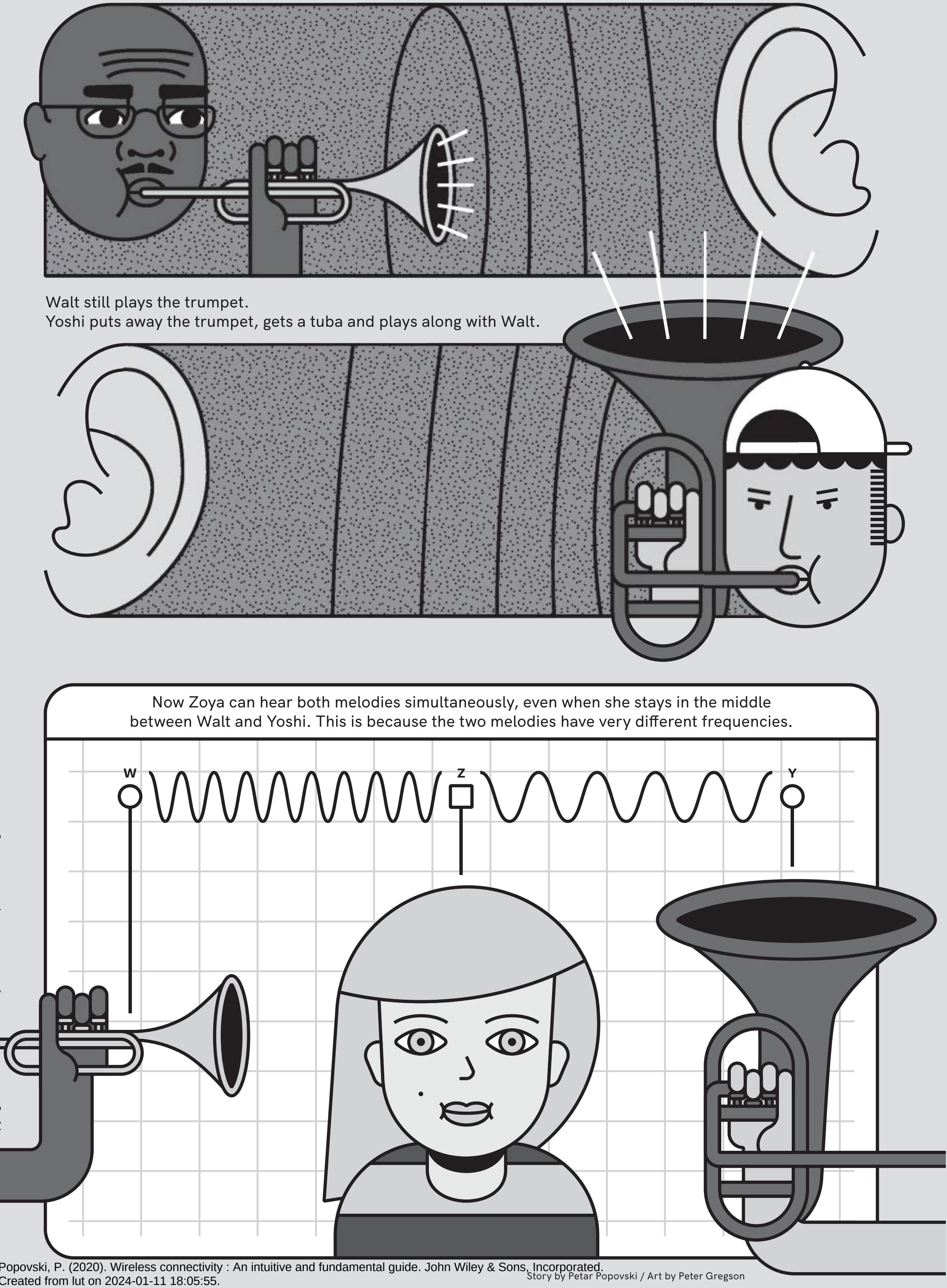

Story by Petar Popovski / Art by Peter Gregson

9

Time and Frequency in Wireless Communications

Medium access control (MAC) protocols use time as an underlying physical variable. The main entity at a MAC layer is a packet, represented by suitably modulated/coded symbols at the physical layer. On the other hand, the fundamental characterization of communication channels is based on the notion of a *channel use* and information theory is concerned with the capacity to transmit information bits per channel use. A channel use should be understood as an abstract transmission opportunity rather than something that occurs at regular time intervals. For example, the pigeon channel, discussed in Chapter 6, has a certain capacity per channel use that is determined independently of the fact that the pigeon flies once per day.

This chapter makes a connection between, on the one side, the channel uses in a communication channel and, on the other side the actual data carrying signals that occur in time. This is instrumental to create the physical basis for the notion of time used in MAC protocols and other protocols running at the higher layers. The physical signals used for communication are *analog signals and waveforms*, which are continuous functions of time. The chapter discusses how a digital communication signal is mapped onto physical signals. In this way, the problem of sending information per an abstract channel use becomes a problem of sending information per time unit. It is thus not immediately clear whether the discrete communication channels that we have described so far are relevant in such a setting. Another important concept introduced in this chapter is the one of a *frequency*, practically central to any communication system. In that sense we will discuss the notions of *signal bandwidth* and *carrier frequency*.

9.1 Reliable Communication Requires Transmission of Discrete Values

We start by establishing the fact that reliable communication is essentially a process of sending discrete rather than continuous values. Indeed, noise is a continuous waveform that, with probability one, has a non-zero value at any time instant. Hence, a transmitted waveform will almost surely be distorted by noise at the receiver. In the best case, the only thing that the receiver can state reliably is that the received waveform belongs to a certain class of waveforms. Here the term "class of waveforms" plays a role that is analogous to the one of the noise circle from Chapter 5: the transmitted waveform corresponds to the

Wireless Connectivity: An Intuitive and Fundamental Guide, First Edition. Petar Popovski.

center of the noise circle, while the noise circle itself corresponds to the set of random waveforms that can be received when the transmitted waveform is polluted by noise. We are thus interested in the largest number of M waveforms that can be reliably distinguished at the receiver. Here "reliably" means that the probability of error is below a predefined value or, ideally, zero. The transmission of one of those waveforms, selected randomly and uniformly, corresponds to sending $\log_2 M$ bits of information. However, the difference with the communication channels discussed before is the fact that now we consider physical signals that consume actual time. Instead of finding how many bits per channel use one can send, the proper question now is: *how many bits per second (bps) can be reliably communicated by using analog signals?*

A way to find the data rate in bits per second, denoted by R, is to first find how many independent channel uses there are per second, denoted by D, calculate the capacity in terms of bits per channel use (bits/c.u.), denoted by C, and then find R through:

$$R = D \cdot C \quad \text{(bps).} \tag{9.1}$$

We have used D to denote the number of real-numbered channel uses per second. In this chapter we will refer to a channel use also as a *degree of freedom (DoF)*, such that in this case there are D real-numbered DoFs per second. As we are going to use the term DoF quite extensively in this chapter, it deserves a short elaboration.

The term DoF bridges the concepts used for abstract communication channels and the ones used to describe analog waveforms. The bridging role of DoF is related again to the idea of layering, recurring throughout this book. Namely, (9.1) assumes that we have decomposed the problem of communication with analog waveforms into two independent sub-problems:

1. Using continuous waveforms, create a certain number of independent channel uses per unit time. This *discrete* set of channel uses is obtained by a certain sampling of the continuous waveform.
2. Apply communication strategies over the discrete set of channel uses, following the strategies developed for a baseband model of a communication system, treated in the previous chapters.

We note that whenever we use the term DoFs we mean real DoFs. If we observe a single channel use in which a complex symbol can be sent, this can be seen as a complex DoF that consists of two real DoFs. Hence, the real DoF can be treated as an atomic transmission unit for analog signals that can be used to build other types of DoF.

As it is always the case in layering, this separation is only a specific instance in which the communication problem can be solved. Nevertheless, as Shannon already stated in his original paper and later Slepian proved rigorously, this strategy is optimal. The reason is that every continuous signal of interest in communication engineering can be almost perfectly synthesized as follows: take a finite set of D basis waveforms, scale each waveform with a real coefficient and add the scaled waveforms. The resulting waveform is a linear combination of the basis waveforms and its shape can be controlled by choosing the D real coefficients. Assuming that the basis waveforms are predefined and known in advance by the transmitter and the receiver, then the information in the resulting waveform is stored only in the D coefficients that determine the linear combination. This explanation provides an idea for the use of the term DoFs to describe the number of independent discrete inputs that preserve the information about a waveform. Throughout the chapter it will become clear how D is determined and the physical constraints that affect it.

9.2 Communication Through a Waveform: An Example

Let us assume that the TXmodule of Zoya is a black box that can accept a real number as an input and produce as an output a waveform that is scaled with that same real number. Taking the specific example from Figure 9.1, if the input value is 1, the output waveform is shown in Figure 9.1(a), while if the input value is −3, then the waveform in Figure 9.1(a) is multiplied by −3. Let us further assume that once the waveform is produced at the output of the TXmodule, it is transferred to the RXmodule of Yoshi without any distortion, such that in the absence of noise Yoshi receives exactly the same transmitted signal. Furthermore, assume that the RXmodule makes its decoding decision based only on the value that is sampled at the instant where the waveform achieves its largest magnitude. This instant occurs at 0.5 s at the output of the TXmodule. Note that in any physical system there is a delay τ, not depicted in the figure, that equals to the time that is needed for the signal to propagate between Zoya and Yoshi. We need to assume that Yoshi knows this τ, such that he can set the timing for making a decision to $\tau + 0.5$ s. All these assumptions are very idealized and are used to provide a pedagogical example, ignoring the difficulties Yoshi needs to go through in order to learn τ, as well as the fact that the wireless channel always introduces an attenuation and other distortions to the signal sent by Zoya.

If Zoya sends the two pulses to the input of her TXmodule in a way that the pulses are separated at least $T = 1.5$ s in time, then there is no time overlap between the waveforms. The RXmodule will output the observed values, sampled at the time instants with highest amplitudes. Hence, the system TXmodule–wireless channel–RXmodule can be treated as a black box that defines a discrete communication channel: every T s a real value is applied at the input and a noisy value is obtained at the output. Figure 9.1(b) depicts the situation in which the pulses are put $T = 1.5$ s apart and there are four possible inputs $z \in \{-3, -1, 1, 3\}$ that represent the bit pairs 00, 01, 11, 10, respectively. The output signal produced at the sampling or observation instants can be written as:

$$y = z + n \tag{9.2}$$

where n is the noise, not depicted on the figure. Once the waveform channel can be mapped into a discrete communication channel, we can use the communication methods described in the previous chapters. An important point about the example in Figure 9.1(b) is that, when the time separation is $T \geq 1.5$ s, then one can be certain that the channel uses can be independently modulated and what is transmitted/received during one channel use is independent of the other channel uses. Assuming that the capacity per channel use is C (bits/c.u.) and the pulses are separated for $T \geq 1.5$ s, the equation (9.1) can be rewritten as:

$$R = \frac{C}{T} \quad \text{(bps)}. \tag{9.3}$$

The previous equation suggests that lowering T increases the data rate. If we start to lower T below 1.5 s, then the uses of the discrete communication channel start to overlap in time

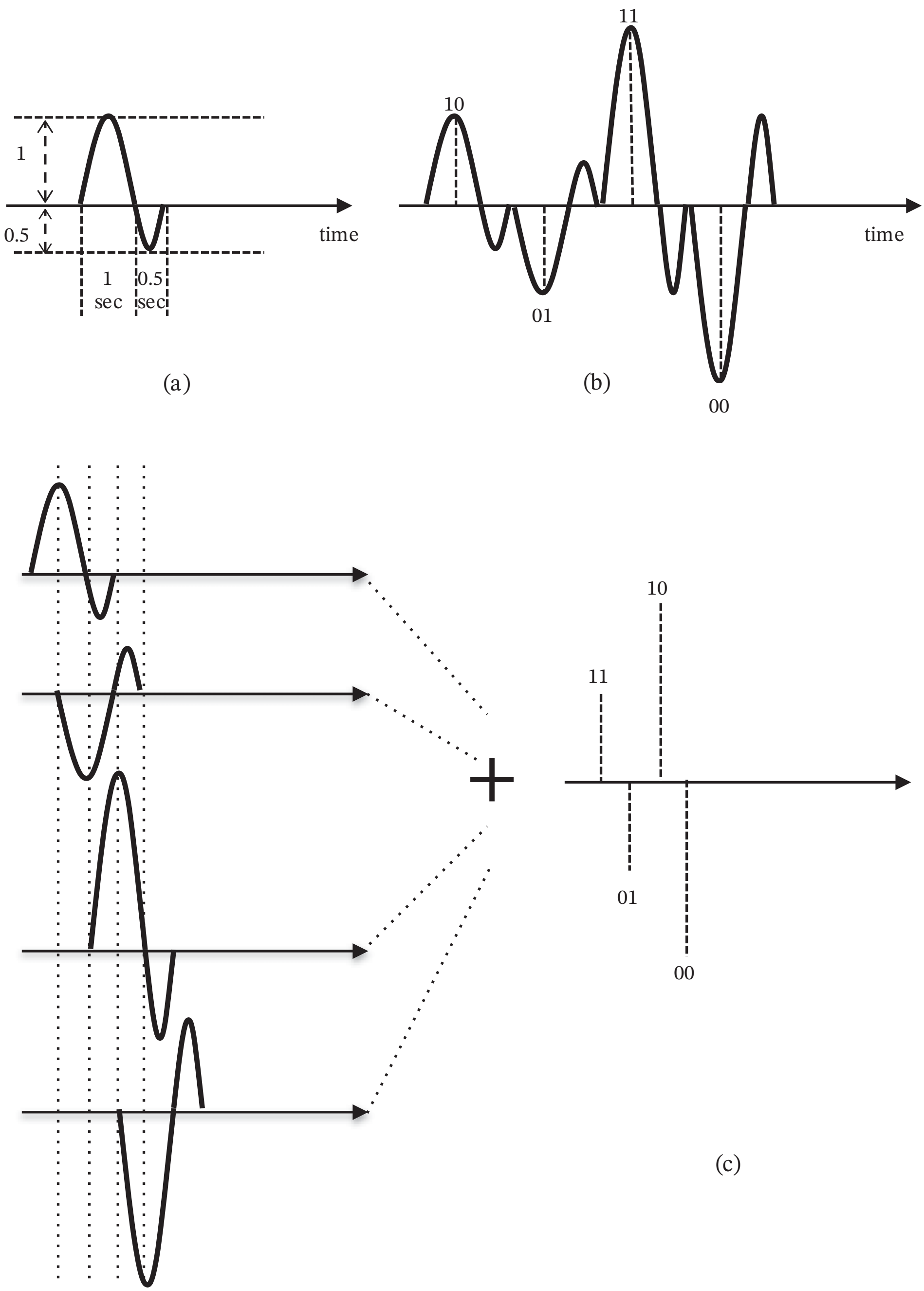

Figure 9.1 A simple example that illustrates how to create a discrete communication channel using continuous waveforms. (a) The elementary waveform. (b) The waveform can be used to carry Gray mapped quaternary symbols: 00 is sent by multiplying it by −3, 01 by −1, 11 by 1, and 11 by 3. The waveforms are sent with a minimal possible separation in time to avoid overlap and there are $D = \frac{1}{1.5} = 0.67$ (c.u/s). (c) Densely packed waveforms with $D = \frac{1}{0.5} = 2$ (c.u/s). The figure only depicts the discrete instants that define the outputs of the communication channel; a full illustration would include a sum of the continuous waveforms contributing to the overall transmitted/received signal.

and can give a rise to *intersymbol interference (ISI)*. With this, the channel uses are no longer independent.

