Chapter 2
Systems Aspects of Millimeter-Wave Power Amplifiers

The primary motivation for operating wireless systems at higher frequencies lies in the fact that an increase in frequency is generally associated with an increase in bandwidth in communication systems, as well as an improvement in resolution in imaging and ranging systems. Alongside these factors, a reduction in the physical size of components and antennas also accompanies all wireless systems that operate at higher frequencies. As one would expect (and as several decades of research have made abundantly clear), there are fundamentally challenging aspects of millimeter-wave systems that complicate the design process and hinder the widespread implementation of such systems.

The characteristics of semiconductor devices are such that the performance of the device decreases with an increase in frequency, and a particular technology typically exhibits lower gain, an increased noise figure and impaired linearity performance as frequency increases. This reduction in performance can generally be attributed to the increased losses that materials experience at higher frequencies. Moreover, free-space propagation of electromagnetic waves exhibit losses that rapidly increase with increasing frequency—apart from distinct windows in the frequency spectrum where attenuation is somewhat reduced, these windows appear around 35, 90, 140 and 250 GHz in the millimeter-wave spectrum. Consequently, the bands listed here are popular for implementing wireless systems in the traditional sense (where maximizing range is a fundamental requirement). On the other hand, one of the oxygen absorption bands exists at 60 GHz, which means that a large portion of the power contained in an electromagnetic wave will be absorbed by oxygen molecules in free space. While this band is not at all suitable for implementing a conventional wireless system, given the significantly reduced propagation distance, it is useful in high-density networks and provides the benefit of aggressive frequency reuse in a relatively small area.

Improvements in process and manufacturing technology in the last decade or so have brought with them a new wave of research focused on system-on-chip (SoC) solutions for complex commercial wireless systems. This is in stark contrast
to the earlier years of millimeter-wave research, where these systems were largely limited to specialized applications and defense systems because of their high implementation costs and the requirement for complex fabrication processes.

2.1 Antennas for Millimeter-Wave Applications

Antennas are a crucial component of any wireless system, since the functioning of such a system would be impossible without an antenna that converts between guided electromagnetic waves propagating along a transmission line and unguided plane waves that propagate in free space, and vice versa. Fundamentally, an antenna is a bidirectional (reciprocal) device, in the sense that an identical antenna may be used for both transmit and receive functions. A transmitter antenna radiates spherical waves that begin to approximate a plane wave once an observer is at a large enough distance from the transmitter. This region is referred to as the far-field of the antenna. On the other end of a wireless system, a receiver would intercept a portion of the power contained in the plane wave and deliver some of this power (or ideally, all of this power) to a load.

2.1.1 Antenna Parameters

2.1.1.1 Power Radiated by an Antenna

Understanding the mechanisms by which an antenna radiates power into free space (and conversely, can receive power from a free-space wave) is a useful exercise. In the far-field region, a general assumption is that localized electromagnetic fields close to the antenna are negligible, and the electric field radiated by an arbitrary antenna in a spherical coordinate system can be expressed as

$$\vec{E}(r, \theta, \phi) = \left[ \hat{\theta} F_\theta(\theta, \phi) + \hat{\phi} F_\phi(\theta, \phi) \right] e^{-jk_0r} \frac{1}{r} \text{V/m}$$  \hspace{1cm} (2.1)

where the electric field vector is denoted by $\vec{E}$, $r$ represents the distance from the origin, $\hat{\theta}$ and $\hat{\phi}$ are unit vectors unique to the spherical coordinate system and finally, $k_0 = 2\pi/\lambda$ denotes the free-space propagation constant, where the wavelength is $\lambda = c/f$ \footnote{1}. The radiated fields that originate from the antenna described in (2.1) also consists of pattern functions unique to the azimuth ($\theta$) and polar ($\phi$) planes, denoted by $F_\theta(\theta, \phi)$ and $F_\phi(\theta, \phi)$, respectively. Furthermore, (2.1) can be qualitatively described as an electric field that propagates with a radial phase variation equal to $e^{-jk_0r}$, while the amplitude of the field reduces by a factor of $1/r$. Since the wave is transverse electromagnetic (TEM), it is possible for the electric
field to be polarized in either the $\hat{\theta}$ or $\hat{\phi}$ direction, but it cannot be polarized in the radial direction. The Poynting vector for this wave can be determined with

$$\vec{S} = \vec{E} \times \vec{H}^* \text{ W/m}^2$$  \hspace{1cm} (2.2)

where $\vec{E}$ and $\vec{H}$ represent the electric and magnetic field vectors associated with the traveling wave. The magnetic and electric fields are related to one another by

$$H_\theta = \frac{E_\phi}{\eta_0}$$  \hspace{1cm} (2.3)

$$H_\phi = \frac{E_\theta}{-\eta_0}$$  \hspace{1cm} (2.4)

where $\eta_0 = 377 \ \Omega$ is the free-space wave impedance. The far-field distance (denoted by $R$) discussed in this section is dependent on the antenna aperture as well as its wavelength, and it is defined as

$$R = \frac{2D^2}{\lambda} \text{ m}$$  \hspace{1cm} (2.5)

with $D$ denoting the maximum dimension of the antenna. It should be noted, however, that in the case of an electrically small antenna (e.g. small loops and short dipoles), (2.5) might result in a far-field distance that is too small. In this situation, a safer approximation is $R = 2\lambda$.

The total power radiated from the antenna can be determined by performing an integration operation on the Poynting vector of (2.2) over the surface of a sphere, which in this case has a radius $r$. It can be shown that this is equivalent to integrating the radiation intensity of the antenna over a sphere with a unity radius, or

$$P_{\text{rad}} = \frac{2\pi}{\phi=0} \int_{\theta=0}^{\pi} \int \vec{S}_{\text{avg}} \hat{r} r^2 \sin \theta d\theta d\phi.$$  \hspace{1cm} (2.6)

An important factor used in analyzing antenna performance is the radiation pattern. Mathematically, the radiation pattern is defined as a magnitude plot of the field strength in the far-field of the antenna versus the angle of observation from the antenna boresight. That is, the distance of the observation point remains constant throughout. The radiation pattern is typically plotted by using the pattern function, $F_\theta(\theta, \phi)$ for an azimuth pattern and $F_\phi(\theta, \phi)$ for an elevation pattern, and using one or the other depends on the polarization of the antenna in question. An example of a radiation pattern plot (often simply referred to as the antenna pattern) is shown in Fig. 2.1.

The radiated power of the antenna is shown for varying azimuth or elevation angles, and it is common practice to normalize this graph to the maximum value. Furthermore, the antenna pattern is seen to exhibit a number of distinct lobes that
peak in different directions as illustrated in Fig. 2.1. The largest of these lobes is known as the main lobe, or main beam, and the remaining ones are referred to as sidelobes. The power level of the sidelobes relative to the main beam is a key performance metric. The antenna pattern can alternatively be plotted in rectangular coordinates, and this is especially useful in pencil beam type patterns such as monopulse arrays.

2.1.1.2 Antenna Directivity

Another fundamental antenna parameter is its ability to concentrate a large portion of its power in a particular direction, and this can be expressed in two ways. First, the 3 dB beamwidth of the antenna is defined as the beamwidth, in degrees, where the power level of the main beam is 3 dB lower than the maximum. It is not required that the reference power level has to be at 3 dB, however, and some applications use different metrics to specify the beamwidth. Rayleigh beamwidth is commonly encountered in radar, and it is defined as the peak-to-null beamwidth, while some applications use the 10 dB beamwidth as a measure of directivity.

Second, the directivity of the antenna is defined as a ratio between the maximum radiation intensity $U_{\text{max}}$ relative to the average radiation intensity $U_{\text{avg}}$ over the entire angular range, and this is given by
\[ D = \frac{U_{\text{max}}}{U_{\text{avg}}} = \frac{4\pi U_{\text{max}}}{P_{\text{rad}}} = \frac{4\pi U_{\text{max}}}{\int_{\theta=0}^{\pi/2} \int_{\phi=0}^{2\pi} U(\theta, \phi) \sin \theta \, d\theta \, d\phi}. \]  

(2.7)

Directivity is dimensionless, and typically expressed on a logarithmic scale as \( D(\text{dB}) = 10 \log(D) \). The relationship in (2.7) makes physical sense when applied to an isotropic antenna, that is, one that radiates power equally in all directions \((U_{\text{max}} = 1)\). The integral identity

\[
\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \sin \theta \, d\theta \, d\phi = 4\pi
\]

may be applied to the denominator in (2.7), reducing the directivity of the isotropic element to 1 (or 0 dB). Seeing that the lowest possible directivity for any antenna is 0 dB, it is often expressed as directivity relative to an isotropic element, dBi. As stated before, beamwidth and directivity provide two methods of quantifying the focusing ability that the antenna exhibits. An antenna with a wider beamwidth will naturally have a lower directivity and vice versa. Contrary to what one might expect, there is no exact relationship between these two quantities, in spite of the intuitive relationship. The main reason for this is the fact that while the beamwidth is merely dependent on the shape and size of the principal beam, directivity is obtained by performing an integration operation on the entire radiation pattern. It is thus not uncommon to see various antennas that have a similar beamwidth, but greatly varying directivities due to the presence of multiple main beams, or perhaps differences in sidelobe levels. A practical approximation to the relationship between directivity and beamwidth for pencil beam antennas is given by

\[ D \approx \frac{32400}{\theta_1 \theta_2} \]

(2.9)

where the angles \( \theta_1 \) and \( \theta_2 \) denote the beamwidths in two orthogonal planes of the principal beam.

