected significantly only by very severe rainfall, while MMW frequencies suffer severe losses for even light-to-medium rain rates.

1.3.2 Radar Waveforms and Transmitters

The radar *waveform* is the term for the signal that is modulated onto the RF carrier for transmission, echo, reception, and demodulation. It plays a major role in determining the sensitivity and range resolution of the radar. There are many different waveforms in common use. They can all be classified as either pulsed or continuous wave. Figure 1.5 illustrates the difference between the two. The CW waveform of Fig. 1.5*a*, as its name suggests, simply radiates a sinusoidal signal at the desired RF continuously. The pulsed waveform in Fig. 1.5*b* transmits a series of finite length pulses. The series is described by the pulse length τ and either the *pulse repetition interval* (PRI) T between pulses or its inverse, the *pulse repetition frequency* (PRF).

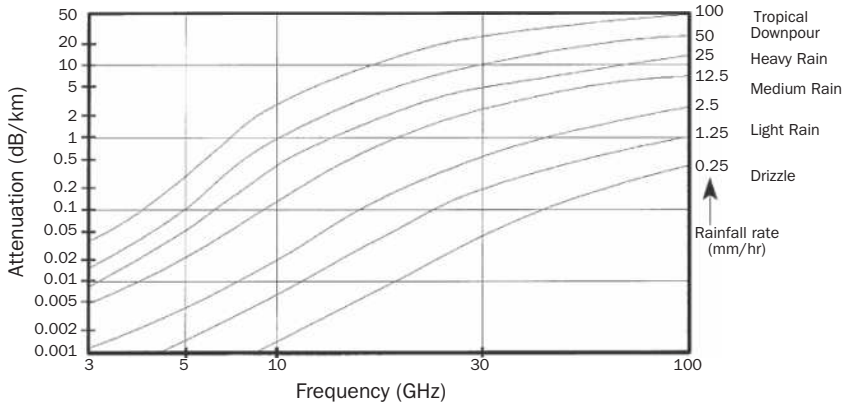


Figure 1.4 Effect of different rates of precipitation on one-way atmospheric attenuation of electromagnetic waves. (*Source: EW and Radar Systems Engineering Handbook*, Naval Air Warfare Center, Weapons Division, http://ewhdbks.mugu.navy.mil/.)

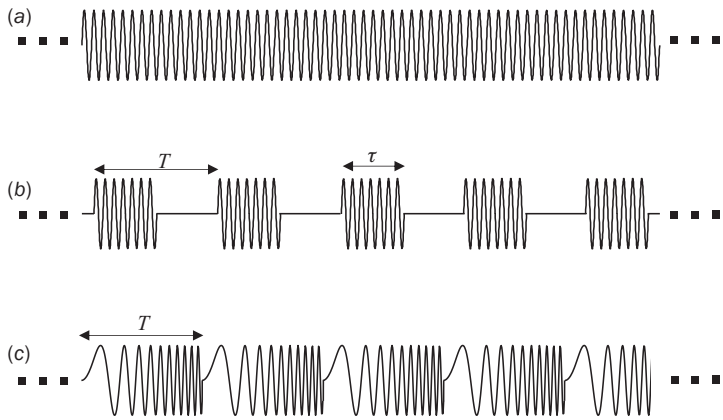


Figure 1.5 Three of the major classes of radar waveforms. (*a*) Continuous wave (CW). (*b*) Pulsed. (*c*) Frequency-modulated continuous wave (FMCW).

The pulsed waveform enables easy measurement of range by observing the time delay between transmission of a pulse and reception of an echo; see Eq. (1.1). Because the CW waveform lacks any distinguishing timing mark such as a pulse edge, another means is needed to measure range. One solution is to modulate the CW waveform with a repeating pattern. Currently the most common example is FMCW. Figure 1.5*c* illustrates an example wherein the frequency of the CW waveform is swept over some bandwidth repeatedly. The duration of one sweep is analogous to the PRI for a pulsed waveform. In the example shown, the sweep rate is constant so that the frequency increases linearly during the sweep interval.

Pulsed waveforms can also exhibit intra-pulse modulation. Both phase- and frequency-modulated pulses are common. Because of their widespread use, in this text the focus will be on pulsed waveforms with phase, frequency, or no modulation, and on linear FMCW waveforms. The details of each are the subject of Chap. 4.

The radar transmitter modulates the waveform to the RF and amplifies it to a useful level. Radar transmitters operate at peak powers ranging from milliwatts to in excess of 10 MW. A wide variety of technologies are used, from solid state sources at lower powers to various vacuum tube devices such as magnetrons and traveling wave tubes at high powers. An excellent survey of transmitter technologies and issues is given in Wallace et al. (2010). High peak power systems are invariably pulsed; CW systems have much lower peak powers. One of the more powerful existing pulsed transmitters is found in the AN/FPS-108 COBRA DANE radar, which has a peak power of 15.4 MW (Brookner, 1988). In pulsed radars the PRF varies widely but is typically between several hundred and several tens of thousands of pulses per second (PPS); in some modern integrated "radar-on-a-chip" systems, the PRF can be several hundred thousand of PPS. The duty cycle of pulsed systems is 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 high 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 or more.

It will be seen in Chap. 6 that the detection performance achievable by a radar improves with the amount of energy in the transmitted waveform. Maximizing energy suggests that a radar waveform should be as long as feasible and be transmitted at maximum power. To satisfy the second condition, radars generally do not use amplitude modulation of the transmitted waveform since that implies that at least a part of the waveform is transmitted at less than full power.

Waveform length is a more complicated issue. It will be seen in Chap. 4 that the nominal range resolution ΔR is determined by the waveform bandwidth β in hertz:

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

For an unmodulated pulse, the bandwidth is inversely proportional to its duration. Fine resolution therefore implies shorter pulses, in conflict with the need for longer pulses to maximize energy. To increase waveform bandwidth for a given pulse length without sacrificing energy, many radars routinely use phase or frequency modulation of the pulse. Similar issues apply to FMCW waveforms. 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. These bandwidths are typically 1 percent or less of the RF. Few radars achieve 10 percent bandwidth, though some achieve bandwidths of 25 percent of the RF or greater, qualifying them as *ultrawideband* (UWB) radars (IEEE, 2017). Nonetheless, most radar waveforms can be considered narrowband, bandpass functions.

1.3.3 Antennas

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

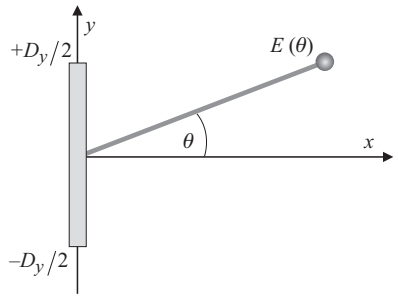


Figure 1.6 Geometry for one-dimensional electric field calculation on a rectangular aperture.

the radiated electric field intensity $E(\theta, \phi)$, known as the antenna one-way *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, 2012):

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

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_t or $2D^2/\lambda_t$ for an antenna of aperture size D. Consider the azimuth (θ) pattern of the one-dimensional linear aperture geometry shown in Fig. 1.6. 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 the inverse Fourier transform[6] of the distribution $A(y)$ of current across the aperture in the azimuth plane (Bracewell, 1999; Skolnik, 2001),

$$E(\theta) = \int_{-D_y/2}^{D_y/2} A(y) \exp[j(2\pi y/\lambda_t)\sin\theta]\, dy \tag{1.5}$$

where the "frequency" variable is the *spatial frequency* or *wavenumber* $(2\pi/\lambda_t)\sin\theta$ and is in units of radians per meter. The idea of spatial frequency is discussed in App. B.

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

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

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

[6]Whether it is the forward or inverse Fourier transform (FT) depends on the FT definition one uses. This text uses the electrical engineering convention, in which the sign of the argument of the exponential in the FT kernel is negative for the forward transform and positive for the inverse.

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

$$A(\lambda_t \zeta) = \int_{-\infty}^{+\infty} E(s)\exp(-j2\pi\zeta s)\,ds \tag{1.7}$$

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, $E(s)$ is taken to be zero outside of this range on s.

Equation (1.7) is a somewhat simplified expression that neglects a range-dependent overall phase factor and a slight amplitude dependence on range (Balanis, 2016). This Fourier transform property of antenna patterns will allow the use of linear system concepts in Chap. 2 to understand the effects of the antenna on cross-range resolution and the angular sampling densities 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 far-field one-way voltage pattern, normalized to its peak, is then the familiar sinc function

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

If the aperture current illumination is separable in the x and y dimensions, the far-field pattern will be the product of two Fourier transforms, one in azimuth (θ) and one in elevation (ϕ).

The magnitude of $E(\theta)$ is illustrated in Fig. 1.7 for the case $D_y = 6\lambda_t$, along with the definitions for three important figures of merit of an antenna pattern. The pattern exhibits the typical structure of a high-gain mainlobe surrounded by low-gain sidelobes. The angular resolution of the antenna is determined primarily by the width of its mainlobe. The mainlobe width is conventionally quantified by the *3-dB beamwidth*, which is the width of the mainlobe

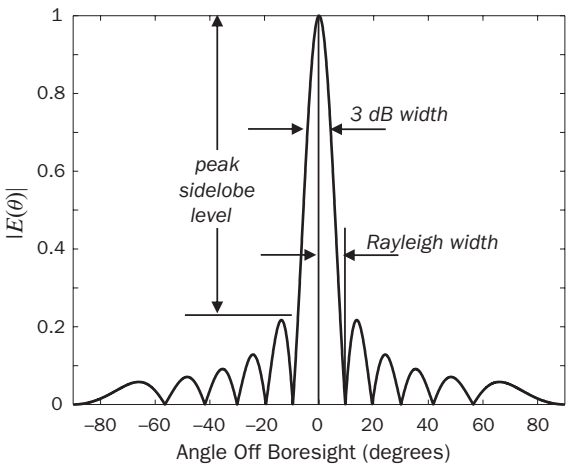


Figure 1.7 One-way radiation pattern of a uniformly illuminated aperture with $D_y = 6\lambda_t$. The 3-dB beamwidth, Rayleigh beamwidth, and peak sidelobe definitions are illustrated.

between the points where $|E(\theta)|^2$ is reduced by 3 dB (approximately one-half) from its peak at $\theta = 0$. This can be determined by setting $E(\theta) = 1/\sqrt{2} \approx 0.707$ and solving for the argument $\alpha = \pi\,(D_y/\lambda_t)\sin\theta$. The answer is found numerically to be $\alpha = 1.4$, which occurs at the angle $\theta_0 = \sin^{-1}(1.4\lambda_t/\pi D_y)$. The 3-dB beamwidth extends from $-\theta_0$ to $+\theta_0$ and is therefore

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

The small-angle approximation used in the last step holds for most radar antenna beamwidths.