As a simple example of ISI, not related to the example in Figure 9.1, the following baseband received signal can be considered:

$$y_k = z_k + 0.5z_{k-1} + n \tag{9.4}$$

where we see that the kth received symbol by Yoshi is dependent both on the $(k-1)$th and kth symbols transmitted by Zoya. However, even if the waveforms of different symbols overlap in time, what matters is whether there is an ISI that ultimately appears in the baseband representation, as in (9.4). As the example in Figure 9.1(c) shows, setting $T = 0.5$ s also produces discrete channel uses that are independent and free of ISI, although their respective waveforms do overlap in time. Using $T = 0.5$ instead of $T = 1.5$ results in having three times more symbols sent per seconds compared to Figure 9.1(b), which results directly in a three times larger data rate, expressed in bits per second.

Nevertheless, T cannot be decreased arbitrarily while still avoiding ISI in the resulting discrete channel. The reader can check that $T = 0.5$ s is the absolutely minimal time separation between two pulses at which one can still get a communication channel that consists of a sequence of independent channel uses, leading to a data rate of $R = 2C$ (bps). However, if we are willing to depart from the model of a communication channel without ISI and make $T < 0.5$ s, then the discrete communication channel starts to have a *memory*, as in the example (9.4), which is an effect not considered in our baseband models so far. This is in contrast to the *memoryless* communication channels, in which the output of a channel use is affected by the input of the same channel use, but not the inputs of the previous channel uses. The example (9.4) illustrates a channel with memory, where the output y_k has a statistical dependency with both y_{k-1} and y_{k+1}. We would like to emphasize that reliable communication and definition of channel capacity is possible despite the occurrence of ISI, but it requires encoding/decoding strategies that are different from the ones used for memoryless channels.

A practical wireless communication channel inherently introduces ISI, as well as other additional distortions to the signal. The common approach to address the ISI caused by the wireless channel is a layered one: at the lower layer, use *equalization* in order to remove ISI and create a memoryless communication channel; at the higher layer, apply the known coding/modulation strategies for that channel. This leads us to the conclusion that a communication channel can be created by either removing the ISI or by keeping it to remain a part of the communication channel. Once the channel definition is fixed, we can compute its capacity.

Nevertheless, that is not the ultimate answer to the question of how much information we can carry through waveforms with certain physical features since "the channel is the part of the system we are unwilling or unable to change", while each possible ISI pattern defines a new channel. Given the interference pattern, we can compute the capacity in bits per second of the respective channel; it thus remain to find the interference pattern whose channel has the highest capacity. As a fundamental question, we would like to know the information carrying properties of the set of all possible waveforms that can be produced in a given communication system.

Consider the following communication method adopted by Zoya. In transmitting information to Yoshi, Zoya may decide to depart from the conventional strategy that avoids ISI and start to send symbols to the TXmodule using arbitrarily low T. In the meantime, Yoshi samples the output of the RXmodule using that same value of T. With this Zoya tries to increase the capacity R from (9.3) by decreasing T. At the same time, she hopes that the capacity C of the channel with ISI will not decrease so much so as to eliminate the gain obtained by sending the symbols faster. Can T be really decreased to be arbitrarily low? This line of thinking needs to take into account the other physical limitations of the system and the notion of frequency characteristics of the waveforms.

9.3 Enter the Frequency

A very simple way to increase the data rate of the system based on the elementary waveform from Figure 9.1(a) would be to compress the waveform in time. For example, if the same waveform is compressed to take in total 0.15 s instead of 1.5 s, then one gets ten times more channel uses per second that, with unchanged number of bits per channel use, leads to a ten times higher data rate. Compressing the waveform to a tenth of the original duration implies that one allows ten times faster changes in the waveform and therefore ten times faster operation of the circuits and modules underlying the communication system. However, there must be physical constraints on how fast the changes in a waveform may occur; otherwise one can make the duration of the waveform infinitely short and thus get the data rate to grow to infinity. Those physical constraints are often expressed through the *frequency characteristics* of the system and the associated waveforms. There is neither space nor ambition to cover the details of the rich area of frequency analysis, Fourier transform, time-frequency diagrams, etc., but we will provide intuitive arguments to highlight the role that frequency plays in communication.

9.3.1 Infinitely Long Signals and True Frequency

Frequency is a quantity that can only be defined precisely for a signal of infinite duration. An example signal of frequency f is:

$$\tilde{z}_f(t) = |A|\cos(2\pi f t + \phi) \quad -\infty < t < \infty \tag{9.5}$$

where we put a "wave" $\tilde{\cdot}$ to denote a signal that is continuous in time[1], $|A|$ is the amplitude, and ϕ is the phase of the sinusoidal signal. Once the frequency is fixed, the amplitude and the phase represent two channel uses or two DoFs that can be used to modulate data. We note here that these two DoFs are not identical in a communication/statistical sense, since noise or other physical distortions affect the amplitude and the phase in two different ways.

A bigger conceptual problem with these two DoFs is their practical significance as channel uses for a communication system. If Zoya wants to use these two DoFs to communicate with Yoshi, then she needs to send the information only once, infinitely back in the past, and the transmitted symbols will last indefinitely. Common statements such as "change the

1 This will only be used when there is a danger of confusion.

frequency between time t_1 and t_2" do not make sense simply because, by the strict definition, a frequency cannot change in time! This paradox and the discrepancy between the "mathematical" frequency and "practical" frequency has been noted already in the early days of communication engineering, as eloquently put forward by Dennis Gabor in his seminal article "Theory of Communication". Nevertheless, for the discussion that follows we embrace the absurdity of the setup in which Zoya sends the symbols infinitely back in the past and let them last indefinitely. On the other hand, we require Yoshi to receive the symbols sent by Zoya by turning his receiver on for a finite duration of time T.

An equivalent way to represent a signal from (9.5) that has a frequency f is:

$$\tilde{z}_f(t) = z_{I,f} \cos(2\pi f t) - z_{Q,f} \sin(2\pi f t) \tag{9.6}$$

where $z_{I,f}$ is the *in-phase (I)* and $z_{Q,f}$ is the *quadrature (Q)* component at the frequency f. Given $z_{I,f}$ and $z_{Q,f}$, the values of $|A|$ and ϕ in (9.5) can be uniquely determined, and vice versa. The representation (9.6) is convenient as $z_{I,f}$ and $z_{Q,f}$ can be seen as two DoFs that are of the same statistical nature, in the sense that if noise is added to $\tilde{z}_f(t)$ it is going to affect $z_{I,f}$ and $z_{Q,f}$ in statistically identical ways. In other words, $z_{I,f}$ and $z_{Q,f}$ are two channel uses of the same kind that can be independently modulated, which will become even clearer when we take into account the impact of the noise. The information about these two DoFs can be compactly represented as a single complex symbol:

$$z_f = z_{I,f} + jz_{Q,f} \tag{9.7}$$

such that the signal $\tilde{z}_f(t)$ (9.6) can be obtained as:

$$\tilde{z}_f(t) = \mathrm{Re}\{z_f e^{j2\pi f t}\}. \tag{9.8}$$

Although $z_{I,f}$ and $z_{Q,f}$ can be seen as two independent DoFs, we note that when Zoya transmits $\tilde{z}_f(t)$, Yoshi observes interference between $z_{I,f}$ and $z_{Q,f}$ as they are both received simultaneously, while the objective is to extract each of them as a separate value. Yoshi can achieve this by using the fact that the signals $\cos(2\pi f t)$ and $\sin(2\pi f t)$ are *orthogonal* in any time interval that is an integer multiple kT where $T = \frac{1}{f}$ is the period of the signal (9.6). This means that if the signal from (9.6) is received during an interval of length kT, where k is an integer, the orthogonality property guarantees that there is a way for Yoshi to extract each of the values $z_{I,f}$ and $z_{Q,f}$ independently, without interference from the other one. This can be written as follows:

$$\begin{aligned} z_{I,f} &= \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{z}_f(t) \cos(2\pi f t) \mathrm{d}t \\ z_{Q,f} &= -\frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{z}_f(t) \sin(2\pi f t) \mathrm{d}t \end{aligned} \tag{9.9}$$

where, for convenience, we have chosen to use the orthogonality on the interval $\left[-\frac{T}{2}, \frac{T}{2}\right)$, but any interval of length T would work. If the signal (9.6) is received during an interval shorter than T, then it is not possible to extract $z_{I,f}$ and $z_{Q,f}$ in such a way that they are free of each other's interference.

The idea of extracting interference-free discrete symbols from continuous waveforms that interfere is appealing and we will generalize beyond the I/Q signals at a single frequency. Let Yoshi set his receiving interval to length T and, without loss of generality, let that interval be $\left[-\frac{T}{2}, \frac{T}{2}\right)$. It is interesting to find out the following: how many independent, non-interfering channel uses (DoFs) can Yoshi extract from the signal sent by Zoya? The orthogonality property extends to signals with different frequencies. Having a fixed interval of length T, then any sinusoidal signal of frequency $\frac{k}{T} = kf$, regardless of its phase, is orthogonal to a sinusoidal signal of frequency vf as long as the integers k and v are different $k \neq v$. Using complex number notation, this can be written as:

$$\frac{1}{T}\int_{t}^{t+T} e^{j(2\pi kft+\phi_k)} e^{-j(2\pi vft+\phi_v)} dt = \delta_{k,v} = \begin{cases} 0 & k \neq l \\ 1 & k = v \end{cases} \tag{9.10}$$

where $\delta_{k,v}$ is the Kronecker symbol, while ϕ_k and ϕ_v are used to emphasize that the phase of the sinusoidal signal does not affect the orthogonality. The latter implies that the I component at the frequency kf is orthogonal to both the I and the Q component at the frequency vf. We note also that when the constant signal, having a frequency of 0, is the only "sinusoidal" signal in which one cannot differentiate the I and Q components such that it can carry only one DoF. Let us denote by $z_{I,kf}$ and $z_{Q,kf}$ the I and the Q components, respectively, sent by Zoya at the frequency kf, and observe the following signal:

$$\tilde{z}(t) = \sum_{k=0}^{\infty} \tilde{z}_{kf}(t) = \sum_{k=0}^{\infty} z_{I,kf}\cos(2\pi kft) - z_{Q,kf}\sin(2\pi kft) \tag{9.11}$$

where we set by definition $z_{Q,0} = 0$, i.e. the quadrature component at frequency 0 does not carry information. By receiving $z(t)$ over an interval of length T and using the orthogonality property, Yoshi can receive an infinite amount of symbols $\{z_{I,kf}, z_{Q,kf}\}$ that are free from interference from each other. The symbols are extracted as follows:

$$\begin{aligned} z_{I,kf} &= \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{z}(t)\cos(2\pi kft)dt \\ z_{Q,kf} &= -\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{z}(t)\sin(2\pi kft)dt \end{aligned} \tag{9.12}$$

where the "−" sign is due to the definition of a quadrature component in (9.6). We note that the interval $-\frac{T}{2} \leq t < \frac{T}{2}$ is not special in any way: Yoshi can choose any interval of length T and use the orthogonality property in that interval. This is because the signal $z(t)$ is periodic and the information about $z_{I,kf}$ and $z_{Q,kf}$ is identical anywhere; recall that the symbols are sent in infinite past and last until infinity. However, it is important to stick to the same length T, as this is determining the fundamental frequency $f = \frac{1}{T}$ and the *harmonics kf* of the orthogonal signals.

By choosing all $z_{I,kf}$ and $z_{Q,kf}$, Zoya synthesizes a periodic waveform with period T, such as the one depicted in Figure 9.2(a). On the other hand, Yoshi receives the signal only during a single, specific interval of length T, as in Figure 9.2(b). In fact, while the transmitter Zoya carries out *synthesis*, Yoshi is faced with the problem of *analysis* of a waveform: given the waveform from Figure 9.2(b), find out the coefficients $\{z_{I,kf}, z_{Q,kf}\}$ that have produced that waveform. This is known as *Fourier analysis* and the representation of the signal $\tilde{z}(t)$ in (9.11) is known as *Fourier series*. If for some k both coefficients are zero $z_{I,kf} = z_{Q,kf} = 0$, then this is often referred to as "the frequency kf is absent from the signal". Otherwise, if at least one of $z_{I,kf}$ or $z_{Q,kf}$ is not zero, it is considered that the frequency kf is present in the signal or, in other words, the signal carries energy at the frequency kf.

