2

Link Performance in Interference Channels

2.1 A Review of Multi-User Access in Mobile Communication Systems

2.1.1 Introduction and Problem Formulation

Classical communication theory deals with point-to-point links disturbed by thermal (Gaussian) noise, adverse propagation conditions, and channel variations that are difficult to predict. Real-life radio systems have to cope with additional problems. The most dominant feature of modern radio communication is that virtually no radio link or system is alone in its allocated frequency band. Other radio transmitters, near and far, constantly cause interference. Interference is, in many cases, the limiting factor to the performance of the system. With the increasing use of wireless communications, the load on the frequency spectrum has increased tremendously since the days of Marconi. A key problem area, as was already noted in Chapter 1, is how to effectively manage the frequency spectrum in order to keep the adverse effects of this interference at a minimum. Can interference be avoided, or are there efficient methods that minimize the loss in performance?

The radio transmission medium is, whether intended or not, a broadcast medium. In a wireless network, a large number of users in a geographical region attempt to communicate as illustrated by Figure 2.1. This feature is in many cases a blessing, since it enables the quick establishment of new connections between a large number of arbitrary users. In Figure 2.1, there

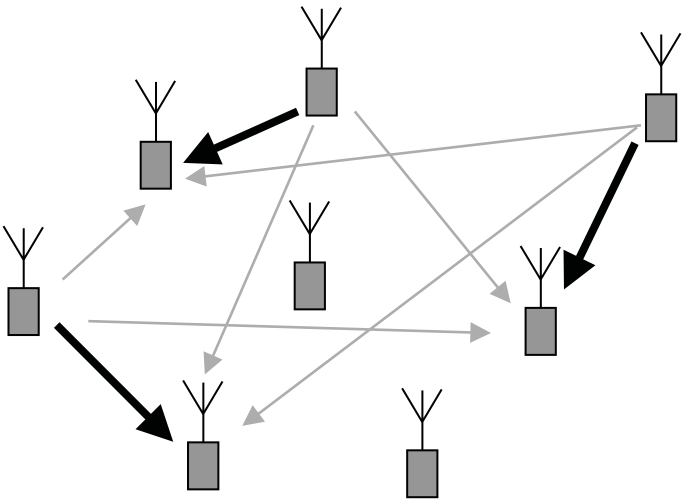

Figure 2.1 A wireless network.

are three transmitters transmitting information to three different receivers indicated by the solid black arrows. These paths are denoted the *active communication links.* However, the communication resource (the radio spectrum) is shared by all users despite the mutual interference this may cause. The transmissions of the three transmitters in Figure 2.1 give rise to interference on the unwanted cross-links as indicated by the light gray arrows. The character of the interference will depend on the waveforms and transmitter powers selected by the interfering transmitters as well as the propagation conditions on the cross-links. The impact on the performance of the active communication link will depend not only on the waveforms, the powers and the propagation conditions in the active link, but also on the detection scheme used in the receiver. In order to simplify the analysis of such a wireless network, a two-step approach is used. First the interference is characterized at the receiver by its received power and the effects of the propagation conditions. In the second step the impact of that interference on the performance of the active link is analyzed. The latter problem is the topic of this chapter.

In order to assess the impact of the interference on the performance, a given link in the network is studied. The detection scenario can be described by the multiple access channel (MAC) model in Figure 2.2. This is a straightforward extension to the M transmitter case of the models of classical communication theory. The receiver in the link of interest is facing the

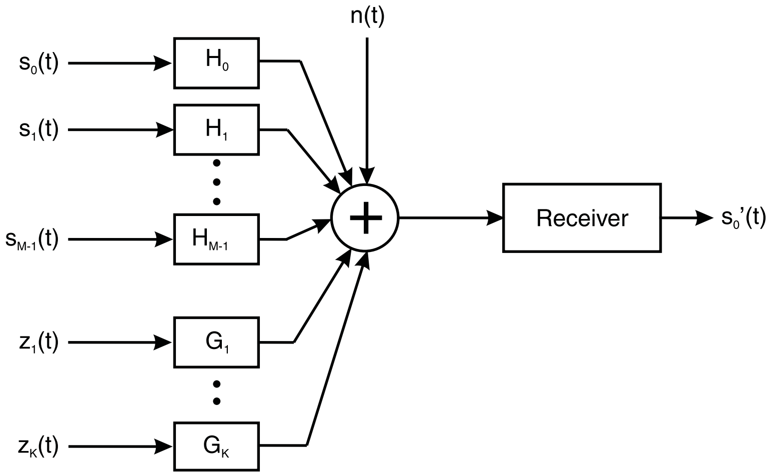

Figure 2.2 Model of signals and interference in multi-user radio system.

problem of detecting one particular transmitted signal $s_0(t)$ from some transmitter in the active link of interest. While this signal is being received, the signals $s_1(t)$, $s_2(t)$ . . . $s_{M-1}(t)$ from the other $M-1$ transmitters in the system are also on air, possibly causing interference at the receiver. As well as the known $M-1$ transmitters in the systems, there may be additional external interference, from other (maybe distant) radio systems using the same frequency range.

Ultimately, the situation where this type of interference is deliberate and aimed at disruption of the communication links will be taken into account. For the time being, assume that there are K of these external signals and that they are denoted $z_1(t)$, $z_2(t)$ . . . $z_K(t)$. Finally, as in the classical communication theory, the receiver is subject to (thermal) noise.

In many situations when the interference dominates and the noise may be neglected, the term *interference-limited* system is used. The converse, *noise-limited* systems are becoming more and more rare and are mostly found in space communications.

Methods to separate and distinguish between different users may be divided into two types: multiplexing techniques and multiple access schemes. Multiplexing describes general methods of choosing signals and combining information from different sources. In general this is done at one location, for example, a telephone switch or a microwave link carrying a large number of calls. The multiplex design problem boils down to choosing a signal constellation $s_0(t), s_1(t) . . . s_{M-1}(t)$ such that the system achieves the required performance. Performance-criteria may differ from system to system, but in principle, they are the same as in classical communication theory. As high a data rate as possible is required at some given, low level of signal or message distortion. In the case of digital communication the bit or message error probability could be used, but also the message delay is used as a performance measure. There is, however, an important additional performance criterion that distinguishes multi-user from conventional point-to-point systems, that is, how many users are allowed simultaneously in the system at some given bandwidth. The performance of a multi-user system will depend on the selection of signals $s_0(t), s_1(t) \ldots s_{M-1}(t)$. This problem is covered in Section 2.2.

Later in the chapter, the focus will be on the multiple access problem. The overlaying of information from a large number of users is not done in the same equipment, but rather in a distributed fashion on air (Figure 2.2). As well as the pure signal design problems, problems like synchronization and coordination of message transmissions will be encountered. Subsequent sections will deal with these problems.

2.1.2 Signal Design in Multi-User Systems

The design of communication signals in multi-user systems differs from the design of conventional point-to-point systems in several respects. To investigate these differences the discussion is started by studying optimal detection strategies for a wide class of signals in a multi-user system. However, ultimately the investigation will be confined to the class of binary, digital communication systems. Similar results can be derived also for analog schemes and for digital systems employing multilevel modulation schemes. A further assumption is that the system uses antipodal signaling, that is, the transmitters are emitting independent signals of the form

$$s_i(t) = a_i u_i(t) \qquad a_i = \pm 1 \tag{2.1}$$

For the sake of simplicity assume that the receiver is linear and that the channel filters have a flat response (i.e., equivalent to multiplication by constant h_i). The received signal $r(t)$ may now be written as

$$r(t) = \sum_{i=0}^{M-1} a_i h_i u_i(t) + \sum_{j=1}^{K} z_j(t) = \sum_{i=0}^{M-1} a_i h_i u_i(t) + z(t) \tag{2.2}$$

The interference vector $\boldsymbol{z}'$ consists of a number of components. The information symbol $a_0 = +1$ is assumed to be transmitted, where $z(t)$ is the sum of all the K interfering signals. The task of our selected receiver will be to determine whether the signal $s_0(t)$ or the signal $-s_0(t)$ was transmitted by transmitter zero. Equation (2.2) can now be rewritten as

$$r(t) = a_0 h_0 u_0(t) + \sum_{i=0}^{M-1} a_i h_i u_i(t) + z(t) = a_0 h_0 u_0(t) + z'(t) \quad (2.3)$$

where $z'(t)$ is the sum of the interference $z(t)$ and all (other) signal components that are independent of a_0. $z'(t)$ is thus independent of the transmitted information and can be interpreted as noise.

Using the standard vector space analogy in communication theory [1], the received signal vector $\boldsymbol{r}$ may be rewritten as the vector sum of one signal component $a_0 h_0 \boldsymbol{u}_0$ and one "noise" component z' (see Figure 2.3). If the information symbols ± 1 are equally probable, one may show [1] that the detector that minimizes the bit error probability is the one that chooses the symbol a that will maximize the probability density of the vector $\boldsymbol{r}$. Such a maximum likelihood (ML) detector will choose $a_0 = 1$ if

$$p_{\boldsymbol{r}}(\boldsymbol{r} | a_0 = +1) > p_{\boldsymbol{r}}(\boldsymbol{r} | a_0 = -1) \quad (2.4)$$

and $a_0 = 0$ otherwise. If the constant h_0 is known, $\boldsymbol{z}'$ would constitute the only remaining stochastic component in the received vector $\boldsymbol{r}$. The fact that $\boldsymbol{z}' = \boldsymbol{r} - a_0 h_0 \boldsymbol{u}_0$ enables us to rewrite the expression above by using the probability density of $\boldsymbol{z}'$ according to

$$p_{\boldsymbol{r}}(\boldsymbol{r} | a_0 = +1) = p_{\boldsymbol{z}'}(\boldsymbol{r} - h_0 \boldsymbol{u}_0) > p_{\boldsymbol{z}'}(\boldsymbol{r} + h_0 \boldsymbol{u}_0) = p_{\boldsymbol{r}}(\boldsymbol{r} | a_0 = -1) \quad (2.5)$$

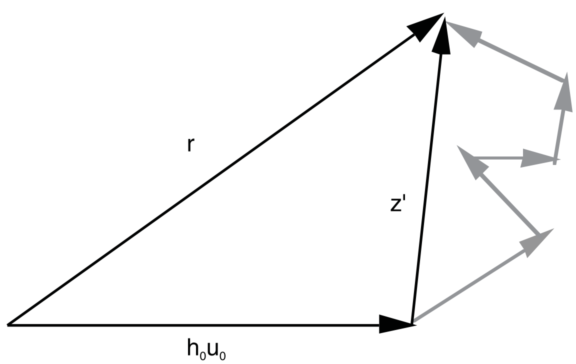

Figure 2.3 Example of vector space representation of signaler in a multi-user environment.