### 2.1.1.3 Radiation Efficiency and Antenna Gain

The radiation efficiency of an antenna is defined similarly to the efficiency of a power amplifier, and it is determined by

\[ \eta_{\text{rad}} = \frac{P_{\text{rad}}}{P_{\text{in}}} = \frac{P_{\text{in}} - P_{\text{loss}}}{P_{\text{in}}} \]

(2.10)

where the power radiated by the antenna is denoted by \( P_{\text{rad}} \), the power applied to the input port of the antenna is \( P_{\text{in}} \), and \( P_{\text{loss}} \) is the sum of losses experienced by the antenna. A large portion of losses will be resistive losses that result from non-ideal dielectrics and metals, but impedance mismatch at the input port and polarization...
mismatch between the transmit and receive antennas also contribute to loss. While resistive losses are mostly out of the control of the designer, impedance mismatch and polarization mismatch can be greatly reduced through improved design, and as a result they are not attributed to the antenna itself. The gain of an antenna is closely related to directivity, and this relationship is based on efficiency,

\[ G = \eta_{\text{rad}} D. \]  

(2.11)

An ideal antenna would thus have equal gain and directivity values as a result of having zero losses (that is, \( \eta_{\text{rad}} = 1 \)), and as such, directivity can be considered as the maximum attainable gain value of the antenna.

### 2.1.1.4 Antenna Aperture Efficiency

An aperture antenna is one in which the radiation originates from a clearly defined aperture (or area), e.g. parabolic reflectors, horn antennas, lens antennas and microstrip patches. A maximum directivity that is physically possible for an electrically large antenna with aperture area \( A \) can be determined by

\[ D_{\text{max}} = \frac{4\pi A}{\lambda^2}. \]  

(2.12)

Since a practical aperture antenna will be adversely affected by factors such as aperture blockage, spillover radiation from the feed pattern, mismatched amplitude or phase characteristics, it is necessary to define aperture efficiency. The reduced directivity is thus defined as a ratio between the actual directivity and the maximum attainable directivity, and can be written as

\[ D = \eta_{\text{ap}} \frac{4\pi A}{\lambda^2}. \]  

(2.13)

where \( \eta_{\text{ap}} \) denotes the aperture efficiency, a quantity which is less than or equal to unity in all situations. Although the definitions provided here are in terms of a transmit antenna, they are equally applicable to receive antennas.

### 2.1.2 Antenna Structures for Millimeter-Wave Systems

Millimeter-wave systems generally require antennas that are quite different in nature to those typically found in radio frequency (RF) and microwave systems. A brief classification of millimeter-wave antennas is presented here, in part for reference but also to highlight the shift in focus from non-planar antennas towards planar integrated techniques [2]. A general observation is that millimeter-wave antennas are designed through dimensionally scaling their low-frequency counterparts (lens antennas, reflectors, and so forth). Dimensional scaling is achieved with advanced micromachining and fabrication techniques such as photolithography and laser
cutting, and the tight tolerances required generally lead to an increase in cost. Another key difference is the fact that antennas that operate in the RF and microwave range consist of metallic geometries, while millimeter-wave radiators are implemented with hybrid materials and dielectric structures.

Historically, the development of substrate-integrated antennas went hand in hand with the emergence of substrate-integrated circuits. Substrate-integrated circuit technology has proven to be exceptionally useful at millimeter-wave frequencies, and alongside integrated antennas, has led to the system-on-substrate approach that facilitates effective integration of antennas and circuitry [3–5]. Some of the most popular antenna structures used at millimeter-wave frequencies are briefly discussed here.

2.1.2.1 Slot Arrays

Applications that require an antenna with a steerable, directional pattern and a reasonable gain typically utilize slot arrays. This antenna is constructed by machining slots into the sidewalls of a waveguide structure, and the slots are shaped according to a particular current manipulation, which controls the shape of the beam. Power delivered to the antenna is thus radiated from the slot openings into free space. The fact that this antenna is suited for implementation in a traditional rectangular waveguide as well as substrate-integrated waveguide (SIW) is advantageous for millimeter-wave systems [6, 7]. A conceptual drawing of an SIW periodically perturbed with rectangular slots is shown in Fig. 2.2.

2.1.2.2 Integrated Horn Antennas

Conventionally, horn antennas are highly directive, high-gain and high-power antennas. An integrated version of the conventional horn antenna is shown in Fig. 2.3. This antenna consists of a probe antenna (usually a dipole) that is suspended on a dielectric membrane in a pyramid-shaped cavity. The cavity is etched into the dielectric substrate.

SIW waveguide technology can also be utilized in the design of horn antennas. A common problem encountered in the integrated version is the mismatch between the radiating aperture and free space, and several attempts have been made to
increase the efficiency of the SIW horn antenna. However, the thickness of the substrate presents a fundamental issue, and it is expected that higher frequencies may lead to improved radiation efficiencies for a particular thickness, which is not necessarily a problem for millimeter-wave applications. In addition, this characteristic has contributed to the popularity of SIW horn antennas at submillimeter-wave and terahertz frequencies [2, 8, 9].

2.1.2.3 Conventional Printed Antennas

Printed antennas, such as the familiar microstrip patch (shown in Fig. 2.4.), generally do not meet the required performance criteria in their dimension-scaled versions when operating at millimeter-wave frequencies. This is primarily a result of extremely high conductor losses that originate from large current densities at patch edges, and this is especially prominent within the feeding network. Furthermore, patch antennas are highly susceptible to the excitation of surface waves at millimeter-wave frequencies, which results in additional losses and greatly reduces the attainable radiation efficiency. Specialized fabrication techniques are generally required in order to suppress these surface waves.

One solution to the surface wave issue is the use of an SIW cavity, where the patch elements as well as the feed network are etched on one substrate, which is stacked on top of a second substrate [10–12]. The SIW cavities are then fabricated onto the bottom substrate, which is generally much thicker to promote bandwidth enhancement.
2.1.2.4 Surface Wave and Leaky Wave Antennas

Spurious radiation from open transmission lines, which is generally a challenging aspect in patch antenna design, can be manipulated in an effective manner. Leaky wave antennas utilize the fact that radiation originates from structures where the first of the higher order modes is known to appear at high frequencies. Closed waveguides can also be used to generate leaky waves, by perturbing the aperture with tapered slots or any type of open aperture. Leaky wave antennas suitable for millimeter-wave applications are dielectric rod antennas, non-radiative dielectric guide antennas, tapered slot antennas, partially reflective surface patches and printed log-periodic dipole arrays (LPDA) or Yagi-Uda antennas [5]. A printed Yagi-Uda antenna is shown in Fig. 2.5.

Fig. 2.5 Printed Yagi-Uda antenna with a dipole antenna as the driving element
2.1.2.5 Dielectric Resonator Antennas

Dielectric resonators have been shown to be a viable option for millimeter-wave systems, mostly because they provide greater radiation efficiencies than microstrip antennas. One disadvantage of implementing large arrays of dielectric resonator antennas is the complicated nature of the fabrication process. Each element of the antenna needs to be placed and bonded according to strict positioning requirements throughout the array, e.g. one method requires inserting a lattice of holes into a dielectric substrate. An improved implementation would remove the need to locate and bond each of the antenna elements, and instead fabricate the entire array from a single sheet of dielectric material.

An example dielectric resonator antenna is shown in Fig. 2.6. Structurally, it is very similar to the conventional microstrip patch antenna, with the resonating element being replaced by a dielectric material. The resonating element can be of rectangular (as in Fig. 2.6), triangular, hemispherical or cylindrical shape, and the resonant frequency of the antenna is determined by the dimensions and dielectric constant of the element.

Dielectric resonator antennas generally yield higher radiation efficiencies as opposed to what could be obtained with microstrip antennas. One approach often encountered for millimeter-wave designs is replacing the microstrip feed with a half-mode SIW [13].