Although 3-dB beamwidths are traditional, for analysis the Rayleigh beamwidth is often simpler to compute. This is the one-sided width of the mainlobe from its peak to its first null, and is given by (see Prob. 7)

$$\theta_R \;=\; \sin^{-1}\left(\frac{\lambda_t}{D_y}\right) \approx \frac{\lambda_t}{D_y}\ \text{radians} \tag{1.10}$$

For the constant aperture illumination case, the Rayleigh beamwidth also happens to equal the 4-dB beamwidth. This is not true for antennas in general. Finally, the null-to-null beamwidth θ_{nn} is simply twice the Rayleigh beamwidth, encompassing the entire mainlobe.

Whichever metric is used, 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 large volume search; these designs are called *fan beam antennas*.

The *peak sidelobe* of the pattern affects how echoes from one object affect the detection of neighboring objects. For the uniform illumination pattern, the peak sidelobe is 13.2 dB below the mainlobe peak. This is often considered insufficient in radar systems because strong unwanted signals entering the antenna from the sidelobe directions may not be attenuated enough to enable detection of weaker target echoes in the mainlobe direction. Antenna sidelobes can be reduced by use of a nonuniform aperture distribution (Skolnik, 2001), sometimes referred to as *tapering, shading, or apodization of* 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 spectrum analysis and 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 (see App. B). Lower sidelobes are possible but may be increasingly difficult to achieve due to manufacturing imperfections and inherent design limitations.

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)

$$\begin{aligned} G &\approx \frac{26{,}000}{\theta_3\phi_3} \qquad (\theta_3, \phi_3 \text{ in degrees}) \\ &\approx \frac{7.9}{\theta_3\phi_3} \qquad (\theta_3, \phi_3 \text{ in radians}) \end{aligned} \tag{1.11}$$

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. Suppose a wave with power density W W/m^2 incident on the antenna results in a power P delivered to the antenna load. The effective aperture is defined as the ratio (Balanis, 2016)

$$A_e = \frac{P}{W} \text{ m}^2 \tag{1.12}$$

Note that A_e is not the actual physical area of the antenna. It is the fictional area such that, if all of the power incident on that area was collected and delivered to the load with no loss, it would account for all of the observed power output of the actual 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, 2016)

$$G = \frac{4\pi}{\lambda_t^2} A_e \tag{1.13}$$

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* (ULA) in one dimension, as shown in Fig. 1.8. Figure 1.9 illustrates examples of real array and aperture antennas. Many more examples are given in Chaps. 1 and 9 in Richards et al. (2010).

The voltage pattern for the linear array is most easily arrived at by considering the antenna in its receive mode. Suppose the rightmost element is taken as a reference point, there are N elements in the array, and the elements are isotropic (constant gain for all θ). The signal in branch n is weighted with the complex weight a_n. The incoming electric field voltage $E_0 \exp(j\Omega_t t)$ at the reference element will also appear at each of the other elements, but delayed in time by an additional $d \sin\theta / c$ from each element to the next due to the increasing propagation distance to each. The total output voltage will be the sum of the time-shifted and weighted outputs of all of the array elements. The amplitude of the output voltage is (Skolnik, 2001; Stutzman and Thiele, 2012; or see Sec. 9.2)

$$E(\theta) = E_0 \sum_{n=0}^{N-1} a_n \exp[j(2\pi/\lambda_t) n d \sin\theta] \tag{1.14}$$

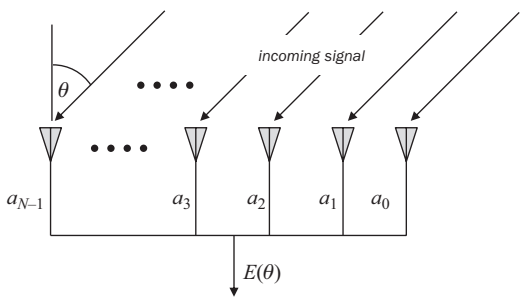
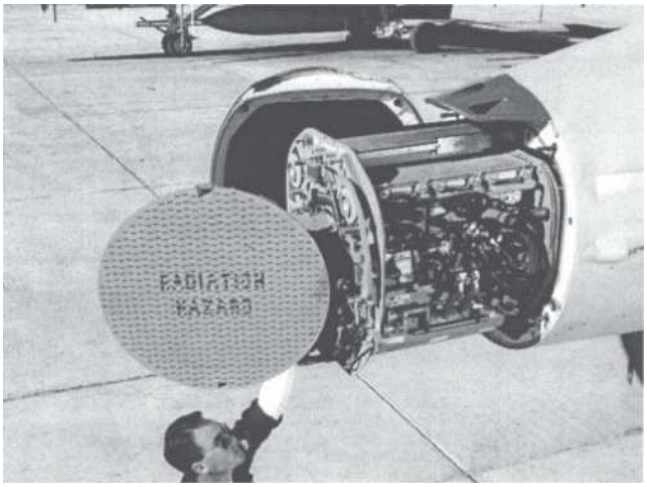
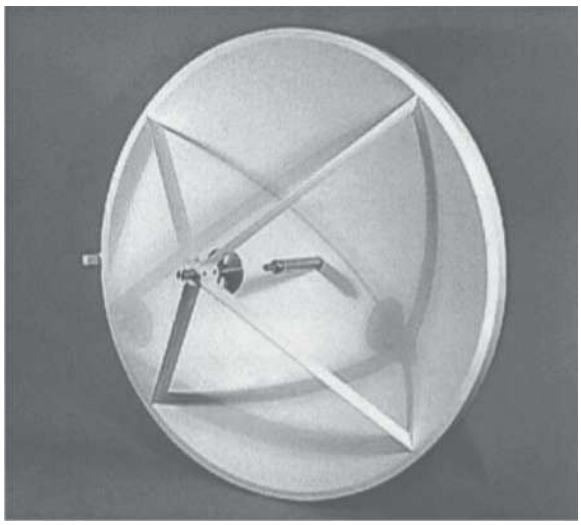


Figure 1.8 Geometry of the uniform linear array antenna.



(*a*)



(*b*)

FIGURE 1.9 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 Technologies. Image (*b*) courtesy of Quinstar Corp. Used with permission.]

This is similar in form to the inverse *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. For the case where all the $a_n = 1$, the pattern is the familiar "aliased sinc" (asinc) function,[7] whose magnitude is

[7]Also called the digital sinc (dsinc) or Dirichlet function. It is the discrete-variable equivalent of the usual continuous-variable sinc function.

$$|E(\theta)| = E_0 \left| \frac{\sin[N(\pi d/\lambda_t)\sin\theta]}{\sin[(\pi d/\lambda_t)\sin\theta]} \right| \tag{1.15}$$

This function is very similar to that of Eq. (1.8) and Fig. 1.7. 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 $0.89\lambda_t/D$, 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. By varying the amplitudes of the weights a_n, it is possible to reduce the sidelobes at the expense of a broader mainlobe and a reduction in output SNR.

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

$$E_{el}(\theta) \approx \cos\theta \tag{1.16}$$

The right-hand side of Eq. (1.15) is then called the *array factor* $AF(\theta)$, and the composite radiation pattern becomes

$$E(\theta) = AF(\theta)E_{el}(\theta) \tag{1.17}$$

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 well 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, 2016). 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 location is the same in both transmit and receive modes.

1.3.4 Virtual Elements and Virtual Arrays

Two more useful antenna concepts are the antenna *phase front* (or *wave front*) and *phase center* (Sherman, 2011; IEEE, 2014; Balanis, 2016; Richards, 2018). 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. 1.6, the phase center would occur at $y = 0$ but possibly $x \neq 0$, depending on the detailed antenna shape.

The phase center idea is especially useful in developing the *virtual array* (VA) concept for analyzing situations where the transmit and receive antennas are not collocated due either to actual physical separation or to platform motion between transmission and reception times. Consider the situation in Fig. 1.10, which shows separate transmit and receive antenna elements at coordinates x_T and x_R and a point scatterer **P** in the far-field. It is straightforward to show that the phase shift of the signal transmitted from x_T and received at x_R (total path length equal to $R_R + R_T$) is, to a good approximation, the same as that of a signal transmitted from and received at the *virtual element* (VE) location **VE** halfway between the two (total path

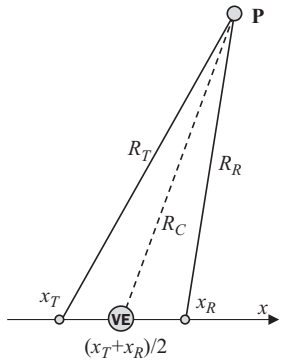


Figure 1.10 Virtual element corresponding to a transmit and receive element pair.

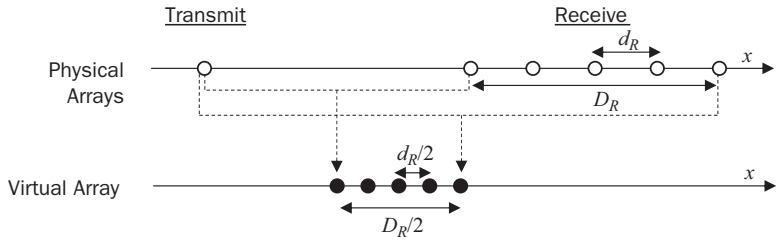


Figure 1.11 Virtual array formed by paring one transmit element with five independent receive elements. (The virtual array has been offset vertically for clarity.)

length $2R_C$) (Richards, 2018). Consequently, for analysis purposes the transmit and receive elements can be replaced by the single VE. This substitution will be useful in the discussion of ground moving target indication in Chap. 5.

Now consider the configuration of Fig. 1.11. The transmit antenna is the single isotropic element on the left. The receive antenna is the five-element uniform linear array of white elements with element spacing d_R and total aperture size D_R on the right. The signal received by the leftmost element of the receive array will be essentially identical to one transmitted and received by a VE located halfway between it and the transmit element. This location is the leftmost of the five black elements forming the VA, as shown by the dotted line. (The VA will actually be located on the same x-axis line as the transmit and receive elements; it is offset vertically in the figure for clarity.) Each of the five transmit-receive element pairings generates a VE. These collectively form the five-element VA shown. Note that the element spacing and overall aperture size are half the physical receive array spacing and size.

The VE and VA concepts can be used to provide a common analysis framework for a variety of problems in antenna design, synthetic aperture analysis (Chap. 8), and emerging techniques such as multi-input, multi-output (MIMO) radar. Much more detail on these concepts is available in Richards (2018) and Richards (2019).

1.3.5 Receivers

It was shown in Sec. 1.3.2 that radar signals are usually narrowband, bandpass (because they are on a carrier frequency), phase- or frequency-modulated functions.

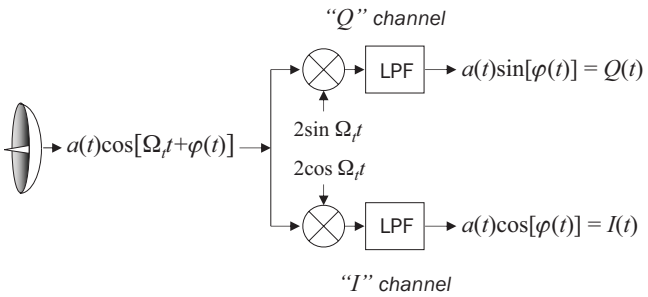


Figure 1.12 Quadrature or "I/Q" channel receiver model.