Now, consider the special case, where $\boldsymbol{z}'$ has a probability density function that

i) Only depends on $|\boldsymbol{z}'|$;

ii) Is monotonically decreasing in $|\boldsymbol{z}'|$;

(for instance, Gaussian noise) it can be seen that the ML receiver opts for the signal alternative that minimizes $|\boldsymbol{r} - a_0 h_0 \boldsymbol{u}_0|$, that is, chose the signal $a_0 h_0 \boldsymbol{u}_0$ which is closest to the vector $\boldsymbol{r}$. In this particular case, only the coordinate of $\boldsymbol{r}$ that is parallel to $\boldsymbol{u}_0$ will be relevant to the detection process. The optimum detector for this case is the well know correlation (matched filter) detector that will determine the sign of the correlation (scalar product) $\boldsymbol{r} \cdot \boldsymbol{u}_0$ yielding the estimate

$$\hat{a}_0 = sgn(\boldsymbol{r} \cdot \boldsymbol{u}_0) \tag{2.6}$$

In the general case, however, the conditions i) and ii) are not satisfied. In particular this will be the case when the number of interfering transmitters is small, or if a small number of interferers dominate the noise component $z'(t)$. The latter case tends to be quite common in many radio communication situations [2]. There are, however, also some interesting situations when the conditions i) and ii) are indeed satisfied. The most common situation is when $z'(t)$ consists of many signal components of roughly comparable energy. In this particular case, due to the Central Limit Theorem z' can be approximated by a zero mean Gaussian vector, thus satisfying conditions i) and ii). The receiver given by (2.6) is, in this extreme case, optimal.

Example 2.1

When detecting a BPSK-signal, the following waveform is used

$$s_0(t) = a_0 \sqrt{\frac{2E_0}{T}} \cos\left(2\pi \frac{t}{T}\right) \qquad 0 \le t < T$$

The reception is disturbed by interference that is dominated by a single PSK-modulated signal

$$s_1(t) = a_1 \sqrt{\frac{2E_1}{T}} \cos\left(2\pi \frac{t}{T} + \phi\right) \qquad 0 \le t < T$$

Other interference and thermal noise can be approximated by additive white Gaussian noise $n(t)$, with spectral density $N_0/2$. The information symbols $a_i \in \{-1, +1\}$ are independent and equally probable. The received signal $r(t)$ may be written as

$$r(t) = s_0(t) + s_1(t) + n(t)$$

a. What is the bit error probability achieved by a correlation detector according to (2.6)?

b. Is the correlation detector optimal in the ML sense?

Solution:

a. The received signal using the vector model is described in Figure 2.4 where it is assumed that the symbol $a_0 = +1$ is transmitted. The correlation receiver will only use the projection of the received vector $\boldsymbol{r}$ on $\boldsymbol{u}_0$ to make its decision. A detection error occurs if $\boldsymbol{r}$ falls in to the "wrong" half plane. If the symbol $a_0 = +1$ was transmitted, the detector will make an erroneous decision if the projection of $\boldsymbol{r}$ on $\boldsymbol{u}_0$ becomes negative.

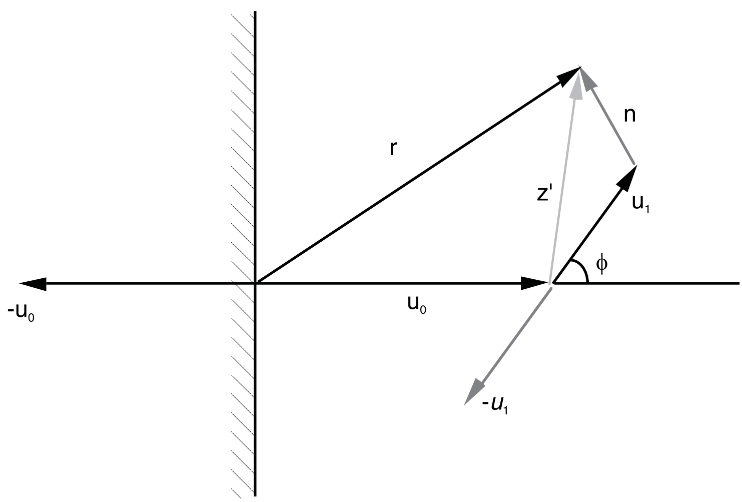

Figure 2.4 Signal constellation in example 2.1.

The projection (scalar product) is given by

$$\boldsymbol{r} \cdot \boldsymbol{u}_0 = a_1 \sqrt{E_1 E_0} \cos \phi + \sqrt{\frac{E_0 N_0}{2}} n_c$$

where n_c is a zero mean Gaussian stochastic variable with unity variance. Dividing by $\sqrt{E_0 N_0 / 2}$ the scalar product becomes negative and an erroneous decision is made if

$$n_c < -\sqrt{\frac{2E_0}{N_0}} - a_1 \sqrt{\frac{2E_1}{N_0}} \cos \phi$$

The probability of this event is given by

$$P(\text{error} \mid a_0 = +1;\ a_1) = Q\left(\sqrt{\frac{2E_0}{N_0}} + a_1 \sqrt{\frac{2E_1}{N_0}} \cos \phi\right)$$

Since a_1 takes the values 1 and −1 each with probability 1/2 and the situation for $a_0 = -1$ is completely symmetric the error probability may be written as

$$P_e = \frac{1}{2} Q\left(\sqrt{\frac{2E_0}{N_0}} + \sqrt{\frac{2E_1}{N_0}} \cos \phi\right) + \frac{1}{2} Q\left(\sqrt{\frac{2E_0}{N_0}} \sqrt{\frac{2E_1}{N_0}} \cos \phi\right)$$

Note here that if the interfering signal is orthogonal to $\boldsymbol{u}_0$ (i.e. $\cos \phi = 0$), it will not have any impact at all and the resulting error probability will be the same as in just the Gaussian noise with no interference present.

The error probability as a function of the signal-to-noise ratio is shown in Figure 2.5 where the signal-to-interference

$$\eta = \frac{E_0}{E_1 \cos^2 \phi} = \frac{E_0}{E_1 \theta_2}$$

is used as a parameter. θ denotes the normalized scalar product or *cross correlation* between the signals, defined as

$$\theta = \frac{\mathbf{u_0} < \mathbf{u_1}}{|\mathbf{u_0}||\mathbf{u_1}|} = \frac{1}{\sqrt{E_0 E_1}} \int_0^T u_0(t) u_1(t) dt$$

$E_1 \theta$ can be seen as the *effective* interference energy.

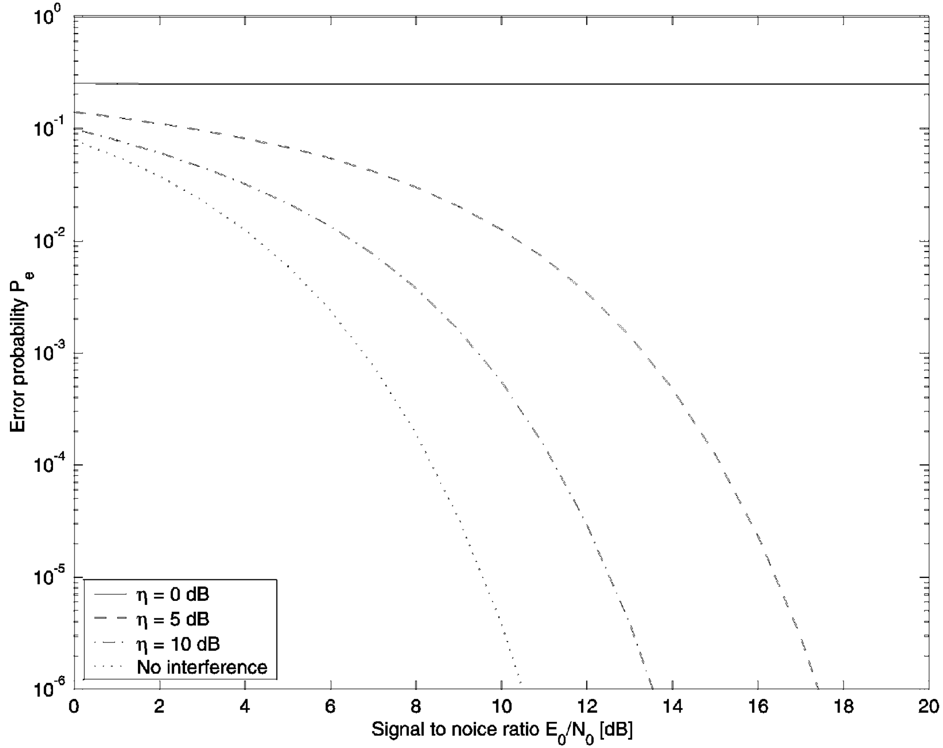

Figure 2.5 Bit error probability for correlation detector in Example 2.1 as function of the signal-to-noise ration with the signal-to-interference ratio η as parameter.

b. Let us study the interference vector

$$\boldsymbol{z}' = \boldsymbol{s}_1 + \boldsymbol{n} = a_1\boldsymbol{u}_1 + \boldsymbol{n}$$

and its probability density function

$$p_{\boldsymbol{z}'}(\boldsymbol{z}') = \frac{1}{2}p_{\boldsymbol{n}}(\boldsymbol{z}' - \boldsymbol{u}_1) + \frac{1}{2}p_{\boldsymbol{n}}(\boldsymbol{z}' + \boldsymbol{u}_1)$$
$$= C\exp(-|\boldsymbol{z}' - \boldsymbol{u}_1|^2/N_0) + C\exp(-|\boldsymbol{z}' + \boldsymbol{u}_1|^2/N_0)$$

Here, $p_{\boldsymbol{n}}$ is the probability density of the noise vector $\boldsymbol{n}$. $p_{\boldsymbol{n}}$ has its extreme value (maximum) at the origin and satisfies the conditions i) and ii) above. $p_{\boldsymbol{z}'}$ thus has two maxima, one around the vector $\boldsymbol{u}_1$, and one around the vector $-\boldsymbol{u}_1$, and cannot satisfy the two conditions. The correlation receiver is therefore not optimal for this case.

The reader is referred to [3] for a more thorough investigation of this example. Also, in the references, the cases where the signals are known up to some parameters, for example, the phase ϕ or the amplitudes are treated. Receiver designs that more or less explicitly exploit knowledge about the properties of the other interfering signals, have in the literature been labeled multiuser detectors. One has to be careful to distinguish between a true multiuser detector and an ML single user receiver. In the first case a whole set of data symbols from different transmitters are to be decoded simultaneously. This is typically the case in the base station of a mobile telephone system. In the second case only one signal is actually detected and the other data symbols are treated as unknown, but irrelevant, parameters. Example 2.1 illustrates this latter case, which is typically found in the terminal in a wireless network. The reader is referred to [4] for a more thorough treatment on advanced detection schemes.