2.2 Millimeter-Wave Wireless Communication Systems

Spectrum crowding and the ever increasing requirement for faster data rates have always been motivators for using higher frequencies of operation, and significant progress in 60 GHz systems has been made in the last decade or so. Millimeter-wave technology implemented for the 60 GHz band is an exciting prospect for antenna, circuit and communication system engineers alike. With the possibility of data rates exceeding several gigabits per second (Gb/s), alongside
several other benefits, such as aggressive frequency reuse and highly integrated multi-standard systems, the vast investment both from industry and the research community can be justified [14].

2.2.1 The Friis Transmission Formula

The Friis formula is a method of solving the fundamental question on the amount of power received by an antenna in a radio system. Consider an arbitrary wireless link that transmits a signal with power $P_T$ through an antenna with gain $G_T$ to a receiver system with gain $G_R$ that intercepts a portion of the transmit power, $P_R$. This scenario is depicted in Fig. 2.7.

The power density that is radiated from an isotropic antenna at a radial distance $R$ is

$$S_{\text{avg}} = \frac{P_T}{4\pi R^2} \text{ W/m}^2. \quad (2.14)$$

For a transmit antenna with a gain that is greater than 0 dB, the radiated power density is obtained from multiplying (2.14) by the directivity, $D$. Including the radiation efficiency from (2.10) and (2.11) accounts for losses in the antenna, effectively converting the directivity parameter into gain. The result is a general expression for the radiated power density, applicable to any antenna,

$$S_{\text{avg}} = \frac{G_T P_T}{4\pi R^2} \text{ W/m}^2. \quad (2.15)$$

When the power density in (2.15) is incident on the receive antenna, it is necessary to define an effective receiving aperture $A_e$ as a proportionality constant.

Fig. 2.7 A generic radio link
that modifies the total power received. Writing an expression for \( P_R \) (which we can define as the power that is received by a conjugately matched load) in terms of the effective aperture results in

\[
P_R = A_e S_{\text{avg}}, \tag{2.16}
\]

Physically, the relationship in (2.16) makes sense, seeing that the units of \( P_R \) are \( \text{W} \), and the units of \( S_{\text{avg}} \) are \( \text{W}/\text{m}^2 \), leaving \( \text{m}^2 \) as the unit for the effective aperture. Relating the effective aperture to the directivity leads to a quantity known as the maximum effective aperture,

\[
A_e = \frac{D \lambda^2}{4\pi} \tag{2.17}
\]

which is nothing more than a manipulation of (2.12). This definition unfortunately does not account for losses in the antenna, but this can be remedied by replacing \( D \) with the antenna gain \( G \). A general observation is that antennas that are electrically large (parabolic reflectors, horns, and so forth) exhibit a value for the effective aperture that is reasonably close to the physical area, whereas electrically small antennas (such as loops and short dipoles) do not have such a simple relationship between \( A_e \) and physical area.

Now that a definition for the effective aperture has been obtained, it may be used to find the received power from (2.15),

\[
P_R = A_e S_{\text{avg}} = A_e \frac{G_T P_T}{4\pi R^2} \text{W} \tag{2.18}
\]

and by replacing (2.17) for \( A_e \), the final expression for the received power becomes

\[
P_R = \frac{G_R G_T P_T}{(4\pi R^2)^2} \text{W}. \tag{2.19}
\]

Note that instead of using \( D \) as in (2.17), the gain was used, which simply means that it represents a maximum attainable value for the received power, and as expected, there are several other factors that are going to reduce this value. The relationship in (2.19) is well known as the Friis Transmission Equation, one of the fundamental equations in radio engineering.

### 2.2.2 Link Budget

The individual terms in the Friis formula of (2.19) can be tabulated individually to formulate what is known as a link budget, a metric to determine the usability of an RF transceiver. The maximum coverage of a transceiver is primarily determined by
the total noise figure, while linearity is the most prominent contributing factor in terms of performance. The received signal power at the input of a wireless receiver is given by

\[ P_R = P_T + G_T + G_R - L_0 - L_A - L_R - L_T \text{ dBm} \] (2.20)

where \( P_T \), \( G_T \) and \( G_R \) were defined earlier in this chapter, and the \( L \) terms denote losses:

- \( L_0 \) is the free-space path loss.
- \( L_A \) is the loss due to atmospheric attenuation.
- \( L_T \) is the sum of line losses in the transmit antenna.
- \( L_R \) is the sum of line losses in the receive antenna.

Path loss can be defined as the reduction in signal power that results from free-space propagation between a transmitter and receiver, and it is given by

\[ L_0 = 20 \log \left( \frac{4\pi R}{\lambda} \right) \text{ dB}. \] (2.21)

Furthermore, (2.20) can be extended to include losses that result from impedance mismatches at either the transmit or receive antennas. The previous chapter in this text covered the concept of impedance matching in sufficient detail, and a non-zero reflection coefficient that results from an impedance mismatch (denoted by \( \Gamma \)) will introduce a loss given by

\[ L_i = -10 \log \left( 1 - |\Gamma|^2 \right) \text{ dB}. \] (2.22)

Practical communication systems are usually specified to provide a minimum acceptable quality of service. While this in itself can be expressed in several different ways (bit error rate being one), it is largely determined by a threshold power level that should be exceeded by the received signal. The ratio between the minimum threshold that translates into acceptable quality of service and the received signal is commonly expressed as the carrier-to-noise ratio, which can be interpreted as a minimum signal-to-noise (SNR) requirement. This allowance in terms of system design is known as link margin (LM), and it is given by

\[ LM = P_R - P_{R(\text{min})} \text{ dB}. \] (2.23)

A reasonable LM provides the communication system with a level of robustness against unavoidable, unpredictable effects that adversely influence system performance. Factors such as a moving mobile user, multipath and signal fading that results from weather cannot be controlled, but the robustness of the system can serve to counteract them.
2.2.3 Digital Modulation

Encoding a sinusoidal carrier signal with information is generally achieved through amplitude, frequency or phase modulation [1, 15]. An analog modulating signal results in a continuous variation in the amplitude, frequency or phase of the carrier, while a digital modulating signal restricts the carrier parameters to two distinct values. In the latter case, the techniques are known as amplitude shift keying (ASK), frequency shift keying (FSK) or phase shift keying (PSK). Digital techniques are preferred over their analog counterparts because of vastly superior performance in the presence of noise, as well as being much more suitable for error correction and encryption. An arbitrary bit stream along with the resulting modulated signals is shown in Fig. 2.8.

Other than the digital modulation schemes depicted in Fig. 2.8, numerous techniques are implemented in wireless communication systems. For example, quadrature phase shift keying (QPSK) uses two data bits to differentiate four different phase angles: 0°, 90°, 180° or 270°. Expanding on this by adding amplitude modulation to the existing phase modulation results in a technique known as quadrature amplitude modulation (QAM). QAM and other higher-order modulation schemes are commonly referred to as M-ary modulation, since $2^M$ phase states are

![Fig. 2.8 A binary sequence and the resulting digitally modulated signals](image-url)
used to encode an M-bit data signal. Higher order schemes do involve more processing complexity, but are capable of achieving higher data rates for a particular channel bandwidth.

At 60 GHz, there are a few modulation schemes that have been experimented with as communication systems and standards in this frequency band evolved; these are orthogonal frequency division multiplexing (OFDM), linear single-carrier modulation and constant-envelope modulation.

### 2.2.3.1 Orthogonal Frequency Division Multiplexing

OFDM is a modulation scheme that transmits several subcarriers in parallel, with each of these occupying a narrow bandwidth [16–18]. The properties of the channel affect the amplitude and phase of each subcarrier signal, but simple equalization serves to compensate for the gain and phase of each of the subcarriers. Generating these carrier signals consists of inverse fast Fourier transform (IFFT) operations on blocks of \( M \) symbols at the transmitter, and they are extracted at the receiver by performing FFTs on blocks comprised of \( M \) discrete samples. A block diagram of an OFDM transmitter is shown in Fig. 2.9.

The length of the FFT blocks is typically chosen to be between 4 and 10 times that of the maximum duration of the impulse response [17]. This minimizes the amount of excessive overhead introduced by the cyclic prefix that is required at the start of each block, and this prefix is discarded at the receiver end. The cyclic prefix serves two purposes: to prevent degradation of a particular block as a result of inter-symbol interference (ISI) from an adjacent block, and to induce a periodicity in the received block of data, where the period would be equal to \( M \). Furthermore, the introduced periodicity emulates the presence of circular convolution, a property that is essential to the effective functioning of the FFT algorithm [15].