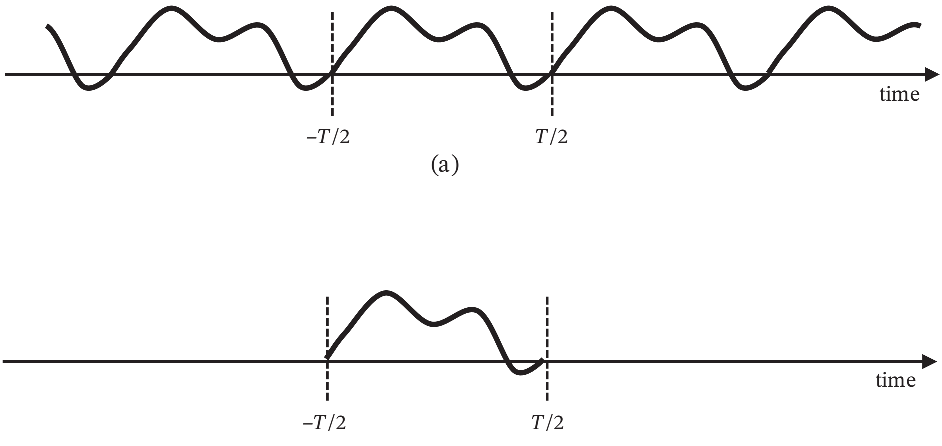

Figure 9.2 (a) The periodic waveform with a period of T sent by Zoya. (b) Yoshi observes only a fragment of the waveform of duration T, which contains all information about the waveform.

9.3.2 Bandwidth and Time-Limited Signals

From the Fourier series representation of a signal in (9.11) one can conclude that there is an infinite number of sinusoidal waveforms that are orthogonal within the finite time interval of length T. In other words, Zoya can potentially send an infinite number of interference-free symbols $\{z_{I,kf}, z_{Q,kf}\}$ to Yoshi, since k can be arbitrarily large. Then Yoshi observes infinite number of symbols within a finite time, which means that the data rate is infinite.

Nevertheless, this is physically not possible due to the following. The energy that the signal carries at the frequency kf within the limited time T is proportional to:

$$E_{kf} = (z_{I,kf}^2 + z_{Q,kf}^2)T \tag{9.13}$$

such that the total energy carried by the signal is:

$$E = \sum_{k=0}^{\infty} E_{kf}. \tag{9.14}$$

To keep the things physically sane, the energy E needs to be finite. One way to ensure this is to put the constraint that Zoya's signal is allowed to contain frequencies that belong to a limited frequency band, such that $E_{kf} = 0$ for all $k \geq k_{\mathrm{H}}$. Here $f_{\mathrm{H}} = k_{\mathrm{H}} f$ is the largest frequency contained in the signal. Alternatively, the energy per frequency can decrease to zero as the frequency grows to infinity, such that the sum (9.14) is kept finite.

One can try to increase the speed of sending new symbols by shortening the duration of the waveforms that represent the symbols. The physical limit on how short the waveform can be depends on the physical limits of the system on how fast it can process the symbols. The frequency representation provides an alternative view of the physical limits: if high frequencies kf are present in the signal, then it means that Zoya and Yoshi communicate with signals that change very quickly. Putting a limit on $k_{\mathrm{H}} f$, the highest frequency in the signal, or requiring that the energy per frequency goes to zero as k goes to infinity, means that we are limiting the speed at which the information carrying signal can change.

This brings us to the notion of a *bandlimited signal*, which has non-zero energy only at the frequencies that lie within a certain band $[f_L, f_H]$. In other words, the signal contains the frequencies within the limited band $[f_\mathrm{L}, f_\mathrm{H}]$. This is a rather simplified definition that will be revised later on, but it is sufficient for our present discussion. Although the notion of frequency can only be defined for a signal of infinite duration, we can meaningfully speak about the frequencies that the receiver can observe in a signal within a time window T, as in Figure 9.2(b). However, one should keep in mind that these are the frequencies that are defined for the signal that is a periodic extension of the observed time-limited signal, as in Figure 9.2(a).

If Zoya uses *lowpass* transmission and includes the frequency $f_\mathrm{L} = 0$, then in a *bandwidth* of size W [Hz] contains at most

$$\mathrm{DoF}_{\mathrm{lowpass}} = 2\frac{W}{f} + 1 = 2WT + 1 \tag{9.15}$$

non-zero coefficients $\{z_{I,kf}, z_{Q,kf}\}$, which can be seen as channel uses that Zoya can modulate with information. The multiplier 2 in (9.15) is due to in-phase and quadrature components for each non-zero frequency and +1 is the DoF of frequency 0. Alternatively, Zoya can create a *passband* signal that contains the frequencies from the range $[f_\mathrm{L}, f_\mathrm{H}]$ with $f_\mathrm{L} > 0$, such that the bandwidth of the signal is $W = f_\mathrm{H} - f_\mathrm{L}$ and the number of available DoFs is

$$\mathrm{DoF}_{\mathrm{bandpass}} = 2\frac{W}{f} = 2WT. \tag{9.16}$$

This is the point where we need to revise our assumption that Zoya communicates with periodic signals of infinite duration. Namely, Yoshi receives the signal from Zoya by observing only a limited interval of time T and what Zoya transmits outside of that interval does not affect the reception of Yoshi in any way. Therefore, Zoya can use any other time outside of that interval to send other symbols to Yoshi, which removes the annoying assumption that she should have started to send sinusoidal signal infinitely back in the past. For example, Zoya picks a frequency kf and in the interval $[-T/2, T/2)$ she modulates it with the symbols $z_{I,kf}(1)$ and $z_{Q,kf}(1)$, where "1" stands for the first modulated symbol. In the next interval $[T/2, 3T/2)$, she sends the second set of symbols $z_{I,kf}(2)$ and $z_{Q,kf}(2)$. In general, during the ith symbol interval $[(2i-3)T/2, (2i-1)T/2)$, Zoya modulates kf with the symbols $z_{I,kf}(i)$ and $z_{Q,kf}(i)$. We therefore refer to T as a *symbol duration* or *channel use duration* or *DoF duration*. The signal obtained in this way, when observed during the entire time axis, is not periodic and therefore cannot be treated as a "signal of frequency kf", understood in a strict sense. However, within each interval $[(2i-3)T/2, (2i-1)T/2)$, which corresponds to a single symbol duration, the signal looks like a portion of a periodic signal and Yoshi can use the orthogonality property to extract the symbols sent by Zoya without any *ISI*.

As long as Yoshi is perfectly synchronized with Zoya, then within each symbol interval T one can consistently speak about symbols carried at a particular frequency $kf = \frac{k}{T}$. However, if Yoshi uses a time window for reception that has a duration different from T, then the orthogonality property is lost. Even more, if Yoshi uses the same symbol duration T, but its reception window is not aligned to the timing of the symbols sent by Zoya, then the orthogonality property is also lost. This is illustrated in Figure 9.3 for a transmission that

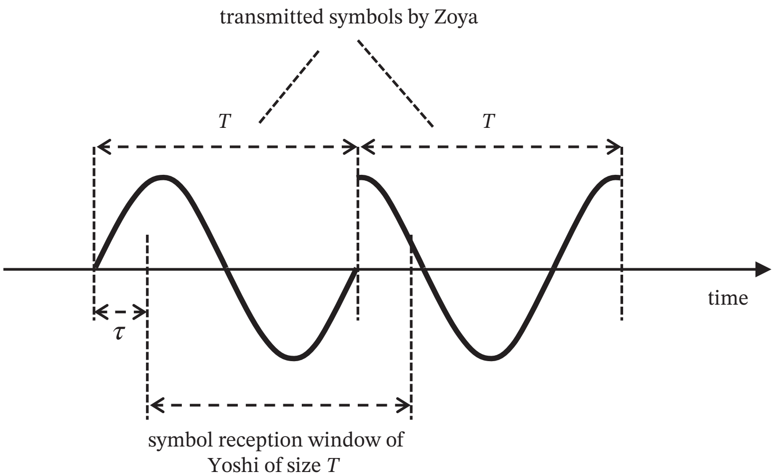

Figure 9.3 Zoya uses the frequency f to transmit the complex symbol $(0, -1)$ at time 0 and the symbol $(1, 0)$ at time T. Yoshi's reception window has the correct duration T, but is misaligned and therefore Yoshi does not observe a sinusoidal signal within his reception window.

occurs at a frequency $f = \frac{1}{T}$. In this example, Zoya and Yoshi have agreed that the symbols should be placed as $[(i-1)T, iT)$ instead of $[(2i-3)T/2, (2i-1)T/2)$. However, Yoshi has the start of his reception window misaligned for a duration of τ. Zoya transmits the symbols $z_{I,f}(0) = 0, z_{Q,f}(0) = -1$ in the interval $[0, T)$ and the symbols $z_{I,f}(1) = 1, z_{Q,f}(1) = 0$ in the next interval. Within his misaligned interval, Yoshi does not observe a sinusoid, that is, a signal that represents a "cut" from an infinite sinusoidal signal with frequency f, and therefore the orthogonality property cannot be applied[2]. Instead, the information carried in the symbol transmitted at time T affects the decoding of the symbol sent at time 0 and Yoshi experiences ISI.

9.3.3 Parallel Communication Channels

The most important quality that the concept of frequency brings to our communication models is the possibility of creating *parallel communication channels*, which can exist simultaneously at different frequencies. Once T and W are fixed, then in each time interval T, Zoya can send $2WT$ different symbols in parallel (ignoring the +1 for lowpass signals). This is the basis for *orthogonal frequency division multiplexing (OFDM)*, a transmission scheme that has become dominant within broadband wireless transmission technologies since the 2000s. In OFDM, each of the sinusoidal signals that is modulated with symbols is called a *subcarrier* and we adopt the same terminology to denote a certain frequency that is modulated with data. The notion of a *carrier* is reserved for the frequency that is usually the central frequency of a passband signal. In this sense, a carrier frequency can be used to

2 To be precise, if it happens that Zoya sends the same symbol consecutively, then $z_{I,f}(1) = z_{I,f}(0)$ and $z_{Q,f}(1) = z_{Q,f}(0)$, such that the signal observed by Yoshi is still sinusoidal. However, this is not true in a statistical sense, as symbols change randomly from one interval to another.

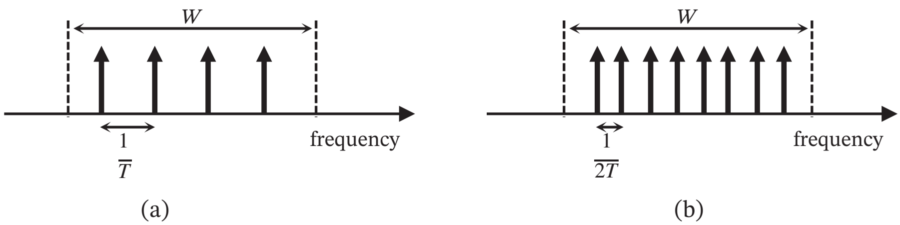

Figure 9.4 Number of channel uses in a given bandwidth W. Each vertical arrow represents a frequency subcarrier. (a) The channel use duration is T and the separation between two neighboring subcarrier frequencies is $\frac{1}{T}$. (b) The channel use duration is $2T$ and the separation between two neighboring subcarrier frequencies is $\frac{1}{2T}$.

shift a lowpass OFDM signal into an equivalent passband signal; then the data-modulated subcarriers are placed around this carrier frequency and hence the name.

Let us assume that the available bandwidth for communication is W and each subcarrier is modulated with Mary symbols, such that each symbol carries $\log_2 M$ data bits. Each channel use has a duration of T, corresponding to the time interval over which the different subcarriers with frequencies $\frac{k}{T}$ are orthogonal. Since now each channel use has a well defined duration, one can calculate the data rate in bits per seconds:

$$R = \frac{2WT\log_2(M)}{T} = 2W\log_2 M \quad \text{(bps).} \tag{9.17}$$

It is interesting to note that the data rate does not depend on the choice of the channel use duration, but it depends on the available bandwidth W. In order to see why this is the case, see the example in Figure 9.4. Each vertical arrow represents a subcarrier that be modulated with information, or, equivalently, two channel uses of DoFs (in-phase and quadrature). When the channel use duration doubles, the number of subcarriers within the bandwidth also doubles; however, this does not change the number of DoFs per second, which remains constant and equal to $2WT$.

Although the data rate in (9.17) does not explicitly depend on T, the whole discussion is based on the assumption that a certain value of T has been fixed in advance and is used as a time reference to modulate all subcarriers.

9.3.4 How Frequency Affects the Notion of Multiple Access

The parallel channels, defined on orthogonal frequency subcarriers, offer the possibility of sharing the medium among several wireless links simultaneously, without using some form of TDMA. We can take the example in which two mobile terminals, Zoya and Yoshi, communicate with a base station Basil. The total available bandwidth is W. All parties (Zoya, Yoshi, and Basil) agree in advance on the symbol period T, during which the symbols that are modulated on the sinusoidal carriers do not change. It is also agreed in advance that the frequencies that lie in the subband that contains the lower half of the frequency band W are allocated for communication between Basil and Zoya, while the ones from the upper half are allocated for communication between Basil and Yoshi. If a third user Xia joins the system to communicate with Basil, then she also needs to use the same symbol period T and Basil needs to re-allocate the bandwidth W by dividing it into three parts.