Orthogonal Signaling

In current radio systems, most signal sets are chosen to be orthogonal, i.e. satisfying the condition

$$\mathbf{u}_j \cdot \mathbf{u}_i = 0 \qquad i \neq j,$$

corresponding to $\theta = 0$. With techniques similar to the one in the example above, one may show that if the interfering signals are orthogonal to the

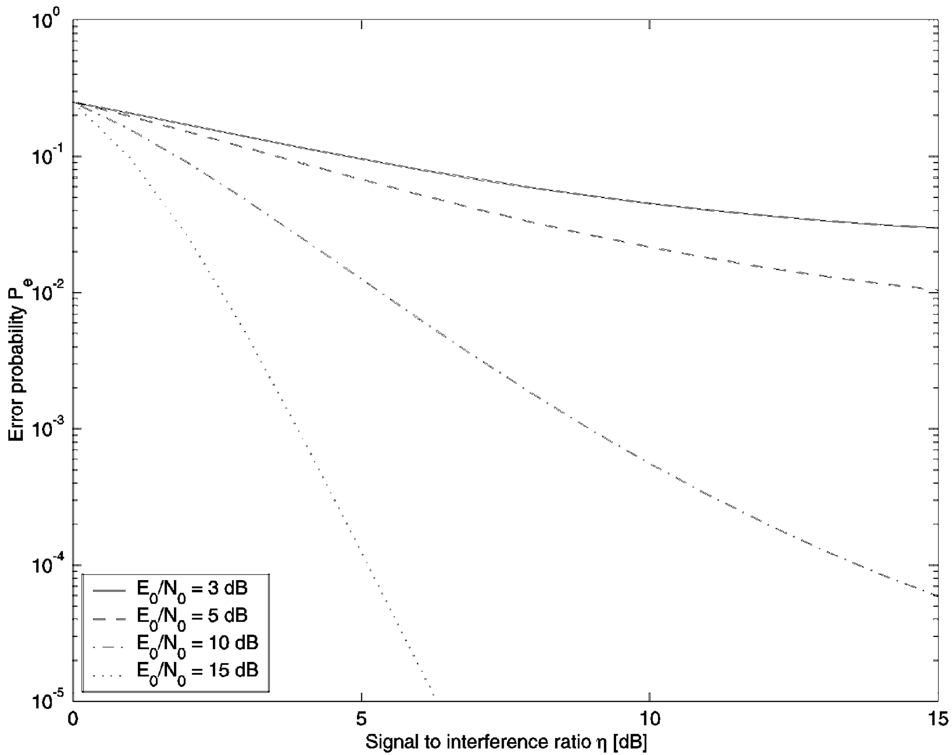

Figure 2.6 Bit error probability for correlation detector in Example 2.1 as function of the signal-to-interference ratio ration with the signal-to-interference ratio η as parameter.

wanted signal, the simple correlation receiver will be the optimum choice in the ML-sense. In fact, the effect of the interference may be completely eliminated. In Example 2.1 above, this is evident from the fact that the effective signal-to-interference ratio grows to infinity as the angle ϕ approaches 90°. The price paid for this lack of interference is that the design of orthogonal signal sets requires considerable bandwidth. It is obvious from the example that adding yet another orthogonal signal to an existing orthogonal set will require one additional dimension in the vector space. The number of dimensions, in turn, is intimately coupled to the required bandwidth. Careful studies of the properties of orthogonal signal sets [5], yield a lower bound on the required bandwidth (for any waveform). The number of orthogonal waveforms N of duration T that can exist in a bandwidth W is limited by

$$N \leq 2WT \tag{2.7}$$

In signal sets of the size given in (2.7) there may be signals that only differ in the carrier phase. Signal sets containing such signal pairs are, of course, not suited for noncoherent detection. If only signal sets that can be distinguished without a phase reference (i.e., possible to detect noncoherently) are considered, the relationship above becomes

$$N_{nc} \leq WT$$

It may not come as a surprise that for a constant data rate $(1/T)$ the minimal required bandwidth is directly proportional to the number of signals. Further, it may be noted that the performance (e.g., bit error probability and number of signals) does not depend on the explicit waveforms, but only on the correlation properties of the signals. Every reasonably, carefully selected set of orthogonal signals will, in principle, exhibit the same communication theoretic performance. The preference for a certain type of waveform is dictated by other reasons, typically of an implementational nature. A few of the most popular waveforms are now investigated.

Maybe the most straightforward signal set design yields the class of Time Division Multiplex (TDM) signal sets. The orthogonality condition gives

$$\mathbf{u}_j \cdot \mathbf{u}_i = c\int u_j(t)u_i(t)dt = 0$$

The simplest way to satisfy this condition is to let the integrand become zero, that is, let

$$u_j(t) = 0 \quad \Rightarrow \quad u_i(t) \neq 0$$
$$u_i(t) = 0 \quad \Rightarrow \quad u_j(t) \neq 0$$

The obvious interpretation of this is that information is transmitted in one of the signals at a time (whereas all other signals are zero). Typically the signals are chosen such that each signal $u_i(t)$ is not zero in one unique time interval—a fraction of the symbol time T. If the time intervals of the different signals (users) are not overlapping, then the orthogonality condition above is satisfied. The receiver of a certain signal may concentrate all its efforts to this particular time interval (slot) and ignore the received signal in the rest of the symbol interval. Figure 2.6b illustrates the waveform and spectrum of some time-multiplex signals.

In a similar fashion, it is possible to utilize signals that occupy disjoint frequency intervals. Using Parsevals relation

$$\mathbf{u}_j \cdot \mathbf{u}_i = c\int u_j(t)u_i(t)dt\ \mathrm{dt} = c\int U_j(f)U_i^*(f)df \tag{2.8}$$

where $U_i(f)$ denotes the Fourier transform of $u_i(t)$ and $U_i(f)$ its complex conjugate. It can again be noted that the signal set is orthogonal if $U_i(f)$ and $U_j(f)$ are nonoverlapping, that is,

$$U_i(f) \neq 0 \qquad U_j(f) = 0 \ \forall \ i \neq j$$

Here, different users use disjoint frequency ranges (channels) to communicate. This class of signal sets, the frequency division multiplex (FDM) is the second basic principle for designing orthogonal signal sets. Figure 2.7a illustrates these waveforms and their spectra.

A problem arising when using orthogonal signals is the impact of the channel filter on the correlation properties of the signal. To achieve interference-free communication, the signals of the signal set have to be orthogonal at the output of the channel filter. In general, this is a difficult problem since the channel affects the relative positions of the signal vectors which in turn may cause interference (cross-talk) even though the originally transmitted waveforms were orthogonal. A simple example of this is when a TDM scheme is used in a channel that has band-limited characteristics. The channel will cause time dispersion in the transmitted pulses, which

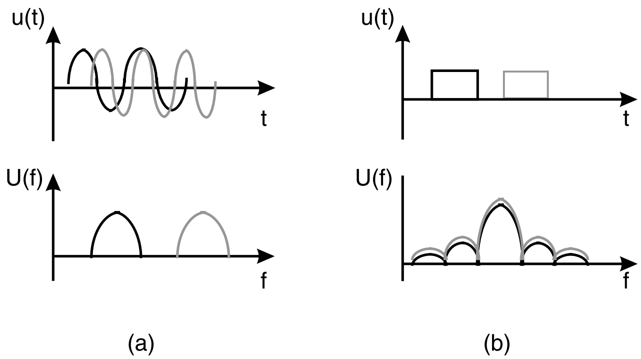

Figure 2.7 Frequency- and time-multiplex signals.

means that a message transmitted in one time-slot will partially overlap with successively transmitted messages from others. An FDM scheme is not that strongly affected by this type of channel unless the channel including the receiver can be described by a linear model. In a linear system, as in our model, no new frequency components are generated, and FDM signals remain orthogonal. However, if the channel is time varying (which is typical) or the receiver contains nonlinearities, new and overlapping frequency components may be generated (intermodulation (IM) distortion).

Using orthogonal signals is the predominant practice in most contemporary radio systems. Results from information-theory, however, indicate that there exist even larger nonorthogonal signal sets that could provide reliable communication despite the resulting interference. Unfortunately these results are not constructive, that is, no indication is given as to how a system should be designed to actually achieve these signal sets with size exceeding the orthogonal bound (2.7). This problem is further addressed in Section 2.1.4.

2.1.3 Basic Orthogonal Multiplex Schemes

Frequency Division Multiple Access (FDMA)

In radio communication history, the most popular multiplexing principle is frequency division multiplex (and frequency division multiple access, FDMA). As has been noted, FDM/FDMA means that signals that have disjoint (nonoverlapping) spectra are used. In practice, the available bandwidth is subdivided into a large number of narrow band-pass channels. If two stations choose to communicate, they (in some way) select a vacant channel. The optimum matched filter receiver consists in principle of a narrow band-pass filter selecting the appropriate signal. After this, the signal may be easily detected since the filtering process eliminates all the adjacent channel interference. All kinds of bandpass modulation techniques, analog as well as digital, with bandwidths small enough to fit into the bandpass channels may be used to convey the information (Figure 2.8).

There are several distinct advantages that have made FDMA techniques immensely popular during the entire radio communication era. The main reason is perhaps that FDMA schemes are well suited for analog circuit technology. The basic operations in FDMA transmitters and receivers are filtering and mixing of high-frequency signals for which passive and active analog circuits are ideal. Another advantage is simplicity. Although FDMA systems need to achieve a reasonable accuracy in frequency, they do not require any time or phase synchronization. The transfer of information in the separate channels occurs independently of each other, which for instance, allows the mixing of analog and digital information in the radio system.

In the early days of radio, when information rates were low compared to the available bandwidths, there was no problem in separating signals well enough in the frequency domain. The result was that the requirements on filters and absolute frequency accuracy were rather moderate. Transmitters and receivers were simple and robust devices.

With the demand for higher data rates and the ever increasing number of users, more and more signals are forced to share a limited bandwidth. These developments made increasing demands for frequency accuracy and receiver selectivity, that is, the capability of the receiver to extract the wanted signal from the multitude of signals on air. In addition, transmitters are

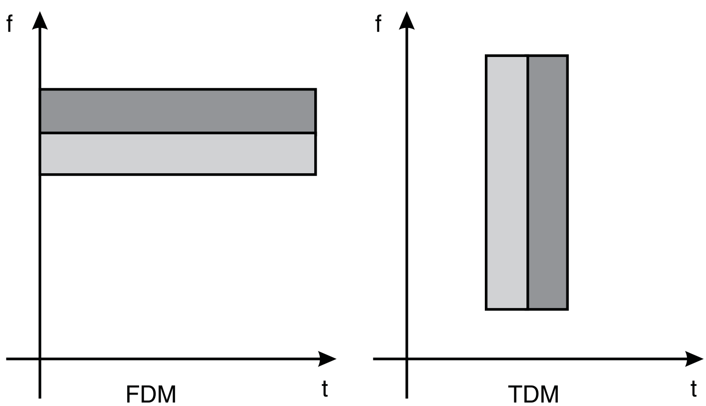

Figure 2.8 Time/Frequency diagram for TDMA and FDMA signals in a system with two active stations.

required to emit very few unwanted signal components, so-called spurious signals (lacking orthogonality), in order not to interfere with adjacent signals. In radio systems, this is a very important problem, since the dynamic range of signals at the receiver, that is, the ratio of powers between the strongest signal and the weakest signal, may be extremely large. As an example of this, consider a receiver receiving a weak signal from a distant station at the same time as it receives a signal (in an adjacent channel) but with a signal power more than 100 dB larger than the weaker signal. Extreme linearity is required in the receiver amplifiers to ensure that the output signal will still contain only the sum of these input signals, without any harmonic and/or intermodulation (IM) products appearing at frequencies other than the original signals. The design of a receiver capable of handling these situations is complex and costly.

