
1

WIRELESS COMMUNICATIONS SYSTEM ARCHITECTURE AND PERFORMANCE

WAYNE STARK

*Department of Electrical Engineering and Computer Science
The University of Michigan, Ann Arbor*

LARRY MILSTEIN

*Department of Electrical and Computer Engineering
University of California—San Diego*

1.1 INTRODUCTION

Low power consumption has recently become an important consideration in the design of commercial and military communications systems. In a commercial cellular system, low power consumption means long talk time or standby time. In a military communications system, low power is necessary to maximize a mission time or equivalently reduce the weight due to batteries that a soldier must carry. This book focuses attention on critical devices and system design for low power communications systems. Most of the remaining parts of this book consider particular devices for achieving low power design of a wireless communications system. This includes mixers, oscillators, filters, and other circuitry. In this chapter, however, we focus on some of the higher level system architecture issues for low power design of a wireless communications system. To begin we discuss some of the goals in a wireless communications system along with some of the challenges posed by a wireless medium used for communications.

RF Technologies for Low Power Wireless Communications
Edited by Tatsuo Itoh, George Haddad, and James Harvey
ISBN 0-471-38267-1 Copyright © 2001 by John Wiley & Sons, Inc.

1.2 PERFORMANCE GOALS AND WIRELESS MEDIUM CHALLENGES

A system level (functional) block diagram of a wireless communications system is shown in Figure 1.1. In this figure the source of information could be a voice signal, a video signal, situation awareness information (e.g., position information of a soldier), an image, a data file, or command and control data. The source encoder processes the information and formats the information into a sequence of information bits $\in \{\pm 1\}$. The goal of the source encoder is to remove the unstructured redundancy from the source so that the rate of information bits at the output of the source encoder is as small as possible within a constraint on complexity. The channel encoder adds structured redundancy to the information bits for the purpose of protecting the data from distortion and noise in the channel. The modulator maps the sequence of coded bits into waveforms that are suitable for transmission over the channel. In some systems the modulated waveform is also spread over a bandwidth much larger than the data rate. These systems, called spread-spectrum systems, achieve a certain robustness to fading and interference not possible with narrowband systems. The channel distorts the signal in several ways. First, the signal amplitude decreases due to the distance between the transmitter and receiver. This is generally referred to as propagation loss. Second, due to obstacles the signal amplitude is attenuated. This is called shadowing. Finally, because of multiple propagation paths between the transmitter antenna and the receiver antenna, the signal waveform is distorted. Multipath fading can be either constructive, if the phases of different paths are the same, or destructive, if the phases of the different paths cause cancellation. The destructive or constructive nature of the fading depends on the carrier frequency of the signal and is thus called frequency selective fading. For a narrowband signal (signal bandwidth small relative to the inverse delay spread of the channel), multipath fading acts like a random attenuation of the signal. When the fading is constructive the bit error probability can be very small. When the fading is destructive the bit error probability becomes quite large. The average overall received amplitude value causes a significant loss in performance (on the order of 30–40 dB loss). However, with proper error control coding or diversity this loss in performance can essentially be eliminated.

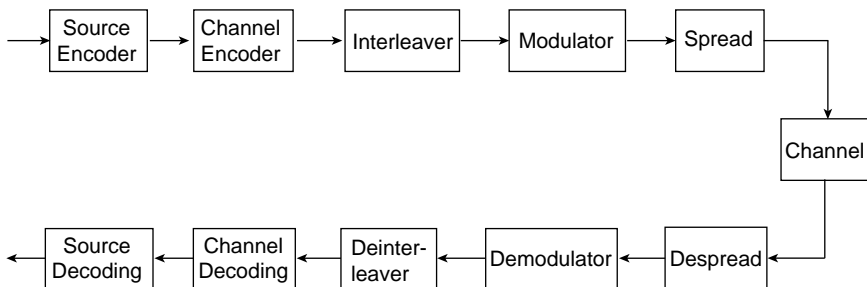


Figure 1.1. Block diagram of a digital communications system.

In addition to propagation effects, typically there is noise at the receiver that is uncorrelated with the transmitted signal. Thermal (shot) noise due to motion of the electrons in the receiver is one form of this noise. Other users occupying the same frequency band or in adjacent bands with interfering sidelobes is another source of this noise. In commercial as well as military communications systems interference from other users using the same frequency band (perhaps geographically separated) can be a dominant source of noise. In a military communications system hostile jamming is also a possibility that must be considered. Hostile jamming can easily thwart conventional communications system design and must be considered in a military communications scenario.

The receiver's goal is to reproduce at the output of the source decoder the information-bearing signal, be it a voice signal or a data file, as accurately as possible with minimal delay and minimal power consumed by the transmitter and receiver. The structure of the receiver is that of a demodulator, channel decoder, and source decoder. The demodulator maps a received waveform into a sequence of decision variables for the coded data. The channel decoder attempts to determine the information bits using the knowledge of the codebook (set of possible encoded sequences) of the encoder. The source decoder then attempts to reproduce the information.

In this chapter we limit discussion to an information source that is random data with equal probability of being 0 or 1 with no memory; that is, the bit sequence is a sequence of independent, identically distributed binary random variables. For this source there is no redundancy in the source, so no redundancy can be removed by a source encoder.

There are important parameters when designing a communications system. These include data rate R_b (bits/s, or bps), at the input to the channel encoder, the bandwidth W (Hz), received signal power P (watts), noise power density $N_0/2$ (W/Hz), and bit error rate $P_{e,b}$. There are fundamental trade-offs between the amount of power or equivalently the signal-to-noise ratio used and the data rate possible for a given bit error probability, $P_{e,b}$. For ideal additive white Gaussian noise channels with no multipath fading and infinite delay and complexity, the relation between data rate, received power, noise power, and bandwidth for $P_{e,b}$ approaching zero was determined by Shannon as [1]

$$R_b < W \log_2 \left(1 + \frac{P}{N_0 W} \right). \quad (1.1)$$

If we let $E_b = P/R_b$ represent the energy used per data bit (joules per bit), then an equivalent condition for reliable communication is

$$\frac{E_b}{N_0} > \frac{2^{R_b/W} - 1}{R_b/W}.$$

This relation determines the minimum received signal energy for reliable communications as a function of the spectral efficiency R_b/W (bps/Hz). The

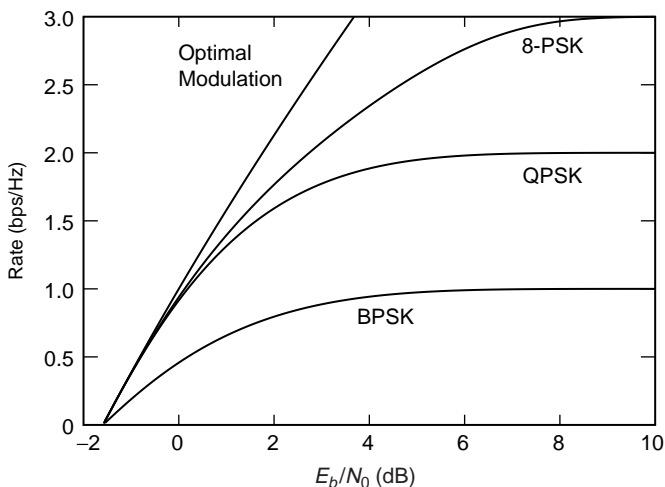


Figure 1.2. Possible data rate for a given energy efficiency.

interpretation of this condition is that for lower spectral efficiency, lower signal energy is required for reliable communications. The trade-off between bandwidth efficiency and energy efficiency is illustrated in Figure 1.2. Besides the trade-off for an optimal modulation scheme, the trade-off is also shown for three modulation techniques: binary phase shift keying (BPSK), quaternary phase shift keying (QPSK), and 8-ary phase shift keying (8PSK).

In this figure the only channel impairment is additive white Gaussian noise. Other factors in a realistic environment are multipath fading, interference from other users, and adjacent channel interference. In addition, the energy is the received signal energy and does not take into account the energy consumed by the processing circuitry. For example, power consumption of signal processing algorithms (demodulation, decoding) are not included. Inefficiencies of power amplifiers and low noise amplifiers are not included. These will be discussed in subsequent sections and chapters. These fundamental trade-offs between energy consumed for transmission and data rate were discovered more than 50 years ago by Shannon (see Cover and Thomas) [1]. It has been the goal of communications engineers to come close to achieving the upper bound on data rate (called the channel capacity) or equivalently the lower bound on the signal-to-noise ratio.

To come close to achieving the goals of minimum energy consumption, channel coding and modulation techniques as well as demodulation and decoding techniques must be carefully designed. These techniques are discussed in the next two sections.

1.3 MODULATION TECHNIQUES

In this section we describe several different modulation schemes. We begin with narrowband techniques whereby the signal bandwidth and the data rate are roughly

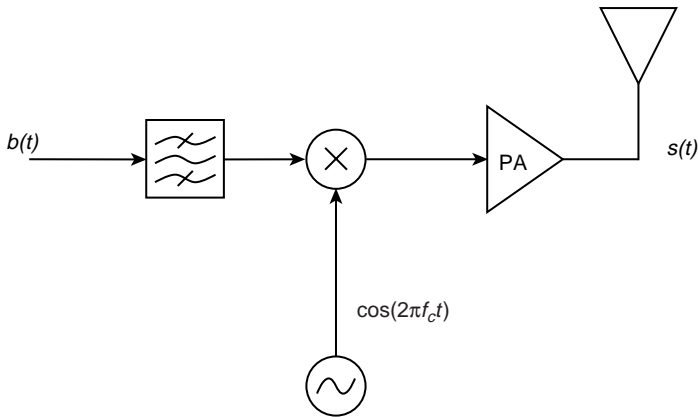


Figure 1.3. Transmitter.

equal. In wideband techniques, or spread-spectrum techniques, the signal bandwidth is much larger than the data rate. These techniques are able to exploit the frequency-selective fading of the channel. For more details see Proakis [2].