The method of channelization in which different users use different frequencies is known as *frequency division multiple access (FDMA)*. In fact, in this case, since we assume that all frequency subcarriers are orthogonal, we can speak about *orthogonal frequency division multiple access (OFDMA)*. Given the bandwidth W and using the same arguments as in Figure 9.4 it can be shown that the number of DoFs, and therefore the capacity, of TDMA and (O)FDMA is identical. In a TDMA transmission scheme, Zoya is allocated the whole bandwidth W within the symbol duration T, thus using the $2WT$ DoFs within that time. If there are M users in the system and they are served in turn using a round robin, then the average number of DoFs available to each user per unit time is $\frac{2WT}{MT} = \frac{2W}{M}$. Alternatively, each user can be served all the time, but using only a fraction of the available bandwidth $\frac{W}{M}$, which again leads to $2\frac{W}{M}$ DoFs per unit time. In the case of FDMA Zoya can apply a frequency filter and receive only the frequencies that are allocated to her. Similarly, if TDMA is used, Zoya can apply a form of time filtering and be active only during the time window allocated to her. In practice, frequency filters in FDMA (though not in OFDMA) are less flexible compared to the time filtering, such that TDMA is the preferred option when the traffic of different links varies and the resources need to be allocated dynamically.

Despite this difference, the framework introduced so far cannot really capture the true difference between TDMA and FDMA, as in both cases we require all users to be perfectly synchronized. In TDMA this is required in order to preserve the synchronization of the time windows allocated to the users, while in FDMA in order to preserve the orthogonality among the subcarriers belonging to different users. On the other hand, the relation (9.17) suggests that the frequency bandwidth W determines the data rate of a communication link, irrespective of the choice of the symbol duration T.

Perfect synchronization is not so much of an issue in a cellular communication scenario where both Zoya and Yoshi communicate with a base station Basil. This is because Basil takes care to ensure that everybody uses the same T and the same synchronization reference. However, one can think of a scenario in which the same bandwidth is used by two links that are belonging to different systems, but are still within a spatial proximity and can interfere with each other. For example, one link is Zoya–Yoshi and the other link is Xia–Walt. If these links use different technology, then there is no reason to impose that they are synchronized and use the same symbol timing T. Instead, it would be desirable to find a way to divide the bandwidth W into two portions of, say $\frac{W}{2}$ (Hz) each, allocate one portion to the link Zoya–Yoshi, the other portion to Xia–Walt, and let each link define its own symbol timing and synchronization reference. This would indeed bring a new quality to our communication models, enabling the coexistence of two or more independent wireless links that are within a spatial proximity, requiring only a pre-allocation of frequency bands, but not any temporal coordination among them. In order to understand how this can be achieved, we first need to see how noise enters into the model for waveform communication.

9.4 Noise and Interference

9.4.1 Signal Power and Gaussian White Noise

Physically, the waveform $\tilde{z}(t)$ transmitted by Zoya represents a voltage or a current such that the instantaneous power is proportional to $\tilde{z}^2(t)$, multiplied by a constant. We can conveniently choose the constant, without losing generality, and define the average power within a given interval of duration T as:

$$P_z = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{z}^2(t)\mathrm{d}t. \tag{9.18}$$

This choice of the constant is motivated by the way the receiver is designed to extract the in-phase and quadrature components in (9.12). Hence, the overall signal power is given as:

$$P_z = \sum_{k=1}^{\infty} z_{I,kf}^2 + z_{Q,kf}^2 = \sum_{k=1}^{\infty} P_{z,kf} \tag{9.19}$$

where $P_{z,kf}$ is the total power at the frequency $kf = \frac{k}{T}$. The physical interpretation of (9.19) is that the total power in the signal is equal to the sum of the total power carried in all the frequencies that are integer multiples of $\frac{1}{T}$.

The waveform $\tilde{y}(t)$ received by Yoshi can be obtained by adding noise $\tilde{n}(t)$ to the transmitted one $\tilde{z}(t)$:

$$\tilde{y}(t) = \tilde{z}(t) + \tilde{n}(t). \tag{9.20}$$

The equation (9.20) is a simplified model of a communication system where the only distortion of the transmitted signal occurs due to the additive noise. It is noted that the channel coefficient h, introduced in the baseband communication model, is absent and we will bring it back to the discussion later on. Since the transmitted signal $\tilde{z}(t)$ is completely specified by the signals carried by its constituent frequencies, we will describe the noise $\tilde{n}(t)$ through its impact on the frequencies observed by the receiver.

The signal that Yoshi gets when he tries to extract, for example, the in-phase component of frequency kf is:

$$y_{I,kf} = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} [\tilde{z}(t) + \tilde{n}(t)]\cos(2\pi kft)\mathrm{d}t = z_{I,kf} + n_{I,kf} \tag{9.21}$$

where

$$n_{I,kf} = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \tilde{n}(t)\cos(2\pi kft)\mathrm{d}t \tag{9.22}$$

is the projection of the additive noise, or a noise sample, to the in-phase component of the frequency kf. The common model used for the noise $\tilde{n}(t)$ is the *additive white Gaussian noise (AWGN)*. Here "additive" is self-explanatory, while "white" means that $\tilde{n}(t)$ affects all frequencies in identical way: $n_{I,kf}$ or $n_{Q,kf}$ is a Gaussian random variable with zero mean and a variance of

$$E[n_{I,kf}^2] = E[n_{Q,kf}^2] = \frac{P_\mathrm{N}}{2} \tag{9.23}$$

which is identical for each frequency kf. The noise components of two different frequencies kf and vf are uncorrelated, which for Gaussian random variables also means independent. Furthermore, the noise components of the I and Q signals at the same frequency are also independent.

AWGN is a theoretical construct of "the most random" noise that varies wildly and affects each frequency in the same statistical way. Since each of the I/Q component of a given frequency kf is affected by a noise with average power $\frac{P_N}{2}$, then the total noise power at the frequency kf is P_N. However, note that $f = \frac{1}{T}$ is a choice made by Zoya and Yoshi at the time when they established their communication system, such that Yoshi's receiver collects only the noise power that occurs at the frequency $\frac{k}{T}$, where k is an integer.

Noise exists in nature independently of this choice made by Zoya and Yoshi. Had Zoya and Yoshi agreed upon a different value T_1, then Yoshi's receiver would have collected noise power at the frequency $\frac{k}{T_1}$. In fact, the model of Gaussian noise assumes that it affects *all* frequencies and is white over a continuum of frequencies. However, since the power of the continuous signal is a sum of the powers that the signal carries at different frequencies, we are arriving at a physical paradox: if $\tilde{n}(t)$ is a white signal that has the same power at all frequencies, then the average power of $\tilde{n}(t)$ is infinite! This paradox is resolved by observing we are not interested in $\tilde{n}(t)$ per se but only in how $\tilde{n}(t)$ manifests itself in the "eye of the beholder", which here is the receiver of the communication system. The receiver processes a limited set of frequencies and therefore gets a limited noise power. By treating the noise samples as independent across frequencies, we are making the worst-case assumption about how our communication system is affected by external random disturbances. On the other hand, this implies that, from a communication perspective, any correlation among the noise samples that affect different frequencies can be useful. The existence of correlation means that, in principle, one can learn something about the noise at one frequency by observing the noise at another frequency.

Taking into account that the received signal of the Q-component has identical form as the one of the I-component in (9.21), we can compactly write the received signal at frequency kf by using a complex notation:

$$y_{kf} = z_{kf} + n_{kf} \tag{9.24}$$

where the complex noise is $n_{kf} = n_{I,kf} + jn_{Q,kf}$ and the baseband transmitted signal $z_{kf} = z_{I,kf} + jz_{Q,kf}$.

We now bring the channel coefficient back in the picture. While noise is a phenomenon present at the receiver, the transmitted signal $\tilde{z}(t)$ also undergoes a distortion when propagating from Zoya's transmitter to Yoshi's receiver. In the simplest case, the I/Q components at each frequency kf are scaled with their respective coefficients, such that Zoya's signal at frequency kf that enters Yoshi's receiver is given by:

$$h_{I,kf} z_{I,kf} \cos(2\pi kft) - h_{Q,kf} z_{Q,kf} \sin(2\pi kft) \tag{9.25}$$

which leads to the received signal:

$$y_{kf} = h_{kf} z_{kf} + n_{kf} \tag{9.26}$$

where $h_{kf} = h_{I,kf} + jh_{Q,kf}$ is the channel coefficient. We have thus arrived at the baseband model from Chapter 5; now we can rely on the principles of modulation and coding that are applicable to the baseband model. Taking that the total transmitted power at the frequency kf is given by $P_{z,kf}$, see (9.19), then from Chapter 5 we find the SNR at the frequency kf as:

$$\gamma_{kf} = \frac{|h_{kf}|^2 P_{z,kf}}{P_{\mathrm{N}}}. \tag{9.27}$$

The context of multiple frequencies that are independently modulated, while each frequency channel is scaled independently, sheds a new light on the notion of water filling. Namely, the average power P available at the transmitter needs to be distributed across the frequencies kf following the principles of water filling. For *frequency selective channels* the channel coefficients change from frequency to frequency, which causes the SNR to be different for a different frequency channel, although the noise power is identical for all frequencies. Specifically, the change in the SNR is due to the change of the *channel gain* $|h_{kf}|^2$, and not the phase of the channel coefficient h_{kf}.

When multiple orthogonal frequencies are independently modulated, the analog waveform that results from the superposition of these frequencies may have very large variations in the signal amplitude. This leads to the problem of *peak-to-average power ratio (PAPR)*. The transmitted signal is a sum of sinusoids, each of the frequencies kf independently modulated with data symbols. During a certain symbol time, it can happen that the combination of modulated data symbols is unlucky and the modulated frequencies superpose to a waveform that has a very high peak value. These high peaks put pressure on the power amplifier at the transmitter, regardless of the fact that the average power stays constant. The problem can be mitigated by avoiding the peaks while keeping the average power constant. However, any method for avoiding the undesired data combinations that lead to high peaks will necessarily restrict the way different frequencies (subcarriers) are modulated. This means that a frequency cannot be modulated independently of the other frequencies, which results in decrease in the available DoFs and therefore the data rate.

9.4.2 Interference between Non-Orthogonal Frequencies

Let Zoya and Yoshi agree upon a symbol time T_1 and communicate with each other using a single frequency f_1, where $f_1 = \frac{1}{T_1}$. Independently of them, let Xia and Walt agree upon a different symbol time T_2 and let they choose $f_2 = \frac{k}{T_2}$ as a frequency to communicate, where k is an integer. Let us assume that T_2 is chosen to be much longer than T_1, such that the symbol transmitted from Xia to Walt does not change while the symbol is sent from Zoya to Yoshi, see Figure 9.5(a). Other than that, we assume that T_2 is chosen independently from T_1, in a sense that $\frac{T_2}{T_1}$ is not necessarily an integer, such that the frequencies f_1 and f_2 are not necessarily orthogonal on the interval of length T_1. Finally, we assume that $f_2 > f_1$ and let $\Delta f = f_2 - f_1$.

The question we want to investigate is: how does the transmission from Xia to Walt interfere with Yoshi's reception of Zoya's signal? We start by observing that Yoshi receives the signal over a multiple access channel, with Zoya and Xia as transmitters:

$$\tilde{y}(t) = \tilde{z}(t) + \tilde{x}(t). \tag{9.28}$$

The signal $\tilde{z}(t)$ sent by Zoya will be referred to as a useful signal, while $\tilde{x}(t)$ sent by Xia as an interfering signal. In (9.28) we have assumed that the channel coefficients are 1 and, for

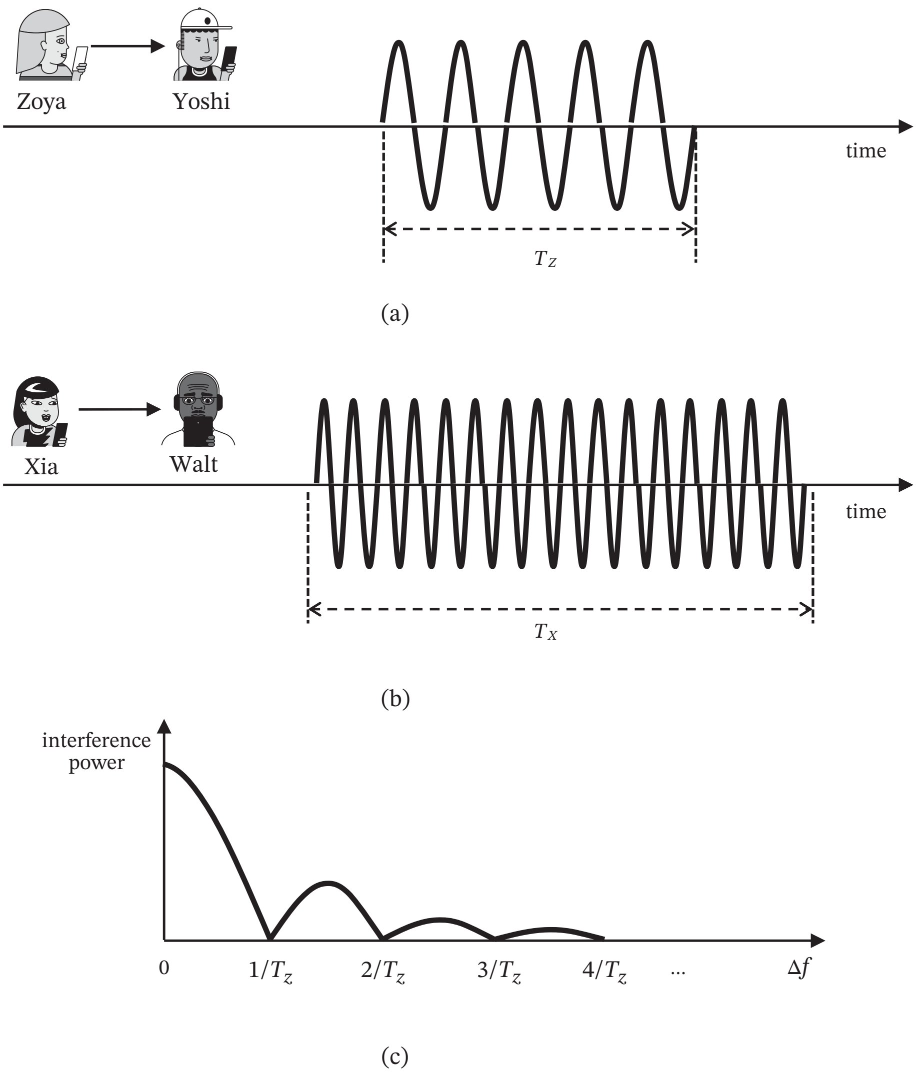

Figure 9.5 Interference between signals with different frequencies. (a) Observed signal transmitted from Zoya to Yoshi. (b) Interfering signal sent from Xia to Walt. (c) Interfering power as a function of the frequency difference.

the moment, ignore the impact of the noise. Both signals are sinusoids with amplitude 1:

$$\tilde{z}(t) = \cos(\omega_1 t) \qquad \tilde{x}(t) = \cos((\omega_1 + \Delta\omega)t + \phi) \tag{9.29}$$

where $\omega_1 = 2\pi f_1$ and $\Delta\omega = 2\pi\Delta f$. Here ϕ denotes the phase of Xia's signal at the start of the observed symbol interval in which Zoya sends to Yoshi. The phase ϕ is another element used in our discussion to indicate the lack of coordination between the links Zoya–Yoshi and Xia–Walt. Specifically, we assume that the phase ϕ of the interfering signal is chosen uniformly randomly within the interval $[0, 2\pi)$.