Ever more narrowband signals introduce ever more stringent requirements on frequency accuracy in both transmitters and receivers. This causes problems mainly at high carrier frequencies (i.e. in the GHz range). The most severe drawback of FDMA technology is that the bulk of the operations in the receiver involves radio-frequency (RF) bandpass filtering, which is inherently an analog process. The production process of such devices requires either expensive high-precision analog components, or costly manual adjustments of every single device. In addition, analog RF technology is less suited for high-density VLSI implementation.

Practical examples of FDMA-based systems can be readily found in day-to-day life. In particular the analog cellular telephony systems could be mentioned. The Nordic Mobile Telephone system (NMT) was the first large-scale, commercial, fully automatic, wireless telephone system (1981). This system, and its similar systems (AMPS, TACS, and so forth) soon to follow in other countries, use a large number of narrowband channels of 25 kHz (in the United States 30 kHz) bandwidth for analog FM transmission between base stations and mobile stations. A truly full duplex voice communication link is provided by these systems and radio transmission takes place in two channels, the mobile-base (up) and base-mobile (down) channel. These duplex channels are generally separated by 10-20 MHz to allow for simultaneous transmission and reception in the mobile stations. Most analog telephone systems initially operated in the 450 MHz range, but soon expanded into the 900 MHz range.

Time Division Multiple Access (TDMA)

A multiple access scheme highly suited for digital transmission is time multiplexing (time division multiple access, TDMA). Instead of assigning only a small part of the available bandwidth to each station, all stations use the entire signal bandwidth but are confined to short, nonoverlapping time-intervals. Usually, the stations are assigned short time slots, which are repeated in a cyclic (round robin) fashion. The modulation schemes that can be used for FDMA systems can be used here as well, provided they are scaled to the larger bandwidth. It must, however, be noted that the transmitted information has to be in some time-discrete representation. The number of time slots in one cycle (or frame) is as large as the maximum number of stations that are capable of communicating simultaneously. The number of slots is equivalent to the number of channels in FDMA systems (Figure 2.8). Provided the same modulation scheme is used in both systems, the same amount of information can be transferred in the same interval in a given bandwidth in both systems.

The requirements for selectivity and frequency stability are considerably lower than for an FDM system. The task of the receiver bandpass filtering is only to eliminate out-of-band interference, and not to distinguish between different transmitters in the band used. Instead, the requirement of time accuracy, that is, synchronization, is considerable. The receivers are required to distinguish their particular time slot from the time slots of other users. This may not be an easy task, in particular if the propagation delays in the systems are comparable to the slot duration. A synchronization scheme has thus to solve two problems:

1. The classical problem of determining in which slot the wanted signal is located (this may even vary from frame to frame in some systems);
2. Transmitting/receiving accurately within the wanted slot in order to avoid overlap with other signals (lacking orthogonality).

Achieving nonoverlapping signals in all the receivers in the system is not always easy. The simplest solution is to leave some fraction at the edges (so-called guard intervals) of the time slot unused, thus allowing for small timing errors. This is analogous to leaving some portion of the spectrum between signals unused in an FDMA system. The obvious drawback of introducing guard intervals is the waste of time, which will lower the effective data rate that can be achieved in the system. When a high-efficiency is required, effective synchronization schemes are needed in order to keep the required guard intervals as short as possible. In star-shaped radio networks where many stations communicate only with a central station (e.g., mobile telephony) there is an effective solution to this problem. Here, the clock of the central station may be used as a time reference. Knowing the propagation delay to the central station would enable the peripheral stations to adjust their clocks. The peripheral stations would then transmit their messages slightly early in order to let the central controller receive the packet exactly with the proper time slot. The required remaining guard interval has now only to be in the order of the timing inaccuracy of the path delay estimates, not the whole path delay. However, as the following example illustrates, this scheme will not work if the network is not star-shaped (hierarchical).

Example 2.2

A radio system consisting of three stations in a network for tactical communication has a topology illustrated by Figure 2.9. The system uses TDMA in a half-duplex mode, that is, transmission and reception of messages cannot be simultaneous. Each station will transmit in its own time slot, and can receive messages from the other two stations in the two other time slots. The stations are using a fixed time reference and begin their transmission exactly at the beginning of a time slot. What is the minimum guard interval τ_g that is required to avoid all message overlapping. Can the starting times of message transmissions be delayed to avoid overlap?

Solution:

The speed of light can be expressed as 300 m/μs. The path delays τ_{ij} between the stations can thus be computed to be 10, 15, and 20 μs. Study the

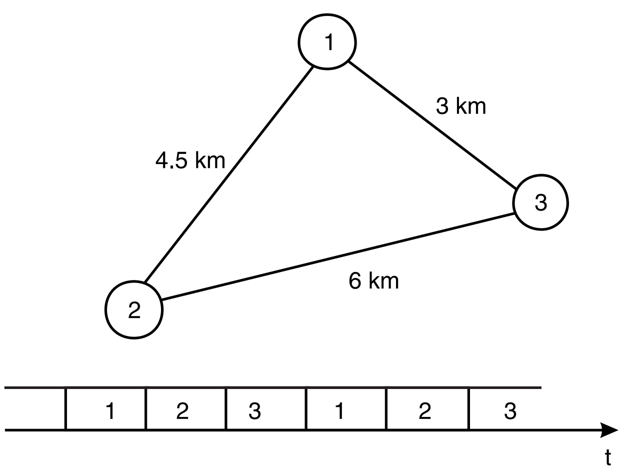

Figure 2.9 Example 2.2.

messages during reception in Figure 2.9. The most severe overlap is found in receiver 3 (time slot 3) while transmitting during the reception of a message from station 2. A message in slot 2 overlaps by 20 μs in slot 3. The guard time at the beginning (or end) of each slot must thus be at least 20 μs. In general τ_g is given by

$$\tau_g \geq \max \tau_{ij}$$

Can better results be achieved by adjusting the starting times of the transmissions? It will be seen that by delaying the starting time of transmissions by 5 μs in slot 3 the maximal overlap, and thus the guard interval can be reduce to only 15 μs.

Due to the synchronization problems, TDMA-based systems had not been common in radio communication, and time multiplexing had been mainly confined to wired transmission systems. The advent of digital signal processing and VLSI technology has radically changed this, and many TDMA-based radio systems have been developed in recent years. Maybe the most spectacular examples are found among the digital mobile telephony systems, for example, the pan-European GSM system (Global System for Mobile communication). The network structure of these systems is quite similar to the structure of their analog counterparts.

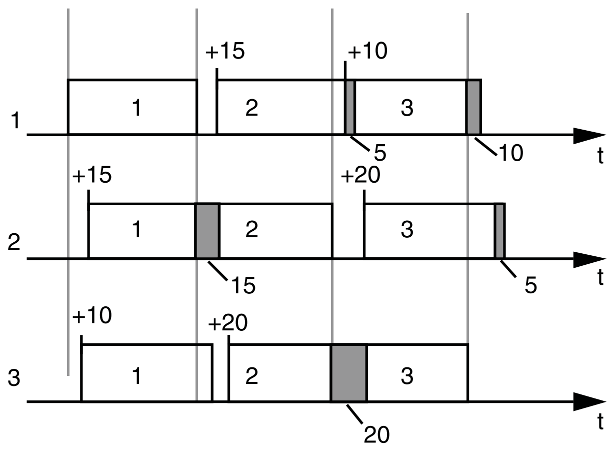

Figure 2.10 Timing diagram for reception of messages at different stations in Example 2.2. The largest message overlap is found at station 3 where two messages from station 2 and 3 overlap by 20 μs.

Regarding channel access, there are, of course, large differences. Both the D-AMPS as well as the GSM-system can be said to be FDMA/TDMA hybrids. The GSM system uses a number of approximately 200 kHz wide FDMA-channel, each of which in turn is subdivided into 8 TDMA channels for speech traffic and control information (Figure 2.11). Frames of slots of 0.577 ms each are repeated about 220 times per second. In each slot, each station transmits a burst containing 114 information bits. The gross data rate in the system is 271 Kbps. The base-mobile and mobile-base transmission occurs on separate frequency channels, even though this, at least in principle, would not be necessary. The system utilizes path delay compensation, keeping the guard interval to a relatively low value. From Figure 2.11 it can be seen that the guard interval corresponds to roughly 8 bits or about 30 μs. The fraction of wasted time is as low as approximately 5%.

2.1.4 Spread-Spectrum and Nonorthogonal Multiplexing

There is a group of multiple access techniques that are not easily classified in terms of time and frequency multiple access. These methods are often, truly or falsely, denoted as spread-spectrum or code division multiple access (CDMA) schemes. These schemes are characterized (similar to TDMA) by signals with a bandwidth much larger than 1/T. The two most popular

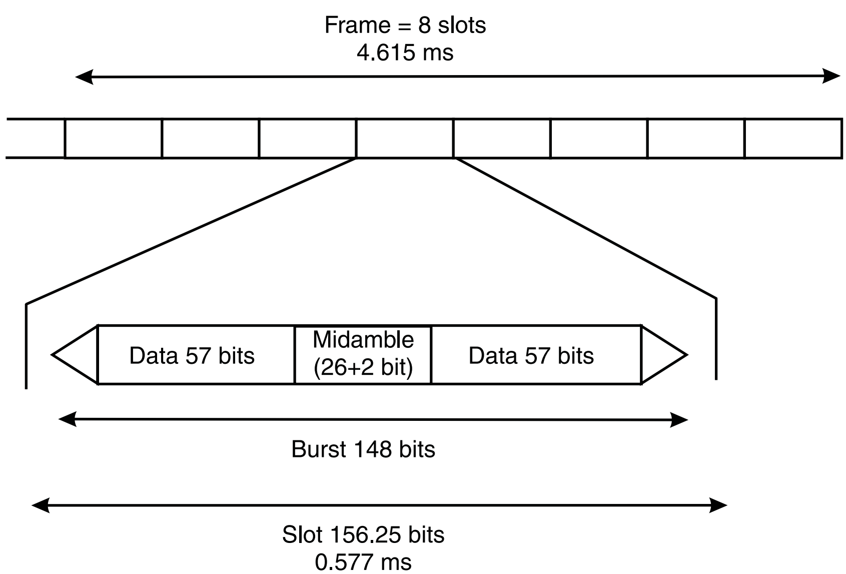

Figure 2.11 TDMA-frame structure in GSM system for mobile telephony system [6].

schemes in this class are denoted frequency hopping (FH) systems and direct sequence (DS) systems. These two techniques will now be studied briefly.