This means that the echo waveform $\overline{y}(t)$ received from a single scatterer can be modeled in the form

$$\overline{y}(t) = a(t)\cos[\Omega_t t + \varphi(t)] \tag{1.18}$$

Here Ω_t is the nominal RF in radians per second; $\varphi(t)$ is the phase modulation (PM) function, which also includes frequency modulation as a special case; and $a(t)$ is the amplitude modulation (AM). In a CW system, $a(t)$ would be a constant. In the simplest pulsed systems, $a(t)$ would be just a rectangular pulse envelope. In a constant frequency pulsed or CW system (no FM or PM), $\varphi(t)$ would be a constant. The major function of the receiver processing is demodulation of the information-bearing part of the radar signal to baseband with the goal of estimating $a(t)$ and $\varphi(t)$.

Figure 1.12 illustrates the signal processor's simplified view of the receiver structure used in most classical radars. The lower channel mixes the received signal with a *local oscillator* at the radar frequency. The output of the mixer is the product of its two input signals. Applying a trigonometric identity, the mixer is seen to generate both sum and difference frequency components at its output:

$$2\cos(\Omega_t t) \cdot a(t)\cos[\Omega_t t + \varphi(t)] = a(t)\cos(\Omega_t t) + a(t)\cos[2\Omega_t t + \varphi(t)] \tag{1.19}$$

The high-frequency sum frequency term is then removed by the lowpass filter (LPF), leaving only the difference frequency, also called the *beat frequency*. Note that the difference frequency is taken to be the received echo frequency minus the reference oscillator frequency. In this case, the beat frequency is zero so the output is just the modulation term $a(t)\cos[\varphi(t)]$. The upper channel mixes the signal with a *quadrature* oscillator having the same frequency but a 90° phase shift as compared to the lower channel oscillator. The upper channel mixer output is

$$2\sin(\Omega_t t) \cdot a(t)\cos[\Omega_t t + \varphi(t)] = a(t)\sin(\Omega_t t) + a(t)\sin[2\Omega_t t + \varphi(t)] \tag{1.20}$$

which, after filtering, leaves the modulation term $a(t)\sin[\varphi(t)]$.

If the input $\overline{x}(t)$ is written as $a(t)\sin[\Omega_t t + \varphi(t)]$ instead, the lower and upper channel outputs are interchanged. Whichever form is used for the input, the receiver channel having an output proportional to the cosine of the input phase is called the *in-phase* or "I" channel; the other is called the *quadrature* phase or "Q" channel.

Both the I and Q channels are needed to unambiguously determine the amplitude and phase of the echo. Suppose only the I channel is implemented in the receiver, giving the single measured value $a(t)\cos[\varphi(t)]$. There are an infinite number of combinations of a and

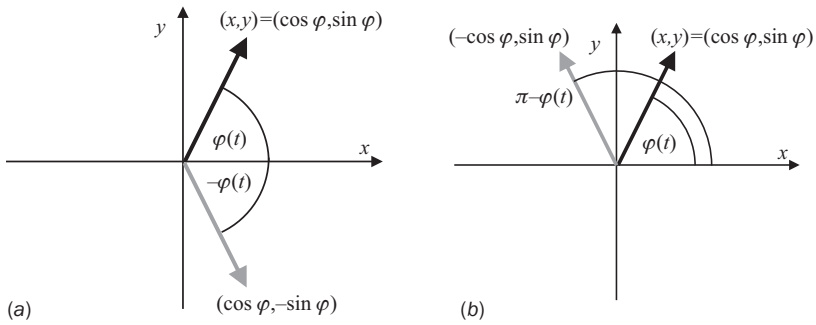


Figure 1.13 (*a*) The I channel of the receiver in Fig. 1.12 measures only the cosine of the phasor $\theta(t)$. (*b*) The Q channel measures only the sine of the phasor.

φ that produce the same product; the I measurement alone is not sufficient to specify both. However, if the Q channel measurement $a(t)\sin[\varphi(t)]$ is also available, the amplitude can be found as $a(t) = \sqrt{I^2(t) + Q^2(t)}$.

Knowing $a(t)$ is still not sufficient to determine $\varphi(t)$ using only one of the I or Q measurements. Figure 1.13 illustrates the problem. In Fig. 1.13*a* the signal phase $\varphi(t)$ is represented as a unit-magnitude black phasor in the complex plane; the amplitude $a(t)$ has been removed. If only the I channel is implemented in the receiver, only the cosine of $\varphi(t)$ will be measured. In this case, the true phasor will be indistinguishable from the gray phasor $-\varphi(t)$. Similarly, if only the Q channel is implemented so that only the sine of $\varphi(t)$ is measured, then the true phasor will be indistinguishable from the gray phasor of Fig. 1.13*b*, which corresponds to $\pi - \varphi(t)$. When both the I and Q channels are implemented, the phasor quadrant is determined unambiguously.[8]

In modern coherent radars, the signal processor will normally assign the I signal to be the real part and the Q signal to be the imaginary part of a new complex signal

$$x(t) = I(t) + jQ(t) = a(t)\exp[j\varphi(t)] \tag{1.21}$$

Equation (1.21) 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 cosine function, an equivalent complex exponential function is used instead.[9] The received echo signal of Eq. (1.18) is thus replaced by

$$\bar{x}(t) = a(t)\exp[\Omega_t t + \varphi(t)] \tag{1.22}$$

The receiver structure of Fig. 1.12 can then be replaced with the simplified model of Fig. 1.14, where the echo is demodulated by multiplication with a complex reference oscillator $\exp(-j\Omega_t t)$.

[8]This is analogous to the use of the two-argument `atan2()` function instead of the single-argument `atan()` function in many programming languages.

[9]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. (1.22) is the *analytic signal* associated with the real-valued signal of Eq. (1.18). The imaginary part of Eq. (1.22) is the *Hilbert transform* of the real part (Papoulis, 1987; Bracewell, 1999).

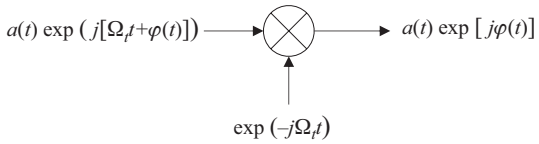


Figure 1.14 Simplified transmission and receiver model using complex exponential signals.

This technique of modeling the transmitted and received real-valued signals by equivalent complex signals using a corresponding complex demodulator produces exactly the same output result obtained in Eq. (1.21) by explicitly modeling the real-valued signals and the I and Q channels, but is much more compact and easier to manipulate. 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 1.12 implies several requirements on a high-quality receiver design. For example, the local oscillator frequencies in the transmitter modulator and receiver demodulator 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. 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, 2017). 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 or CW sweeps. Consider a pulse transmitted at time *zero* of the form $a(t)\cos(\Omega_t t + \varphi)$, where $a(t)$ is the pulse shape. In a coherent system, a pulse transmitted T seconds later will be of the form $a(t - T)\cos(\Omega_t t + \varphi)$. Note that both pulses have the same argument for their cosine term. Only the envelope term is delayed, shifting the pulse location on the time axis while keeping the same underlying sinusoid. An example of a noncoherently related second pulse would be $a(t - T)\cos[\Omega_t(t - T) + \varphi] = a(t - T)\cos[\Omega_t t + (\varphi - \Omega_t T)]$, which is nonzero over the same time interval as the coherent second pulse $a(t - T)\cos(\Omega_t t + \varphi)$ and has the same frequency, but has a different phase at any instant in time.

Figure 1.15 illustrates the difference. In the coherent case, the two pulses appear as if they were excised from the same underlying continuous, stable sinusoid; in the noncoherent case, the second pulse is not in phase with the extension of the first pulse. Another type of noncoherency arises when the starting phases φ of successive pulses are random, which occurs with some types of transmitters such as magnetrons (see Wallace et al., 2010). Because of the phase ambiguity discussed earlier, coherency also implies a system having both I and Q channels.

Another receiver 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). 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, not 89.9°. Of course, real receivers do not have perfectly matched channels. The effect of gain and phase imbalances will be considered in Chap. 3.

In the receiver structure shown in Fig. 1.12, the information-bearing portion of the signal is demodulated from the carrier frequency to baseband in a single mixing operation. This simple model is adequate to capture the receiver characteristics most important to radar signal processing. While convenient for analysis, radar receivers are virtually never

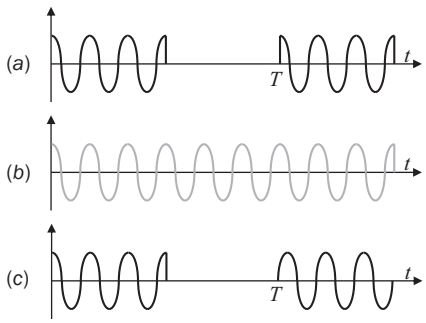


Figure 1.15 Illustration of the concept of a fixed phase reference in coherent signals. (*a*) Coherent pulse pair generated from the reference sinusoid. (*b*) Reference sinusoid. (*c*) Noncoherent pulse pair.

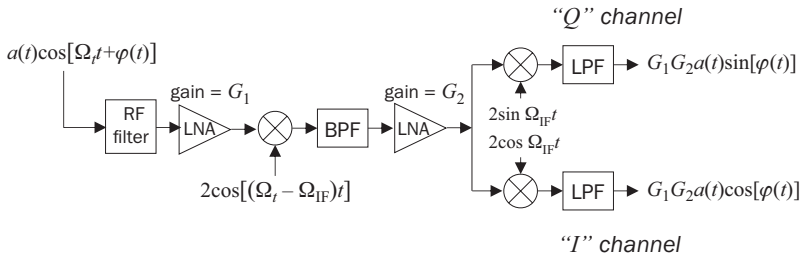


Figure 1.16 Structure of a superheterodyne radar receiver.

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 (Scheer, 1993)*. 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 1.16 shows a more representative *superheterodyne* receiver structure. The key feature of the superheterodyne receiver is that the demodulation to baseband occurs in two or more stages. The received signal is amplified immediately upon reception using a *low-noise amplifier* (LNA). The LNA, more than any other component, determines the *noise figure* of the overall receiver. It will be seen in Sec. 2.4 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 signal is next demodulated to an IF. The bandpass filter following the first mixer eliminates the sum frequency term, passing only the difference frequency term at the IF frequency Ω_{IF}. The RF filter eliminates any input signals at the *image frequency* $\Omega_t - 2\Omega_{IF}$; if present, these signals would generate undesired difference frequencies at the IF frequency. The signal is then amplified further. Amplification at IF is easier than at RF 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* (power loss in the mixer), improving the receiver sensitivity. 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 local oscillator to track changes in the transmitted frequency. More information on radar receivers is in Bruder (2010).