From the received signal Yoshi obtains:

$$y = \frac{2}{T_1}\int_0^{T_1} \tilde{y}(t)\cos(\omega_1 t)\mathrm{d}t = 1 + x \tag{9.30}$$

where 1 is the useful symbol and x is the interference that the signal at frequency f_1 receives from the signal at frequency $f_1 + \Delta f$:

$$x = \frac{2}{T_1}\int_0^{T_1} \cos(\omega_1 t)\cos((\omega_1 + \Delta\omega)t + \phi)\mathrm{d}t. \tag{9.31}$$

Clearly, the interfering signal x plays a role that is similar to the role of the noise, but the statistical characteristics are quite different. In our model, x draws its randomness from the random phase ϕ. Similar to the way we have treated noise, we look at the variance of the random variable $E[x^2]$, which measures the interfering power received by Yoshi. The expectation is taken with respect to the uniform random choice of ϕ. Differently from the Gaussian noise, the interfering power of x depends on the difference between the frequencies $\Delta f = f_2 - f_1$, as well as the choice of the symbol durations. More precisely, the average power of the interference depends on $\Delta f T_1$ and this dependency is plotted in Figure 9.5(b). It is seen that when $\Delta f T_1$ is an integer, then the interference power is zero; this is because in that case $f_2 = \frac{l}{T_1}$ and the two signals are orthogonal over the interval of length T_1.

Recall that the Gaussian noise is chosen such that it affects the signal in the "worst possible" way, without any correlation between different noise samples. In this sense, the behavior of the interference term is quite different. To illustrate this fact, let us denote by x_1 the interference term that affects the observed symbol from Zoya in Figure 9.5(a). Let us assume that, once that symbol is finished, Zoya sends another symbol of duration T_1 (not depicted in the figure), and let us denote the interference term that affects this subsequent symbol as x_2. In the considered example, the random phase ϕ of the interfering signal from Xia is selected at the start of the observed symbol. Hence, the same random choice in Xia's signal affects both symbols that Zoya sends to Yoshi. In other words, x_1 and x_2 are both dependent on the same value of ϕ and, considering that the values of T_1, T_2,f_1, and f_2 are fixed, we can conclude that the interference terms x_1 and x_2 are correlated. This is clearly different from the uncorrelated noise samples that affect the neighboring symbols. Hence, modeling the interference samples as independent Gaussian samples should be done with caution.

Returning to Figure 9.5(b), it can be seen that when two uncoordinated links use different frequencies, the impact of the interference decreases as the separation Δf between the frequencies of the two links increases. This decrease is not monotonous, as there are zeros and bumps, but the tendency is clear. When the frequencies become sufficiently separated, then the interference power becomes negligible compared to the noise power, such that the interference can be neglected. In this way the concept of frequency brings a qualitative novelty to the multi-user communication: *a sufficient separation in frequency makes the interference negligible and the links can operate simultaneously without synchronizing/coordinating the links*. This feature makes it possible to divide the frequency spectrum into different frequency channels/bands and allocate different frequency bands to different communication systems that are not coordinated. In order to get a full frequency characterization of an information carrying, random signal, there is a need to explain the concept of power spectrum.

9.5 Power Spectrum and Fourier Transform

The frequency representation based on Fourier series results in frequencies that are present in the signal (contain energy) and, consequently, frequencies that are absent from the signal (zero energy). Following this line of thought, Xia can cause interference to Zoya only if Xia puts energy at the frequency that Zoya uses for communication. On the other hand, we have seen from the example in Figure 9.5(b) that Xia can cause interference to Zoya even though they are not using the same frequency. In fact, the difference in the frequency Δf between the two interfering signals can be an arbitrary real number. Hence, there must be something contradictory about the frequency representation based on a Fourier series.

Let us look again into how we have arrived at the frequency representation. If a signal sent by Xia is limited to an interval of duration T_X, then we *can* choose the interval for the Fourier series to be T_X. Based on that, one may conclude that Xia's signal contains energy at the frequencies that are integer multiples of $\frac{1}{T_X}$. The key to the contradiction about frequencies present in the signal and interference caused to other signals lies in the choice of T_X. Recall that the we have defined the frequencies featured in a signal of duration T_X by assuming that the same signal is periodically repeated outside of the interval T_X. It is important to understand that, when the signal is not intrinsically periodic, then the choice of the interval length T_X is not unique. That is why it was stated above that the interval used in the Fourier series *can* be chosen to be T_X, but this is not the only choice.

Let us look closer in this issue and consider the signal of finite duration T depicted in Figure 9.6(a). One possible representation with Fourier series is to take $T_X = T$, from which it will follow that the signal contains energy at the frequencies $\frac{1}{T}$. Alternatively, we can choose $T_X = T_1 > T$, which will result in frequency representation that contains the frequencies $\frac{k}{T_1}$.

Hence, the contradiction vanishes by observing that the set of frequencies that are contained in a signal represented through Fourier series depend on the choice of T_X, which, for a finite-length signal, can be arbitrary. In the example in Figure 9.6(a) we only require that $T_X \geq T$. On the other hand, Figure 9.5(b) shows the statistical impact of a random interfering signal, seen as the distribution of the power across different frequencies. This naturally raises the question: can we find an objective statistical description of the way in which a certain signal interferes with random victim signals at different frequencies? Here by *objective* we mean that the description is not tailored to a specific victim signal and can be applied to arbitrary victim signals.

This kind of description is given by the *power spectrum*, also known as the *power spectral density* of the random, information carrying signal. In order to relate the notion of power spectrum to the example from the previous section, let us assume that Zoya transmits to Yoshi a periodic, infinitely long sinusoidal signal of frequency f. On the other hand, Xia sends to Walt a modulated signal created in the following way. Xia sends a sinusoidal signal of frequency $f + \Delta f$ and divides the time into slots of duration T_X, where each slot corresponds to the transmission of a single data symbol. At the beginning of the ith time slot, Xia modulates data by randomly choosing the phase of the sinusoidal signal according to the ith data symbol she wants to transmit. Then the average power of the disturbance that Yoshi experiences as a function of the frequency difference $f + \Delta f$ looks like Figure 9.5(b) and it represents the *power spectrum* of Xia's signal.

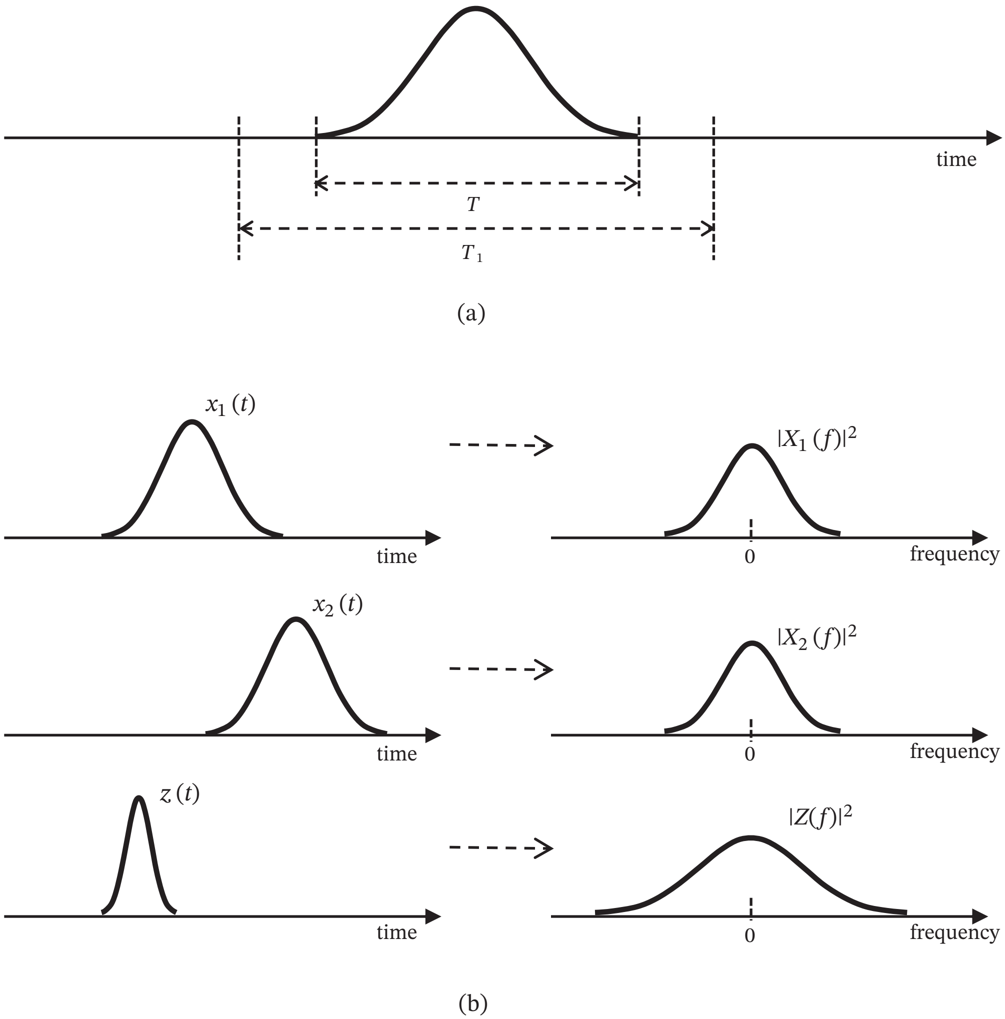

Figure 9.6 Signals with finite duration and the Fourier Transform. (a) Possible choice of two different intervals for Fourier series; in both cases the information about the useful signal is preserved. (b) Two signals $x_1(t)$ and $x_2(t)$ with identical energy spectrum, but they occur, and can cause interference at different time instances: here $z(t)$ gets interference from $x_1(t)$, but not from $x_2(t)$.

Although we have arrived at the notion of the power spectrum of a signal by treating the interference it creates to another sinusoidal signal, the power spectrum is an intrinsic property of a signal. Xia's signal is a specific type of an information carrying signal, represented by a cyclostationary statistical process, whose statistical properties vary cyclically in time. Any random signal that can be represented as a continuous function in time is also information carrying signal and one can calculate its power spectrum, provided that the random signal satisfies certain statistical properties, such as stationarity (see below). For the signals that usually occur in communication engineering, these properties are satisfied and one can safely use the power spectrum to describe how the power of a given signal is distributed across frequencies.

The precise understanding the notion of power spectrum requires introduction of the *Fourier transform*. Consider again the deterministic finite-length signal $x(t)$ of Xia in Figure 9.6 and assume that the interval T_X used in Fourier series increases its value. When T_X goes to infinity, then the interval for obtaining the Fourier coefficients approaches a correct description of the signal: it captures the fact that the signal is aperiodic and has value zero outside the interval of length T. However, the problem with the Fourier series is that infinite T_X will make all coefficients equal to zero due to the $\frac{1}{T_X}$ factor. The Fourier transform is given as:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \tag{9.32}$$

which is a (complex) continuous function of the frequency f rather than a discrete set of coefficients associated with frequencies in Fourier series. The *energy spectrum* of the deterministic signal $x(t)$ is given by $|X(f)|^2$ and the energy contained in $x(t)$ can be calculated using Parseval's theorem:

$$E_X = \int_{-\infty}^{\infty} \tilde{x}^2(t)dt = \int_{-\infty}^{\infty} |X(f)|^2 df \tag{9.33}$$

from which it is seen that the unit of $|X(f)|^2$ is joules per hertz.

Now let us assume that the signal $x(t)$ of Xia has the same finite duration as the signal in Figure 9.6(a), but the signal is random. If $x(t)$ is well behaved and has only finitely high values within that interval, then $X(f)$ exists, but it is also random and the same is valid for $|X(f)|^2$. We remind the reader that we have arrived to this point by trying to statistically characterize the interference that Xia causes to the signal that Yoshi receives from Zoya. Due to the properties of the Fourier transform, the energy spectrum $|X(f)|^2$ does not depend on the exact placement of the signal $x(t)$ on the time axis. Consider the example in Figure 9.6(b), where $x_1(t)$ and $x_2(t)$ are two possible signals that can be sent by Xia. The power spectra of $x_1(t)$ and $x_2(t)$, denoted by $|X_1(f)|^2$ and $|X_2(f)|^2$, are identical and they overlap with the power spectrum of Zoya's signal $z(t)$, denoted by $|Z(f)|^2$. However, we can see that $x_1(t)$ overlaps in time with $z(t)$ and therefore, if Xia sends $x_1(t)$, she causes interference at Yoshi's receiver. On the other hand, if Xia sends $x_2(t)$, Yoshi receives $z(t)$ free of interference. This phenomenon is not captured by the energy spectrum of the signals $x_1(t)$ and $x_2(t)$, as their energy spectra are identical.