Frequency Hopping Systems

An FH system is, in principle, a combined time and frequency multiple access scheme. Similar to the traditional FDMA systems the available bandwidth is divided into a number of narrow channels. In addition, similar to TDMA, time is also divided into slots. The stations transmit narrowband signals in one of the channels during a time slot, a chip. In the subsequent time slot the station keeps transmitting, but on a new frequency channel. The station thus "hops" from frequency to frequency (Figure 2.12). The sequence of frequencies used by the transmitter is denoted as the hop sequence. All transmitters use unique but predetermined hop sequences. The (narrowband) receiver follows the same hop sequence, thus tracking the transmitter in every time slot. Since the signal in every slot is of a narrowband character, detection is done with conventional (FDMA) techniques. If the hop sequences are chosen such that no chips will overlap, it is obvious that the FH signals are orthogonal and such a system will be capable of transferring the same amount of information as an FDMA or TDMA system occupying the same bandwidth.

One usually distinguishes between fast frequency hopping (FH) systems and slow frequency hoppers (SH). In a fast hopping scheme, only one symbol (or less) is transmitted in every time slot. The hopping rate is thus equal (or larger) than the data rate. In a slow hopping system, the hopping rate is less than the data rate and several symbols, or even whole messages are transmitted in each chip. An example of the latter is GSM where the specifica-

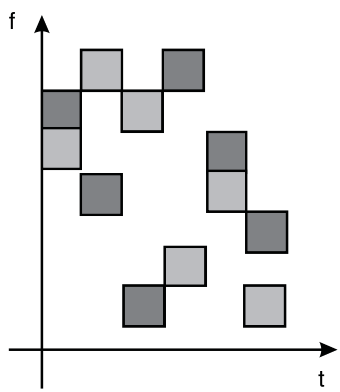

Figure 2.12 Time/frequency diagram for an FH system with two active transmitters.

tion allows for frequency hopping, transmitting one entire burst (>100 bits) on each frequency.

Unfortunately, the FH system combines the two largest drawbacks of the time and frequency multiplexing schemes. Here, both frequency selectivity (to distinguish between users) as well as very accurate synchronization are required. The latter is imperative in order to be able to track the transmitter with any success. Additional complexity is due to the rapid change of frequency required. What are the advantages then that make these systems so interesting despite the implementational problems? One could say there are basically two reasons, or applications, where FH is of great importance—resistance to adverse propagation conditions, and its capability to withstand larger amounts of interference (accidental or intentional).

Using a large number of frequencies makes a well-designed FH system highly resistant to narrowband frequency selective fading. In such a fading environment, certain frequency channels will be exposed to deep fades, whereas most of the other channels will work well. A frequency hopper will be subject to these deep fades now and then, but will never stay long in such a fade. In fact, in a Raleigh fading environment with reasonably average power, a vast majority of the frequencies will provide adequate signal power. Combining frequency hopping with error correction coding will be highly capable of correcting the errors occurring when the system hits the fading minimum. This technique is extremely useful, in particular, when considering mobile communication systems with slowly moving stations. In a narrowband system, when the receiver becomes stationary and happens to find itself in a fading minimum, error correction techniques described in the previous chapter are practically useless. The decoder will hardly receive any symbols of adequate quality to correct the erroneous ones. Here, frequency hopping will make a stationary receiver "move around" in the standing wave pattern around him. In addition, if, in a fast FH system, the chip duration is small compared to the delay spread (e.g., in FH systems for the HF range), the receiver will hop to another frequency before the delayed multipath components have the chance to reach the receiver. In this case, the receiver effectively "hops away" from the intersymbol interference.

Example 2.3 Frequency Hopping in Rayleigh Fading

A binary digital radio link utilizes DPSK modulation and frequency hopping. The transmission is disturbed by Gaussian noise and successive bits can be assumed to be received with independent Rayleigh fading amplitudes. A bit error probability of 10^{-4} is required.

a. What is the required SNR if no coding is used?
b. What is the required SNR if a simple single error correcting Hamming (15,11) code is used?

Solution:

a. The received power (SNR) in a Rayleigh fading channel [7] is exponentially distributed, that is,

$$p(\gamma) = \frac{1}{\gamma_0} e^{-\gamma/\gamma_0}$$

where γ_0 is the average SNR. Now, the bit error probability for a DPSK link given a constant SNR γ, is

$$P_e(\gamma) = \frac{1}{2} e^{-\gamma}$$

Combining these two allows removing the conditioning on γ:

$$P_e = \int P_e(\gamma) p(\gamma) d\gamma = \int \frac{1}{2} e^{-\gamma} \frac{1}{\gamma_0} e^{-\gamma/\gamma_0} d\gamma = \frac{1}{2(1 + \gamma_0)} \approx \frac{1}{2\gamma_0}$$

Requiring a P_e of 10^{-4} yields an average SNR $\gamma_0 = 5000$ (37 dB)

b. The code can correct one error in 15 transmitted bits. The code word error probability for low error probabilities becomes:

$$\begin{aligned} P_{cw} &\approx \Pr[\leq 1 \text{ error in codeword}] \\ &= (1 - P_e)^{15} + 15 P_e (1 - P_e)^{14} \approx 14 \cdot 15 P_e^2 \end{aligned}$$

Using the common approximation [7]

$$P_e' \approx \frac{d_{\min}}{n} P_{cw} \approx \frac{3}{15} 14 \cdot 15 P_e^2 = 42 P_e^2$$

Using the result from the a) part and noting the fact that only 11/15 of the energy is spent on the transmission of information bits we get

$$P_e' \approx 42P_e^2 = 42\left(\frac{15}{2 \cdot 11\gamma_0}\right)^2 \approx \frac{20}{\gamma_0^2}$$

Requiring a P_e of 10^{-4} yields an average SNR $\gamma_0 = 450$ (26 dB)—a gain of more than 10 dB.

The powerful error correction that can be used in the FH system also has another application. It is possible, for instance, to allow more users than there are frequencies into the available bandwidth. By doing so, collisions, that is overlapping chips, are inevitable. The signals in such a system will no longer be orthogonal (2.7). However, if the excess number of transmitters is moderate, a particular receiver will be hit only now and then, and error correction coding may still be able to recover the original message transmitted. The advantages of such a technique are obvious: We could allow more users into the systems at the price of a moderate performance degradation. Unlike the orthogonal schemes, this multiplexing technique has no definite upper limit on the number of users. Instead the maximum number of transmitters will be determined by the required reception quality.

This capability to withstand interference has been a feature of great interest in military communication systems. In these applications, a hostile party in several ways threatens a communication link. In particular, the enemy may choose to deliberately transmit signals, so-called *jamming signals*, with the explicit purpose of disrupting the transfer of information in the link. Consider the scenario in Figure 2.13 where a link has been established between the transmitter A and the receiver B. A hostile jammer, J, observes the signals from A, and based on these observations tries to transmit signals in order to make the reception in B of the wanted signals as difficult as

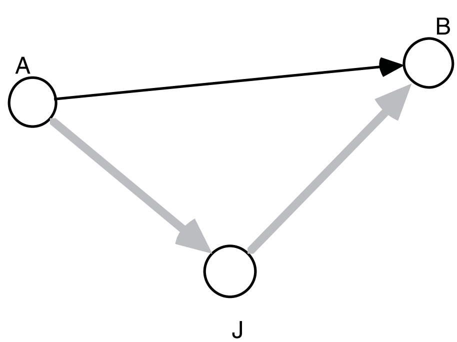

Figure 2.13 Jamming scenario.

possible. Clearly, if A uses an FDM or TDM scheme, or even slow frequency hopping, J would be able to detect the signals from A and immediately transmit a signal on the same frequency. However, if the FH system hops fast enough, the receiver B will already be at some other frequency when the jamming signal hits it. Let τ_{AB}, τ_{AJ} and τ_{JB} denote the path delays between the stations in Figure 2.13, we see that the condition for this to happen is

$$(\tau_{AJ} - \tau_{JB}) - \tau_{AB} \geq T_c$$

where T_c denotes the chip duration. In this case the jammer cannot rely on the observations of the signals from A but has to guess where A will be transmitting the next time. If the hop sequence is such that it appears to be random with every frequency equally probable, the jammer is completely in the dark and may as well randomly jam as many frequencies as possible. This is illustrated in the example below.

Example 2.4 Partial Band Jamming

A frequency hopping system hopping over random L frequencies is being jammed by a so called *partial band jammer*. This jammer randomly selects a fraction q of the frequencies and transmits a jamming signal concentrating all its jamming power on these frequencies. For the sake of simplicity, assume that if the signal to interference ratio (SIR) at the receiver drops below γ_0, the chip is lost (bit error probability $P_b = 1/2$), otherwise the chip is received perfectly ($P_b = 0$). Assume that the wanted signal energy per bit is E_b and that the jammer has energy E_J at its disposal. Estimate the bit error probability as function of the energies and q. Which value q will achieve the maximum bit error probability?

Solution:

Since the jammer distributes its energy evenly over qN frequencies, the SIR at the receiver becomes

$$\Gamma = \frac{qNE_b}{E_J}$$

If Γ is below γ_0 no errors will occur, otherwise a fraction of q of the symbols will be hit and received with $P_b = 1/2$. We can express this as

$$P_b = \begin{cases} \frac{1}{2}q & \frac{qNE_b}{E_J} = \gamma_0 \\ 0 & \frac{qNE_b}{E_J} < \gamma_0 \end{cases}$$

We can clearly see that the jammer should choose q such that the SIR falls just barely below threshold,

$$q^* = \min\left(1, \frac{\gamma_0 E_J}{NE_b}\right)$$

and the corresponding error probability becomes

$$P_b^* = \min\left(\frac{1}{2}, \frac{\gamma_0 E_J}{2NE_b}\right)$$

The bit error probability decays inversely proportional to the wanted signal energy (Raleigh fading). This result holds also when using a more detailed model to describe the bit error probability as a function of the SIR.

The bandwidth expansion factor N is usually called the processing gain of the system. It can be seen from the final expression in the example that the system achieves the same performance as a single channel system with a transmitter power that is N times larger than in the frequency hopping system.