1.3.1 Narrowband Techniques

A very simple narrowband modulation scheme is binary phase shift keying (BPSK). The transmitter and receiver for BPSK are shown in Figures 1.3 and 1.4, respectively. A sequence of data bits $b_i \in \pm 1$ is mapped into a data stream and filtered. The filtered data stream is modulated onto a carrier and is amplified before being radiated by the antenna. The purpose of the filter is to confine the spectrum of the signal to the bandwidth mask for the allocated frequency. The signal is converted from baseband by the mixer to the desired center or carrier frequency (upconversion). The signal is then amplified before transmission. With ideal devices (mixers, filters, amplifiers) this is all that is needed for transmission. However, the

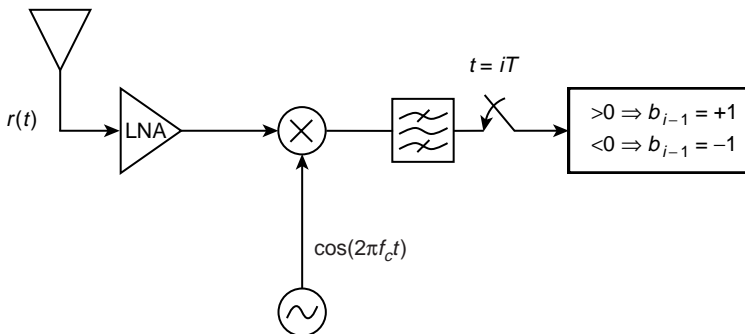


Figure 1.4. Receiver.

mixers and amplifiers typically introduce some additional problems. The amplifier, for example, may not be completely linear. The nonlinearity can cause the bandwidth of the signal to increase (spectral regrowth), as will be discussed later.

For now, assume that the filter, mixer, and amplifier are ideal devices. In this case the transmitted (radiated) signal can be written as

$$s(t) = \sqrt{2P} \sum_{l=-\infty}^{\infty} b_l h(t - lT) \cos(2\pi f_c t), \quad (1.2)$$

where P is the transmitted power, T is the duration of a data bit or the inverse of the data rate R_b , f_c is the carrier frequency, and $h(t)$ is the impulse response of the pulse-shaping filter. There are various choices for the pulse-shaping filter. A filter with impulse response being a rectangular pulse of duration T seconds results in a constant envelope signal (peak-to-mean envelope ratio of 1) but has large spectral splatter, whereas a Nyquist-type pulse has high envelope variation and no spectral splatter. The disadvantage of high envelope variation is that it will be distorted by an amplifier operating in a power efficient mode because of the amplifier's nonlinear characteristics. Thus there is a trade-off between power efficiency and bandwidth efficiency in the design of the modulation.

The simplest channel model is called the additive white Gaussian noise (AWGN) channel. In this model the received signal is the transmitted signal (appropriately attenuated) plus additive white Gaussian noise:

$$r(t) = \alpha s(t) + n(t). \quad (1.3)$$

The noise is assumed to be white with two-sided power spectral density $N_0/2$ W/Hz.

The receiver for BPSK is shown in Figure 1.4. The front end low noise amplifier sets the internal noise figure for the receiver. The mixer converts the radio frequency (RF) signal to baseband. The filter rejects out-of-band noise while passing the desired signal. The optimal filter in the presence of additive white Gaussian noise alone is the matched filter (a filter matched to the transmitter filter). This very simplified diagram ignores many problems associated with nonideal devices. For the case of ideal amplifiers and a transmit filter and receiver filter satisfying the Nyquist criteria for no intersymbol interference [2], the receiver filter output can be expressed as

$$X_l = \sqrt{\bar{E}} b_{l-1} + \eta_l,$$

where \bar{E} is the received energy ($\bar{E} = \alpha^2 PT$) and η_l is a Gaussian distributed random variable with mean zero and variance $N_0/2$. The decision rule is to decide $b_{l-1} = +1$ if $X_l > 0$ and to decide $b_{l-1} = -1$ otherwise. For the simple case of an additive white Gaussian noise channel, the error probability is

$$P_{e,b} = Q\left(\sqrt{\frac{2\bar{E}}{N_0}}\right),$$

where $Q(x) = \int_x^\infty (1/\sqrt{2\pi}) \exp(-u^2/2) du$. This is shown in Figure 1.5.

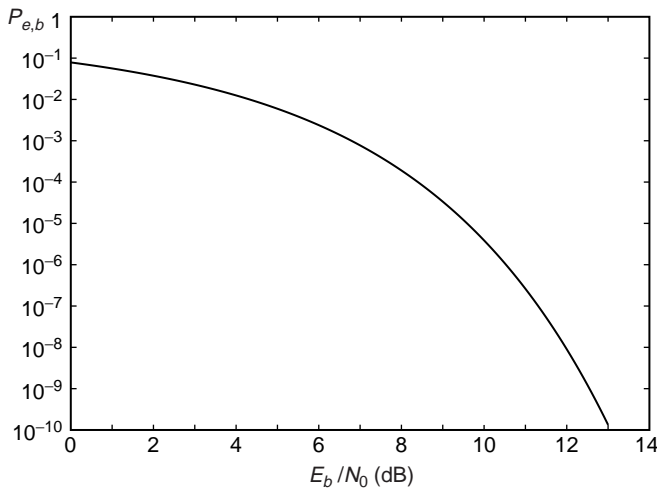


Figure 1.5. Bit error probability for BPSK in AWGN.

From Figure 1.5 it can be seen that in order to provide error probabilities around 10^{-5} it is necessary for the received signal-to-noise ratio to $\bar{E}/N_0 = 9.6$ dB. The capacity curve for BPSK in Figure 1.2, however, indicates that if we are willing to lower the rate of transmission we can significantly save on energy. For example, it is possible to have a nearly 0 dB signal-to-noise ratio if we are willing to reduce the rate of transmission by 50%. Thus about 9.6 dB decrease in signal power is possible with a 50% reduction in transmission rate.

The above analysis is for the case of additive white Gaussian noise channels. Unfortunately, wireless channels are not accurately modeled by just additive white Gaussian noise. A reasonable model for a wireless channel with relatively small bandwidth is that of a flat Rayleigh fading channel. While there are more complex models, the Rayleigh fading channel model is a model that provides the essential effect. In the Rayleigh fading model the received signal is still given by Eq. (1.3). However, α is a Rayleigh distributed random variable that is sometimes large (constructive addition of multiple propagation paths) and sometimes small (destructive addition of multiple propagation paths). However, the small values of α cause the signal-to-noise ratio to drop and thus the error probability to increase significantly. The large values of α corresponding to constructive addition of the multiple propagation paths result in the error probability being very small. However, when the average error probability is determined there is significant loss in performance. The average error probability with Rayleigh fading and BPSK is

$$\bar{P}_{e,b} = \int_{r=0}^{\infty} f(\alpha) Q\left(\sqrt{\frac{2E\alpha}{N_0}}\right) d\alpha = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\bar{E}/N_0}{1 + \bar{E}/N_0}}, \quad (1.4)$$

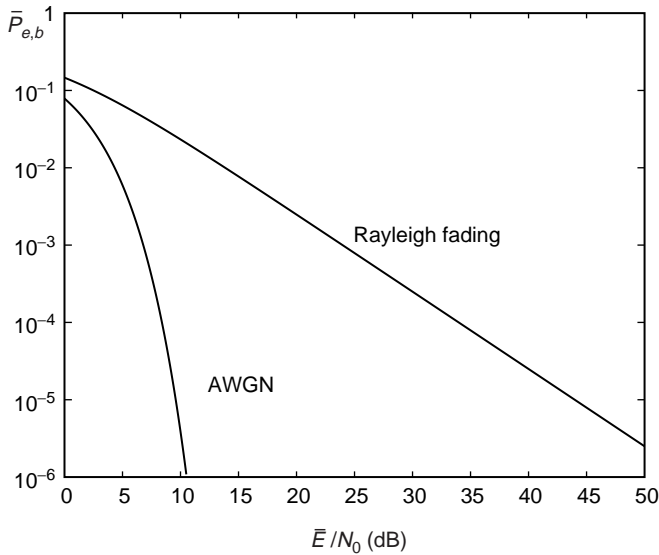


Figure 1.6. Bit error probability for BPSK with Rayleigh fading.

where $f(r)$ is the Rayleigh density and \bar{E} is the average received energy. The average error probability as a function of the average received energy is shown in Figure 1.6. Included in this figure is the performance with just white Gaussian noise. As can be seen from the figure, there is a significant loss in performance with Rayleigh fading. At a bit error rate of 10^{-5} the loss in performance is about 35 dB. In early communications systems the transmitted power was boosted to compensate for fading. However, there are more energy efficient ways to compensate for fading.

For a Rayleigh fading channel the fundamental limits on performance are known as well. The equations determining the limits are significantly more complicated [3]. Nevertheless, they can be evaluated and are shown in Figure 1.7. By examining the curve for BPSK we can see that it is possible to reduce the loss in performance to about 2 dB (rather than 35 dB) if proper signal (coding) design is used. Thus by reducing the data rate by 50% with proper coding a 33 dB savings in energy is possible. One method of signaling that reduces this performance loss is by the use of wideband signals, as discussed in the next section.