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. Improving SIR, resolution, or both is the major goal of many 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*, *region of support* (ROS) *expansion*, and *maximum likelihood estimation*. The remainder of this section defines SIR and resolution, and then gives simple examples of integration, phase history modeling, and ROS 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

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

$$x[n] = s[n] + w[n] \tag{1.23}$$

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 square of the peak signal amplitude 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 $\mathbf{E}\{|w[n]|^2\}$, where $\mathbf{E}\{\cdot\}$ represents the expected value of a random variable or process. If the interference is zero mean, as is very often the case, then the power also equals the variance σ_w^2 of the interference. If the desired signal is also modeled as a random process, 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

$$\chi = \frac{A^2}{\sigma_w^2} \tag{1.24}$$

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 $s[n]$ to the average noise power:

$$\chi = \frac{E_s}{\sigma_w^2} \tag{1.25}$$

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, $E_s = N{\cdot}A^2$. It will be seen in Chap. 6 that when matched filters are used, the peak SIR at the filter output equals the energy SIR of the original signal, so higher energy signals result in higher SIRs.

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 model of the target echo characteristics and radar detection algorithm, P_D is related to P_{FA} according to

$$P_D = (P_{FA})^{-1/(1+\chi)} \tag{1.26}$$

which shows that $P_D \longrightarrow 1$ as $\chi \longrightarrow \infty$ for fixed P_{FA}. As another example, the limit on precision (standard deviation of measurement error) 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. 7. 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 and Region of Support

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.[10] 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, for example range, and be unresolved in another, perhaps velocity.

Figure 1.17 illustrates the concept of resolution, in this case in frequency. The Fourier transform of a T-second cosinusoidal pulse $x(t) = A\cos(\Omega_0 t), -T/2 \le t \le +T/2$ is

$$X(\Omega) = \frac{A}{2}\left\{\frac{\sin[(\Omega-\Omega_0)T/2]}{(\Omega-\Omega_0)T/2} + \frac{\sin[(\Omega+\Omega_0)T/2]}{(\Omega+\Omega_0)T/2}\right\} \tag{1.27}$$

This function has peaks at positive and negative Ω_0 rad/s as expected for a real-valued signal. Considering just the positive frequency peak, the Rayleigh width (peak to first null) is easily seen to be $2\pi/T$ rad/s or $1/T$ Hz. Figure 1.17*a* shows a portion of the positive frequency spectrum of the sum of two unit-amplitude cosine functions with a frequency separation of $\delta F = 500$ Hz, one at 1000 Hz and one at 1500 Hz, with both having zero initial phase. The observation time T is 10 ms so that the Rayleigh width of each mainlobe is $\Delta F = 1/T = 100$ Hz. The two vertical dotted lines mark the two cosine frequencies. This signal could represent the Doppler spectrum of two moving targets with the same echo strength but different radial velocities.

These two signal components are considered well-resolved: there are two distinct, well-separated peaks in the spectrum. The frequency of each peak is perturbed slightly from the actual frequency by the sidelobes of the other sinusoid. Figure 1.17b to d repeats this measurement with the frequency spacing reduced to 100, 75, and 50 Hz. At $\delta F = 100$ Hz the two

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

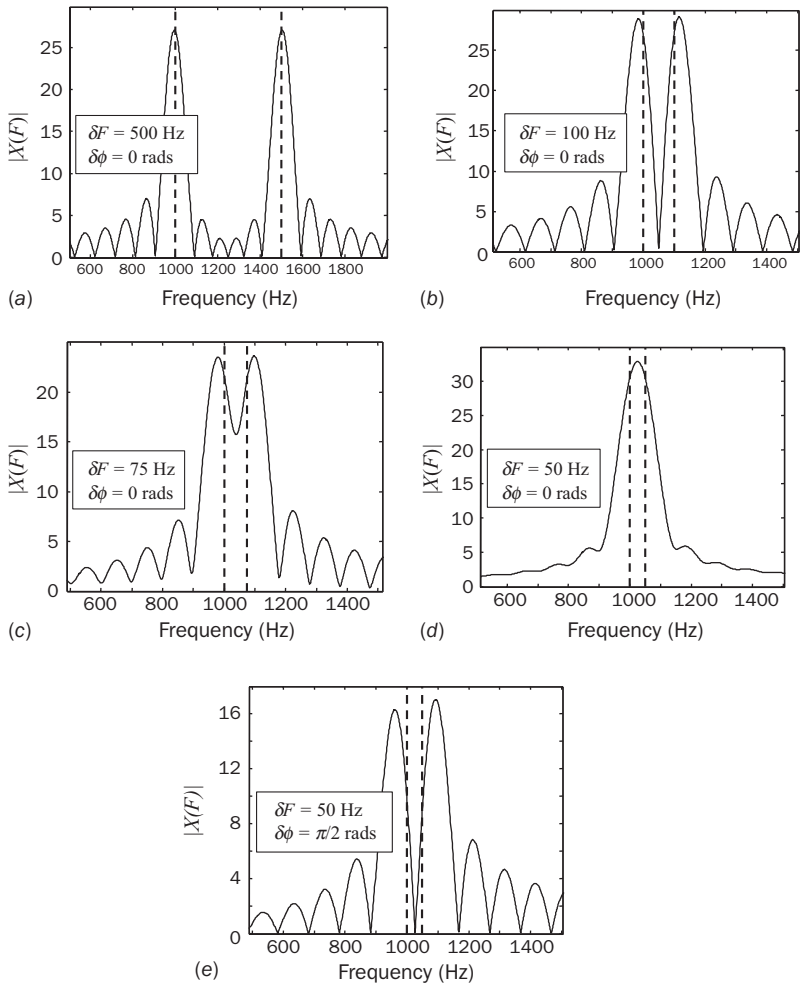


Figure 1.17 Resolution of two sinusoids in frequency, each having a Rayleigh width of 100 Hz and a 0° initial phase. (*a*) Well resolved at 500 Hz spacing. (*b*) Well resolved at 100 Hz spacing. (*c*) Marginally resolved at 75 Hz spacing. (*d*) Unresolved at 50 Hz spacing. (*e*) Well resolved at 50 Hz spacing and a phase difference of 90°.

spectral peaks are still well resolved, though with more perturbation of the apparent frequencies, but as the separation drops below the Rayleigh width ($\delta F < \Delta F$) to 75 and then to 50 Hz, they start to blur into a single spectral peak. At $\delta F = 50$ Hz they are no longer resolved; the spectrum measurement does not show two separate signals. At $\delta F = 75$ 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. Although resolution is defined in terms of pairs of scatterers, this demonstration suggests that the mainlobe width of the signature of a single isolated target is the major determinant of the system's resolution.

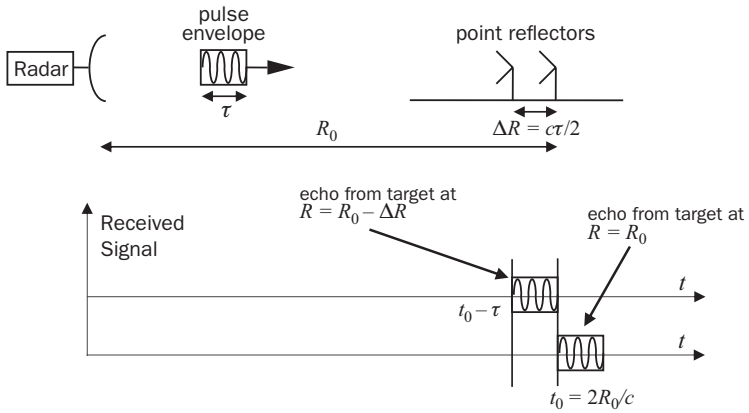


Figure 1.18 Geometry for describing conventional pulse range resolution. See text for explanation.

Figure 1.17*e* shows that phase also has an effect. In part € the sinusoid frequencies are the same as in part (*d*) but the initial phases differ by $\pi/2$ radians, whereas in part (*d*) the initial phases were the same. Although their locations are perturbed significantly and their amplitudes are imbalanced, nonetheless there are two clearly separated peaks even though the frequencies are separated by less than the Rayleigh width. As will be seen, echo phases are extremely sensitive to range and therefore effectively random in many cases. Consistent resolution of two signals independent of their relative phasing requires that they be separated by at least the Rayleigh width of the single-signal signature.

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 1.18 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 measured voltage at time t_0. Thus, scatterers distributed over an interval of $c\tau/2$ 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 range, azimuth angle, and elevation angle.

This description of range resolution applies only to unmodulated, constant frequency pulses. As will be seen in Chap. 4, both pulsed and CW waveforms, when combined with matched filtering, achieve range resolution determined by their bandwidth as in Eq. (1.2), namely $\Delta R = c/2\beta$ meters.

Angular resolution in the azimuth and elevation dimensions is determined by the antenna beamwidths in those 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 one-way 3-dB beamwidth

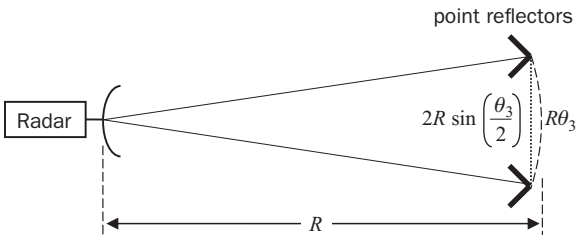


Figure 1.19 The angular and cross-range resolutions are determined by the 3-dB antenna beamwidth θ_3.

θ_3 of the antenna. Thus, the two point scatterers in Fig. 1.19 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* ΔCR to denote resolution in a dimension orthogonal to range. The arc length between the two scatterers, indicated by the curved dashed line, is $R\theta_3$ meters. The cross-range resolution is the length of the straight dotted line in Fig. 1.19 and is given by

$$\Delta CR = 2R\sin\left(\frac{\theta_3}{2}\right) \approx R\theta_3 \text{ m} \tag{1.28}$$

The approximation error is less than 1 percent when the 3-dB beamwidth is less than about 14°, which is usually the case for pencil beam antennas.

Three details bear mentioning. First, the literature frequently fails to specify whether one- or two-way 3-dB beamwidth is required or given. The one-way pattern appears to be most commonly used for quantifying angular resolution in monostatic radar. Second, note that cross-range resolution increases (degrades) linearly with range, whereas range resolution was independent of range. Finally, as with range resolution, it will be seen later (Chap. 8) that signal processing techniques can be used to improve angular or cross-range resolution well beyond the conventional $R\theta_3$ limit and to make it independent of range as well.

The radar resolution cell volume ΔV 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

$$\begin{aligned} \Delta V &= \pi\left(\frac{R\theta_3}{2}\right)\left(\frac{R\phi_3}{2}\right)\Delta R = \frac{\pi}{4}R^2\theta_3\phi_3\Delta R \\ &\approx R^2\theta_3\phi_3\Delta R \text{ m}^3 \end{aligned} \tag{1.29}$$

The approximation in the second line of Eq. (1.29) 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 angular spreading of the beam at longer ranges.