The choice of hop sequence clearly depends on the application. In the jamming example, the hop sequence has to appear randomly, that is, be impossible to predict for a jammer or eavesdropper. For the civilian application, as a countermeasure against fading, this does not seem to be critical, as long as all frequencies are used regularly. In a system with N frequencies and a hop sequence length of L time slots, there are

$$M = (N!)^L \tag{2.9}$$

different orthogonal sequences. The number of feasible sequences grows very fast with the sequence length. In civilian systems with no hostile interference short sequences that are repeated cyclically are well suited to do the job. Such a short sequence may be easily detected, and future frequencies can be readily predicted which make the synchronization process fast and reliable.

In military applications however, the sequences have to be long and hard to decipher. This requires a large L. If such a sequence is chosen and communicated secretly to the receiver, it would be virtually impossible for the jammer to predict the next chip-frequency. There are two practical problems involved with this. First, synchronization becomes slow and complex. In the worst case, we would have to wait one full hop sequence cycle (or in noise even more!) before transmission could start. The other problem involved with long sequences is that we would need a compact way of describing which one of the M sequences that we have chosen. Just enumerating all M sequences requires a number of $L \log_2 (N!) = L(N/2) \log_2 N$ bits. To give a realistic example, if $L = 1000$ and $N = 100$ there would be 330 000 (!) bits required to fully specify the hop sequence. For obvious reasons, in practice only those subsets of all feasible hop sequences that have compact descriptions are used. These are called *pseudo noise* (PN) sequences that are generated by linear feedback shift registers (LFSRs). We will study some of the properties of these sequences in the next section.

Direct Sequence Systems

Direct sequence (DS) modulation represents the other classical spread spectrum technique. DS systems use long and complex, but usually binary waveforms typically of the shape,

$$u_i(t) = \sum_{k=1}^{N} c_{ik} p(t - k\tau) \qquad c_{ik} \in \{+1, -1\} \tag{2.10a}$$

where

$$p(t) = \sqrt{E_0 / N} \qquad 0 \le t \le \tau \tag{2.10b}$$

that is, a rectangular pulse of duration τ. The name "direct sequence" is derived from the fact that the data is directly and antipodally (PSK) modulated on the pulse train, the code or code sequence $\boldsymbol{c}_i = \{c_{ik}\}$.

The transmitted signal $s_i(t)$ could thus, according to (2.1), be written as

$$s_i(t) = a_i u_i(t)$$

The time interval $\tau = T/N$ is also here denoted a "chip" (Figure 2.14). N is usually a large number, which yields a short chip duration and thus a

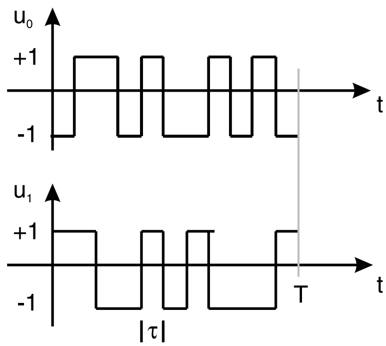

Figure 2.14 Example of signals in DS systems, N = 11.

large bandwidth. The signal in (2.10) has a *chip rate* $1/\tau$ that exceeds the information rate $1/T$ by a factor N (i.e., bandwidth expansion factor N).

Deriving the scalar product of two such signals yields

$$\mathbf{u}_i \cdot \mathbf{u}_j = \sum_{\kappa=1}^{N} c_{ik} c_{jk} p^2(\tau - kt) = \frac{E_0}{N} \sum_{\kappa=1}^{N} c_{ik} c_{jk} \tag{2.11}$$

The last sum is the cross correlation of the two codes $\boldsymbol{c}_j$ and $\boldsymbol{c}_i$ (Example 2.1). To achieve orthogonal signaling, this sum has to be zero for different i and j.

In many DS systems with high chip rates and long code lengths, the synchronization is difficult since variations in propagation delay make it virtually impossible to synchronize all users. In addition, delayed versions of the original signal may be received due to multipath. To avoid the multipath problem and to ease synchronization, one aims at designing sequences that are self-orthogonal, that is, where all cyclically shifted version of the sequence are orthogonal. Synchronization now in principle becomes simple—the optimum receiver would consist of a bank of correlators, each comparing the received signal with a delayed version of the signal. The code delay in the correlator with the winning output would be a measure of the time delay. When synchronization is established, only the output of the winning correlator will be used for the demodulation of subsequent symbols. This fact leads us to a simpler synchronization scheme—the sliding correlator indicated in Figure 2.15. Here a single correlator is used in which the sequence time offset (delay) is slowly varied (slid) between receiver and transmitter. Running the sequence generator at a slightly higher or lower chip rate can do this. Due to the self-orthogonality property, the correlator

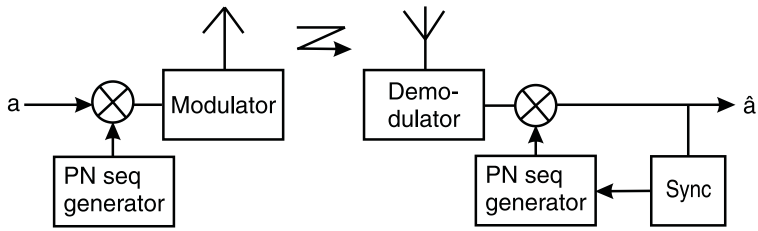

Figure 2.15 DS spread spectrum system.

will not produce any output (but noise) until the correct delay is reached. When a signal of sufficient strength is reached, the correlator is locked.

Being very convenient for synchronization, the self-orthogonality property of a code has a serious drawback. From the definition of the $u_i s$ in (2.10) it is clear that all the signals of this type can be described in a vector space of (at most) N dimensions (e.g., use $\phi_i(t) = p(t - i\tau)$ as base functions). Obviously there cannot be more than N orthogonal wave forms in this vector space. By using a self-orthogonal code, that is, a code where all N cyclic shifts are orthogonal, we effectively use up all dimensions. Shifted versions of the other signals $u_i(t - j\tau)$ can thus not be orthogonal to u_i. In practice one will therefore have to trade off synchronization properties against interference (cross-correlation) properties.

Going towards practical code design, let us first study the interference properties. Let r_{ijl} denote the cross-correlation between signal $u_i(t)$ and $u_i(t - l\tau)$, that is, at time shift l. Then

$$r_{ijl} = \frac{1}{N} \sum_{\kappa=1}^{N} c_{i(k+l)} c_{jk} \tag{2.12}$$

The target is to make the r_{ijl} to be as small as possible. As was noted above, there are no codes with $r_{ijl} = 0$ (except when $i = j$ and $l = 0$), that is, codes that are both orthogonal and self-orthogonal. In fact, we have the following bound

$$\max r_{ijl} \geq \sqrt{\frac{M-1}{MN}} \approx \frac{1}{\sqrt{N}} \qquad \text{(Welsh bound)} \tag{2.13}$$

where the last approximation holds for large signal sets (i.e., large M). The largest cross correlation coefficient thus decays only as the square root of the code length. There are in fact classes of sequences (so called Kasami and

Gold sequences) for which the bound (2.13) is also an upper bound on the cross correlation. This cross correlation can, however, still be quite high. In a system with $N = 100$, we would have max $r_{ijl} \geq 0.1$, i.e. $0.1^2 = 1\%$ of the signal power will hit some other user.

A class of sequences with more favorable properties would be the purely *random sequences*. Such (binary) sequences $X_i = \{X_{i1}, X_{i2}, \ldots X_{iN}\}$ would be generated by a sequence of uncorrelated, balanced coin tosses. Every chip in this sequence would take the values +1 and −1 with equal probability independent of other chip values. Calculating the autocorrelations yields

$$R_{ijl} = \frac{1}{N}\sum_{\kappa=1}^{N} X_{(k+l)}X_{jk}$$

Note that R_{ijl} is a random variable with

$$E[R_{ijl}] \approx 0 \qquad \text{(unless } i = j \text{ and } l = 0\text{)}$$

$$Var[R_{ijl}] \approx 1/N \qquad \text{(unless } i = j \text{ and } l = 0\text{)}$$

For long sequences, due to the law of large numbers, these random sequences will have the wanted properties. Due to the central limit theorem the R_{ijl} will have roughly Gaussian distribution. A reasonable interference model is that the users affect each other as Gaussian noise with a relative power that is N times lower than the wanted signal. However, for moderate N, the variance of the cross correlation is large, and a coin tossing process may produce a sequence with bad properties.

Random sequences are of course not very practical to generate. The amount of information describing the sequence that has to be communicated to the receiver is simply too large (actually the entire sequence would have to be shared by the receiver and transmitter). A class of sequences that retains the favorable correlation properties, but are easy to describe, are the so-called Pseudo Noise (PN) sequences generated by Linear Feedback Shift Registers (LFSR) (Figure 2.16).

The LFSR is a (binary) shift register where the delay element outputs are weighted and summed. The result is fed back to the input of the register. Note that the delay element contents, weight coefficients, and the sum are taken over the field of binary numbers, that is, $g_i \in \{0, 1\}$, and the sum is a modulo-2 summation. An LFSR is an autonomous state machine that can be described by the state diagram in Figure 2.17. Since the network has no input signals, the next state is determined by the previous state. Typically

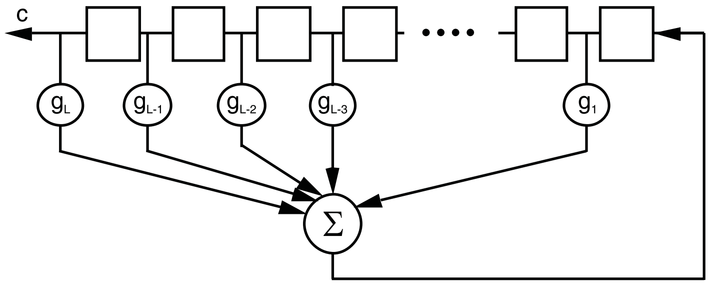

Figure 2.16 Linear feedback shift registers (LFSR).

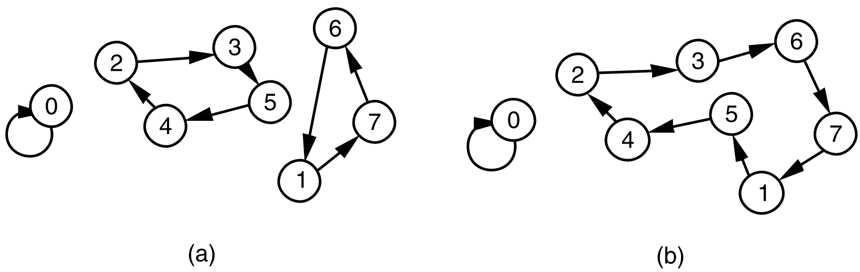

Figure 2.17 Sample state diagrams of LFSR: a) Multiple sequences, b) maximal length sequence.

the LFSR has a state diagram that contains multiple loops as in Figure 2.17a. However, there are settings of the feedback coefficients that yield a state diagram as in 2.17b, where there are only two cycles—one containing the zero state and one containing all the other states. Given that the register is initially in a nonzero state, it will run through all states exactly once. That the output (consisting of a sequence of register contents) will run through all L-tuples (except 0000..000) of binary symbols may not come as a surprise. Such a sequence is termed a *maximal length sequence* or just *m-sequence*.