1.3.2 Wideband Techniques

Wideband signals have the potential of overcoming the problem of fading [4]. This is because the fading characteristics are frequency dependent. Different frequencies fade differently because the phase relationships of different paths change as the frequency changes. In addition, wideband techniques are able to handle interference from jammers or from other users. There are several techniques that are employed for wideband communications systems. Two popular techniques are direct-sequence

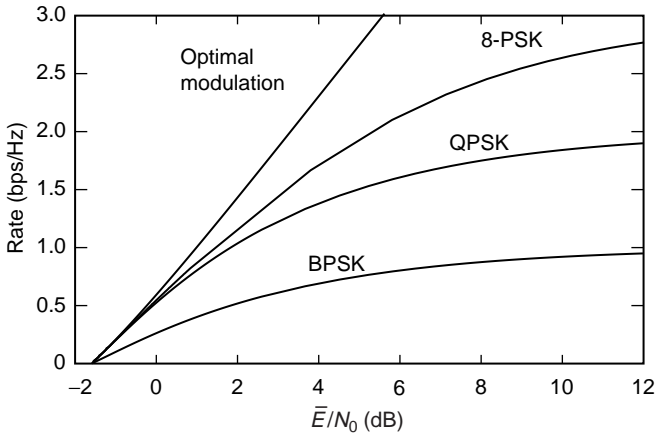


Figure 1.7. Fundamental limits for Rayleigh fading.

(DS) spread spectrum and frequency-hopped spread spectrum. We first discuss direct-sequence spread spectrum.

1.3.2.1 DS Spread Spectrum Conceptually, a direct-sequence spread spectrum works as shown in Figure 1.8. A data sequence is first encoded for error protection. The encoded waveform is then modulated using a standard narrowband modulation technique (e.g., BPSK). The narrowband waveform is then spread over a wider bandwidth with a spreading code as shown in the figure. At the receiver the received signal is despread by mixing with an identical spreading code followed by a narrowband demodulator. The result is then used for decoding or error correction. If there are multiple users using the same bandwidth at the same time but with

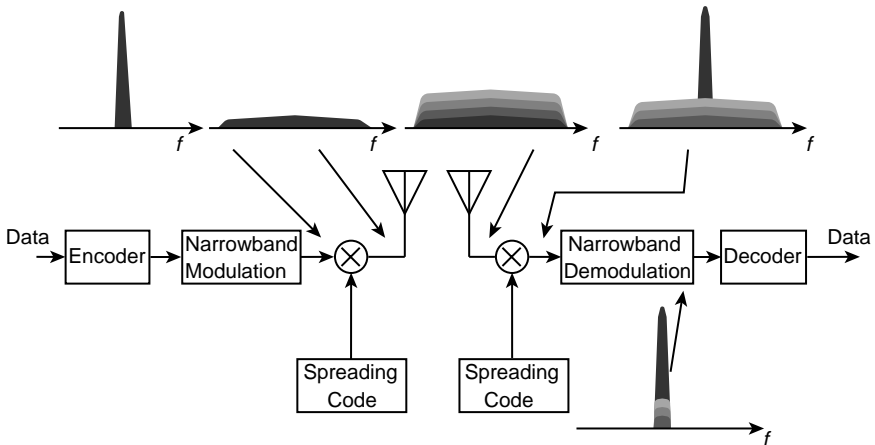


Figure 1.8. DS code division multiple access (CDMA) system.

different spreading codes then the received waveform consists of the sum of the signals of the different users (as shown in the figure). However, after mixing with the spreading code of the desired user, only the signal of the desired user becomes narrowband. The other users' signals, since they have a different spreading code than that used by the receiver, remain wideband signals. The narrowband demodulator then removes much of the interference. The amount by which the interference signals are reduced in power is roughly equal to the ratio of the bandwidth of the wideband signal to the narrowband signal. This is often referred to as the processing gain of the system.

The situation for a jamming signal is similar. Consider a jammer that transmits energy in a narrow bandwidth directly in the band of the desired user. The received signal will consist of a sum of the desired signal and the jamming signal, as shown in Figure 1.9. The receiver processes the signal by first mixing with a replica of the user's spreading code (assumed not available to the jammer). The desired signal gets despread while the jamming signal becomes spread. After demodulation the jamming signal power gets reduced by a factor equal to the processing gain.

The above are conceptual descriptions of the way in which a direct-sequence spread-spectrum system rejects interference. In both cases the receiver depicted is not the optimal receiver. The receivers depicted are optimal only if the interference is white Gaussian noise. Better receivers exist for multiuser interference and for jamming interference. In addition, for jamming signals, we have not considered the worst possible jamming signal of a given power. Power consumption occurs mainly in two places in this conceptual diagram. The first is at the transmitter in amplifying the signal. If the signal is not constant envelope, then nonlinearities of the amplifier cause distortion. When the amplifier is operating in the linear region the power efficiency of the amplifiers is small. If the envelope is constant, then, while the nonlinearities do not affect the signal, the signal has higher sidelobes than a nonconstant envelope signal. Significant power is also needed for the despreading operation. For example, if the despreading is done digitally then an analog-to-digital converter is needed with dynamic range equal to that of the jamming signal and

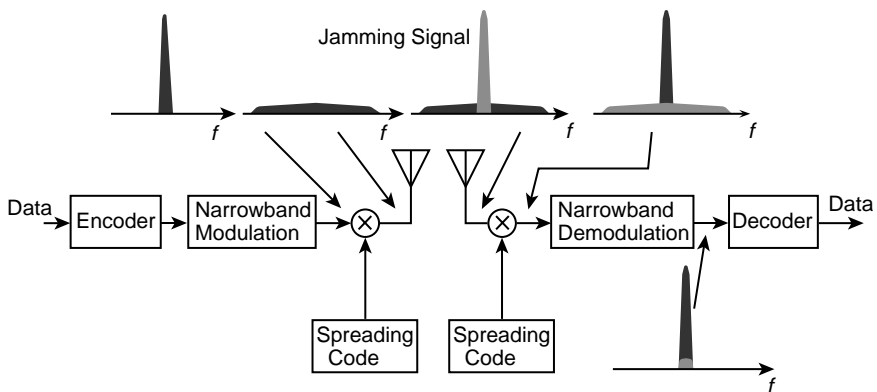


Figure 1.9. DS spread spectrum with jamming.

bandwidth of the desired signal. This combination of high dynamic range and bandwidth leads to high power consumption for the digital circuitry needed to despread. Thus despreading a spread-spectrum signal and thereby rejecting unwanted interference requires circuitry with significant power consumption.

Below we specify a simple direct-sequence spread-spectrum system. We analyze the performance in the presence of jamming, in a fading channel, and in multiple-access interference. The transmitter for a direct-sequence system is shown in Figure 1.10. The data sequence $b(t)$ consists of a sequence of data bits of duration T . The data sequence is multiplied with a binary spreading sequence $a(t)$, which has N components called chips per data bit.

In Figure 1.10, $b(t)$ represents the data and can be expressed as

$$b(t) = \sum_{l=-\infty}^{\infty} b_l p_T(t - lT), \quad b_l \in \{+1, -1\},$$

where $p_T(t)$ is a rectangular pulse of unit amplitude and duration T beginning at $t = 0$. Similarly, the spreading code $a(t)$ is written as

$$a(t) = \sum_{l=-\infty}^{\infty} a_l p_{T_c}(t - lT_c), \quad a_l \in \{+1, -1\},$$

where a_l is a binary symbol $\in \pm 1$ and $T_c = T/N$. In this case it is useful to model a_i as a sequence of independent, identically distributed binary random variables equally likely to be ± 1 . The number of chips per bit is often referred to as the “processing gain.” It is the factor by which the signal is spread. The transmitted signal is then

$$s(t) = \sqrt{2Pa}(t) b(t) \cos(2\pi f_c t).$$

The transmitted signal has power P . In Figure 1.11 we show a data signal and the result of multiplying by a spreading signal with 7 chips per bit.

The receiver consists of a mixer followed by a filter matched to the spreading code of the transmitter as shown in Figure 1.12. A typical filter output is shown in

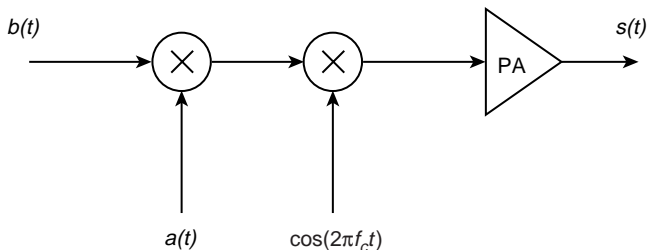


Figure 1.10. Block diagram of direct-sequence spread-spectrum transmitter.

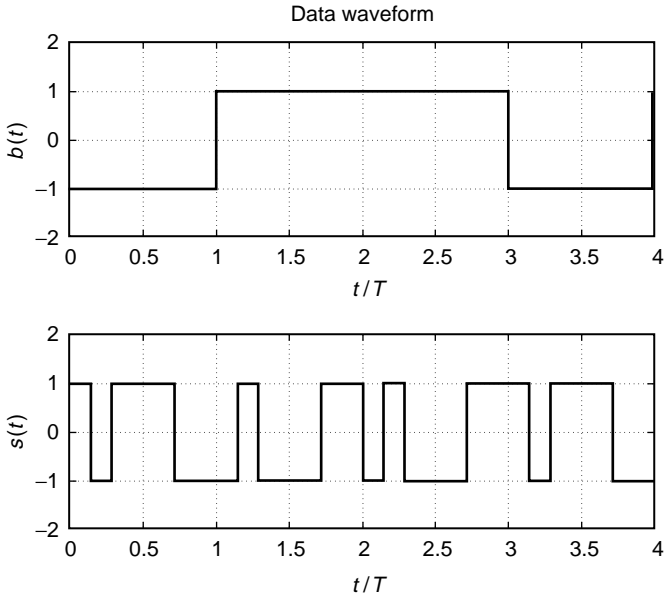


Figure 1.11. Waveforms $b(t)$ and $a(t)b(t)$.