The previous discussion leads us to conclude that, if we want to provide statistical characterization of the interference made by Xia, then we need to assume that Xia's signal has an infinite duration. In this case, regardless of where on the time axis Zoya's transmission takes place, it will experience statistically the same interference from Xia's signal. One technical requirement for this to be true is to require that Xia's random signal is stationary (in fact, wide-sense stationary). Now that we require that Xia's signal is of infinite duration, it cannot be guaranteed that the energy calculated in (9.33) is finite. On the other hand, the signal power still needs to be finite in order to put a realistic constraint on Xia's transmitter. Recall that, in defining the information-theoretic Gaussian communication channel, we have limited the power, but not the energy.

In order to arrive at the precise definition of the power spectrum, we take a cut of length T from Xia's infinite signal $x(t)$ and denote it by $x_T(t)$, such that $x_T(t) = x(t)$ for $-\frac{T}{2} \le t \le \frac{T}{2}$ and $x_T(t) = 0$ outside of that interval. Since $x_T(t)$ is of finite duration, its Fourier transform exists and is denoted by $X_T(f)$. The average power P of $x(t)$ can be found by averaging over an increasingly larger cut and we can establish the relation:

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{-\frac{T}{2}} x^2(t)\mathrm{d}t = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} x_T^2(t)\mathrm{d}t$$
$$\stackrel{(a)}{=} \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} |X_T(f)|^2\mathrm{d}f = \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{1}{T}|X_T(f)|^2\mathrm{d}f = \int_{-\infty}^{\infty} S_X(f)\mathrm{d}f \tag{9.34}$$

where (a) follows from (9.33) and $S_X(f)$ is the power spectrum. The latter equation is valid for a deterministic signal, while the desired statistical characterization of the power spectrum can be given by taking the expected value as follows:

$$S_X(f) = \lim_{T\to\infty} \frac{1}{T} E[|X_T(f)|^2]. \tag{9.35}$$

Further analysis shows that $S_X(f)$ is a Fourier transform of the autocorrelation function of $x(t)$. This means, for example, if the values of $x(t)$ are weakly correlated with the nearby values $x(t+\tau)$, where τ is small, then $x(t)$ exhibits fast variations and therefore the power density $S_X(f)$ will not be zero for high values of the frequency f. When $x(t)$ is only correlated with itself, and not with any other $x(t+\tau)$, then $S_X(f)$ has equal power density across all frequencies and thus infinite power. This is the case of the Gaussian noise, which plays the role of the worst-case uncorrelated random disturbance.

An important property of the Fourier transform is that a signal of finite duration does not have a limited spectrum and there is no frequency above which the energy density of the signal is strictly zero. This would imply that two finite-duration signals that overlap in time will, highly likely, not be separable in frequency and will therefore cause interference to each other. The statistical power spectrum is calculated for a signal that is stationary. On the other hand, no finite duration signal can be considered stationary, since after some time, the finite duration signal will be identically equal to zero.

Assessing the interference requires that we observe the power spectrum and make assumptions about how the signals overlap in time. For example, take one symbol of duration T sent by Zoya and placing it randomly in time such that this symbol overlaps fully with a transmission from Xia. From the perspective of this symbol of Zoya, the signal of Xia satisfies the statistical properties of an interfering signal that has infinite duration and we can use the power spectrum of Xia to assess the power (variance) of the interference caused to Yoshi's receiver. When the power spectrum of Xia is far below the noise level at Yoshi's receiver, and therefore far below the level of Zoya's signal, then we can assume that the signals of Zoya and Xia are separated in frequency.

We have thus arrived at the answer to the question posed earlier in this chapter: *how do you separate in frequency the signals from two communication links that are not synchronized in time?* The separation in frequency is rather approximate: there will always be an interference between the signals, as a signal that is strictly bandlimited has an infinite duration.

9.6 Frequency Channels, Finally

In the first chapter we used the concept of a frequency channel and transmitters/receivers that are tuned to a given channel, but at this point we have a better picture about the physical meaning of a frequency channel. A waveform, modulated with data, is said to occupy a certain frequency band if the energy contained outside of that band, i.e. the values of its power spectrum outside of that band, are negligible. A frequency channel is band of contiguous frequencies, associated with a certain *filter mask*. The role of a filter mask is to provide bounds that limit the amount of energy radiated outside a given frequency band. In this way, the interference created towards the signals transmitted within the adjacent frequency bands is low or, at least, acceptable. Furthermore, the filter mask also limits the energy of the signal within the frequency band of interest, which is related to the limit on the maximal transmission power that can be used in a given frequency band. The analog signal that conforms to the filter mask can, for all practical purposes, be treated as a bandlimited signal.

9.6.1 Capacity of a Bandlimited Channel

Let W be the frequency band in which the communication signal should contain its energy, or at least most of it. The noise is white, having a flat spectral density of N_0 over the entire bandwidth W. Hence, the total noise power that affects the received signal is $P_{\mathrm{N}} = N_0 W$. Let T be a very long time interval in which we observe the communication signal. Then, assuming that the power of the received information carrying signal is P, what is the capacity of the channel in bits per seconds?

Since the signal is constrained to have most of its energy to the bandwidth of size W, the total number of real DoFs that are present in the signal observed through a duration T is $m = 2WT$. In deriving the capacity, we will assume that all m DoFs are statistically identical: the received power of the useful signal at each DoF is identical and each transmitted symbol at a given DoF is affected by an identical random process that represents the noise. The signal power per DoF can be calculated as:

$$P_{\mathrm{DoF}} = \frac{P}{2WT} \tag{9.36}$$

while the noise power per DoF can be calculated as

$$P_{\mathrm{DoF,N}} = \frac{P_{\mathrm{N}}}{2WT} = \frac{N_0 W}{2WT} = \frac{N_0}{2T}. \tag{9.37}$$

The m DoFs can be treated as m channel uses of the same AWGN channel in which the SNR is equal to:

$$\gamma = \frac{P_{\mathrm{DoF}}}{P_{\mathrm{DoF,N}}} = \frac{P}{N_0 W} \tag{9.38}$$

and the capacity per one real channel use (DoF) is:

$$C^{\mathrm{DoF}} = \frac{1}{2}\log_2(1+\gamma) = \frac{1}{2}\log_2\left(1 + \frac{P}{N_0 W}\right) \quad \text{(bit/DoF)}. \tag{9.39}$$

For fixed W, this capacity is achieved when T goes to infinity, such that n also goes to infinity and the capacity in bits per second is given as:

$$C_W = \frac{2WTC^{\mathrm{DoF}}}{T} = W\log_2\left(1 + \frac{P}{N_0 W}\right) \quad \text{(bps)} \tag{9.40}$$

which is the well known expression by Shannon for the capacity of bandlimited signals.

At the beginning of this chapter, we stated that it is not immediately clear why it is optimal to communicate with analog waveforms through a layered (separation) approach, which consists of modulating discrete samples and then using them to synthesize an analog waveform. The intuitive answer comes from the sampling theorem by Nyquist–Shannon–Kotelnikov, which states that a bandlimited signal can be completely described by the sampling process with a sufficiently high frequency. One concern comes from the fact that the communication signal is not strictly bandlimited, but it also contains some energy outside of its prescribed band. In fact, one may aim to squeeze more bits through the channel by modulating directly the analog waveforms while still satisfying the constraints of the filter mask. However, as rigorously proved by Slepian, the separation into sampling and synthetizing waveforms is optimal in sense of communication capacity, when the transmitted data packets are asymptotically long.

By normalizing the data rate with the bandwidth W, we get the *spectral efficiency*, denoted by ρ. The maximal spectral efficiency can be obtained by normalizing (9.40) with W:

$$C = \log_2(1+\gamma) = \log_2(1+\text{SNR}) \quad ((\text{bit/s})/\text{Hz}) \tag{9.41}$$

which corresponds to a capacity of a channel in which each channel use is a complex value. Hence, one should always choose the spectral efficiency to be $\rho \leq C$.

The energy spent per transmission of a single complex symbol can be calculated as

$$E_\text{s} = \frac{PT}{WT} = \frac{P}{W}. \tag{9.42}$$

We have divided by WT instead of $2WT$, as we want to find the energy per one complex DoF. On the other hand, the energy of the noise per one complex DoF is $\frac{N_0 WT}{WT} = N_0$, such that we get:

$$\text{SNR} = \frac{P}{N_0 W} = \frac{E_\text{s}}{N_0}. \tag{9.43}$$

For given spectral efficiency ρ, expressed in bits per symbol, one can calculate the normalized measure of the *signal-to-noise ratio per information bit*:

$$\frac{E_\text{b}}{N_0} = \frac{E_\text{s}}{\rho N_0} = \frac{\text{SNR}}{\rho}. \tag{9.44}$$

The normalized SNR is sometimes denoted by E_b/N_0. In order to get intuition about it, the inequality that needs to be satisfied by the spectral efficiency can be written as:

$$\rho \leq \log_2\left(1+\rho\frac{E_\text{b}}{N_0}\right) \tag{9.45}$$

which means that the minimal E_b/N_0 that is required to attain a certain spectral efficiency ρ can be found from:

$$\frac{E_\text{b}}{N_0} \geq \frac{2^\rho - 1}{\rho} \tag{9.46}$$

which is termed the *Shannon limit* and is used as a reference to determine how efficient a given modulation/coding scheme is. In order to send reliably a single bit via each complex channel symbol, the minimal E_b/N_0 can be calculated to be 1 or 0 dB. An interesting case is the minimal required E_b/N_0 to support reliable communication with spectral efficiency that approaches zero $\rho \to 0$, which is $E_\text{b}/N_0 = -1.59$ dB. The interpretation of the case $\rho \to 0$

can be obtained by writing $\rho = \frac{C_W}{W}$ and letting the bandwidth of the signal W go to infinity. Hence, the minimal energy per bit is achieved by using a very wide bandwidth to transmit the data signal.

9.6.2 Capacity and OFDM Transmission

We have derived the capacity by considering bandwidth W and a duration T in which the total number of DoFs is $2WT$. Additionally, we have assumed that all DoFs are statistically identical. In Section 9.3.3 we introduced the parallel frequency channels in an OFDM transmissions, which is one possible way to use the DoFs offered by the bandwidth W. There are $m = WT$ subcarriers, each carrying 2 DoFs corresponding to the I and the Q components, respectively. The simplest analog waveform that can be used to modulate the information on a subcarrier is a rectangular pulse of duration T. This means that, at the start of a symbol, we select the I/Q components of each subcarrier according to the data that needs to be modulated and keep them constant until the next symbol, when a new selection is made.

Let us observe l consecutive symbols, each of duration T. The total number of channel uses (DoFs) during these symbols is $2lm = 2lWT$. If all subcarriers are statistically identical, such that the received power at each subcarrier is identical, then all DoFs are statistically identical and the capacity of an OFDM transmission is given by the capacity derived in the previous section. However, in frequency selective channels the received power across subcarriers differs, as explained in Section 9.4.1. In this case, the capacity of an OFDM transmission should be computed as a capacity of a water filling transmission over parallel, non-identical Gaussian channels. Note that, if all DoFs are statistically identical, then a single codeword is spread across all subcarriers and consumes all $2lm$ channel uses. On the other hand, if each subcarrier is modulated independently, then there are m different codewords, each codeword consuming l channel uses.

The power spectrum of an OFDM signal for which each subcarrier is modulated with rectangular pulses is, clearly, not confined to a bandwidth W, and part of its energy is leaked outside of that band. The shape of the power spectrum can be changed by changing the shape of the pulse in an OFDM transmission, while being careful not to destroy the orthogonality. Another way to shape the power spectrum is to depart from the OFDM paradigm and not consider a symbol-per-symbol transmission, but treat the set of $2WT$ DoFs as a general set of channel uses, not necessarily statistically identical, that can be used to create suitable codewords and send data.

9.6.3 Frequency for Multiple Access and Duplexing

One way to use the availability of different frequency channels is through FDMA, where each frequency channel is allocated to a different transmitter or different link. If the link Zoya–Yoshi uses a frequency channel that does not overlap with the frequency channel of Xia–Walt, then the interference between these two links is negligible, practically zero. The central frequency f_0 of a channel is called a *carrier*. This term reflects the traditional way in which a data-carrying signal is placed within a frequency channel. The signal is modulated in a lower band and then multiplied/mixed with a sinusoidal signal of frequency f_0 to be carried into the frequency channel centered at f_0. Finally, the signal is sent through a filter centered at f_0 in order to limit its bandwidth to the constraints set by the frequency channel. In the sequel we will refer to a frequency channel through its center frequency, for example "channel f_0".

The separation of the signals in frequency removes the need for spectrum sharing and/or time synchronization between two unrelated links, such as Zoya–Yoshi and Xia–Walt. There is still need for coordination to avoid interference, as one link should choose the channel f_0 and the other one the channel f_1. Furthermore, there is also a requirement on synchronization, as the devices need to be capable of generating the frequencies f_0 and f_1. Imperfectly tuned frequencies do not guarantee that the interference between the two signals will follow the levels that are prescribed by the filter mask. Yet, speaking in terms of interference avoidance, acceptable frequency synchronization is not as restrictive as time synchronization. Recall from the Fourier series representation that time synchronization requires that both links use identical symbol time, which is not required in FDMA.