The primary disadvantage of OFDM is the amplitude fluctuations that are often erratic and result in high peak-to-average power ratios (PAPRs). This makes OFDM particularly sensitive to nonlinear distortion that is caused by the power amplifier in the transmitter chain. Without a sufficiently high level of power backoff, the system would be subject to several adverse effects, e.g. intermodulation distortion and

![Fig. 2.9 Block diagram of an OFDM transceiver](image-url)
spectral broadening, resulting in a degradation of the overall performance. Simply increasing the backoff power is one method of addressing this problem, but this in turn reduces the efficiency of the amplifier, which is of particular concern in mobile systems where power is drawn from a battery. Other challenges associated with OFDM are the sensitivity to Doppler mismatch and absence of multipath diversity in uncoded OFDM [19].

### 2.2.3.2 Constant Envelope Modulation

In a constant envelope modulation (CEM) scheme, just as the name implies, the information content is restricted entirely to the phase of the transmitted signal. In terms of power efficiency, CEM is an ideal option, seeing that the baseband signal is not affected by nonlinear distortion [20]. As a result, CEM signals are able to operate in the saturation region of the transmitting power amplifier. A popular variation of CEM is continuous phase modulation (CPM), a scheme in which the phase is selected as part of a continuous time signal, increasing bandwidth efficiency.

As with all modulation schemes, CEM and CPM do suffer from their own unique disadvantages. For reasonably high SNR values, CEM and CPM systems are able to achieve a comparatively lower throughput. Moreover, since CPM signals require differential encoding at the transmitter, the equalization process at the receiver can end up being overly complex. One method to reduce this complexity is to implement the equalization in the frequency domain, although this is not sufficient to alleviate the rate at which the complexity increases once the constellation size is increased.

### 2.2.3.3 Single-Carrier Modulation Schemes

Single-carrier (SC) modulation has traditionally been the format of choice in digital communications since the earliest days of wireless systems [21]. It can be considered a predecessor to multi-carrier schemes (such as OFDM) used in several applications today. A block diagram of such a transceiver is shown in Fig. 2.10.
SC systems that are implemented with linear modulation techniques (e.g. QAM) provide a reasonable compromise between power and spectral efficiency, and the complexity of transmitter and receiver hardware. SC systems are also capable of achieving better PAPR values in comparison to OFDM [20, 22]. From the discussions presented here, it can be concluded that there is no modulation scheme that will perform optimally in all scenarios, and it is therefore a primary requirement of standards boards to ensure that future systems are tailored to the strengths of each particular approach.

### 2.2.4 Wireless Communication Standards

Innovation in wireless communication systems is often stimulated by the allocation of new portions of the electromagnetic spectrum that support new systems and applications. For example, the Federal Communications Commission allocated the industrial, scientific and medical (ISM) band in the mid-1980s, permitting spread-spectrum and unlicensed use of the 900 MHz, 2.4 GHz and 5.7 GHz bands, a change that essentially made the widespread use of wireless local area networks (WLANs) and Wi-Fi possible in the first place. The hardware required to implement wireless systems in the 1–5 GHz range was excessively expensive at the time. An agreement between spectrum authorities across the world on the allocation of the ISM band meant that a very large market need was created for wireless systems that operate in this frequency range. Moreover, the evolvement of semiconductor processes provided the technology required for low-power wireless modules that could be mass-produced at low cost. Without these two changes in industry, the wireless networks that everyone is familiar with today might have turned out quite differently.

Since the 1980s, wireless communication systems have almost exclusively been designed to operate somewhere between 800 MHz and 5.8 GHz, while the technology required to implement wireless systems has advanced significantly. As a result, the traditional paradigm of wireless networks has gradually been shifting in the past decade towards predominantly broadband systems operating in 60 GHz bands and above [23–25]. The performance demonstrated by several 60 GHz transceiver modules (developed both industrially and by dedicated research facilities) has been proven to be compliant with the IEEE 802.11ad standard, the first wireless standard published by the IEEE that is related to the 60 GHz band [24, 25]. At present it seems likely that the market will see a surge in products operating at 60 GHz, similar to the rapidly increasing widespread use of Wi-Fi devices and cellular networks in the 1–5 GHz bands observed in the early 2000s. Furthermore, it is likely that the continuous reduction of transistor gate lengths will lead to inexpensive implementation of sub-terahertz wireless systems in the not-so-distant future.
Communication standards authorities worldwide have already agreed to the unlicensed use of the 60 GHz band for wireless personal area networks, and some of these allocations are shown in Fig. 2.11.

Considering the fact that currently implemented fourth-generation (4G) mobile systems provide users with around 50 Mb/s bandwidth each, as well as the widespread adaptation of internet-equipped smartphones, it makes sense that carrier companies are pushing for sub-terahertz spectrum use to become a reality. Agreement on a wireless standard and a spectrum allocation are the first two steps towards accomplishing this goal.

### 2.2.5 Millimeter-Wave Cellular Networks

A typical mobile network will consist of a geographically arranged series of base stations. The base stations are positioned so as to maximize coverage and improve service quality. Millimeter-wave operation of cellular networks offers numerous potential benefits, but their successful implementation does pose significant challenges. In order to realize the true potential of these systems, the feasibility of such networks should be thoroughly understood for system designers to be able to combat the associated challenges. For example, the high omnidirectional path loss that is often encountered in millimeter-wave radios can be accounted for by implementing beamforming networks at the transmit and receive chains. While this reduces the path loss significantly, beamforming systems are still vulnerable to shadowing, resulting in inconsistencies in the quality of the channel [26, 27].

Conventionally, millimeter-wave technology in communication systems has been reserved for satellite communications and cellular backhaul systems [28–30]. Nowadays, the 60 GHz band is being used for high data rate applications in wireless networks and personal area networks [3, 22]. These networks generally operate at very short ranges, but are capable of data rates exceeding 1 Gb/s [25]. However, utilizing millimeter-wave bands for non-line-of-sight mobile communications seems to be the logical next step, but whether or not such systems are feasible has been the topic of considerable debate. While we are all familiar with the
vast bandwidths available in the millimeter-wave spectrum, the propagation of
signals in this band is much less favorable in comparison with current cellular
bands.

Two trends in millimeter-wave mobile systems have sparked reconsideration of
the viability of such systems [26]. The first of these is the advances made in RF CMOS
technology and digital signal processing capabilities, which have been a huge enabler
in millimeter-wave systems by providing small, low-cost integrated circuits suitable
for mobile devices in the commercial sector. Power amplifiers and array-combining
techniques have made significant progress, and because of the minute wavelengths,
large antenna arrays can be fabricated in an area of less than 2 cm². This allows
multiple arrays to be placed in a single device, providing path diversity.

Secondly, cellular networks have been evolving in recent years towards smaller
cells, with support for so-called femtocell and picocell networks that are integral
parts of emerging wireless standards [31–34]. In many densely populated urban
areas, cells are often 100 m or less in radius, which could place them within range
of millimeter-wave signals. Increasing the capacity of existing cells (in terms of
users per cell) would be a necessity in the absence of newly allocated spectrum, but
it might be a costly affair due to setup and rollout costs. A common (and reasonably
conservative) estimate is that backhaul infrastructure forms about 30%–50% of the
operating costs involved in cellular networks [35]. In high-density areas,
millimeter-wave systems may aid cost reduction as a result of their wide band-
widths, which provide an alternative and effective method to increase cell capacity.

2.2.6 Wireless Communication Algorithms

2.2.6.1 Multiple Input, Multiple Output

Multiple input, multiple output (MIMO) antenna systems have been one of the
default proposed solutions to achieving higher data rates and increasing the capacity
of a communication system.

Figure 2.12 shows a generic MIMO configuration.
The potential of multiple antenna systems has resulted in proposed implementations in a number of application areas, from wireless broadband links in future cellular networks to short-range WLANs and local data service overlays [14]. A crucial attribute required in multiple antenna systems is the ability to provide efficient and reliable channel state information (CSI), mostly because it has a significant effect on the ability of the system to realize the capacity gains that result from operating several antennas in a MIMO configuration. The sheer number of parameters that have to be estimated in order to produce useful CSI complicate the process in MIMO channels, when compared to single-antenna systems. The complications are exacerbated by the data rates required, since they are usually much higher than in traditional wireless links.

The performance limits of a MIMO system with and without reliable CSI has been characterized [36, 37], and there have been numerous attempts to relate these fundamental limitations to the parameters of CSI estimation algorithms [38, 39]. Some factors that result in loss of performance are feedback delays, channel correlations and inefficient power allocation. In addition to the actual availability and reliability of CSI, corrections need to be performed quickly and frequently in higher data rate applications.