Figure 1.13. In this figure the filter output is sampled every T seconds. At these times the filter impulse response is completely correlated with the incoming desired signal and produces a large amplitude (positive or negative depending on the sign of the data bit) signal. The sample output corresponds to a data sequence consisting of two positive polarity data bits followed by two negative polarity data bits followed by a positive polarity bit. For the case of an additive channel (no fading) with interference and a lowpass filter matched to the transmitted pulse shape (rectangular assumed here), the output of the filter at time iT can be expressed as

$$Z(iT) = \sqrt{E}b_{i-1} + I + \eta_i,$$

where I represents the contribution due to the interference and η_i is the output at time iT due to background noise (e.g., thermal noise).

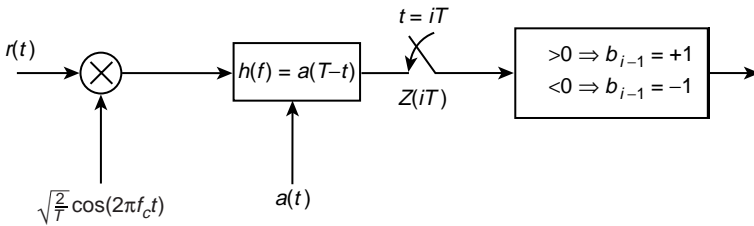


Figure 1.12. Block diagram of direct-sequence spread-spectrum receiver.

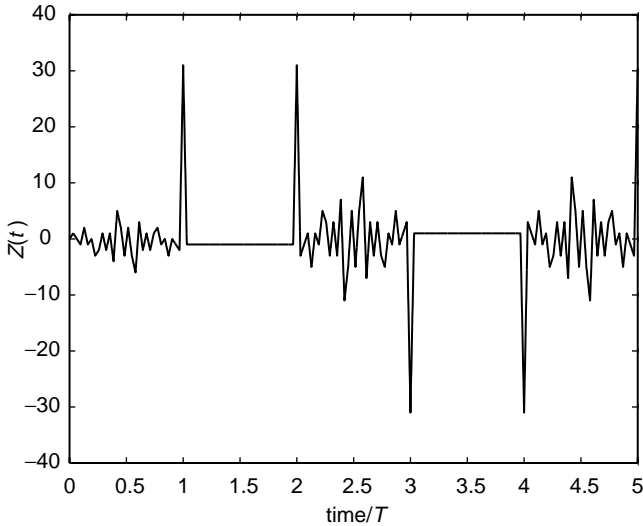


Figure 1.13. Output of matched filter for a sequence of length 31 in the absence of noise.

The component of the output of the receiver due to interference will depend on the nature of the interference (jamming, multiuser, etc.) as well as the spreading code $a(t)$ used. For simplicity in what follows we will assume that the spreading code consists of a sequence of independent and identically distributed random variables with equal probability of being $+1$ and -1 .

One simple measure of performance of a communications system is the signal-to-noise ratio at the output of the receiver. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{(E[Z(iT)])^2}{\text{Var}[Z(iT)]}.$$

Consider a direct-sequence system operating in the presence of a jamming signal $j(t)$, which is a pure tone operating at the exact carrier frequency as the desired signal (and with the same phase):

$$j(t) = \sqrt{2J} \cos(2\pi f_c t).$$

Then the signal-to-noise ratio at the output of the receiver is given by

$$\text{SNR} = \frac{P}{J/N}.$$

From this we can see the effect of spreading is to reduce the effective jamming power by a factor of N , the processing gain. If the jamming signal has a random

phase offset from the desired signal then the SNR would be increased by a factor of 2. The error probability can be approximated from the SNR by assuming the output of the receiver due to the interfering signal has a Gaussian density. With this assumption,

$$P_{e,b} \approx Q(\sqrt{\text{SNR}}).$$

Now consider the case of multiuser interference. Assume that there are K users with unique spreading codes (modeled as a random sequence). In addition, assume there are random delays and phases between the users. With these assumptions the signal-to-noise ratio for user 1 is given by

$$\text{SNR} = \frac{E_1}{\sum_{k=2}^K E_i / (3N) + N_0/2},$$

where K is the number of users, E_i is the received energy per bit for user i , and $N_0/2$ is the two-sided power spectral density of the background noise. Again it is clear how the spreading reduces the effect of the interference. The factor of 3 arises due to the random phase and delays between users.

Finally, consider the case of multipath interference. Because multipath is not additive but rather multiplicative the model needs updating. Consider a simple model whereby the received signal is a sum of delayed versions of the transmitted signal plus additive white Gaussian noise. That is,

$$r(t) = \sum_{l=1}^L \alpha_l s(t - \tau_l) + n(t).$$

For simplicity, assume that the delays are separated by at least T_c seconds. In this case the paths are said to be resolvable. This implies that the output of the matched filter consists of peaks that are nonoverlapping and thus can be resolved. Consider the ideal case of a spread-spectrum signal with zero sidelobes. In this case the output of the filter consists of a peak for each multipath present. If we assume that the correlations are ideal (so that the output consists of just impulses) and that the amplitude of each path is independent and Rayleigh distributed, then the performance improves dramatically compared to a single-path system as shown in Figure 1.14.

1.3.2.2 Frequency-Hopped Spread-Spectrum Frequency-hopped spread spectrum works by pseudorandomly changing the center frequency of the carrier over a set of frequencies. The sequence of frequencies used is called the frequency-hopping pattern. In a jamming environment this can force the jammer to spread its power over a very wide bandwidth in order to guarantee that the transmitted signal is disrupted (to some extent). When the jammer spreads its signal over the whole bandwidth the amount of power in each frequency slot is a small fraction of the total

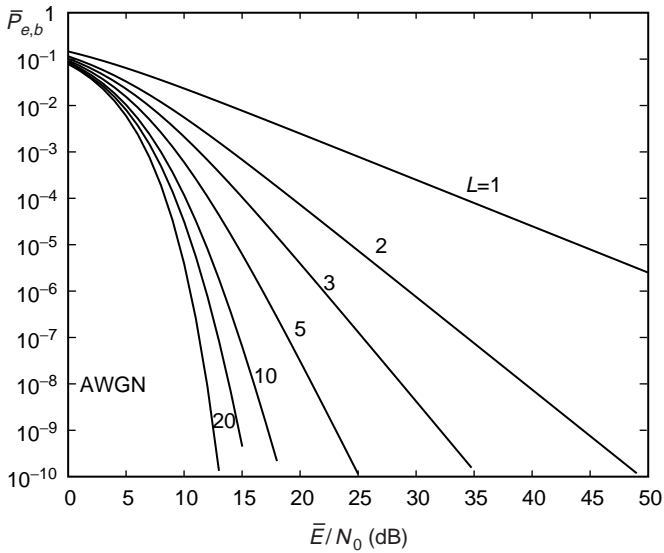


Figure 1.14. Bit error probability for BPSK with diversity and Rayleigh fading.

power. Thus the effectiveness of a wideband jammer is reduced in proportion to the bandwidth over which the signal is spread. If the spreading is large enough the jammer is not effective at disrupting communications. If the jammer only jams a fraction of the band but with high power in certain slots then the performance can be severely degraded (as opposed to a wideband jammer with the same total power). However, in this case a proper error control code and decoding algorithm that corrects errors in slots jammed can change the optimal jamming strategy from narrowband to broadband and thus regain the advantage of spreading the spectrum.

For multiple access, different users have different frequency-hopping patterns and different users will collide occasionally. These collisions can be handled with the use of appropriate error control coding. A code with a very low rate is needed for a large number of interferers while for a small number of interferers a large rate code should be employed. To maximize the information throughput per unit bandwidth an optimal code rate and number of users can be found.

For fading channels, provided the frequency separation between slots is larger than the coherence bandwidth, different frequencies will fade independently. If the bandwidth within a slot is small compared to the coherence bandwidth then the fading will be nonselective within a hop. An error control code will be able to correct errors from a badly faded hop. It is interesting to note that the (uncoded) performance in a fading environment is actually worse than the performance in a jamming environment. In both cases the performance degradation (without coding) is on the order of 30–40 dB compared to additive white Gaussian noise. This can be reduced somewhat in the partial-band jamming case by spreading over a very large bandwidth. However, with the proper combination of coding and spreading

a jammer’s optimum strategy is to jam the whole bandwidth. With this optimum jamming strategy the performance of a spread-spectrum system with worst-case jamming becomes the same as an additive noise channel with the same average power. In this sense then, a jammer can be defeated with the right combination of spreading and coding. In fact, the required signal-to-noise ratio with proper spreading and coding with a partial-band jammer is lower than the required signal-to-noise ratio for an uncoded spread-spectrum system with just additive noise of the same average power.

1.3.2.3 Multicarrier Techniques Multicarrier modulation techniques have recently gained significant popularity in the United States and Europe. In the United States multicarrier is used for digital subscriber loop (DSL) applications, while in Europe it is used in digital audio broadcasting (DAB). Multicarrier works by employing more than one carrier simultaneously. Consider an encoded data stream $b(t)$ consisting of data bits with rate $R_b = 1/T_b$ bits per second. The data stream is converted from a single stream into M separate streams via a serial-to-parallel converter as shown in Figure 1.15. After the serial-to-parallel converter the data streams are mixed onto M different carriers before being combined and amplified. At the receiver the inverse process is used. The received signal is mixed down to baseband using M different carriers. After mixing down to baseband the signal is filtered before a decision is made regarding each bit.

There are several motivations for considering multicarrier modulation techniques and several disadvantages of multicarrier techniques. For high data rate (relative to the inverse delay spread of the channel) applications, single-carrier systems suffer from severe intersymbol interference. This interference can be mitigated by sufficient equalization but requires significant complexity at the receiver to do this.