The region of support of a function is the portion of the independent variable axis (typically time, frequency, angle, or spatial coordinate in radar) over which the function is nonzero. Thus, a 100-μs pulse has a larger ROS in time than a 10-μs pulse. As seen above, antenna beamwidths are inversely proportional to the aperture length D and the mainlobe of the frequency spectrum of a sinusoid is inversely proportional to the length of the sinusoid in time. This reciprocal behavior of the signal domain and Fourier domain regions of support is illustrated in Fig. 1.20 and is sometime referred to as "reciprocal spreading." 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

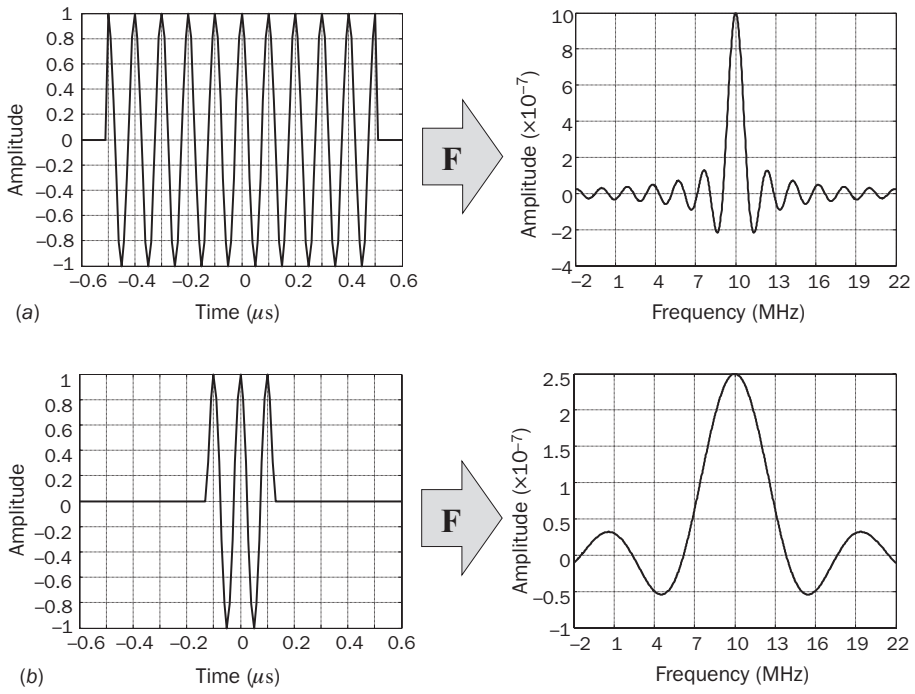


Figure 1.20 Illustration of reciprocal spreading property of Fourier transforms. (*a*) A sinusoidal pulse and the main portion of its Fourier transform. (*b*) A narrower pulse has a wider transform. See text for details.

sinc function centered on 10 MHz having 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 4.

Improving resolution in one domain requires increasing the size of the ROS in the complementary Fourier domain. For example, improving range resolution for simple pulses requires using shorter pulses, as was seen in Sec. 1.4.2; and Fig. 1.18 shows that a shorter pulse implies a larger ROS in frequency (wider spectrum), that is, more bandwidth. Conversely, improving resolution in the frequency domain requires a narrower spectrum mainlobe and thus a longer observation (larger ROS) in the time domain. This behavior holds for any two functions related by a Fourier transform: finer resolution in one domain requires a wider ROS in the complementary domain.

Radar designers have developed techniques for increasing the appropriate ROS to obtain improved resolution in various dimensions. For example, the increased waveform bandwidth required to improve range resolution has led to the use of wideband phase- and frequency-modulated waveforms in place of the simple pulse (Chap. 4). Improving cross-range resolution in radar imagery 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 frequency) resolution requires a longer time observation and is accomplished with multipulse or multisweep waveforms. Because the antenna far field pattern is the Fourier transform of the aperture current distribution, improved angular resolution requires larger apertures, that is, bigger antennas.

1.4.3 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 (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 signal is transmitted, reflects off a target, and at the appropriate time the receiver output voltage signal is measured, yielding a complex echo amplitude $A\exp(j\varphi)$ 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

$$\chi_1 = \frac{\text{signal power}}{\text{noise power}} = \frac{A^2}{\sigma_w^2} \tag{1.30}$$

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 complex sum operation, retaining the phase information, is a coherent integration:

$$\begin{aligned} z &= \sum_{n=0}^{N-1} \{A\exp(j\varphi) + w[n]\} \\ &= NA\exp(j\varphi) + \sum_{n=0}^{N-1} w[n] \end{aligned} \tag{1.31}$$

The power in the integrated signal component is its amplitude squared, $(N{\cdot}A)^2$. Provided the noise samples $w[n]$ are independent of one another and zero mean, the power in the noise component is easily shown to be 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

$$\chi_N = \frac{N^2A^2}{N\sigma_w^2} = N\left(\frac{A^2}{\sigma_w^2}\right) = N\chi_1 \tag{1.32}$$

Coherently integrating N measurements has improved the SNR by a factor of N compared to that of one individual measurement. 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 complex-valued (amplitude and phase) data samples are added. The hope is that the signal components will add in phase, that is, constructively. This is often described as adding on a *voltage* basis, since the amplitude (voltage) of the integrated signal component ~~is~~ increased by a factor of N, with the result that signal power increased by N^2. The noise samples, whose phases varied randomly, will add 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 and create an integration gain.

Sometimes the data must be preprocessed to ensure that the signal component phases align so that the maximum 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. (1.31) would instead become

$$z = \sum_{n=0}^{N-1} A\exp[j(2\pi f_D n + \varphi)] + w[n] \tag{1.33}$$

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 N^2A^2. However, if the Doppler shift is known or can be estimated, the phase progression of the signal component can be compensated before summing by pre-multiplying the data by a countervailing phase progression:

$$\begin{aligned} z' &= \exp(-j2\pi f_D n) \cdot z = \sum_{n=0}^{N-1} \exp(-j2\pi f_D n)\{A\exp[j(2\pi f_D n + \varphi)] + w[n]\} \\ &= NA\exp(j\varphi) + \sum_{n=0}^{N-1} \exp(-j2\pi f_D n) w[n] \end{aligned} \tag{1.34}$$

The phase compensation aligns the signal component phases so that they add constructively. The noise phases remain random with respect to one another. The integrated signal power is again N^2A^2, 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 target samples add in phase is an example of using *phase history modeling*: if the sample-to-sample pattern of target echo phases can be predicted or estimated (at least 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.

Coherent integration carries a risk. If the target echo phases are not constant or are not compensated successfully, they will not add in phase and the integration gain will be reduced. Worse, they may even combine destructively, actually reducing the integrated SNR relative to the single-measurement SNR. This risk can be avoided by considering noncoherent integration, in which the phases are discarded and the magnitudes, magnitudes-squared, or log magnitudes of the measured data samples are added. If the magnitude-squared is chosen, then z is formed as

$$\begin{aligned} z &= \sum_{n=0}^{N-1} \left|A\exp(j\varphi) + w[n]\right|^2 = \sum_{n=0}^{N-1} \{A\exp(j\varphi) + w[n]\}\{A\exp(j\varphi) + w[n]\}^* \\ &= \sum_{n=0}^{N-1} \left|A\exp(j\varphi)\right|^2 + \sum_{n=0}^{N-1} \left|w[n]\right|^2 + \sum_{n=0}^{N-1} 2\,\mathrm{Re}\{A\exp(j\varphi)w^*[n]\} \\ &= NA^2 + \sum_{n=0}^{N-1} \left|w[n]\right|^2 + \sum_{n=0}^{N-1} 2\,\mathrm{Re}\{A\exp(j\varphi)w^*[n]\} \end{aligned} \tag{1.35}$$

The important fact is that phase information in the received signal samples is discarded so the result is not sensitive to phase alignment.

The first line of Eq. (1.35) 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 defined as simply as it was for the coherent case.

However, it is possible to define a noncoherent gain indirectly. 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. The coherent integration gain for this example would be a factor of $N = 10$.

Computing the noncoherent integration gain typically requires derivation of the probability density functions of the noise-only and signal-plus-noise cases. 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 ($\alpha = 1$). Thus, noncoherent integration is less efficient than coherent integration. This should not be surprising, since not all the signal information is used.

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 verified that maximizing SNR also optimizes detection performance. Until the detector is considered, it is hard to see precisely how performance depends on SNR. Having seen the major components of a typical coherent radar system, the most common signal processing operations in the radar signal processing chain are now described heuristically. Sketching out this preview of the "big picture" from beginning to end may make it easier to understand the motivation for and interrelation of many of the processing operations to be described in later chapters.

Figure 1.21 illustrates one possible sequence of operations in a generic radar signal processor. Each major category of processing has a few more specific representative operations listed adjacent. The left branch is for search and track radars intended mainly to detect and locate discrete targets and track them over time. The right branch is for imaging radars. Many other types of radars, for example, weather radars, follow flows with many similarities to the search and track column. The sequences shown and their ordering are not unique or 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, though the trend in modern systems is to digitize closer and closer to the antenna. Radar signal *phenomenology* must also be considered.

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 orders of magnitude or more. Different classes or levels of operations tend to operate on significantly different time scales. Figure 1.22 illustrates one possible association of operations and time scale.

Operations that are applied to data from a single pulse or FMCW sweep occur on the shortest time scale, often referred to as *fast time* because the PRI or FMCW sweep rate, which limits the time available for processing before data from the next pulse or sweep arrives, is relatively short. The PRI or sweep rate is typically on the order of tens of microseconds to hundreds of milliseconds. Furthermore, the sample rate of the fast time data, determined by the instantaneous pulse bandwidth or FMCW sweep bandwidth (see Chap. 2), is on the order of hundreds of kilohertz to as much as a few gigahertz in some cases. Corresponding sampling intervals range from a few microseconds down to a fraction of a nanosecond. Typical fast time operations are digital I/Q signal formation, beamforming, matched filtering, sensitivity time control, and some forms of jammer suppression.