Since there are 2^L states, the m-sequences have length $N = 2^L - 1$. For reasonable register lengths, there are quite a large set of such sequences, all with correlation properties similar to the random sequences. The number of parameters to describe this sequence is just $2L$, the feedback coefficients g_i and the initial register state.

The autocorrelation properties are shown in Figure 2.18. After the correlation peaks corresponding to $l = 0$, the correlation drops to $-1/N$. The sequence is thus for large N almost self-orthogonal. The cross correlation properties resemble those of random sequences.

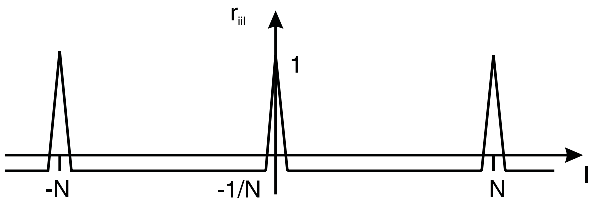

Figure 2.18 Autocorrelation of m-sequence as function of time shift *l*.

The favorable synchronization accuracy of direct-sequence systems have other interesting applications. In fact, the path delay (or to be precise, changes in the path delay) can be measured with high accuracy, that is, within ±1 chip. Several modern navigation systems, including the satellite based Global Positioning System (GPS) utilize this technique.

The synchronization accuracy can also be used to resolve multipath propagation and to estimate the channel impulse response. In a multipath environment, there will not only be one signal component present, but several delayed versions. If the path delay differences are larger than the chip duration, our synchronization scheme could lock onto any of these paths, thereby completely eliminating any multipath fading. The output of the sliding correlator may look something like Figure 2.19. As the time shift is varied, several peaks in the output signal, each proportional to the signal envelope corresponding to that multipath delay, will occur. In fact, Figure 2.19 represents an estimate of the (magnitude of the) instantaneous impulse response of the channel sampled at the chip intervals. To be more precise, the sample points represent the (vector) sum of all the signal components integrated over the chip symbol period. If the number of multipath components is large in every chip bin (poor resolution), this sum will exhibit a rather large variation—a popular model is to assume that the distribution

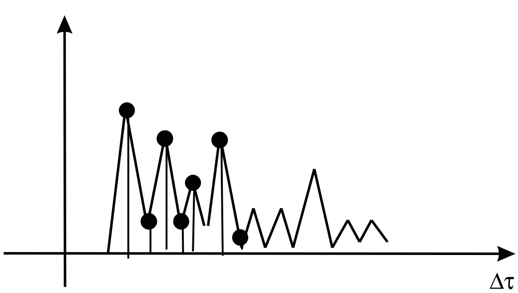

Figure 2.19 Sliding correlator output in synchronizing circuit of DS system.

of the sum amplitude is a Rayleigh distribution. If the number of components is very small (typically only one or no component in each bin) the variations are very limited. There are now several ways of utilizing this property. If we simply lock onto the strongest signal components, we will effectively use the self-orthogonal property of the code to suppress the other components. For m-sequence multi-path components falling into adjacent bins will in this case suffer a $1/N$ power reduction. The effect is a path diversity receiver—in this case using selection diversity [7, 1]. Clearly, instead of locking onto only one of these peaks, one could make use of several multipath components. The effect would be a combining diversity receiver, for example, of the maximum ratio combining type. Since such a receiver would be designed using a tapped delay line, it has been coined a *rake* receiver due to the visual similarity of the block diagram to the garden tool.

As can be seen, DS systems are obviously well suited for digital signal processing and implementation with digital logic. Up until a few years from now, the main use of this technique was in the military sector. Systems with long sequences that are hard to predict are also very hard to jam. The jammer has to rely on rather blunt jamming techniques such as transmitting wideband noise. Another advantage is the noise-like structure and the very low spectral density of the signal (approximately N times lower than the original signal). Such signals are hard to detect, in particular with narrow band receivers (so-called low probability of intercept (LPI) systems).

Signal management is very easy in DS-CDMA systems. Users can simply pick their codes at random. Since the code set is usually very large, the risk of a collision is low. The main practical drawback of the DS-systems emanates from the same feature. Since signals are not orthogonal, a small number of users will interfere with each other. Although the relative cross correlations can be made small, the large dynamic range of the radio channel can cause severe problems. Assume for instance that a sequence of lengths $N = 100$ is used. In such a system users will create interference to each other that is 100 times or 20 dB lower than the wanted signal. However, if the interfering transmitter is very close to the receiver, whereas the wanted signal comes from far away, the interferer may get a power advantage that may be by far more than these 20 dB, maybe up to 100 dB. In this case, reception of the weaker signal is not possible. This problem is treated further in Chapter 9.

2.2 Link Performance Models

As was seen in Section 2.1, the actual calculation of a transmission quality measure such as the bit-error probability in a multi-user scenario may be quite complex. The performance will depend on many parameters such as the waveforms, the instantaneous amplitudes, and the phases of the interfering signals. When analyzing wireless systems with a large number of terminals and complex propagation conditions, the exact analysis will pose a formidable task. Thus, some simplified model or procedure is necessary. In Example 2.1, however, it was seen that the signal-to-interference ratio, E_0/E_1, played a key role in determining the bit error probability. Using this quantity will be the approach in this book. One transmitter-receiver pair, one link, at a time will be studied and all the interfering signals will be characterized by their aggregate, or sum power, that is,

$$\text{Quality} = f(E_0/E_1) = f(\Gamma)$$

where Γ denotes the signal-to-interference ratio. In those cases where the receiver noise cannot be neglected, the Noise energy has to be included in the interference term, and Γ is referred to as the signal-to-interference + noise ratio (SINR). There are cases where this is indeed a reasonable assumption, as we have already seen in Section 2.1. One such case is when the number of interferers is large, and when all these inteferers have similar received powers. In this case the central limit theorem will ensure that the total interference signal approaches a zero mean Gaussian vector, which is characterized only by its energy N_0. If the discussion is confined to the bit-error probability as performance measure, this quantity can straightforwardly be computed by means of classical results from Gaussian detection theory.

This is illustrated by Figure 2.20, which shows the bit-error probability for a coherent BPSK reception as function of the total SIR if the interference is composed of N BSPK signals of identical magnitudes, but random (uniform) phases. As can be seen, already when there are three interferers, the Gaussian approximation is very good.

More complex is the situation when the transmission channel exhibits frequency selective multipath propagation. Clearly, the degree of multipath (delay spread, Doppler frequency, and so forth) will influence the performance and, for example, the SINR will not be the sole quantity that determines the bit-error probability. On the other hand, most communication systems designed for mobile communications will have some means of combating multipath fading. This could be a combination of error control coding, frequency hopping (FH), or equalization techniques, or a Rake-receiver in the DS-spread spectrum case.

In Figure 2.21, one example from the GSM system is shown (which uses a combined FH/equalizer/coding scheme). It is notable that, although

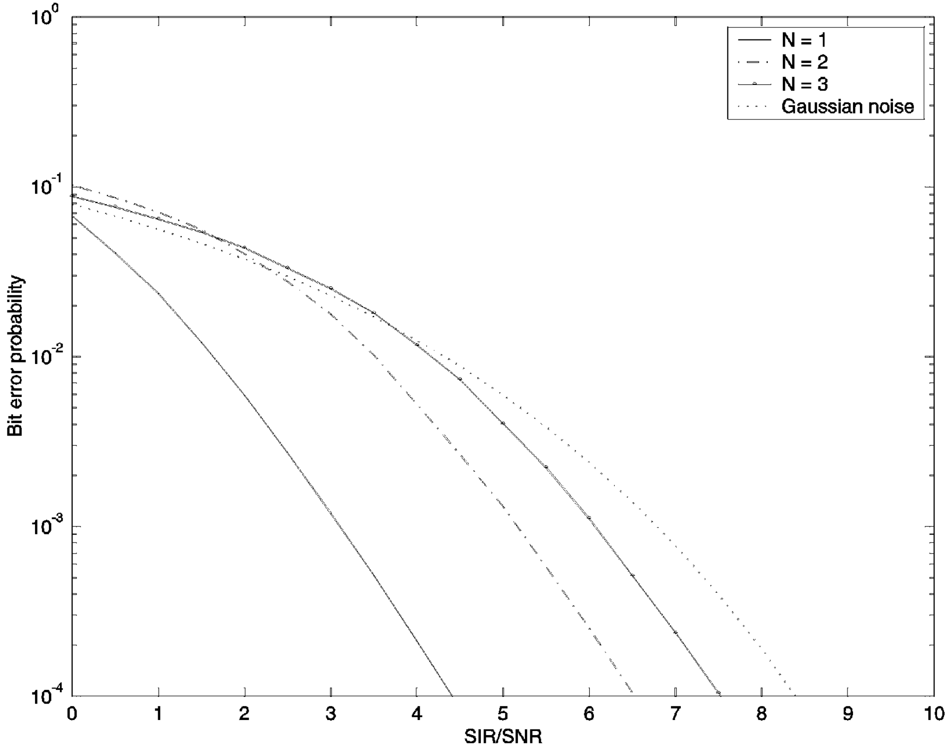

Figure 2.20 Bit-error probability for BPSK as function of signal-to-interference ratio when interference is composed of N = 1, 2, 3 identical BPSK signals with random phase (see Example 2.1). Dashed line BPSK performance in Gaussian noise with corresponding SNR.