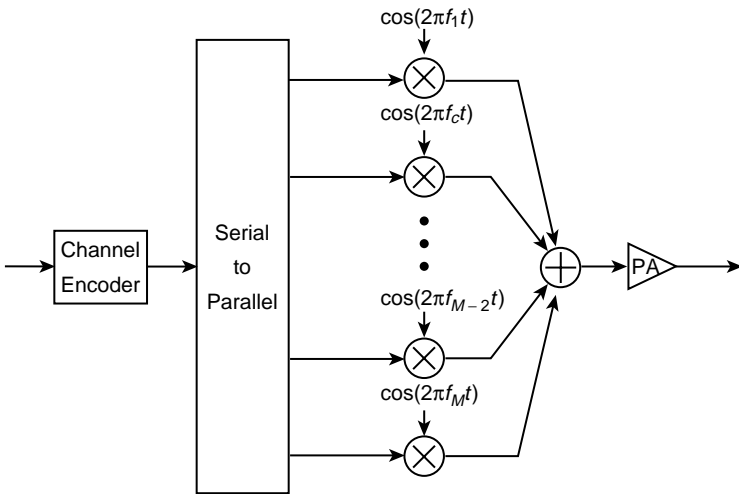


Figure 1.15. Multicarrier modulation.

By using multiple carriers the data rate on each carrier is reduced by a factor M . This means that the intersymbol interference is reduced by a factor of M and the receiver complexity is also reduced. However, the peak-to-mean power ratio at the input to the power amplifier now becomes large. For amplifiers operating with high backoff (not near saturation) this is not a problem. However, operating an amplifier with high backoff is typically not very efficient. If the amplifier is operating with low backoff then the large envelope variations cause distortion because of intermodulation products that fall in band or in adjacent channels. Another advantage of multicarrier modulation is that the frequency occupancy can be rather flexible. That is, we can build a system that occupies noncontiguous frequency bands. If we consider a linear amplifier and additive white Gaussian noise alone then the performance on multicarrier modulation is identical to a single-carrier system. Multicarrier techniques can also be applied to direct-sequence spread-spectrum systems. In such a system the data stream on each carrier is spread with a spreading code. This gives a multicarrier system the advantage with respect to interference that a direct-sequence system has. Later we will show the performance of multicarrier direct-sequence systems with nonideal amplifier characteristics.

1.4 CODING TECHNIQUES

Coding techniques are crucial to reducing the power consumption of digital communications systems. The basic idea of coding is to add redundancy to the transmitted data. For example, for every four information bits we might want to transmit four information bits and three parity bits or redundant bits. In this way the four information bits are encoded into seven coded bits. If we represent the information bits by b_0, b_1, b_2, b_3 , which are 0 or 1, then the coded bits are determined by $p_4 = b_0 + b_1 + b_2, p_5 = b_0 + b_1 + b_3$, and $p_6 = b_1 + b_2 + b_3$, where the equations are interpreted to mean mod 2 addition. The transmitted codeword is $(b_0, b_1, b_2, b_3, p_4, p_5, p_6)$ and is said to have block length 7. At the receiver the information bits can be determined even if a bit is received in error by recomputing the parity equations. For example, if the third bit (b_2) is received in error then the first and last parity equations are not satisfied. Because b_2 is the only bit that participates in only those two parity equations, it is found to be the bit in error and the decoder can correct the error. The code described above is called the Hamming code and can correct any pattern of a single error. If the energy used to transmit a single bit of information is denoted by E_b and the energy used to transmit an encoded bit is E , then for this example $4E_b = 7E$ or $E = 4E_b/7$. So each coded bit actually has less energy than what is allocated for an information bit. Because of this, the signal-to-noise ratio for each coded bit is reduced by a factor of $4/7$ from what could be used in an uncoded system. In spite of this, the error correction capability of the code makes up for this loss when the signal-to-noise ratio is reasonably large.

In practice, convolutional codes are used in many communications systems because of their excellent performance. In Figure 1.16 we show the bit error

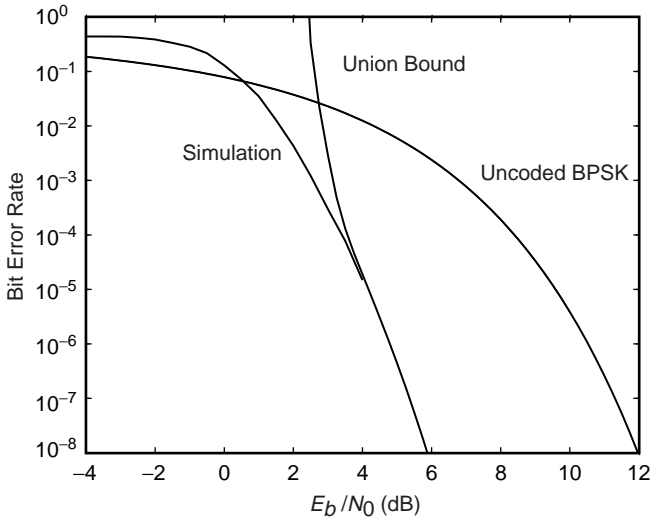


Figure 1.16. Error probability of BPSK with convolutional coding in additive white Gaussian noise.

probability for a typical convolutional code compared to an uncoded system. In this example the modulation is BPSK and the channel has just additive white Gaussian noise. This code has rate 1/2 meaning that for each information bit two coded bits are transmitted. From this figure we can see a 5 dB power performance improvement compared to an uncoded system. The disadvantage is that the bandwidth efficiency has been reduced by a factor of 2.

In Figure 1.17 the performance in Rayleigh fading (a more realistic environment) is shown. In this case the improvement is about 37 dB compared to an uncoded system. This illustrates the benefits in energy efficiency of coding relative to an uncoded system. The penalty paid for this improved energy efficiency is a decrease in bandwidth efficiency plus additional complexity at the receiver. Recently, another coding technique called turbo codes was invented. These codes achieve even better performance provided their block length is sufficiently long.

1.5 EFFECT OF NONLINEAR AMPLIFICATION ON DIRECT-SEQUENCE MULTICARRIER WAVEFORMS

1.5.1 Introduction

Multicarrier (MC) direct-sequence (DS) signaling, as described in Section 1.3, has certain desirable properties relative to single-carrier DS, such as flexibility in deploying the waveform over a noncontiguous bandwidth. That is, if certain segments of a given frequency band are occupied with narrowband signals, the

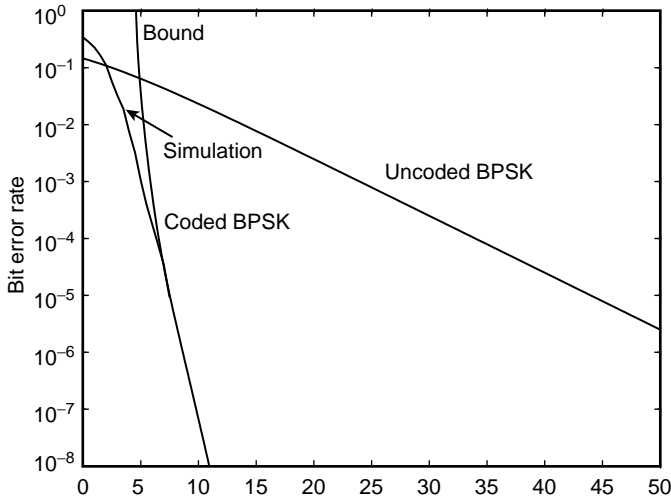


Figure 1.17. Error probability of BPSK with convolutional coding in additive white Gaussian noise and Rayleigh fading.

subcarriers that constitute the MC waveform can be interspersed between the narrowband signals so as not to cause any interference. However, MC signaling is not without its drawbacks, and arguably first among them is the fact that the composite envelope is not constant. Indeed, one can have a very high ratio of peak-to-average power and thus be susceptible to degradation due to nonlinear amplification.

In this section, we attempt to quantify the extent of this degradation; we also present a technique that can regain some of the lost performance by suppressing, at the receiver, some of the intermodulation products that were generated at the transmitter. In particular, we consider a binary communications system operating over a frequency-selective Rayleigh fading channel. The communications system employs convolutional coding with soft decision decoding.

1.5.2 System and Channel Description

A block diagram of the transmitter is shown in Figure 1.18 (see also Xu and Milstein [5]); it consists of a rate- $1/M$ convolutional encoder, followed by an interleaver, a serial-to-parallel converter, an MC modulator, and a power amplifier. The input-output characteristic for the power amplifier when the input is a single sine wave is shown in Figure 1.19. Note that this particular amplifier is assumed to exhibit only AM/AM conversion (i.e., it has no AM/PM conversion); this type of characteristic is typical of a solid state amplifier, as opposed to a traveling-wave tube.

The receiver is shown in Figure 1.20; the incoming waveform is demodulated, deinterleaved, and finally decoded.

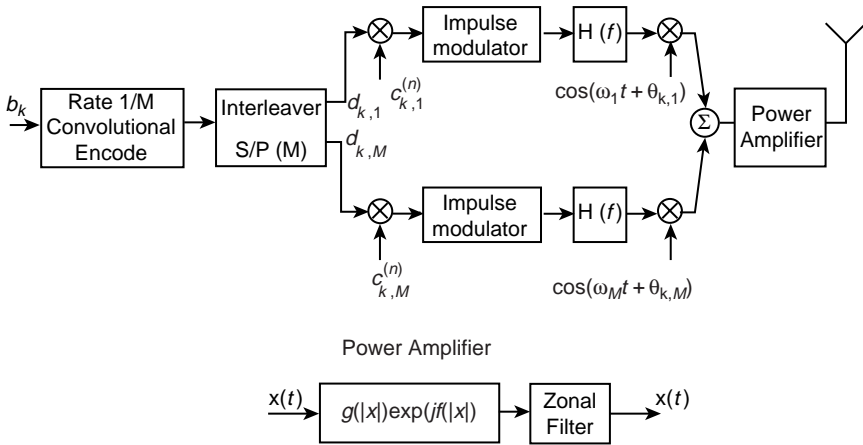


Figure 1.18. Transmitter block diagram for user k .