Expanding on the concept of MIMO, several proposals for so-called massive MIMO systems have been developed in recent years [40–42]. A typical MIMO configuration would employ at most about 10 antennas, which would result in modest spectral efficiency improvements. Massive MIMO is an ambitious attempt at further improving spectral efficiency, and aims to do so by increasing the number of antennas by orders of magnitude (around 100, perhaps even more). Extrapolating the improvements offered by MIMO into the $N_t \rightarrow \infty$ range (where $N_t$ denotes the total number of antennas) has led to some asymptotic arguments, establishing that the adverse effects from uncorrelated noise disappear, throughput becomes independent of cell size, spectral efficiency becomes independent of bandwidth and the transmitted energy required per bit tends to zero [43].

A system utilizing a large number of millimeter-wave antennas would occupy a much larger footprint, in comparison to such an extension (in terms of $N_t$) being applied in current systems and frequencies. Moreover, the increase in available bandwidth would also greatly benefit a multi-antenna system. The high path loss that results from atmospheric absorption on top of rain and foliage attenuation becomes less of an issue when cell radii are in the order of 50–200 m. This means that signal attenuation in a worst-case scenario would only be a few dB, and not completely impair system functionality [42].

### 2.2.6.2 Cooperative Communications

Modern radios are capable of highly efficient power, time and bandwidth management in order to be able to share the spectrum with improved efficiency. Many cooperative algorithms can be implemented independently, and it is even possible to employ them competitively, where each device attempts to maximize its own
performance [14]. Alternatively, cooperation between devices (or terminals) can be regarded as another method of realizing system diversity. Cooperative communications can include techniques such as cooperative coding, collaborative signal processing algorithms, forwarding as well as relaying [44, 45]. Perhaps it is unsurprising that the cooperative approach has the potential to lead to higher performance gains, compared to independent and competitive approaches. A major downside that would potentially hinder the implementation of cooperative techniques is the significant cost, seeing that end users might not be willing to bear larger initial costs in order to receive speculative benefits.

### 2.2.6.3 Dynamic Spectrum Access

Since the earliest days of wireless communication, spectrum sharing has been a crucial part of effective regulation, realized by tighter channels and more efficient frequency reuse. Evolving technologies such as WLAN and next-generation communications are generally limited to contiguous channel allocations ranging up to a maximum in the tens of megahertz range. Such technologies do not provide the flexibility that will allow dynamic adaptation of spectrum, whereas cognitive radio does enable dynamic spectrum access. Cognitive radios recognize the available systems and accordingly adjust their operating frequency, protocols and waveforms to access said systems efficiently. This is a description of perhaps the simplest form of cognitive radio, but one such radio is by no means simple to implement effectively. In theory, cognitive radios are intended to extend the software defined radio (SDR) framework further, allowing it to include model-based reasoning and additional domains of knowledge, as well as negotiation. Knowledge and reasoning include aspects of wireless systems, such as protocols, propagation models and typical spectrum allocations. Negotiation refers to the agreement among industry peers regarding spectrum, waveforms and protocols, with the goal of generating standards that support cognitive radios.

### 2.3 Millimeter-Wave Radar

The concept of radar was introduced some time before World War II, but did not enjoy any real development until the war, when the ability to detect hostile aircraft at distances beyond line of sight was desperately needed. Since then, radar systems have become staggeringly sophisticated, and a large portion of this growth can be attributed to the size of budgets allocated to defense programs. Worldwide defense expenditure in 2014 amounted to a total of US$ 1775.6 billion, broken down into percentage contributions in Fig. 2.13.

While the graphic in Fig. 2.13 encompasses all defense-related spending, including training, manufacturing, research and deployment of any type of defense system and the related personnel, a sizeable portion of budgets is being allocated to
radar and electronic warfare (EW). A recent trend analysis in the EW market has projected that global expenditure on EW-related systems will grow to over US $18.7 billion in the next decade [46].

The rapid and continued growth of EW and radar is largely a result of the availability of these technologies and a growing number of global entities have access to such technologies. For example, remote-controlled improvised explosive devices have been a significant threat to human life in areas across the globe, hence the continued investment in systems that limit the functioning of such devices.

It should be noted that radar is often regarded as a separate entity in the EW domain, but a significant portion of EW technology is directly related to radar. The remainder of this chapter will not discuss EW and the relevance of millimeter-wave technology in EW systems, and instead focus on radar. The interested reader is referred to several high-quality EW texts [47–50].

Modern radar systems consist of sophisticated transducer and computer systems that are capable of tracking, identifying and classifying targets in the presence of adverse effects such as clutter and jamming signals, and have grown tremendously since the rudimentary detection systems used in World War II.

As the vehicle and aeronautic industries have illustrated in the last few decades, radar technology is extremely useful in civil applications as well, on which most of this discussion will be focusing. Development of millimeter-wave radars was initially focused on military applications, and the cost of systems and components in the 1970s meant that exploring civil applications would be infeasible [51]. Several important developments led to mainstream implementation of millimeter-wave
systems. Some of these developments are transistors with cutoff frequencies that exceed 100 GHz, automatic assembly of planar circuits, reliable and low-cost monolithic microwave integrated circuits for millimeter-wave and multilayer, multifunctional circuits.

One system involved the use of a 77 GHz radar as a guidance sensor for autonomous navigation in land vehicles, using an extended Kalman filter to optimally combine range and angle measurements with a vehicle control system [52]. An improved version of this system was later developed that enabled identification of natural features in the environment by using the reflected polarization [53]. The use of synthetic aperture radar (SAR) has been investigated for aircraft landing assistance systems [54]. The so-called synthetic vision system consisted of a scanning radar operating at 35 GHz, a heads-up display (HUD) and real-time digital signal processing hardware. The use of millimeter-wave radars in airport surface surveillance systems has also been explored [55].

### 2.3.1 Radar Fundamentals

A radar system is a collection of electronic and digital sub-systems that transmits modulated RF signals into a particular region of space, in the hope of detecting the presence of a desired target [56–58]. Figure 2.14 shows a basic radar system.

While the details of each radar and its functioning varies, the transmitter, antenna, receiver and signal processor are all common between different types of radars. The radar transmitter is tasked with modulating a pulsed or continuous waveform (which defines two distinct classes of radar), amplifying it, whereafter the antenna directs the signal into the region of interest. Radar signals reflect off

![Fig. 2.14 Block diagram of a radar system](image-url)
most obstacles in the environment (generally referred to as scatterers) and it is inevitable that at least some of the reflected energy will be intercepted by the antenna.

While the radar, depicted in Fig. 2.14, uses a single antenna and therefore switches between transmit and receive modes using a T/R switch, it is not uncommon to use separate antennas for transmit and receive operations in certain applications. The received signal—which has been modulated by the properties of the target and the environment—is mixed down to an intermediate frequency (IF) and passed through a detector, which removes the carrier signal and allows the signal to be passed to the signal processor. Not all radars employ digital signal processing, but it is commonplace in modern systems, and the signal processor performs a variety of tasks, some of which are discussed later in this section.

A fundamental part of radar theory is the radar range equation, which quantifies the power received from a target. Consider a situation in which a target object is illuminated by the signal from a radar antenna. The incident wave will be scattered in several directions, but it is generally accepted that at least some of the energy is reflected back into the coverage area of the radar antenna. This is because the illumination induces time-varying currents on the surface of the target, in such a manner that it becomes a source of EM waves. Some of these waves, which appear as reflections of the illuminating signal, will propagate back in the direction of the radar [57]. The actual power that will be reflected from the target is a function of both the power contained in the incident signal and the radar cross-section (RCS, measured in m² or dBsm) of the target. An expression can be formulated to determine the reflected power from a target (denoted by $P_{\text{REFL}}$), and it is given by

$$P_{\text{REFL}} = \frac{G_T P_T}{4\pi R^2} \sigma = S_{\text{avg},i} \sigma W.$$ \hspace{1cm} (2.24)

Note that this equation has a similar form to (2.15), with the addition of the RCS (denoted by $\sigma$). Furthermore, the distinction between incident and reflected power density is indicated by a change in subscript: $S_{\text{avg},r}$ and $S_{\text{avg},i}$ denote the reflected and incident values, respectively.