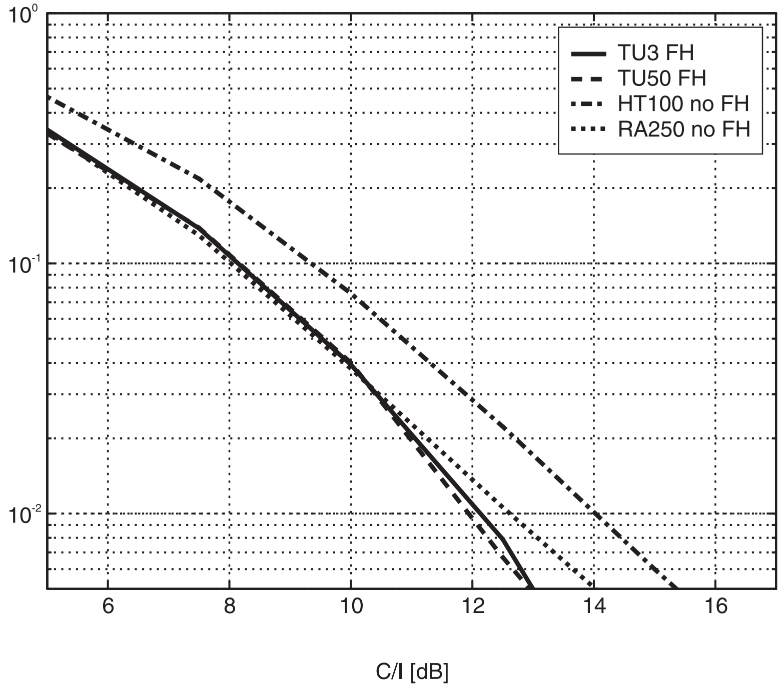

Figure 2.21 Block-Error probability GSM/EDGE system as function of signal-to-interference for various multipath propagation conditions (courtesy Ericsson Radio Systems AB).

the multipath channel characteristics are varying considerably between these cases, the general trend is the same. The same type of behavior can be found also after equalization and error control decoding. Figure 2.22 gives an example of such a model for the packet data communication mode in GSM, GPRS. Here, four different coding schemes may be used, each of them providing a certain transmission quality (here the data block error probability) as a function of the instantaneous SIR.

Graphs like 2.21 also illustrate very well the strategy for analyzing complex wireless networks that will be utilized in the following chapters. In such a network, a large number of transmitters may be active and their signals propagate to the receivers (intended and unintentional) over diverse propagation paths. The detailed analysis of such networks is a formidable task. Instead, the network analysis will be simplified to computing only the wanted signal-to-interference ratios at the receiver (Figure 2.23). We will then map these SIR values on the quality measures of interests using graphs

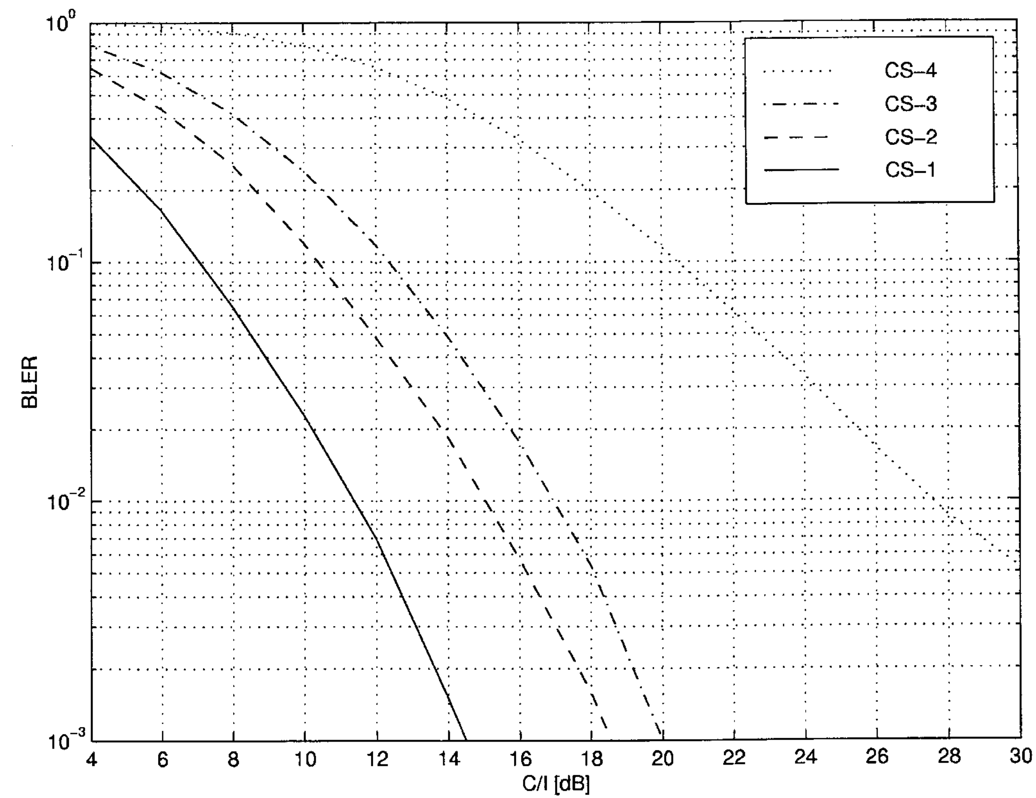

Figure 2.22 Block-error rate as function of SIR in GPRS for 4 different coding schemes [9].

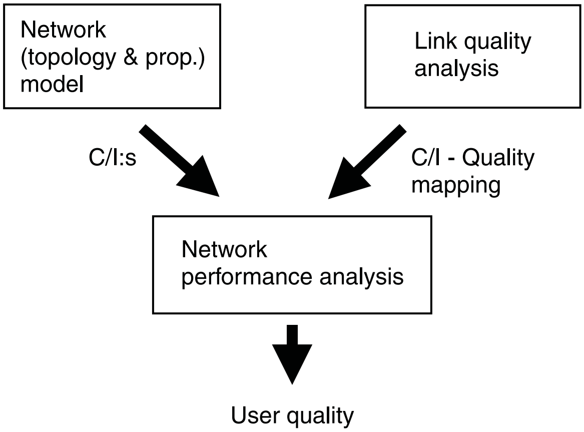

Figure 2.23 Simplified network analysis strategy.

as shown in Figure 2.22, which are derived from the analysis of a single link. This link analysis per se may be quite complex and may include extensive simulation studies. The latter is not a problem since once we have done the calculations and the mapping is known, evaluating the performance given the SIRs becomes a simple table reference.

Problems

2.1 A two-user system utilizes BPSK signaling. The two users are using the waveforms

$$s_0(t) = a_0 \sqrt{\frac{2E_0}{T}} \cos\left(2\pi \frac{t}{T}\right) \qquad 0 \leq t < T$$

and

$$s_1(t) = a_1 \sqrt{\frac{2E_1}{T}} \cos\left(2\pi \frac{t}{T}\right) \qquad 0 \leq t < T$$

The data symbols $a_i \in \{-1, +1\}$ are independent and equally probable. The signal $s_0(t)$ reaches our receiver on a direct path, whereas the signal $s_1(t)$ is subject to multipath propagation. The received signal may be written as

$$r(t) = s_0(t) + s_1(t) - s_1(t - \tau)$$

Study the system for the signal-to-interference ratios $E_0/E_1 = 2$ and $E_0/E_1 = 1/2$.

a) Assume that a conventional correlation detector is used to detect a_0. Determine the error probability as function of the delay spread τ.
b) The communication link is also disturbed by additive white Gaussian noise with spectral density $N_0/2$. Determine again the error probability as a function of the delay spread τ if a conventional correlation detector is used to detect a_0.

2.2 A cellular telephone in the NMT900 system allows that the signals can be received at the same time the transmitter is in operation. This is made possible by the use of a Duplex-filter, provided the signals have to be separated in frequency by 10 MHz. Assume that the phone is in operation at the cell border, 10 km from the base station. The phone may operate properly down to SIR of 20 dB. Assume that free space propagation prevails and that the base station uses the same power as the mobile. Both stations use antennas with an antenna gain of 3 dB. Determine the minimum attenuation of the duplex filter in order to make reception possible at the mobile. What would happen if the propagation loss would be 20 dB higher than free space?

2.3 A binary shift register, according to the figure below is used to generate a synchronization sequence. The binary {0,1} output of the register is coded antipodally according to the mapping in the figure below.

a) Show that this register generates a maximum-length sequence, provided it is not initially in the all-zero state.
b) Determine the autocorrelation of the output sequence!

2.4 Show also that the LFSR in the figure below generates an m-sequence and determine the cross-correlation with the sequence in Problem 7.3.

2.5 A TDMA system uses an m-sequence of length 15 bits for synchronization purposes. In order to find the sync sequence and to receive the frame correctly, all 15 sync bits have to be received correctly. The radio channel can be modeled as a BSC with bit error probability $p = 5\%$.

a) Compute the "miss-probability" of the synchronizer, P_m, i.e., the probability that the receiver will not find the sync sequence due to bit errors even though the system is in sync.

b) Estimate the probability of false sync, P_f (i.e. the probability that some random data or a randomly shifted version of the sequence appears as a correct sequence). Assume that data bits appear correctly with probability 1/2.

c) Repeat a) and b) and compute P_m and P_f if the receiver finds sync if there are at most 3 errors in the sync sequence.

2.6 An FH-system with N = 100 consecutive frequency channels using DPSK-modulation is subject to jamming. The jammer decides to use a partial band jamming technique and selects K frequency channels in which it transmits white Gaussian noise. The jammer can generate 10 times more signal energy at the receiver than the legal transmitter. The same jamming energy is used in all K selected channels. Determine the K that maximizes the bit error probability of the FH system and calculate this bit error probability.

2.7 A DS cellular system has rather long chip duration compared to the multipath profile duration such that the received powers in each chip bin can be assumed to be independent and exponentially distributed. Assuming that the receiver considers three chip sync positions and can choose between either i) selecting the strongest component, or ii) the use of three branch rake receivers using maximum ratio combining. Compare the resulting average SNR for the two receiver strategies i) and ii) if the average branch-SNRs are

a) 10, 10, 10 dB;

b) 15, 10, 5 dB.

References

[1] Proakis, J. G., *Digital Communications*, New York: McGraw-Hill, 1995.

[2] Hagerman, B., "Downlink Relative Co-Channel Interference Powers in Cellular Radio Systems," *Proc. IEEE Veh Tech Conf., VTC '95,* Chicago, July 1995.

[3] Hagerman, B., "Strongest Interferer Adaptive Single-User Receivers in Cellular Radio Environments," *Proc. Globecom '95*, Singapore, Nov. 1995.

[4] Lupas, R., and S. Verdu, "Near-Far Resistance of Multi-User Detectors for Asynchronous Gaussian Multiple-Access Channels," *IEEE Trans. Comm.*, Vol. COM 38, April 1990.

[5] Woozencraft, J. M., and I. M. Jacobs, *Principles of Communication Engineering*, New York: Wiley, 1965.

[6] Mouly, M., and M. B. Pautet, *The GSM System for Mobile Communications*, published by the authors, 1992.

[7] Ahlin, L., and J. Zander, "Principles of Wireless Communication," Student Literature, 1997.

[8] Wigard, J., et al., "BER and FER Prediction of Control and Traffic Channels for a GSM Type of Air-Interface," *48th IEEE Vehicular Technology Conference*, 1998, Vol. 2, pp. 1588–1592.

[9] Schramm, P., et al., "Radio Interface Performance of EDGE, a Proposal for Enhanced Data Rates in Existing Digital Cellular Systems," *48th IEEE Vehicular Technology Conference*, 1998, Vol. 2, pp. 1064–1068.