From Chapter 1 we have assumed that *time division duplexing (TDD)* has been used: Zoya and Yoshi choose a single frequency channel and they take turns in using it as transmitters/receivers, respectively. For example, this has been used in a setting with multiple terminals connected to Basil and the use of suitable frames for downlink and uplink transmission. In the same chapter we also described the possibility of full duplexing, which is simultaneous transmission and reception over a *single* frequency channel. The availability of multiple frequency channels brings the opportunity to introduce *frequency division duplexing (FDD)*: Zoya and Yoshi agree upon two frequency channels, f_1 and f_2. Zoya uses the channel f_1 to transmit to Yoshi, such that Yoshi's receiver is tuned to f_1. Conversely, Yoshi's transmitter and Zoya's receiver are tuned to f_2. Hence, Zoya and Yoshi can simultaneously transmit to each other, without causing interference, or more precisely, *self-interference* from its own transmitted signal to the received signal.

Neither time division nor frequency division has a definitive advantage over the other. For example, one advantage of TDMA over FDMA is that the amount of time resource allocated to a user can be dynamically adjusted. In Chapter 1 we have described a TDMA system with multiple terminals and a base station, in which the latter had the flexibility to allocate resources dynamically from frame to frame. This is not the case in FDMA, since the frequency filters do not have the same flexibility and cannot be easily and dynamically adjusted to allocate a different bandwidth. On the other hand, the advantage of an FDMA system is that it can be used by multiple non-sychronized links.

In terms of duplexing, the cost in a TDD system is the turnaround time needed to switch from transmitting to receiving and vice versa. Analogously, the cost in an FDD system is the gap, unused bandwidth, which separates the band for transmission from the band for reception. Full duplexing is appealing as it seems to integrate the advantages of both TDD and FDD. However, while we treat TDD and FDD as interference free; the price paid in full duplexing is that the receiver needs to suppress the self-interference at the same frequency. Another important difference between TDD and FDD occurs when we take into account the spatial dimension and the fact that the propagation of wireless signals takes actual time, as discussed in the next chapter.

We conclude this section by noting the fundamental equivalence of TDD and FDD in terms of data rates. Let us first look at FDD, fix the time slot for a transmission to be T and let one frequency channel have a bandwidth W. The total bandwidth used in FDD is $2W$ and the total number of DoFs in the observed time-frequency chunk is $2WT$. For fair comparison with the TDD system, it has to operate with a bandwidth of $2W$, such that the total number of DoFs is again $2WT$. However, the switching between uplink and downlink happens after a time $\frac{T}{2}$, but that is irrelevant for the total data rate achieved in both directions. In this idealized comparison, we neglect the DoFs that are used due to the frequency gap in FDD and turnaround time in TDD.

9.7 Code Division and Spread Spectrum

9.7.1 Sharing Synchronized Resources with Orthogonal Codes

Time division and frequency division are examples of orthogonal multiplexing, where each transmission occupies exclusively a time slot or a frequency band and in that way interference in avoided. When the transmitters or the links are not synchronized in time, strict orthogonality in frequency is impossible, as there is always residual interference. However, this interference is negligible when the frequency bands are sufficiently separated.

We now address the question whether the two signals can occur simultaneously in time, occupying the same frequency bandwidth, but the respective receiver can still extract the desired signal with little or even no interference. In fact, we have already stated that this is possible if the two interfering links, Zoya–Yoshi and Xia–Walt are synchronized in time. Alternatively, in an uplink scenario all terminals are synchronized to Basil, a common base station. Let us treat the simpler case, in which the mobile devices MD_1, MD_2, MD_3 and MD_4 communicate with Basil in the uplink. Basil determines the symbol duration T_0 and the timing of the slot symbols, such that the four terminals are synchronized to it. Furthermore, Basil determines how to allocate the communication resources to the terminals. Since the symbols of all terminals are synchronized, they can orthogonally share the frequencies $\frac{k}{T_0}$, where k is an integer. Those frequencies appear through the Fourier series representation of the symbols. Let W_0 be the width of the frequency band that is available for communication and is shared by all terminals. Then the total number of orthogonal resources available for communication, each of them regarded as a DoF, within the duration of a single symbol is $2W_0T_0$.

To make the example more specific, we assume that $W_0 = \frac{4}{T_0}$, such that there are four available frequencies in the band W_0. One way in which Basil can decide to share the DoFs among the terminals is to use frequency division and allocate each frequency exclusively to one terminal. However, frequency division is *only one possibility* for orthogonally sharing the available $2W_0T_0 = 8$ DoFs among the four terminals. Recall that each frequency carries two DoFs, associated with cos(I-component) and sin(Q-component), respectively. Basil can thus allocate the I component of a given frequency to one terminal and the Q component to another. In fact, one can check that there are $\frac{8!}{2!2!2!2!}$ possible orthogonal allocations and each of them is structurally analogous to a time or a frequency division.

The main feature of the orthogonal allocations described so far is that each DoF is exclusively used by one user. We are now interested in a conceptually different type of division that is based on a code. Basil allocates a *code sequence* to each of the four terminals, each sequence consisting of 4 symbols. The allocation is given in Table 9.1. We will call these sequences *spreading sequences* or *spreading codes* as they will be used to spread the

Table 9.1 Orthogonal codes allocated to the terminals.

Terminal	Code			
MD_1	1	1	1	1
MD_2	1	−1	1	−1
MD_3	1	−1	−1	1
MD_4	1	1	−1	−1

same data symbol across multiple DoFs. The scalar product between the sequences of, for example, the mobile devices MD_1 and MD_3 is:

$$1 \cdot 1 + 1 \cdot (-1) + 1 \cdot (-1) + 1 \cdot 1 = 0. \tag{9.47}$$

The reader can check that the scalar product between two different sequences is always zero, while the scalar product of a code sequence with itself leads to the value 4.

The sequences are orthogonal in a sense that the scalar product of two different sequences is zero. Recall that the orthogonality of two frequencies on a given time interval enables the receiver to extract the data that is modulated onto one frequency without an interference from the data modulated onto the other frequency. Here the idea is to modulate the data onto a code instead of a frequency and use the orthogonality among the sequences to be able to extract the desired data without interference from the other data transmissions.

Let b_i denote the complex data symbol sent by the terminal MD_i, where $i = 1, 2, 3, 4$. In order to make the presentation of the main ideas easier, let us at first assume that the channel coefficient from each terminal to Basil is always equal to 1, regardless of which DoFs are used for transmission. For example, if all terminals send their symbol using the jth of the four available frequencies, then the complex signal that Basil receives is at this frequency is

$$y_j = b_1 + b_2 + b_3 + b_4 + n_j \tag{9.48}$$

where n_j is the complex noise at the jth frequency. We use a convenient notation for the codes from Table 9.1: let c_{ij} denote the jth element of the code for the terminal MD_i. Then MD_i transmits its symbol b_i at the frequency j by multiplying it by c_{ij}, such that the received signals at the four frequencies are:

$$\begin{aligned} y_1 &= b_1 + b_2 + b_3 + b_4 + n_1 \\ y_2 &= b_1 - b_2 - b_3 + b_4 + n_2 \\ y_3 &= b_1 + b_2 - b_3 - b_4 + n_3 \\ y_4 &= b_1 - b_2 + b_3 - b_4 + n_4. \end{aligned} \tag{9.49}$$

In order to decode the signal from MD_1, Basil uses the code of MD_i, multiplies y_j by c_{ij} and sums them up, which leads to:

$$r_i = \sum_{j=1}^{4} c_{ij} y_j = 4b_i + \sum_{j=1}^{4} c_{ij} n_j. \tag{9.50}$$

As already indicated above, the code orthogonality removes the interference among the users and Basil needs to make a decision on the transmitted symbol b_i based on r_i, affected only by the noise. Let $P_{\rm N}$ be the power of each of the noise samples n_j. Since $|c_{ij}|^2 = 1$ the total noise power is $4P_{\rm N}$. Similar to maximum ratio combining, discussed in Chapter 5, the desired signals are combined coherently, unlike the noise. If the average power of the transmitted symbol b_i is P, then the total power of the desired signal $4b_i$ is $16P$, such that the SNR at which the desired symbol is decoded is:

$$\gamma_i = \frac{16P}{4P_{\rm N}} = \frac{4P}{P_{\rm N}} = 4\gamma_T \tag{9.51}$$

where γ_T is an SNR calculated for each individual transmission of a complex symbol. In other words, the resulting SNR is identical as if only the terminal MD_i repeats its transmitted symbol b_i four times, without any transmission from the other users. This form of multiple access is called *code division multiple access (CDMA)* and it is different from TDMA and FDMA in that no DoF is allocated exclusively to a user.

In our example we have used a code to spread the data symbols across four frequencies, such that y_j is the received signal at the jth frequency. However, from the equation (9.49) it does not follow that y_j should be a DoF associated with a frequency: it can be any subset of two DoFs from the eight available ones. As an example, instead of spreading the symbols in frequency, as done above, we can spread the symbols across time. Let us look at the same setup in which the data symbol is transmitted during time T_0, using bandwidth $W_0 = \frac{4}{T_0}$, but a different way to use the eight available DoFs. Each terminal transmits four times, the duration of each transmission is $\frac{T_0}{4}$ and this will be referred to as the *chip duration*, since the symbol duration remains as T_0. Now y_j can be interpreted as the jth received chip. Note that the noise power per DoF must be invariant with respect to the way the DoFs are used, i.e. frequency or time. Hence, the noise n_j retains the same statistical characterization as in (9.49).

Here we introduce some standard terminology. The variant in which spreading is done with chips is called the *direct sequence spread spectrum*. The reason it is called a spread spectrum is that, considering that the duration of the symbol is T_0, a sufficient bandwidth to send one complex symbol is $\frac{1}{T_0}$. However, in order to repeat the same symbol four times during the same symbol interval, the bandwidth should be spread to $\frac{4}{T_0}$. The spreading factor G, in this case $G = 4$, is also called a *processing gain*, since the way the received signal is processed, see (9.51) is equivalent to maximum ratio combining in which the number of received instances of the symbol is equal to G. For simplicity, we use spread spectrum to denote any form of spreading symbols across multiple DoFs, not only spreading in time.

9.7.2 Why Go Through the Trouble of Spreading?

We could show that what spreading can offer in multiplexing four terminals is basically the same as if each user has been exclusively allocated orthogonal resources. Indeed, the total power used per symbol by each terminal is $4P$ such that the same SNR can be achieved when the symbol is transmitted only once, over exclusively allocated DoFs, with a power of $2\sqrt{P}$. Why then go through the trouble of spreading when the same thing can be achieved in a simpler way?

One reason why the advantage cannot be appreciated is that we have not taken into account the constraints that come from the hardware side. Namely, transmitting at power

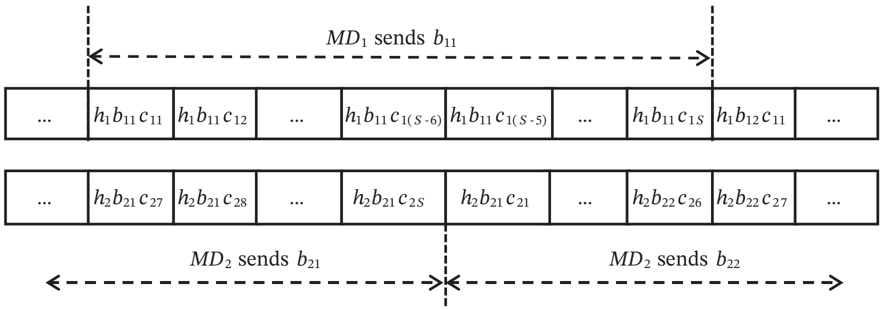

Figure 9.7 Illustration of chip synchronous, but not symbol synchronous, transmissions of two mobile device MD_1 and MD_2.

P over l channel uses is better from the perspective of a power amplifier compared to transmitting over a single channel use with power lP. Another reason why we cannot see any advantage in using a spread spectrum is that we have made the assumption that all the terminals are perfectly synchronized to Basil and all the channel coefficients are equal to 1. We now remove these assumptions and investigate the consequences.

Let the number of chips in the spreading sequence be S, but now we select S to be much larger than one. We consider spreading in time, such that each symbol spreads through S complex channel uses. The channel coefficient of the mobile device MD_i to Basil is h_i and does not change across all channel uses. In the first step we assume a light form of asynchronism (this will be briefly revised afterwards): the terminals are assumed to be chip synchronous, but not necessarily symbol synchronous. This means, for example, that the kth chip of a symbol sent by MD_1 occurs at the same time as the vth chip of a symbol sent by MD_2, but k and v are not necessarily equal. We use b_{ij} to denote the jth symbol of the ith terminal. The notation can easily get complicated, such that we use Figure 9.7 to obtain some insight.

Assume at first that Basil knows the chip at which a new symbol of MD_1 starts. He also knows the (different) chip at which a new symbol for MD_2 starts. For each transmitted chip, Basil receives a complex value that is a sum of what is sent by MD_1 and MD_2, with the appropriate channel coefficients, plus noise (not depicted in Figure 9.7). Thus, the received values that correspond to the chips associated with MD_1 transmission of b_{11} are:

$$\begin{aligned}
r_{11} &= h_1 b_{11} c_{11} + h_2 b_{21} c_{27} + n_{11} \\
r_{12} &= h_1 b_{11} c_{12} + h_2 b_{21} c_{28} + n_{12} \\
&\qquad\qquad\qquad\qquad \cdots \\
r_{1C} &= h_1 b_{11} c_{1S} + h_2 b_{22} c_{26} + n_{1S}.
\end{aligned} \tag{9.52}$$

In order to decode the transmission from MD_1, Basil correlates the received chips with the spreading code of MD_1:

$$\begin{aligned}
r_1 &= \sum_{i=1}^{S} r_{1i} c_{1i} \\
&= S h_1 b_{11} + \left[h_2 b_{21} \sum_{i=1}^{S-6} c_{1i} c_{2(i+6)} + h_2 b_{22} \sum_{i=S-5}^{S} c_{1i} c_{2(i-S+6)} \right] + \sum_{i=1}^{S} c_{1i} n_{1i}.
\end{aligned} \tag{9.53}$$

This operation is sometimes referred to as *despreading* of the signal, as it effectively collects the contributions spread over multiple DoFs into a single one. The second member of the sum in (9.53), put in brackets, is the interference experienced from MD_2 and the last member is the noise collected from all chips. The asynchronous shift between MD_1 and MD_2 for the example is arbitrarily chosen to be 5. Note that it is not possible to choose the spreading sequences such that orthogonality is preserved and interference vanishes for every possible asynchronous shift between MD_1 and MD_2.