Rearranging (2.24) to isolate the power density that is reflected back to the radar results in

$$S_{\text{avg},r} = \frac{P_{\text{REFL}}}{4\pi R^2} W.$$ \hspace{1cm} (2.25)

Substituting in (2.24) for $P_{\text{REFL}}$ completes the expression for the received power density,

$$S_{\text{avg},r} = \frac{G_T P_T \sigma}{(4\pi)^2 R^4} W.$$ \hspace{1cm} (2.26)
The range term in the denominator of (2.26) is raised to the fourth power, the reason for this being the two-way propagation path between the radar antenna and the target (recall that one-way propagation diminishes power by a factor of $1/R^2$). This is a significant decrease, considering that doubling the target range reduces the received power density by a factor of 12 dB [57]. The amount of reflected power that is actually intercepted by the radar antenna depends on the aperture of the antenna. Including this multiplying factor in (2.31) leads to

$$P_R = A_eS_{avg,r} = \frac{G_TP_T\sigma}{(4\pi)^2R^4}W. \quad (2.27)$$

Solving for the effective aperture in (2.13) and inserting the result into (2.32) results in the following expression for the received power:

$$P_R = \frac{G_RG_T\lambda^2\sigma}{(4\pi)^2R^4}W. \quad (2.28)$$

This equation is widely known as the radar range equation, and it is found in this form in most standard radar texts [56–58]. The variables in (2.35) are summarized below for convenience,

- $G_R$ is the gain of the receiving antenna.
- $G_T$ is the gain of the transmitting antenna.
- $P_T$ is the peak value of the transmitted power, in Watts.
- $\lambda$ is the system wavelength, in meters.
- $\sigma$ is the mean of the target’s RCS, in square meters.
- $R$ is the distance between the radar and the target, in meters.

The gain values in (2.35) are supplied in linear units, and a conversion to a logarithmic scale for the gain, power and RCS values would result in the decibel value of the received power.

### 2.3.1.1 Radar Measurements

Modern radars are capable of measuring several parameters related to a target simultaneously. These include:

- Range, $R$.
- Range rate, $\dot{R}$.
- Azimuth angle, $\theta$.
- Elevation angle, $\phi$.
- Polarization.

These parameters are discussed with reference to Fig. 2.15.
The range of the target is determined by the round-trip delay $\Delta T$ of the EM wave, from when the signal leaves the transmitter to the sampling window where the target is declared. The range is computed as

$$R = \frac{c \Delta T}{2}$$

where $c$ denotes the propagation speed in free space, and the approximate value of $3 \times 10^8$ m/s is commonly used.

Relative motion between the radar system and the target will create a difference in frequency between the transmitted and received signals. This phenomenon is known as the Doppler effect. Measuring this frequency difference is the first step in determining the radial velocity of the target. A common approach to measuring Doppler shift in a pulsed radar is to perform spectral analysis on samples of the received signal at subsequent ranges. In the case where the received signal at a particular range only consists of a single sample, spectral analysis would be pointless, and several pulses are consequently transmitted at specific time intervals. The receiver is designed in such a manner that each of the transmitted pulses would correspond to a set of received samples, starting from the minimum range up until the maximum range. Therefore, these sets of received samples allow the signal processor to form a matrix of samples, where one dimension is range and the second dimension is cross-range. The cross-range dimension is then used in the spectral analysis procedure, seeing that in one set of samples in the cross-range dimension all correspond to the same range.

Performing $N$-point discrete Fourier transforms (DFTs) on the cross-range signal results in a distinct frequency domain signal at each range interval that reveals information about the Doppler shift of the target. This shift in frequency ($\Delta f$) can then be converted to a velocity difference ($\Delta v$), using
\[ \Delta f = f_c \frac{\Delta v}{c} \]  \hspace{1cm} (2.30)
Note that the diagram in Fig. 2.17 only covers a single dimension, and an additional receiver channel is required if the system needs to determine an angular error in both the azimuth and elevation planes. In order to determine the angular offset of the incident signal, a monopulse radar uses an additional receiver channel. The two channels are dedicated to the sum (\(\Sigma\)) and difference (\(\Delta\)) signals, and these are derived from the output ports of a 180° hybrid coupler. The angle error is then derived from the monopulse ratio, which is given by

\[
r = \frac{\Sigma}{\Delta} \cos \theta.
\]  

(2.31)

Aside from providing a highly reliable, single-pulse angle measurement, monopulse radars are also extremely resistant to jamming, which is the main reason why this technique has triumphed over older angle estimation techniques [59–62].

The last radar measurement that will be discussed here is polarization. Polarization is intertwined with the vector nature of the EM waves that are received and transmitted by the radar antenna. A target consists of multiple scattering elements, seeing the geometry of most real-world targets are fairly complex, and the RCS varies with both the system wavelength and the exact viewing angle. EM signals that are reflected from the surface of a complex target are subject to alterations in their polarization, and this change in polarization can be used to determine certain properties regarding the geometry of the target. Furthermore, this information can be useful in distinguishing between return signals from wanted targets and clutter (e.g. from rainfall), and may even be used to identify various targets of interest [57, 63–65].

The most complete measurement of polarization is the process of obtaining the polarization scattering matrix of a particular target, and this matrix is described as

\[
\bar{S} = \begin{bmatrix}
\sqrt{\sigma_{11}} e^{i\phi_{11}} & \sqrt{\sigma_{12}} e^{i\phi_{12}} \\
\sqrt{\sigma_{21}} e^{i\phi_{21}} & \sqrt{\sigma_{22}} e^{i\phi_{22}}
\end{bmatrix}.
\]  

(2.32)
As (2.32) shows, each of the terms consists of an amplitude and phase angle. Moreover, the polarizations 1 and 2 are orthogonal, which means that when polarization 1 is left-hand circular, polarization 2 will always be right-hand circular, and so forth. Completing this polarization measurement requires the radar to have a dual-polarized receiver (where the two channels are orthogonal), on top of having a polarization-agile transmitter. The measurement begins by transmitting a wave with a known polarization, and the reflected signal is captured by the receiver. The receiver measures the amplitude and relative phase of the signals in the two orthogonal channels. Thereafter, the transmitter emits a wave with a polarization that is orthogonal to the original wave (i.e. if the original wave was horizontally polarized, this one will be vertical), and the receiver performs the same measurement. Ideally, the two transmitted waves should be pulsed out at the same time, but this requires a much more complex architecture and the general approach is to use two closely spaced pulses with different polarizations. The limiting factors are thus the width of each pulse and the time it takes for the system to switch transmit polarizations.

### 2.3.1.2 Radar Functions

Regardless of the variation in radar systems that is being produced for very different applications and environments, all functions can be roughly divided into three categories: search, track and image. Almost all radars have a search function, but tracking and imaging operations are limited to more specialized systems. Search radars are tasked with continuously scanning a large area in space by changing the direction in which the antenna boresight is fixated. Mechanically steered antennas will cycle through a number of beam positions, while phased arrays will electronically switch beam positions. Phased arrays allow for faster (and generally, more precise) beam cycling, but their scanning angle is limited to an approximate total of $60^\circ$ [66, 67]. For this reason, a compromise approach is generally followed in modern radars, where a mechanical steering mechanism will augment the phased array antenna, extending the angular coverage. However, this does add significant complexity to the system. It is often the case that a search radar will provide basic information to a second radar, which is generally a tracking radar. In such a case, it is the task of the search radar to narrow down a region of interest inside the search volume and forward these data to the tracking radar, which will then continue to track priority targets.

The tracking process begins with the information relayed from the search radar relating to the state of the target—its position, relative angles in both elevation and azimuth planes, and radial velocity. The tracking radar then continues to measure the state of a target over time, and multiple measurements are often combined to improve the state estimation. This scenario is depicted in Fig. 2.18.

Figure 2.18 illustrates that, once the radar has determined that the target is still being tracked, a new state measurement is performed to determine where the radar should “look” next. Fundamentally, tracking implies that the state of the target is
measured with a level of accuracy that is better than the resolution of the radar itself, which can be achieved in a number of ways. For example, the monopulse technique discussed earlier can be used to obtain angular error signals that exhibit much greater resolution than the beamwidth of the individual antennas. The contaminating effect that noise and interference exert on the state measurement has led to the development of track-filtering algorithms, enabling the system to obtain much more reliable track data. More advanced systems use multiple variations of the Kalman filter as part of their tracking algorithms.

The last major function of radar that will be discussed here is imaging. In radar terms, imaging is a general term that can refer to one of multiple methods used to gather detailed information on a scene or a discrete target [56–58]. The process of generating a radar image of a scene comprises two steps: creating a range profile of the target scene in high resolution and repeating the measurement for the

---

**Fig. 2.18** Conceptual functioning of a tracking radar. The dotted line represents discrete measurement intervals.

**Fig. 2.19** SAR image of the Teide volcano on the island of Tenerife. Image courtesy of the National Aeronautics and Space Administration [83]
cross-range (or angular) dimension. In an SAR, a two-dimensional image of the terrain is developed with resolutions that are in the order of 100 m all the way down to 1 m in more sophisticated systems (Fig. 2.19).