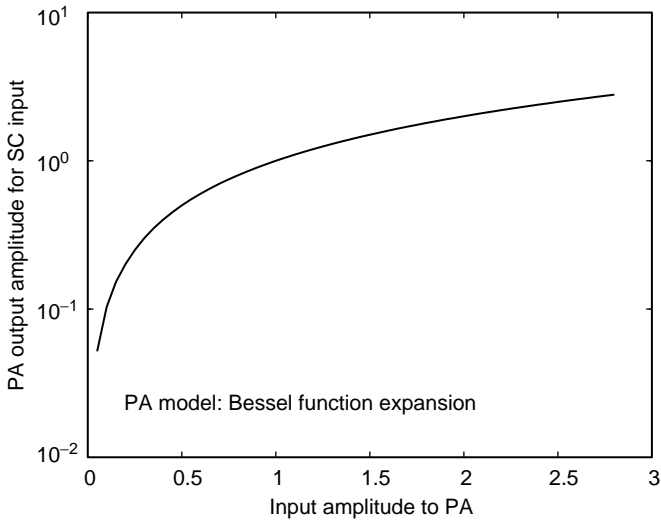


Figure 1.19. Power amplifier transfer function for single carrier.

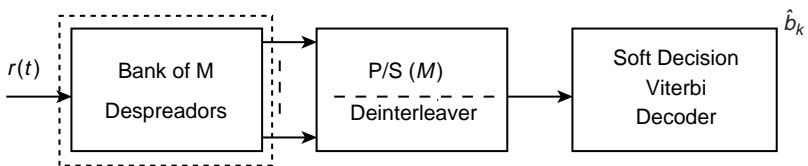


Figure 1.20. Receiver block diagram for user k .

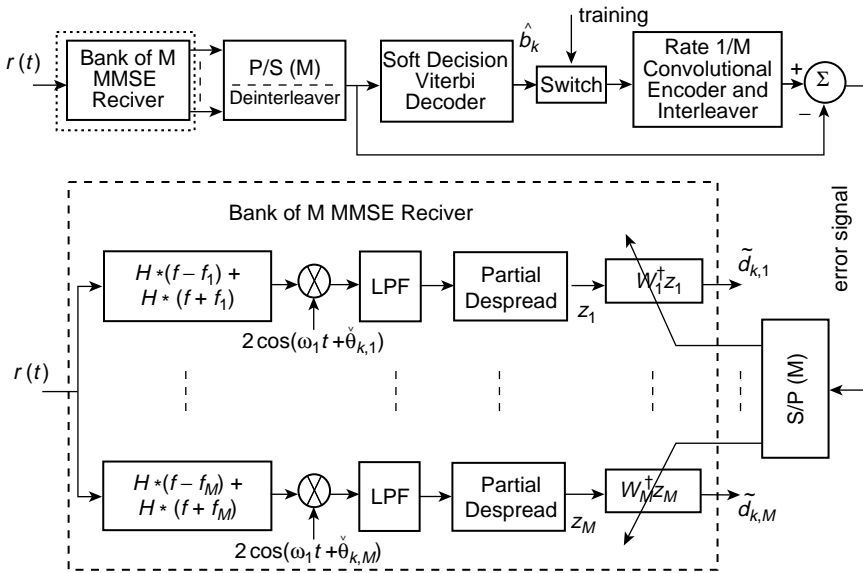


Figure 1.21. Receiver block diagram for user k [5].

As indicated above, an enhanced receiver will also be considered, whereby the enhancement will come from appropriate signal processing at the receiver so that the intermodulation distortion caused by the nonlinear amplifier at the transmitter is reduced. A block diagram of this enhanced receiver is shown in Figure 1.21, and it is seen that the only change from the original receiver is the insertion, after each of the M despreaders, of an interference suppression filter designed via a minimum mean-squared error criterion.

Finally, the channel introduces additive white Gaussian noise (AWGN), as well as flat Rayleigh fading on each subcarrier. The fading is assumed to be sufficiently slow so that it remains essentially constant over several symbols.

1.5.3 Performance Results

In the results that follow, the number of subcarriers, denoted by M , is taken to be 4. Figure 1.22 shows the levels of both third order and fifth order intermodulation (IM) products as a function of the drive level into the amplifier. Also shown is the desired signal response at each input level. It is seen that when the amplifier is driven heavily into saturation, the output levels of both third and fifth order intermodulation products can approach the level of the desired output.

When the above amplifier output is transmitted across the channel, and the receiver of Figure 1.20 is used, the resulting average probability of error is shown in Figure 1.23. It is seen that as the level of the signal into the amplifier is continually increased, a point is reached whereby the performance experiences rapid

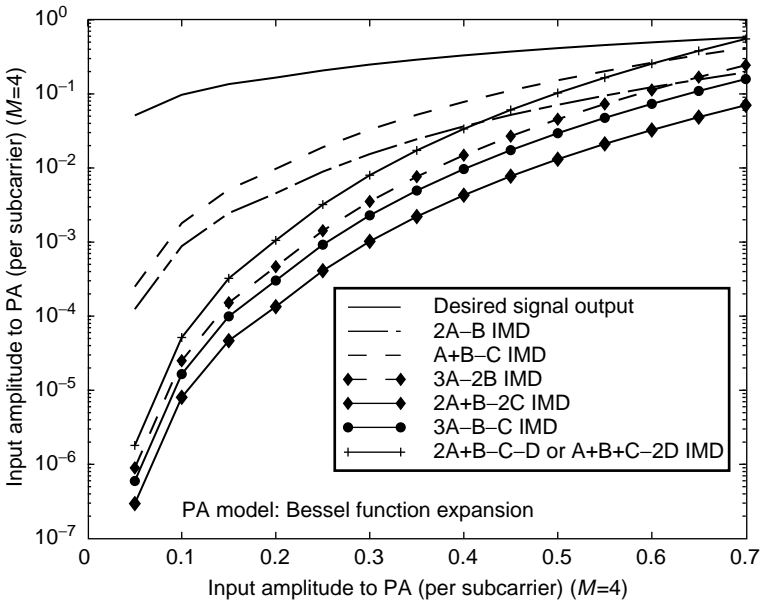


Figure 1.22. Desired signal and IM term amplitudes for multiple carriers ($M = 4$).

degradation. In order to reduce that degradation, the receiver of Figure 1.21 can be used, and its performance is also shown in Figure 1.23. Note that when the initial degradation is small, the enhanced receiver provides no noticeable benefit. However, when the initial degradation is large, the enhanced receiver reduces the average probability of error by about an order of magnitude.

As another illustration of the effect of the nonlinear amplifier, consider the use of the MC DS waveform in a code division multiple access (CDMA) environment. Assume that each user in the system transmits asynchronously using identical power amplifiers. The signal at any receiver then is the sum of all the intermodulation products coming out of each of the transmitters. In Figure 1.24, average probability of error curves are shown plotted against the number of simultaneously active users in the system. There are curves corresponding to two distinct drive levels (per subcarrier) into the amplifier: 0.5 and 0.6. Also, the performances for the two receivers of Figures 1.20 and 1.21 are shown. System performance corresponding to a perfectly linear amplifier is also presented.

It can be seen that, as the number of users in the system increases, the curves corresponding to the presence of intermodulation distortion diverge from the linear amplifier curves. To see what is happening, it is necessary to understand the operation of the suppression filters. Each filter is implemented as a tapped delay line, whereby the number of taps has been set equal to the processing gain per subcarrier; in this particular example, the processing gain is 32. When the total number of interfering waveforms is less than the number of taps, the suppression filter has a sufficient number of degrees of freedom to provide some measure of attenuation to

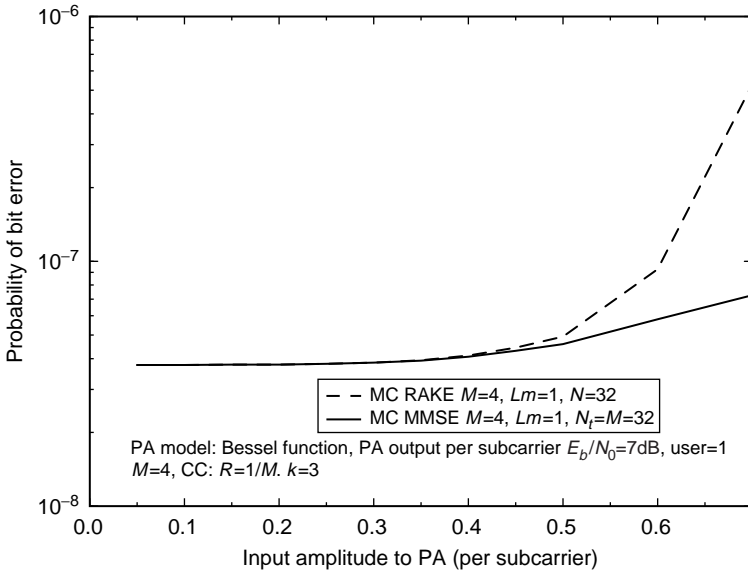


Figure 1.23. Probability of bit error in the presence of intermodulation distortion (IMD) (different PN (DPN) code for each subcarrier).