On the other hand, assuming S is large, let us assign to MD_1 and MD_2 random spreading sequences. Before the communication starts, each chip is generated by flipping a fair coin and selecting 1 or −1, according to the flipping outcome. As such, the product of two chip values $c_{1i}c_{2j}$ is a binary random variable that gets value 1 or −1 with equal probability. When S is large and the central limit theorem is put to work, the sum in the brackets from (9.53) will converge towards a Gaussian distribution, such that the interference becomes equivalent to an additional Gaussian noise. On the other hand, since it is always $c_{1i}^2 = 1$, the contributions of the desired signal from all chips are coherently combined, leading to the processing gain. The SINR at which the signal is decoded is:

$$\mathrm{SINR} = \frac{S^2 P}{SP_I + SP_\mathrm{N}} = \frac{SP}{P_I + P_\mathrm{N}} \tag{9.54}$$

where P is the received power for the desired signal from a single chip, P_N is the power of the noise affecting each chip and and P_I is the power of the interference contributed through each chip. The value of P_I is not zero and it depends on the *cross-correlation* of the spreading sequences. The cross-correlation can be controlled through the selection of a suitable set of spreading sequences. Suitable choices include *pseudorandom* or *pseudonoise (PN)* sequences, and Gold and Kasami sequences. The price for getting better cross-correlation properties is the decrease in the size of the set of available spreading sequences, which practically means decreasing the number of terminals that can stay connected to Basil.

We now remove the assumption that the transmissions of different users are chip-synchronous at Basil's receiver. Note that treating the chips as complex values in (9.52) assumes that a sampling process has already taken place and it is perfectly aligned to the transmissions of both users. In practice, the waveforms sent by different users are misaligned in arbitrary ways and the receiver needs to find the timing information about each of the transmitted spread spectrum signals. This is *timing acquisition*, which can be carried out by letting the terminals transmit *training sequences*. For these sequences, the data symbol is known a priori by the receiver and the received symbols are used to detect the start of the spreading sequences and estimate the channel. For the example from (9.52), MD_1 can obtain a training sequence by setting the symbol value to $b_1 = 1$ and using its spreading sequence, where the latter is known a priori by Basil. Then Basil applies synchronization methods to find out the timing of the chip transmission for MD_1. However, in order to be able to align to the first chip of the sequence and not to another one, the spreading sequence should posses good *autocorrelation* properties and be decorrelated with the shifted versions of itself. Finally, the interference from the MD_2 is again determined by the cross-correlation, but also by the misalignment of the chip timing for the two terminals.

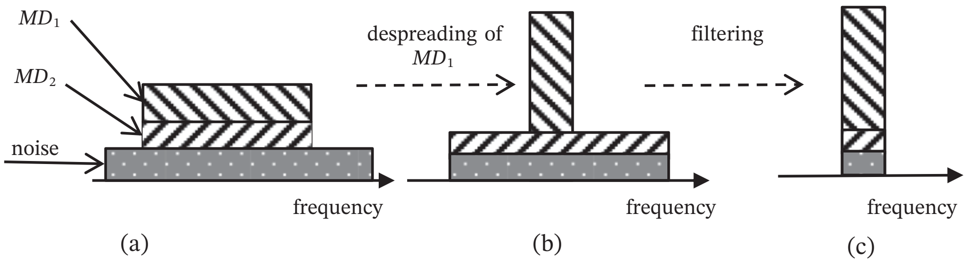

Figure 9.8 An illustration how CDMA works and why the effect of a spread spectrum signal is similar to that of a white noise. (a) The combination of the spread signals of MD_1 and MD_2 and the noise. It should be noticed that noise has a bandwidth that is wider than the bandwidth of the spread spectrum signals. (b) Despreading of the received signal by using the code sequence of MD_1. (c) Filtering out the narrowband, information carrying signal of MD_1.

9.7.3 Mimicking the Noise and Covert Communication

Figure 9.8 shows how CDMA works. The two wideband (spread) signals, see are added up, as in Figure 9.8(a), over the multiple access channel. The receiver uses the code of MD_1 to "unlock" or despread the signal from the mix of multiple signals. After despreading with the code sequence of MD_1, see Figure 9.8(b), the receiver filters out the frequencies that are not necessary to decoded the narrowband information-carrying signal from MD_1, see Figure 9.8(c). In this way, the decoding of the data sent by MD_1 is disturbed by interference power from MD_2 that is only a fraction of the total power that MD_2 contributes to the received signal. Note that here we have used the term "narrowband" information carrying signal to indicate that the bandwidth necessary to carry the signal from MD_1 before being spread is (much) lower than the bandwidth occupied after spreading.

The same figure brings more important insight. Before the spread spectrum signal is correlated and despread with the correct spreading sequence and correct timing, it appears as a signal that is similar to the noise. Indeed, the data is inherently random and the unknown spreading sequences can also be treated as random. Adding to this the fact that the chips are changing S times faster compared to the changes in the data symbol, then we arrive at a characterization similar to the one of the Gaussian noise, with wild and unpredictable changes. This feature has been noticed already from the early days of spread spectrum and it is useful to carry out *covert communication*. To elaborate, since a spread spectrum signal is mimicking the noise, a potential eavesdropper has a hard time determining that any communication is going on. In other words, spread spectrum signals can be used to achieve low *probability of interception*. Furthermore, if the eavesdropper does not know the spreading sequence of the signal he wishes to intercept, then even if he detects that there is an ongoing communication, he cannot decode it. Thus, the communication can, in principle, be kept secret and the secrecy level is proportional to the length S of the spreading sequence.

Speaking about secrecy, there is another form of spread spectrum, namely *frequency hopping*. This was originally invented by the actress Hedy Lamarr to enable secret radio communication. The main idea is to have a transmission that does not use the same frequency channel, but changes the channel at predefined time intervals. The change of the frequency channel is done according to a code that is known by the transmitter and the receiver, but not the eavesdropper. This technique can be treated as a spread spectrum, since the total spectrum that contains all the frequency channels that are hopped over is much larger than the bandwidth of the single frequency channel. Recall that, in a direct sequence spread spectrum, the information carrying signal is spread simultaneously over many DoFs. On the other hand, in frequency hopping only part of the DoFs is used at a given time. Those DoFs correspond to a frequency channel whose bandwidth is lower than the total bandwidth used for communication.

Depending on the time that the signal spends at a given frequency channel before changing to another, frequency hopping can be slow or fast. In slow frequency hopping, the frequency is changed only after a symbol, a group of symbols or a full packet is transmitted. On the other hand, in fast frequency hopping, the frequency channel may be changed multiple times even within a symbol duration.

Besides frequency hopping, there is also the concept of *time hopping*, associated with *ultrawideband (UWB)* transmissions, also called *impulse radio*. Here data is sent using very short pulses, which implies very large bandwidth. An interesting feature of UWB is that there is no carrier that is modulated, the bandwidth is determined by the pulse bandwidth itself. UWB uses time hopping in order to support coexistence/mitigate interference among multiple links.

9.7.4 Relation to Random Access

We have presented spread spectrum/CDMA as a method for non-orthogonal multiple access that enables mitigation of the interference among multiple transmissions. One can argue that random access protocols can also be seen as a class of communication mechanisms that are based on non-orthogonal access. Indeed, in the classical collision model, a random transmission choice may lead to collision or interference and, after randomly timed re-attempts, it eventually leads to a successful, non-interfered transmission. However, random access operates at a *packet level* and its objective is to arrive to a state of non-interfered transmission. On the other hand, CDMA *embraces* the interference and suppresses it through chip/symbol level processing.

A step that bridges the differences between the concept of a random access protocol and CDMA can be taken by using a model for random access that is based on a capture and successive interference cancellation (SIC). Upon receiving the set of collided/interfered signals, Basil applies correlation with all possible spreading sequences and finds the sequence, say the one of MD_1, that leads to the highest received power. Then Basil attempts to decode the signal of MD_1, treating all the remaining interference as a noise. If the decoding is successful, then a capture occurs, the signal of MD_1 can be canceled and Basil can continue to look for the spreading sequence that results in the next strongest signal. In other words, CDMA enables capture and intra-collision interference cancellation.

Random access can also be related to randomized frequency hopping. Namely, instead of choosing the randomized retransmissions only in the time dimension, one can also use the frequency dimension and, upon retransmission, also choose a different frequency channel. However, the receiver needs to know which frequency channel it needs to be tuned to in order to receive this transmission. Hence, if the transmitter needs to have a full freedom in selecting the frequency channel for retransmission, as it has in the time dimension, then the receiver needs to monitor all the frequency channels simultaneously, which brings additional costs in terms of receiver architecture and energy expenditure.

9.8 Chapter Summary

The concepts of frequency, Fourier analysis, and representation of signals are usually the prerequisites that all students need to have before going to the first course in communication engineering. Here the objective has been to provide a different perspective to those concepts, after being primed by the abstract communication models in the previous chapters. This chapter has touched upon very elementary questions, such as the following paradoxical observation: all signals encountered in practice are bandlimited, but then the Fourier transform implies that they have to have an infinite duration. It has been shown under which statistical conditions the power spectrum is defined and therefore the actual band where a certain signal resides. It has also been shown that frequency separation among different communication systems is the way to ensure absence of mutual interference without synchronization and continuous coordination among the systems. Finally, we have introduced the idea of spread spectrum and CDMA and related it to the previous discussions on random access protocols.

9.9 Further Reading

Two classical papers that treat the fundamental questions of frequency analysis and bandwidth are Gabor [1946] and Slepian [1976]. Interestingly, as pointed out in Slepian [1976], the road to the rigorous proof of the result on the capacity of channels with analog waveform was not easy and somehow missed by Shannon in his original paper Shannon [1948]. Two classical books that treat these issues in depth are Wozencraft and Jacobs [1965] and Gallager [1968]. The reader can find further discussion on multiple access (TDMA, FDMA, CDMA), OFDM and spread spectra in standard textbooks on wireless communication, such as Goldsmith [2005] and Molisch [2012].

9.10 Problems and Reflections

1. *System design for FDMA*. In Chapter 1 we presented several frames for single-channel wireless communication systems based on TDMA. Propose a similar frame based design for a system with multiple frequency channels. Focus on a downlink transmission and take into account the fact that signaling information should be received by all users. Consider the possibility for flexible allocation of resources, e.g. multiple frequency channels to a user.
2. *Co-channel interference*. Figure 9.5(c) shows an example of inter-carrier interference that decreases as the carrier frequencies become more separated. This could be used as a basis for modeling interference among different frequency channel. In that model, the communication performance of one channel is not independent of whether another frequency channel has an active transmission or not.

(a) Build a communication model that takes into account interference from other frequency channels. This is often called adjacent channel interference, but strictly speaking, the model should also take into account interference from non-adjacent channels.

(b) If these frequency channels are used as parallel channels from Zoya to Yoshi and Zoya has a limited power, investigate how the power allocation changes compared to the water filling solution.

3. *Wireless slicing with multiple frequencies*. In Chapter 4 we introduced slicing of wireless resources and illustrated it for two services, broadband and reliable low latency, respectively. The study was done in a single channel and two transmitters. Propose a similar system design, but now with more than two transmitters and with multiple available frequency channels.

4. *Frequency hopping and collisions*. Consider two wireless links that are in proximity, one from Zoya to Yoshi and the second one from Xia to Walt. Let W be the total bandwidth available for communication. The link Zoya–Yoshi divides the bandwidth W into F_1 channels, such that the bandwidth of each channel is $\frac{W}{F_1}$. The link Xia–Walt divides the bandwidth W into F_2 channels, such that the bandwidth of each channel is $\frac{W}{F_2}$. The transmissions Zoya–Yoshi are using are slots of size T_1, while the transmissions Xia–Walt are using are slots of size T_2. Both links are based on slow frequency hopping: for example, after Zoya transmits a packet at a channel $f_{1,1}$, she selects the next channel $f_{1,2}$ uniformly randomly among the F_1 channels and, during the next slot of duration T_1 she transmits a packet at the frequency $f_{1,2}$. Xia operates in a similar way, using F_2 channels. Assume a collision model, such that if the transmissions of Zoya and Xia overlap even only partially in time *and* frequency, then a collision occurs and neither Yoshi nor Walt will receive the packet.

(a) Find the probability of collision when $F_1 = F_2$, $T_1 = T_2$ and the networks are synchronized in time, such that the slot for the link Zoya–Yoshi starts at the same time with a slot for the link Xia–Walt.

(b) Find the probability of collision when $F_1 = F_2$, $T_1 = T_2$ and the networks are NOT synchronized in time.

(c) Find the probability of collision when $F_1 > F_2$, $T_1 < T_2$.

5. *Spread spectrum and random access*. The base station Basil receives over a frequency band of size W from a set of K devices. The devices transmit to Basil through a random access protocol. Devise and analyze a random access protocol that is combined with spread spectrum for the following cases:

(a) The spread spectrum is based on frequency hopping. Study the protocol both for the collision model (as in the previous problem) and for a model with capture.

(b) The spread spectrum is based on direct sequence spread spectrum.