2.3.1.3 Radar Resolution

Radar resolution is usually separately defined for each of the radar measurements: range, velocity and angle. While the quality of one of the aforementioned measurements is expressed in terms of its precision and accuracy, resolution refers to the upper limits of radar performance in a particular scenario. Range resolution is defined as the minimum spatial separation between two targets that is required for the radar to distinguish them, and it is largely dependent on the bandwidth occupied by the radar waveform (denoted by $B$). Range resolution is given by

$$\Delta R = \frac{c}{2B}. \quad (2.33)$$

Angular resolution is determined by the specifications of the antenna, particularly the beamwidth (in terms of whichever beamwidth specification the design conforms to). In the Doppler spectrum, resolution is slightly more involved. Earlier in this section, it was mentioned that radars often use multiple transmit pulses to measure Doppler shift. The theoretical resolution in the Doppler domain is dependent not only on the number of pulses used in the transmit waveform (denoted by $K$) but also on the temporal separation of these pulses (known as the pulse repetition interval, denoted by PRI). It can be explained by considering the Doppler spectrum as being limited at its lower and upper boundaries by the bandwidth $1/PRI$, which is known as the pulse repetition frequency (PRF).

It is well known that the process of sampling a discrete signal will cause the resulting spectrum to repeat itself at intervals equal to the sampling frequency [68]. Consider the Doppler spectrum as a continuous entity, which is sampled by each pulse in the transmit waveform (therefore, at a sampling rate equal to the PRF). Similar to the time domain case, sampling the Doppler spectrum will also cause the time domain signal to repeat itself, at intervals equal to the PRI [57, 68]. Signals with a Doppler shift that fall inside this spectrum can be reliably measured by the radar, and any Doppler shifts that fall outside this spectrum would alias to a lower frequency. Therefore, if the PRF represents the spectral coverage, and the sampling points are the individual pulses, the Doppler resolution in a pulsed radar can be written as

$$\Delta f_d = \frac{\text{PRF}}{K}. \quad (2.34)$$
2.3.2 Automotive Radar

In recent years, the market for driver assistance systems that rely on millimeter wave radar sensors has grown rapidly. Any new generation of cars introduced in the last three decades has promised to improve driving safety and make the experience more convenient for the driver. The functionality that we now refer to as driver assistance systems is tasked with relieving drivers from the combination of split-second decision-making in complex scenarios and monotonic tasks such as basic driving [69]. Driver assistance systems can be categorized into four groups: active comfort (e.g. adaptive cruise control), passive comfort (e.g. parking assistance), active safety (e.g. automatic braking) and passive safety (e.g. airbags). While passive systems only react to certain scenarios and are not able to influence the vehicle motion, active systems can indeed influence vehicle dynamics such as braking and accelerating. In order to realize these concepts and implement them in a practical system, a range of different radar sensors is required to make the vehicle aware of its surroundings.

2.3.2.1 Frequency Regulation

Given that radar systems can measure radial distance, velocity and in some systems, angle, they are a key component in driver assistance systems. Furthermore, radars are generally quite robust in adverse weather conditions and poor lighting, making them all the more attractive for these systems. Systems that use millimeter-wave radar technology first appeared in the 1970s [69]. The International Telecommunications Union’s decision at the 1979 World Administrative Radio Conference to support sensor applications in millimeter-wave bands above 40 GHz has been one of the key driving factors [70].

Clutter returns from ground reflections, buildings, adjacent vehicles and guard rails constitute a large portion of the return signal captured by an automotive radar sensor. While sophisticated signal processing can serve to reduce the effect of the clutter response, or separate desired targets from clutter, using higher frequencies can also be beneficial. The ability of an antenna to produce extremely narrow beams at millimeter-wavelengths can provide effective spatial filtering against background clutter and greatly increase the angular discrimination ability of the radar. Moreover, since automotive radars operate at very short ranges, it would seem that operating in millimeter-wave bands would yield optimal performance.

Two frequency bands primarily used in automotive radar are centered at 77 and 24 GHz. Frequencies outside of this range, i.e. below 10 GHz and above 100 GHz, have been explored, but with little practical relevance. The 77 GHz band offers greater possibilities to implement high-performance sensors, but it involves greater difficulty in system design from an engineering perspective. Nonetheless, there are
several factors that motivate the continued use of this frequency band. The size of a radar sensor is determined by the antenna aperture. Operating the radar at 77 GHz allows for a physically small antenna capable of achieving a very narrow beam-width. Using the Rayleigh criterion to determine angular resolution, one system demonstrated that for the same antenna aperture, a resolution of 5.4° is obtainable at 77 GHz, while only 17.5° is possible at 24 GHz [69]. Reducing the size of the sensor also aids integration challenges, since it allows the design to contend with greater size and weight constraints. Significant improvements in fabrication technology in the last two decades have brought the manufacturing costs of 24 and 77 GHz systems relatively close to one another.

Improving range resolution almost invariably requires larger bandwidth wave-forms, and in terms of percent-bandwidth, the 77 GHz system is the better option. Furthermore, emission regulations do not permit high-power (greater than −40 dBm), high-bandwidth (larger than 250–300 MHz) radiation at 24 GHz, while this combination is not restricted at 77 GHz. The only concern for sensor operation in the 77 GHz band has been focused on behind-bumper integration, where high-permittivity metallic based paints could cause significant reflections at the radome. Pfeiffer and Biebl have suggested a narrowband solution using inductive strips and demonstrated the effectiveness of this technique [71].

Mercedes-Benz was the first automobile manufacturer to introduce radar-based autonomous cruise control (ACC) in 1999 [72]. Since then, 77 GHz radar sensors have been used in collision mitigation and pre-crash sensing systems, among others. Newer generations of sensors will improve field of view, range and angular resolution, as well as minimum and maximum range. Nonetheless, aside from technical advancements, frequency regulation will continue to play a pivotal role in the development of millimeter-wave radars. Difficulties arise from the fact that countries generally have their own regulations regarding spectrum usage, and in rare cases these allocations conflict with those of neighboring countries. In the millimeter-wave region, the two major allocations are the 76–77 GHz and 77–81 GHz bands, the latter having replaced ultra-wideband automotive radars in the 24 GHz band in Europe [69]. Moreover, the 76–77 GHz band permits higher maximum transmitter power levels (55 dBm EIRP), while the 77–81 GHz band provides increased bandwidth and lower permitted power levels [73].

The maximum allowable power spectral density in the latter band is specified as −3 dBm/MHz, along with a peak limit of 55 dBm EIRP [73]. In addition, the mean power density outside a vehicle is limited to −9 dBm/MHz, which accounts for the attenuation that results from installing a sensor behind a painted bumper. In the US, emission limits vary based on several vehicle parameters, but it is expected that the regulations will soon adapt to similar values to those used in Europe.

---

1Effective isotropically radiated power.
2.3.2.2 Classification of Automotive Radar Systems

There are three principal groups of automotive radar, distinguished by the range at which they are intended to measure target parameters:

- **Short-range Radar**—sensing in close proximity of the vehicle, such as parking assistance and obstacle detection systems.
- **Medium-range Radar**—sensing at medium distance and with an average speed profile, such as cross-traffic alerting systems.
- **Long-range Radar**—sensing at longer distance where a narrow antenna beam is required in a forward-looking direction, such as an ACC system.

As discussed earlier, resolution is an indication of the ability of the radar to distinguish between targets, whether in range, angle, or Doppler shift (radial velocity). On top of the achievable resolution, there is still some level of uncertainty about the accuracy of the measurement, which is generally much smaller than the resolution itself. A key challenge in the design of automotive radar systems is the separation of closely spaced targets with vastly different RCS values (such as a bus and a motorbike), that may be traveling at the same distance and at the same velocity relative to the radar. This separation can be achieved by designing a system with high dynamic range and small resolution in any one of the measured parameters.

2.3.3 Military Radar

The fact that radars were originally developed for military use has resulted in a significant presence of radar systems of all types in modern combat vehicles and stationary platforms. Versatile radars are quite common in more sophisticated systems, e.g. search and track radars that are capable of maintaining tracking data on several targets simultaneously, which are implemented on a single platform [74]. Fighter aircraft are often equipped with so-called integrated defense systems, which combine several radar functions on top of an array of self-protection jamming systems. The recruitment of millimeter-wave sensors for a variety of tasks in land vehicles (particularly those operating in a combat environment) has been quite limited in comparison [75]. An example system integrates an active protection (AP) radar, a surveillance radar for both ground-based and airborne targets, a high data rate ad hoc communications network and a combat identification unit [75].

First, the AP radar system has three functions: detecting the muzzle flash from an anti-tank weapon using an optical sensor, tracking the threat projectile and firing an interceptor projectile to destroy the incoming missile. Anti-tank weapons most commonly encountered are rocket-propelled grenades (RPGs) fired from handheld weapons, and of course kinetic energy rounds that are fired from main battle tanks. Aside from increasing the thickness of external armor attached to a combat vehicle,
there are few countermeasures against these types of weapons. Moreover, additional armor plating cannot be added indefinitely, since it becomes excessively expensive and interferes with the vehicle’s mobility. Moreover, it is not possible to cover every area of the vehicle with armor. Tank tracks are a prime example, as these are among the weakest spots on the vehicle exterior, which make them primary targets for shoulder-launched RPGs. The small apertures required in millimeter-wave bands means that it is advantageous to operate AP systems at these frequencies.