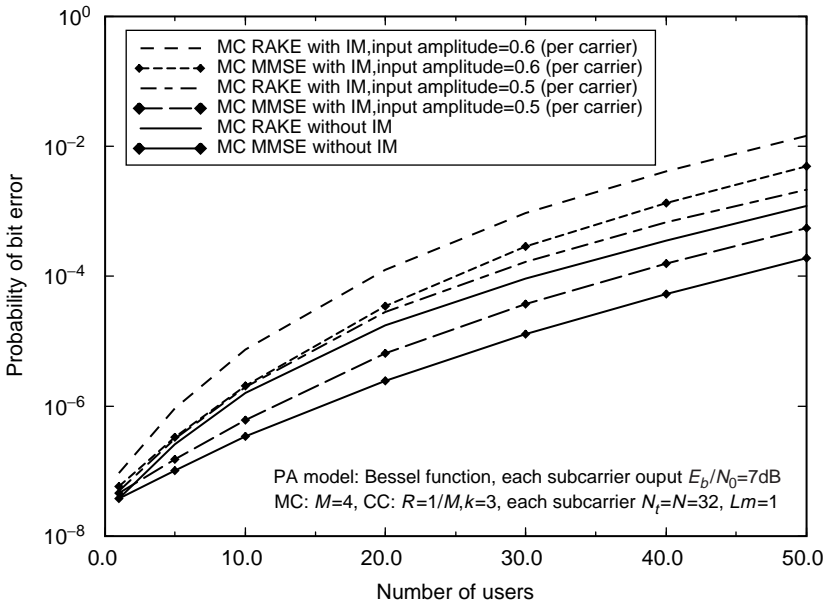


Figure 1.24. Probability of bit error in the presence of different strengths of IMD.

each of the interferers. However, when the number of such interferers exceeds the number of taps, the suppression filter is typically overwhelmed.

Now consider the curves shown in Figure 1.24. As can be seen, when the number of users is small, the receiver of Figure 1.21 can suppress a reasonable percentage of the intermodulation products from all the users and thus results in better performance than does the receiver of Figure 1.20. However, the large number of intermodulation products present as the number of transmitters grows presents too many terms for the suppression filter to handle. That is, the number of terms that the suppression filter must attenuate is not only the total number of first order intermodulation products (i.e., the desired signal term out of the amplifier of each interfering user), but all the significant higher order terms as well (e.g., those corresponding to third and fifth order products). Thus, while the use of the suppression filter still results in enhanced performance over the simple correlation receiver, its use does not result in performance comparable to that when intermodulation products do not exist.

As a final example we consider a scenario where 30 users in a cellular system are communicating with a base station. The different users might be at different distances from the base. Power control is assumed so that the received power is the same for all users. We will assume that the user's received power decreases with distance as d^4 . Users close to the base station will transmit with lower power. Because of this the amplifier will be operating in the linear range. However, users far away will need to transmit substantially more power in order for all the signals to arrive with the same power level. Thus distant users will be operating closer to the nonlinear range of the amplifier. These users then will create more intermodulation products that interfere with the signals of all the other users.

The model for the channel we employ is flat Rayleigh fading on each of the subcarriers. The amplifier output is assumed to saturate for some input power level. The output power level at which we operate the amplifier is reduced from the saturation output power level by some amount (in dB) called the output backoff (OBO). In addition, a rate 1/2 constraint length 7 convolutional code is considered with spreading by a factor of 80. In the numerical results that follow there are 30 users in a cell and the users are at distance r_1 or r_2 from the base station where $r_2/r_1 = 20$. The performance is evaluated for a particular user called user A.

In Figure 1.25 the bit error probability for 30 users each using 10 carriers is shown. In this figure all the users are at location r_2 . The receiver in this case is the conventional matched filter for each carrier without any interference suppression filter. The degradation due to the nonlinearity is seen to be small for large output backoff, while for small output backoff the amplifier is operating closer to the nonlinearity and thus the degradation in performance is larger. The performance for the case where users are at different locations is shown in Figure 1.26. In this figure we plot the additional E_b/N_0 required to achieve the same error probability (e.g., $P_b = 10^{-4}$) due to the nonlinearity. The degradation when the user-of-interest (user A) is located at the closer location (r_1) while the 29 other users are located at the further location (r_2) is quite minimal. In this case the intermodulation products from users at distance r_2 cause only a very small degradation (less than 1 dB). Also

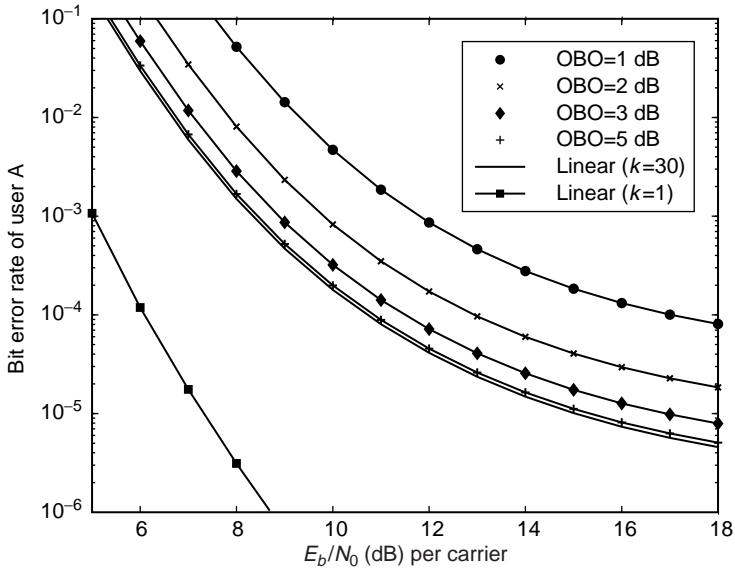


Figure 1.25. Bit error probability for a system with 30 users with 10 subcarriers and spreading 80 for different amplifier operating points.

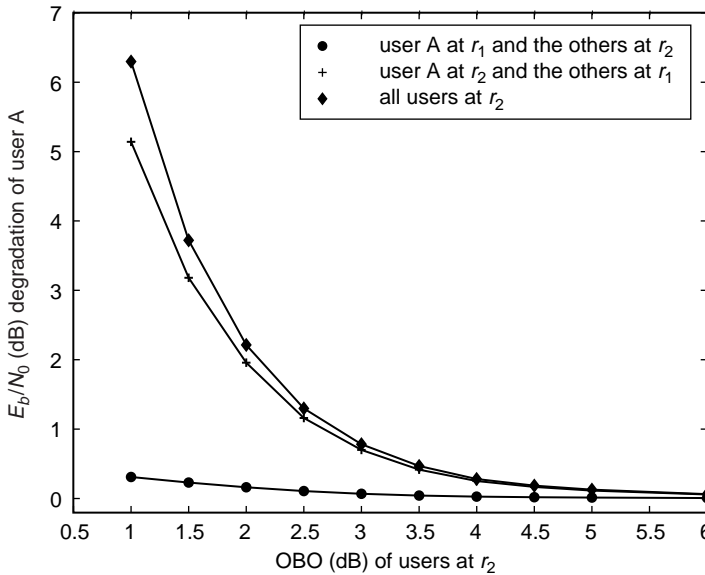


Figure 1.26. Degradation in performance due to nonlinearities for 30 users and spreading 80 for different amplifier operating points.

the desired user, because of power control, has a very significant backoff and does not generate any substantial intermodulation products. However, when the desired user is at distance r_2 the intermodulation products cause a degradation in performance by more than 5 dB at very low backoff. For additional details see Jong and Stark [6].

In the preceding discussion the performance curves were plotted as a function of the received energy per bit. However, in amplifying a signal the power amplifier, besides generating intermodulation products due to the nonlinearity, also has a varying power added efficiency. Power added efficiency is a measure of the amount of power converted from the dc power supply to output power. When the output backoff is large the power added efficiency is very small, while when the backoff is small the power added efficiency is much larger. Thus for large backoff the amplifier does not generate significant intermodulation products but is inefficient in converting dc power to output power. When the backoff is small the amplifier generates intermodulation products that distort the signal but, on the other hand, is able to convert the dc power to output power more efficiently. As a result of this, there is an optimal operating point in order to minimize the energy needed for a given performance (bit error probability). To quantify the overall performance we define the total degradation to be the sum of the output backoff (dB) plus the E_b/N_0 (dB) degradation due to the nonlinearity as shown in Figure 1.26. The output backoff is an accurate measure of the inefficiency of the amplifier (large output backoff implies large inefficiency), while E_b/N_0 degradation is a measure of the distortion due to the nonlinearity. This is shown in Figure 1.27 for the scenario of 30 users all operating at

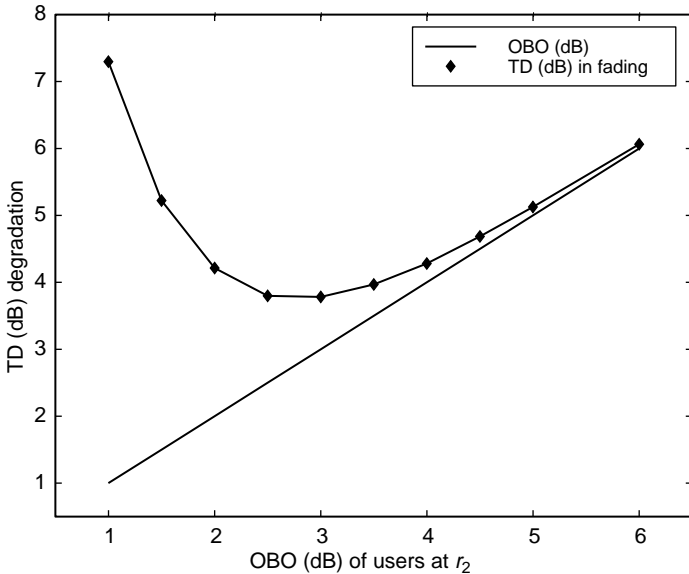


Figure 1.27. Total degradation in performance due to nonlinearities and amplifier inefficiency.

a distance of r_2 using multicarrier modulation and convolutional coding. As can be seen there is an optimal backoff that minimizes the total degradation in performance. This optimal backoff is about 3 dB and results in a performance that is degraded from the optimal by a little less than 4 dB. This incorporates both the degradation due to the intermodulation products and the inefficiency of the amplifier. This topic will be discussed in further detail in Chapter 7.