The next level up in signal processing operations acts on data from multiple pulses or sweeps. The available processing time is simply the number of pulses or sweeps M times the interval between them and is much longer than the single-interval time scale of the

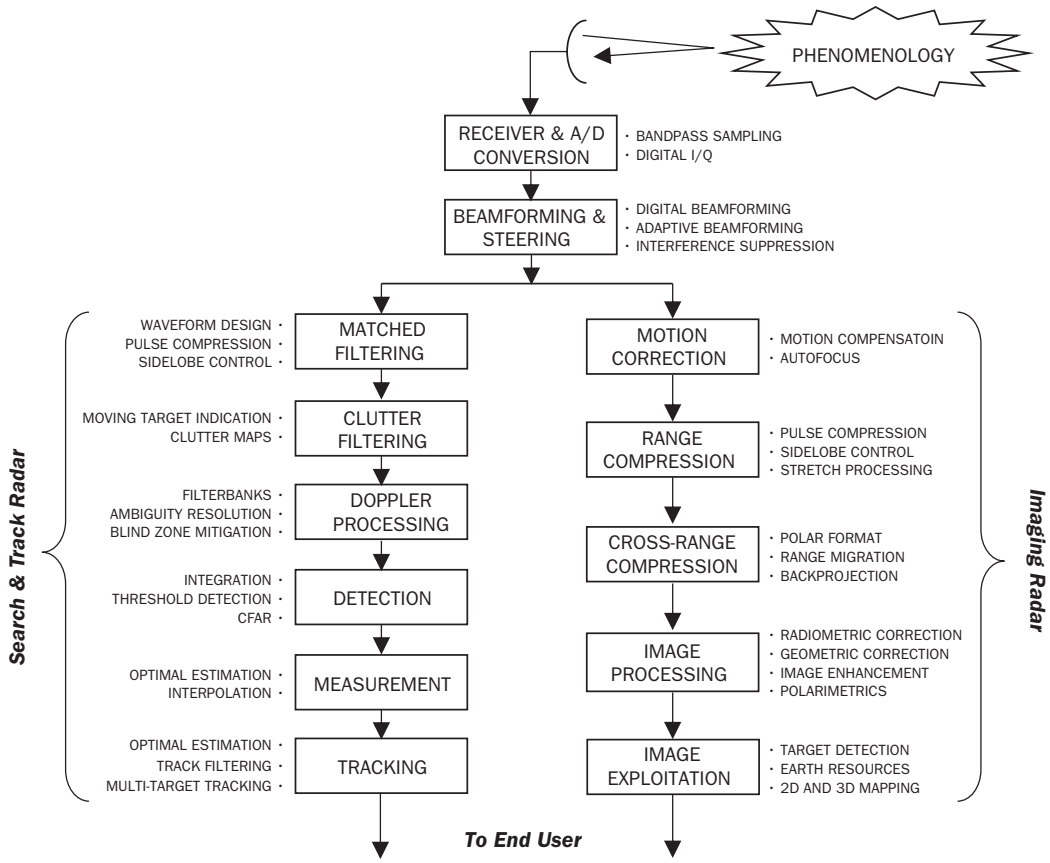


Figure 1.21 One example of a generic radar signal processor flow of operations.

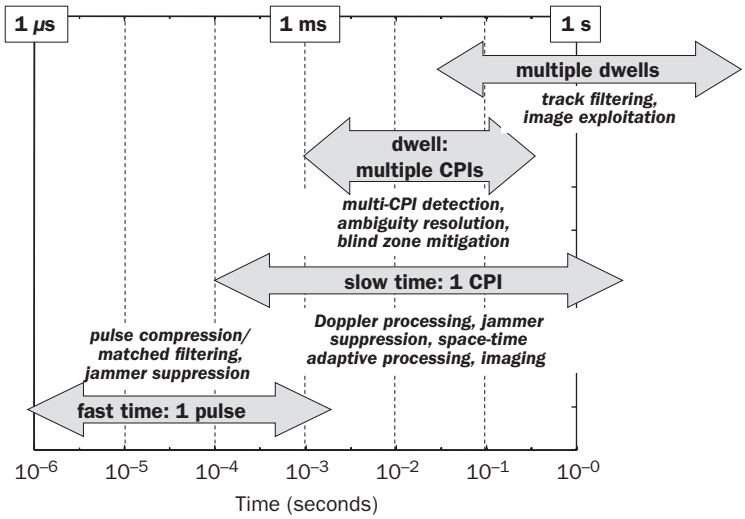


Figure 1.22 The range of time scales over which radar signal processing is performed.

fast time operations. Typical values of M range from single digits to a few tens of pulses or sweeps. Combined with the typical interval values mentioned above, the time scale for most operations at this level is hundreds of microseconds to tens or hundreds of milliseconds.

These multipulse or multisweep operations are said to act in *slow time*. Note that the repetition interval becomes the sampling rate of the slow time data and is much lower than the fast-time sampling rate. Typical slow-time operations include coherent and noncoherent integration, Doppler processing of all types, and space-time adaptive processing. Synthetic aperture imaging is also a slow-time operation; however, the number of pulses involved may be hundreds to many thousands, so that the total time interval can be hundreds of milliseconds to many seconds. 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 or sweeps 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, range and Doppler ambiguity resolution, blind zone mitigation, and multilook SAR imaging.

So long as a target stays within a radar's detection range, the radar track can be continued. Some radars may track detected targets for many seconds or minutes using data from multiple dwells. Track filtering operates in this open-ended regime. Some imaging radars, especially in earth resources and strategic intelligence applications, may monitor an area over days, months, or even years.

1.5.2 Phenomenology

To design a successful signal processor, one must understand how the target or scene information of interest is encoded in those signals. *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, because of the details of its shape, a target may reflect more power to the radar when it is illuminated from the front than from the side.

Models of the behavior of typical measured signals that are relevant to the design of signal processors will be developed in Chap. 2. The radar range equation will give a means of predicting received 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 the amplitude of radar echoes; 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 an idealized arbitrarily fine resolution signal with the radar waveform (in the range dimension) or its antenna pattern (in the azimuth or elevation dimension, both also called the 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 the search and track radar column of Fig. 1.21 can be considered as signal conditioning operations whose purpose is to improve the SIR of the

data prior to detection, parameter measurement, or tracking 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*, *matched filtering*, *integration*, *clutter filtering*, and *Doppler processing*.

Beamforming as a signal processing operation is applicable when the radar antenna is an array, that is, 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. 1.7. 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 sample 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. One can then design a set of weights for combining the channels such that the antenna not only has a high-gain steerable 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 1.23 illustrates interference suppression using STAP, allowing a previously invisible target signal to be seen and perhaps detected. The figure shows the distribution of received energy in an airborne sidelooking radar in angle of arrival and Doppler frequency coordinates at a particular range. The two vertical bands in Fig. 1.23*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 spread across all Doppler frequencies observed by the radar. The diagonal band in Fig. 1.23*a* is due to ground clutter, for

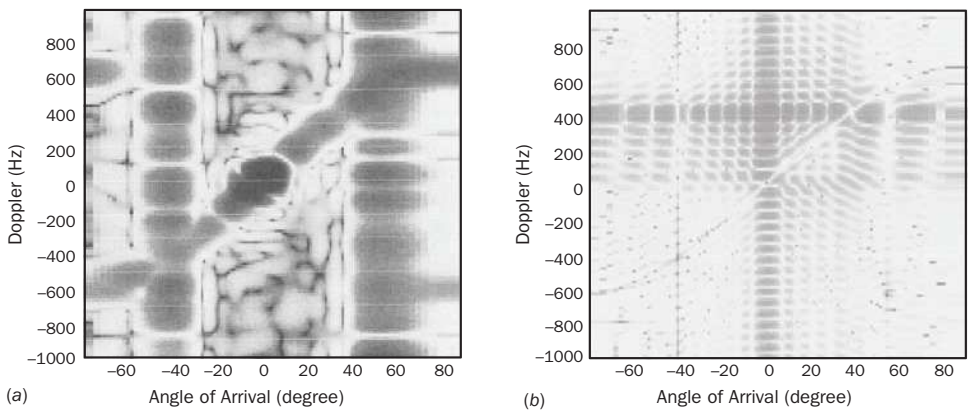


Figure 1.23 Example of effect of adaptive beamforming. (*a*) Map of received signal power as a function of angle of arrival and signal Doppler shift. (*b*) Angle-Doppler map after adaptive processing. A target is now visible at 400 Hz and 0°; the horizontal and vertical bars are the sidelobes of the target response. (Images courtesy of Dr. W. L. Melvin. Used with permission.)

which the Doppler shift depends on the angle from the radar to the ground patch contributing energy. Figure 1.23*b* shows that the adaptive filtering has created nulls along the loci of the jammer and clutter energy, allowing the target peak at 0° angle of arrival and 400 Hz Doppler shift to be seen. (The vertical and horizontal bars are sidelobes of the target signature.) Adaptive interference suppression will be introduced in Chap. 9.

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 a pulsed 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. Sensitivity and fine range resolution appear to be conflicting goals.

Pulse compression provides a way out of this dilemma by decoupling the waveform bandwidth from its duration, 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. The instantaneous frequency of an LFM pulse is swept linearly 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. Figure 1.24*a* shows the real part of a complex LFM chirp that sweeps over 20 MHz in 1 *m*s (black curve); the imaginary part is similar except for a phase shift (gray curve).

The matched filter is by definition a filter in the radar receiver designed to maximize the SNR at its output. Chapter 4 will show that 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. In pulsed radar, pulse compression is the process of designing a modulated 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 1.24*b* shows the output of the matched filter corresponding to the LFM pulse of Fig. 1.24*a*. Note that the mainlobe of the response, measured by its Rayleigh width, is only 1/20th the duration of the 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 $\Delta R = c/2\beta$, as noted earlier in Eq. (1.2).

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* (MTI), which is simply pulse-to-pulse or FMCW sweep-to-sweep highpass filtering of the radar echoes at a given range to suppress constant components, which are assumed to be due to non-moving clutter. Extremely simple, very low-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 Fourier transform, or occasionally some other spectral estimation technique, to explicitly compute the spectrum of

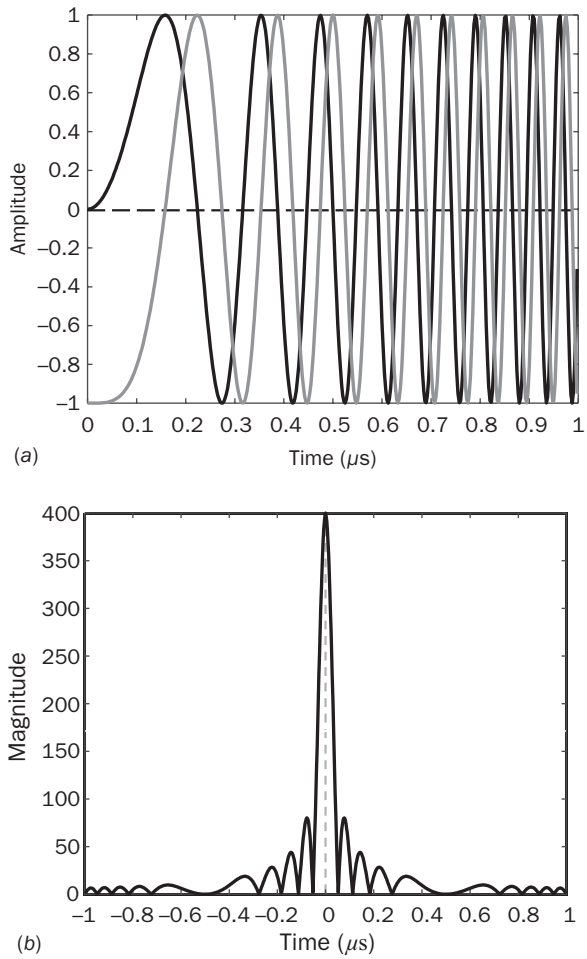


Figure 1.24 (*a*) Complex linear FM waveform modulation function. The instantaneous frequency sweeps from 0 Hz to 20 MHz at a constant rate over a 1-μs interval. The black curve is the real part, the gray curve is the imaginary part. (*b*) Magnitude of the output of the matched filter for the LFM waveform of (*a*).

the echo data for a fixed range across multiple pulses or FMCW sweeps. Due to their different Doppler shifts, energy from moving targets concentrates in different parts of the spectrum from the clutter energy, allowing separation and detection of the targets. Doppler processing obtains more information from the radar signals than does MTI filtering, such as the number and approximate velocity of moving targets. The cost is more required radar pulses or sweeps, consuming more 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 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 jamming or unintentional electromagnetic interference

(EMI), for example from other radars or cellular phone services. The signal processor must somehow analyze the total received signal and determine whether it contains a target echo and, if so, at what range, angle, and velocity.