Second, surveillance radars are a staple part of most defense platforms, since they provide a reasonably detailed image of the surroundings that stretch beyond line of sight. As mentioned earlier, these radars often feed information to secondary systems, and additional radar coverage might be needed to detect vertically incident projectiles (such as those fired by the FGM-148 Javelin man-portable anti-tank weapon).

Third, utilizing the increased bandwidth and data rate offered by millimeter-wave communication networks could be an important step towards replacing legacy wireless hardware (which operates almost exclusively below 2 GHz). High-resolution images and similar large files need to be relayed to several mobile platforms in as little time as possible, and the highly limited data rates available in legacy systems do not meet this requirement.

Lastly, combat identification functions were first implemented in an attempt to reduce friendly fire incidents. Coded signals are exchanged between friendly vehicles before a firing sequence can commence, and they act as a handshaking mechanism between friendly entities. It is imperative that the identification systems are extremely resistant to jamming signals, since it would be possible for an enemy vehicle to evade fire by reproducing coded signals that would normally originate from a friendly vehicle.

2.4 Imaging

Security concerns at heavily trafficked public places—train stations, airports, etc.—are constantly required to adjust to new threats and thus require new techniques to ensure the safety of the public. Technologies such as X-ray imaging to inspect hand luggage and metal detectors at airports have been used for some time to great effect; however, there are several shortcomings that need to be addressed to enhance the effectiveness of these systems [76, 77]. Metal detectors can, for example, only actually detect metallic objects, and the probability of detection is often dependent on factors that are not directly controlled through the design of the system. Another shortcoming is that metal detectors have no ability to distinguish between harmful items and harmless, everyday luggage contents such as glasses, belt buckles, and so forth.

Newly developed systems that provide high resolution imaging are one step towards a solution of this problem. Even though X-ray imaging systems have proven to be extremely effective, acceptance and widespread implementation of
such systems might be slowed down by the perceived adverse health effects. Contrary to X-rays, millimeter-waves are non-ionizing and will not pose health risks even when operated at moderate power levels. Millimeter-wave imaging systems are capable of penetrating clothing and other obscuring materials and have been shown to be capable of extremely high-resolution images. Sensor technology plays a large role in the improvement of imaging systems and newly developed sensors specific to millimeter-wave applications have enabled image production at very high data rates [78].

### 2.4.1 Millimeter-Wave Radiometry

In the millimeter-wave domain, objects in the environment tend to reflect and emit radiation, similar to what would happen in visible and infrared (IR) domains. A polarization-dependent quantity known as emissivity is used to quantify the degree to which the reflection or emission occurs, and in imaging systems this is generally denoted by $\varepsilon$ [78]. A radiating object with an emissivity that is equal to unity is considered a perfect radiator, and often referred to as a blackbody. Conversely, a perfect reflector will have an emissivity value that is equal to zero. Surface roughness, the angle from which the object is being observed and the dielectric properties of the various materials that form part of the object are the main properties that determine the emissivity. The radiometric temperature (often referred to as the surface brightness temperature) of an object is given by

$$T_S = \varepsilon T_0$$

(2.35)

where $T_0$ is known as the physical thermodynamic temperature of the object, and this value will always be unique to the object that is being observed. Emissivity is an important quantity that plays a pivotal role in the generation of scene imagery, seeing that power radiated from different objects vary based on their emissivity. However, if the emissivity were the only influential factor in the formation of scene images, the actual procedure would involve mapping measured $T_s$ values throughout the scene. The second part of the equation lies in the method by which the scene is illuminated. A highly reflective metal plate, for example, may have $\varepsilon = 0$ and thus $T_s = 0$, but the high reflectivity will cause the plate to appear to have a surface brightness temperature equal to that of the illumination rays being reflected by the plate in the first place. We can thus define a surface scattered radiometric temperature, $T_{SC}$, to capture this effect, and it is given by

$$T_{SC} = \rho T_{ILLUMINATOR}$$

(2.36)

where $\rho$ denotes the reflectivity of the object being illuminated and the second term on the right-hand side of (2.36) denotes the radiometric temperature of the
illuminator. The quantities $T_s$ and $T_{SC}$ can be added to yield the effective radiometric temperature, given by

$$T_E = T_s + T_{SC} = \varepsilon T_0 + \rho T_{ILLUMINATOR}. \quad (2.37)$$

In an outdoor scene, down-welling radiation that originates from the sky is the dominating source of radiation. When a radiometer, designed for the detection of thermal radiation, is pointed towards the zenith, it will detect residual radiation originating from deep space as well as down-wells from the atmosphere. This yields a brightness temperature around 60 K at 94 GHz [78]. A metal object with $\rho = 1$ and $\varepsilon = 0$ will thus have $T_E \approx 60$ K, and as a result it will appear to be very cold in a thermal image. Note, however, that this scenario is greatly simplified and does not account for many factors that play into the imaging process. An image can thus be formed by measuring $T_E$ as a function of position around the captured scene, generating a 2D map.

### 2.4.2 Millimeter-Wave Imaging Systems

Imaging systems that operate in the millimeter-wave regime have been implemented to address a large variety of problems and augment the functioning of several distinctly different systems.

#### 2.4.2.1 Aircraft Guidance Assistance

An increasing number of millimeter-wave imaging systems are being designed to provide assistance to pilots when landing aircraft. One system developed recently involved a passive camera operating at 94 GHz [79]. Technological advances in millimeter-wave sensor and circuit technology are highly influential in the implementation of systems like this one. The first of these is the advent of high-bandwidth, low-noise millimeter-wave receivers, which greatly enhances dynamic range and, as a result, image contrast.

The receiver used was designed for 94 GHz operation with an 8 GHz bandwidth and 5.5 dB noise figure. The worst case integration time per pixel was found to be 2 ms, which corresponds to a temperature resolution of about 0.3 K, comparable to commercial IR cameras available at the time. The second technological advancement that especially influenced real-time image acquisition was focal plane array receivers. With focal plane arrays in 1D or 2D configurations, images could be displayed at up to 30 Hz. This was an enormous improvement over single-receiver systems that required mechanical scanning over the area of a scene in order to form a complete picture. Finally, newly developed and enhanced signal processing algorithms were influential in improving the optical resolution of a scene, thereby reducing the required size of optics associated with passive sensors.
Another system implemented primarily in low-flying aircraft (e.g., search and rescue helicopters and small passenger planes) utilizes active millimeter-wave sensors to detect power lines [80]. Power lines running on high voltages and their accompanying distribution towers create an incredibly dangerous flight environment for helicopters. Power lines are barely visible even in high-visibility conditions, and their appearance from the cockpit significantly deteriorates in bad weather conditions.

Millimeter-wave imaging techniques have also been employed in synthetic vision systems as an alternative method of pilot assistance [54, 81]. Such a system generally consists of a scanning type millimeter-wave radar sensor, signal-processing hardware and a HUD. The signal-processing chain is intended to enhance the captured radar image, as well as provide automatic analysis of the image, which is relayed to the pilot. This results in a system that can perform real-time integrity checks on the flight data available from one or more onboard computers, in addition to detecting runway obstacles and collecting other terrain data.

### 2.4.2.2 Airport Security

The EM spectrum below 30 GHz has limited application for imaging systems because of the large physical size of sensor apertures that is required. Above 30 GHz, several systems have been demonstrated that are capable of imaging the contrast in a scene. The relatively small physical size of imaged scenes in security applications, as well as the ability to penetrate many obscuring materials, makes millimeter-wave imaging especially useful in these scenarios. Several systems have been developed for security purposes, such as contraband and concealed weapon detection systems [76, 77, 82].

### 2.5 Closing Remarks

This chapter touched on several important millimeter-wave wireless systems that are actively being developed and researched. While this was by no means intended to be a reference to the systems discussed, it should provide the interested reader with a good starting point for exploring the immensely populated spectrum of wireless systems further. Throughout this chapter, several key publications are highlighted that should give the reader a reasonable background on many developments in millimeter-wave systems that have occurred in the last few decades, in the hope of painting a picture of what could be expected from these systems in the future. Readers are strongly encouraged to review the fundamental texts that expand on the concepts discussed in this chapter, although the information provided here should serve as sufficient context for the subsequent chapters.
References


78 2 Systems Aspects of Millimeter-Wave Power Amplifiers
References