1.6 RECEIVER ARCHITECTURE

Besides containing the desired signal, the received signal contains unwanted interfering signals in adjacent frequency bands. The adjacent channel interfering signals may be much higher in power (by as much as 150 dB in certain systems) than the desired signal as seen in Figure 1.28. So one goal of the receiver is to select the desired signal and to reject the adjacent channel signals. Because the bandwidth is typically much smaller than the center frequency, rejecting the interference usually occurs in stages. First a wideband bandpass filter removes signals that are significantly distant (in frequency) from the desired signal. Then the signal is down-converted to a lower intermediate frequency (IF), where another filter with narrower bandwidth removes signals adjacent to the desired signal. Since the IF is typically much lower than the carrier frequency it much easier to design a filter of a given bandwidth centered at the IF than a filter of the same bandwidth at the carrier frequency.

Another goal in designing a receiver is to minimize the thermal noise of the receiver. For this reason a low noise amplifier is usually incorporated into the receiver very close to the antenna to set the noise figure of the receiver. Another goal is to minimize the receiver size. Filters are generally hard to incorporate into

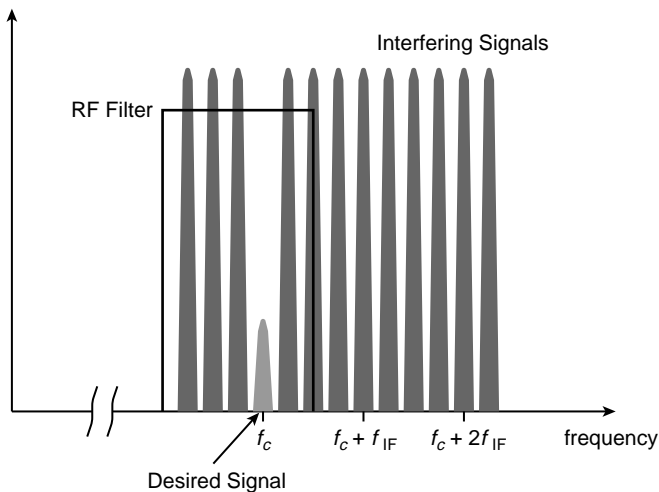


Figure 1.28. Received signal.

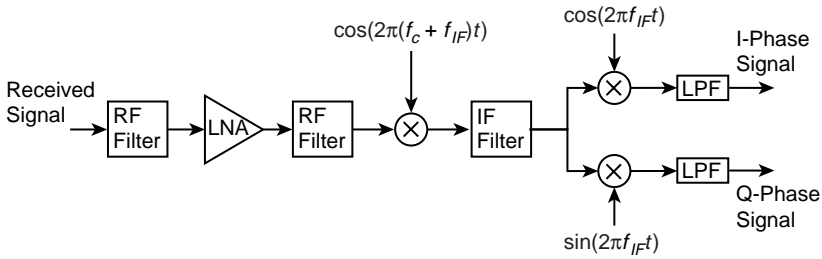


Figure 1.29. Superheterodyne receiver.

integrated circuits and thus are usually off-chip. A key question then is how to reject the interfering signals and have a low noise figure with minimal energy used by the receiver and minimal size. In the remainder of this chapter we explore various options and discuss the advantages and disadvantages of each.

The conventional design of a receiver is known as a superheterodyne receiver and is shown in Figure 1.29. The desired signal at frequency f_c is filtered, amplified, and filtered. The goal of the first filter is to reject the signal at frequency $f_c + 2f_{IF}$. This is known as the image frequency. A signal at this frequency when mixed with a local oscillator frequency of $f_c + f_{IF}$ will be translated to frequency f_{IF} . The desired signal will also be translated to this frequency so it is necessary to first remove the image frequency before mixing down to the intermediate frequency. The second RF filter is used to reject any signals generated by a nonlinearity of the low noise amplifier. The output of the second RF filter is then mixed down to a much lower frequency (f_{IF}) known as the intermediate frequency. It is easier at this (lower) frequency to design a narrow bandwidth filter that rejects the adjacent frequencies than it is at the higher RF frequency. Finally, the signal is mixed either to baseband by a pair of mixers (as shown) or to a very low frequency before being sampled. After sampling, digital processing is used to recover the desired information as described earlier. With conventional filters the superheterodyne architecture has the disadvantage that the filters are not “on-chip” filters and thus the size of the receiver is limited by these filters (which for certain frequency bands can be a limiting factor in the size of the receiver). The low noise amplifier consumes energy in order to amplify the signal. As with the transmitter amplifier, the low noise amplifier needs to operate well within the linear region in order not to generate intermodulation products of the input signal. However, the power efficiency of the amplifier is very small when the backoff is very large. So there is a trade-off here of performance versus energy. Better performance can be achieved if the amplifier operates with larger backoff but at the expense of increased energy consumption. Better performance and lower energy consumption are possible if the RF filters could do a better job at removing all the interference before amplifying. Such a filter may be possible with the advent of microelectromechanical system (MEMS) filters. In addition many MEMS filters can be incorporated into an integrated circuit with the same technology. The design of such filters is the topic of Chapter 12.

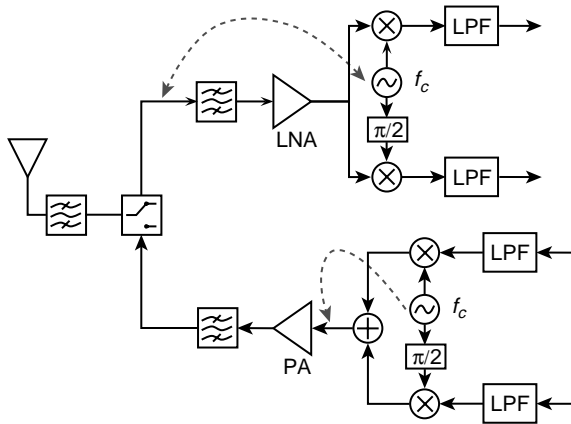


Figure 1.30. Direct conversion transmitter/receiver.

Another potential architecture shown in Figure 1.30 is a direct conversion architecture. In this receiver there is no intermediate frequency. Because of this, all of the components associated with the IF section are eliminated and thus the power consumption should be smaller than that of a superheterodyne architecture. However, the drawback to this architecture is that the signal out of the local oscillator (LO) may leak into the input to the low noise amplifier (LNA). Because the LO signal has a large amplitude relative to the received signal, the part of the LO signal that is coupled into the amplifier input could dominate the desired signal. Because the LO signal and the desired signal are at the same frequency the leakage can cause significant interference to the desired signal. In addition, a strong interfering signal at a frequency different from the carrier might leak into the LO signal input to the mixer. If this happens the interfering signal will get mixed to the same frequency as the desired signal by the part of the interfering signal that has leaked into the LO input to the mixer.

Another architecture is shown in Figure 1.31 and is called direct sampling. In direct sampling the output of the RF section is sampled at a sufficiently high rate and then digital filtering is used to filter out the undesired signals.

A direct sampling architecture requires a large dynamic range for analog-to-digital conversion. Large dynamic range is necessary as an interferer might be 50–150 dB stronger than the desired signal. Either very high Q filters are needed at the front end so that the interference is rejected or very high dynamic range conversion is necessary so that the jammer does not saturate the analog-to-digital converter (ADC) and the digital processing can filter out the interference. In addition, a sampling rate proportional to the carrier frequency is needed. This combination of high number of bits of quantization and high sampling rate is very power hungry and thus not a suitable choice for a low power design. However, for a cellular system this might be a good choice for the front end of the base station receiver since only one front end is needed for all users and power consumption is not as critical as in a mobile.

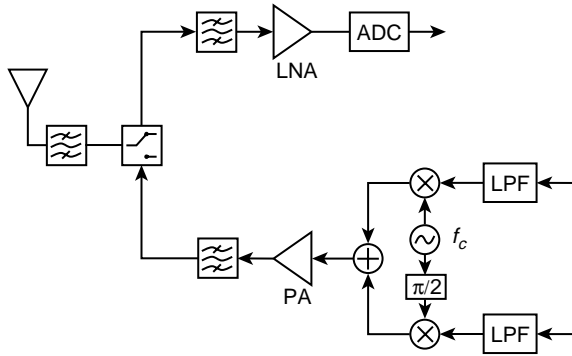


Figure 1.31. Direct sampling transmitter/receiver.

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of Je-hong Jong and Weiping Xu in generating some of the numerical results presented here. In addition, the support of the Department of Defense under the MURI programs on Low-Power Electronics and Low Energy Electronics Design of Mobile Platforms DAAH04-96-0001, DAAH04-96-0005, and DAAH04-96-0377 is gratefully acknowledged.

REFERENCES

1. T. M. Cover and J. A. Thomas, *Information Theory*, Wiley-Interscience, New York, 1991.
2. J. G. Proakis, *Digital Communications*, 3rd ed., McGraw-Hill, New York, 1995.
3. S. G. Wilson, *Digital Modulation and Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1996.
4. R. L. Peterson, R. E. Ziemer, and D. E. Borth, *Introduction to Spread Spectrum Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
5. W. Xu and L. Milstein, "On the use of interference suppression to reduce intermodulation distortion in multicarrier CDMA systems" *IEEE Transactions on Communications*, Vol. 49, No. 1, pp. 130–141 (2001).
6. Je-hong Jong and W. E. Stark, "Power-controlled MC-CDMA in the presence of nonlinearities—Part II: Coded system performance," in *Proceedings of the 1999 IEEE Military Communications Conference*, 1999.