Because the complexity of radar signals will lead to the use of statistical models, detection of target echoes in the presence of competing interference signals is a problem in statistical 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, after any 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 due to the presence of a target echo in addition to the interference, and a detection or "hit" is declared. In essence, the detector decides 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 1.25 illustrates the concept. The "clutter + target" signal might represent the variation in received signal strength versus range (fast time) for a single transmitted pulse or CW sweep. It crosses the threshold at three different times, suggesting the presence of three targets at different ranges.

Because they are the result of a statistical process, threshold detection decisions have a non-zero 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, that is, 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. It will be seen in Chap. 4 that the matched filter maximizes the SIR, thus providing the best threshold detection performance. Furthermore, the achievable SIR increases monotonically with the transmitted signal energy E, thus encouraging the use of longer pulses or FMCW sweeps to get more energy on the target. Since longer pulses or sweeps degrade range resolution, the frequency modulation will also be 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 1.25 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.

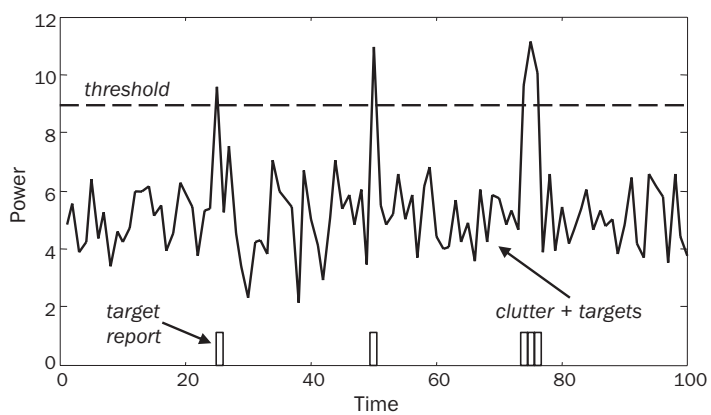


Figure 1.25 Illustration of threshold detection.

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 a selected acceptable rate. However, in many 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.

1.5.5 Measurements and Track Filtering

Radar systems employ a wide variety of processing operations after the point of target detection. One of the most common post-detection processing steps is *tracking* of targets, an essential component of many radar systems. Tracking comprises multiple *measurements* of the position of detected targets followed by *track filtering*.

The radar signal processor detects the presence of targets using signal conditioning and 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 position by using signal processing methods to more precisely estimate the time delay after signal transmission at which the threshold crossing occurred, the angle of the target relative to the antenna mainbeam direction, and its radial velocity. Individual measurements will always still 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 combining a series of such measurements to estimate a complete trajectory of the target over time. It is often categorized 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 new measurements to associate with which targets already 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 1.26 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

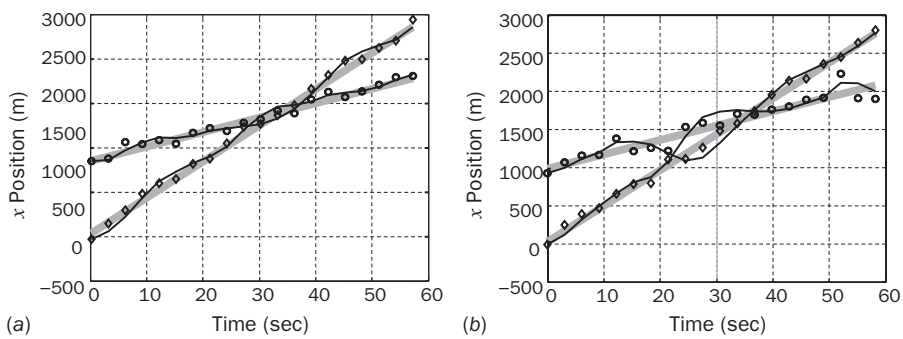


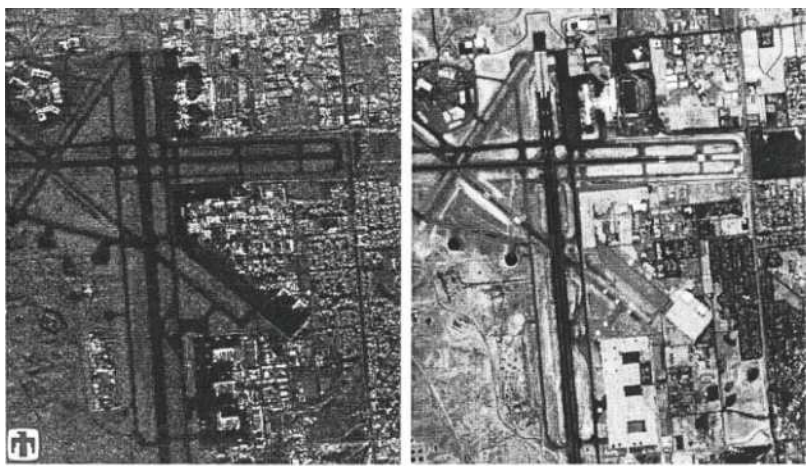
Figure 1.26 Track filtering of noisy measurements for two targets in one dimension using an alpha-beta filter. Markers show individual measurements. Gray lines are the actual position, black lines are the filtered position estimates. (*a*) Tracks follow the current targets in low measurement noise even when the trajectories cross. (*b*) Tracks incorrectly switch targets in higher measurement noise.

filter, to be discussed in Chap. 7. The actual position in the x dimension versus time for each target (called the *ground truth*) is shown by the gray lines. The two targets are moving at different velocities along the x axis and one passes the other at around time 33 seconds. The circle and diamond markers 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. 1.26*b* the noise variance is higher, causing the filter to make multiple association errors in the 30 to 40 second time interval. From time 42 seconds onward the tracks are swapped, each following a different target from when they started, . This represents an error in measurement-to-track *data association*. A variety of techniques attempt to address association problems; a few basic ones are discussed in Chap. 7.

1.5.6 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 1.27 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 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 those observed by the human visual system. In contrast, the radar image, while remarkable, is monochromatic, offers less detail, and exhibits a "speckled" texture, some 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 or ground sources for



(*a*) (*b*)

Figure 1.27 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.)

illumination; they provide their own "light" via the transmitted signal. If the example of Fig. 1.27 were repeated in the middle of a rainy night, the SAR image would not be affected in any noticeable way, but the optical image 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 modulated waveforms, usually linear FM. A waveform that is swept over a large enough bandwidth β and processed using a matched filter can provide very good range resolution in accordance with Eq. (1.2). 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 non-imaging 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. (1.28). Realistic antenna beamwidths for narrow-beam antennas are typically on the order of 0.5° (8.7 mrad) for a low earth orbit satellite at a range of 700 km, or 3° (52 mrad) for an airborne system at 10 km range. The cross-range resolutions that result would be 520 m for the airborne system and about 6 km for the satellite, orders of magnitude worse than typical range resolutions and far 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 associated with moving airborne or space-based radars rather than with fixed ground-based radars. Figure 1.28 illustrates the concept. By transmitting pulses or

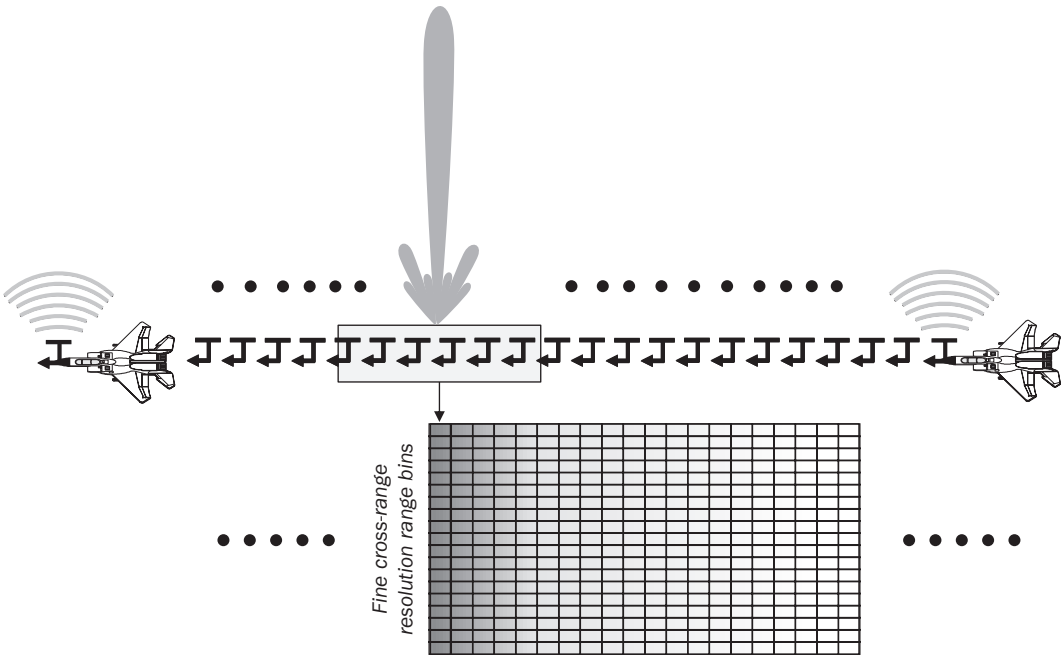


Figure 1.28 The concept of synthetic aperture radar.

FMCW sweeps at each indicated location, collecting the fast time (range) data, 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. These "synthetic arrays" may be hundreds of meters to a few kilometers long. As suggested by Eq. (1.10) (though some details differ in the SAR case), a very large aperture size produces a very narrowly focused effective antenna beam, making possible very fine cross-range resolution. The SAR concept is explained more fully in Chap. 8.

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 2020.

1.6.1 Introductions to Radar Systems and Applications

One of the newest and possibly best "radar 101" introductions is volume 1 of the three-volume *Principles of Modern Radar* (POMR) series, which addresses a full range of basic radar principles, phenomenology, components, and signal processing (Richards et al., 2010). The second volume by Melvin and Scheer (2012) addresses advanced radar techniques, providing more depth to the basic signal processing methods of volume 1 as well as surveying numerous advanced modern processing techniques. An up-to-date survey of a broad range of traditional and modern radar applications is given in the third volume, also by Melvin and Scheer (2013). It shows how many of the techniques discussed in these texts and those below are brought together into complete systems. Collectively, the series provides a comprehensive textbook and reference for studying modern radar systems. At this writing a significantly updated second edition of all three volumes is in preparation, with publication expected in 2022 and 2023.

There are several other good introductory books on radar systems. A classic introduction to radar systems, now in its third edition, is by Skolnik (2001). The comprehensive text by Peebles (1998) is more academic in its approach. Mahafza (2013) provides a number of useful MATLAB® files to aid in simulation and experimentation. Unique in the field, *Stimson's Introduction to Airborne Radar* (Stimson et al., 2014) is a very accessible introduction to radar with a strong emphasis on airborne pulsed Doppler systems and featuring extensive illustrations and examples. Alabaster (2012) provides another excellent introduction to airborne pulsed Doppler systems and signal processing. Morris and Harkness (1996) is also a good introduction to these 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 areas of synthetic aperture imaging and adaptive processing. There have been few books that address the middle ground of basic radar signal processing, such as matched filtering, Doppler filtering, and CFAR detection. Such books are needed to provide greater quantitative depth than is available in the radar system books while at the same time covering a range of topics found in most radar systems, instead of being restricted to in-depth coverage of a single advanced application area. This text aims to fill that gap.

There are a few texts that fit somewhat into this middle area. Nathanson (1991) purports to cover 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 and very concise analyses of many basic signal processing functions. The text by Levanon and Mozeson (2004) addresses the widening variety of radar waveforms in detail. 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 based on experience with the Canadian RADARSAT satellite SARs. The spotlight SAR mode received considerable development in the 1990s, and two major groups published competing texts in the mid-1990s. Carrara et al. (1995) represented the work of the group at the Environmental Research Institute of Michigan (ERIM, ultimately becoming a part of MacDonald Dettwiler and Associates); 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. The newest SAR book is by Jansing (2021) and features many more modern examples and algorithms.

STAP, one of the most active radar signal processing research areas, began in earnest in 1973. Klemm (1998) wrote the first significant open literature text on the subject. The book by Guerci (2014) 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 text by Nitzberg (1999), which discusses several forms of sidelobe cancellers.

New signal processing methods continue to find applications in radar. Volume 2 by Melvin and Scheer (2012) of the POMR series mentioned above surveys a wide range of these and other advanced radar signal processing techniques, including such new topics as MIMO radar and *compressive sensing* (CS). For example, MIMO radar may (or may not) prove to have advantages in many areas of radar search, tracking, imaging, and interference rejection. A good primer is Bergin and Guerci (2018), while more in-depth coverage is available in Li and Stoica (2008). The first book on CS in radar is by De Maio et al. (2019). As in many other fields of technology, applications of artificial intelligence-based ideas such as "deep learning" and "machine learning" to radar are active areas of current research, but no book-level treatments of this area have appeared as of this writing (2021).

1.6.4 Radar Applications

The preceding sections have cited several 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. Volume 3 of the POMR series by Melvin and Scheer (2013) provides an excellent survey of and introduction to a wide range of applications in a single text.

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.* Another good source of imaging radar research is the various publications of SPIE (formerly the Society of Photo-Optical Instrumentation Engineers, now just SPIE).

References

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

Balanis, C. A., *Antenna Theory,* 4th ed. Harper & Row, New York, 2016.

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

Bergin, J., and J. R. Guerci, *MIMO Radar: Theory and Application*. Artech House, Boston, MA, 2018.

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.

Bruder, J. A., "Radar Receivers," Chap. 11 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.

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*. Wiley, New York, 1991.

De Maio, A., Y. C. Eldar, and A. M. Haimovich, *Compressed Sensing in Radar Signal Processing*. Cambridge University Press, Cambridge, UK, 2019.

EW and Radar Systems Engineering Handbook, 4th ed., Oct. 2013. Naval Air Warfare Center, Weapons Division. Available at https://apps.dtic.mil/docs/citations/ADA617071.

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

Guerci, J. R., *Space-Time Adaptive Processing for Radar*, 2d ed. Artech House, Norwood, MA, 2014.

Institute of Electrical and Electronics Engineers (IEEE), "IEEE Standard Definitions of Terms for Antennas," Standard 145-2013, Mar. 6, 2014.

Institute of Electrical and Electronics Engineers (IEEE), "IEEE Standard Radar Definitions," Standard 686-2017, Mar. 23, 2017.

Institute of Electrical and Electronics Engineers (IEEE), "IEEE Standard Letter Designations for Radar-Frequency Bands," Standard 521-2019, Nov. 7, 2019.

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.

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

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

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

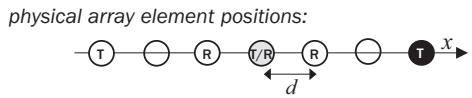
Li, J., and P. Stoica, *MIMO Radar Signal Processing*. Wiley-IEEE Press, Hoboken, NJ, 2008.
Mahafza, B. R., *Radar Systems Analysis and Design Using MATLAB®*, 3d ed. Chapman & Hall/CRC, New York, 2013.
Melvin, W. L., and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Raleigh, NC, 2012.
Melvin, W. L., and J. A. Scheer (eds.), *Principles of Modern Radar: Radar Applications*. 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, no. 3, pp. 1234–1239, Jul. 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 ed. McGraw-Hill, New York, 1991.
Nitzberg, R., *Radar Signal Processing and Adaptive Systems*, 2d ed. Artech House, Boston, MA, 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., "Virtual Arrays, Part 1: Phase Centers and Virtual Elements," technical memorandum, Mar. 2018. Available at http://radarsp.com.
Richards, M. A., "Virtual Arrays, Part 2: Virtual Arrays and Coarrays," technical memorandum, Feb. 2019. Available at http://radarsp.com.
Richards, M. A., J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.
Scheer, J. A. (ed.), *Coherent Radar Performance Estimation*. Artech House, Boston, MA, 1993.
Sherman, S. M., *Monopulse Principles and Techniques*, 2d ed. Artech House, Boston, MA, 2011.
Skolnik, M. I., *Introduction to Radar Systems*, 3d ed. McGraw-Hill, New York, 2001.
Soumekh, M., *Synthetic Aperture Radar Signal Processing with MATLAB Algorithms*. Wiley, New York, 1999.
Stimson, G. W. et al., *Stimson's Introduction to Airborne Radar*, 3rd ed. SciTech Publishing, Raleigh, NC, 2014.
Stutzman, W. L., "Estimating Gain and Directivity of Antennas," IEEE *Transactions on Antennas and Propagation*, vol. 40, no. 4, pp. 7–11, Aug. 1998.
Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*, 3d ed. Wiley, New York, 2012.
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*. Wiley, New York, 2002.
Wallace, T. V, R. J. Jost, and P. E. Schmid, "Radar Transmitters," Chap. 10 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.

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 30 to 300 GHz. Only certain frequencies in this band are widely used for radar; others are deliberately avoided. 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 and assuming longer range capability is preferable, identify a frequency in the MMW band that might *not* be a good choice for a MMW radar operating frequency. Explain.
5. Consider two equal-power transmitters, one at 10 GHz and the other at 30 GHz. The power received from each is measured at a distance of 10 km and compared. Based on Figs. 1.3 and 1.4, estimate how many decibels weaker than the 10 GHz signal will the 30 GHz signal be in sea level clear air conditions. Repeat for a medium rainfall of 2.5 mm/h and a tropical downpour of 50 mm/h.
6. Compute the bandwidth β needed to achieve range resolutions of 1 m, 1 km, and 100 km. What is the length of a constant-frequency rectangular pulse having this Rayleigh bandwidth for each value of resolution?
7. Derive Eq. (1.10) from Eq. (1.8).
8. Using the result from Prob. 7, find the value of D_y/λ_t for which the error in the small-angle approximation of the Rayleigh beamwidth is 10 percent. What are the exact and approximate Rayleigh beamwidths for that value of D_y/λ_t?
9. How large must a uniformly illuminated aperture antenna be (value of D_y) in terms of wavelengths so that its one-way 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. (1.11)?
10. The adjoining figure shows a physical one-dimensional uniform linear array comprising two transmit elements, two receive elements, and one element that both transmits and receives. This array is "sparse" in that not all elements needed to form a conventional "filled" ULA are present; the two dashed circles represent missing elements. Construct the virtual array and show that it is filled. What is the spacing and overall size of the VA elements compared to those of the physical array?



11. 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.
12. Continuing Prob. 11, what is the actual cross-range resolution in feet at each RF if the antenna width is 6 inches?
13. Starting from Eq. (1.14) and setting $a_n = 1$, derive Eq. (1.15). *Hint*: You will need the finite geometric sum formula.

14. 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. (1.28) has an error of no more than 1 percent?
15. Determine the cross-range resolution ΔCR in meters at ranges of 10, 100, and 1000 km for a 3-dB beamwidth $\theta_3 = 3°$.
16. Determine the approximate size of a volume resolution cell in cubic meters, ΔV, for $R = 20$ km, $\Delta R = 100$ m, and $\theta_3 = \phi_3 = 3°$.
17. Equation (1.2) is a general result applicable to any waveform provided a matched filter is used. Specialize this to a common formula for the range resolution of a simple pulse waveform in terms of the pulse length τ.
18. Equation (1.35) showed that the magnitude-squared of signal-plus-noise data cannot be expressed as the sum of only a signal portion and a noise portion; cross products of the signal and noise samples were also present, complicating the definition of post-integration SNR and integration gain. Show that this is also true for the magnitude (not magnitude-squared) of the data. It is sufficient to use $N = 1$, that is, a single sample. *Hint*: Apply a Taylor series to the square root of the magnitude-squared result.
19. Repeat Prob. 18 for a log-magnitude detector.
20. How many signal-plus-noise samples must be coherently integrated to achieve an integration gain of at least 3, 10, and 20 dB? The result N must be an integer.
21. Repeat Prob. 20 for the case of noncoherent integration assuming an integration gain of $N^{0.7}$. Summarize the main advantage and disadvantage of noncoherent integration relative to coherent integration.
22. In a threshold detection scheme like that pictured in Fig. 1.25 and with a fixed SNR, would the probability of false alarm P_{FA} be expected to increase, decrease, or be unchanged if the threshold level was raised? What about the probability of detection, P_D?
23. Suppose that a system for which Eq. (1.26) applies requires $P_D = 0.9$ and $P_{FA} = 10^{-6}$. What is the required SNR in linear units and in decibels?