Chapter 3

Discrete-Time Communication

Most digital communication systems operate by converting digital data into continuous waveforms that can be conveyed through some physical medium to a receiver. For example, digital communication through wires (e.g., Ethernet or USB) is based on moving electrons back and forth in the wire. In contrast, underwater wireless communication uses acoustic transmission through the water. While radio communication relies on the propagation of electromagnetic waves through air.

The process by which a string of bits is converted into a waveform suitable for transmission is known as **modulation**. The reverse operation, called **demodulation**, is performed at the destination and involves extracting the information symbols from the received signal. The mapping between the transmitted waveform and the received waveform is known as the **channel**.

Precise models of the physical channel can be very complicated and many of the key ideas in digital communication do not depend on the exact details. For this reason, one can model and design communication systems based a simplified model that separates the communication problem from the physical models. In this chapter, we develop some basic concepts of digital communication using this simplified model. The goal is to build some intuition about how things work without getting lost in the mathematical details.
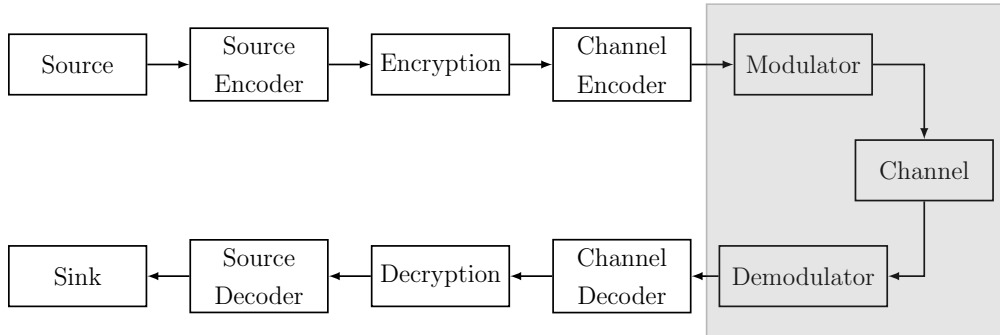
Figure 3.1: The block diagram of a digital communication system where the blocks comprising the discrete-time channel are shaded.

3.1   A Simple Channel Model

In this section, we introduce the **discrete-time channel model** for digital communication systems. We will see later that, for some communication systems, this model is exactly equivalent to the more complicated waveform model. The first channel model we discuss is where the nth output of the channel, Y_n, is equal to the nth input to the channel, x_n, corrupted by an additive noise term Z_n so that

$$Y_n = x_n + Z_n.$$

For this model, it is typical to assume that the noise sequence $Z_1, Z_2, \ldots$ consists of independent identically distributed (i.i.d.) zero-mean Gaussian random variables with variance σ^2. This implies that each Y_n is a Gaussian random variable with mean x_n and variance σ^2, so that

$$f_{Y_n}(y_n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_n - x_n)^2/(2\sigma^2)}.$$

This model is commonly referred to as discrete-time communication in **additive white Gaussian noise** (AWGN) noise.

As an example, consider a system where the transmitter sets the voltage on end of a wire and the receiver measures the voltage on the other end of the wire. It turns out that the thermal agitation of electrons causes voltage fluctuations known as Johnson noise. So, the receiver ends up measuring the

transmitted voltage corrupted by noise fluctuations. In fact, Johnson noise is quite well approximated by AWGN. So, the simple model described above already gives a relatively accurate picture of reality.

3.2 A Simple Modulation Scheme

Once a channel model has been defined, the next step is choosing how to transmit digital data through the channel. A common approach is to choose a small set $\mathcal{U}$ of information symbols and represent each one by a distinct channel input in the set $\mathcal{X}$. This is the discrete-time model of a scheme known as **pulse-amplitude modulation** (PAM). Let $u_n \in \mathcal{U}$ be the information symbol transmitted during the nth time interval and $x_n = M(u_n)$ be the nth input to the channel, where $M : \mathcal{U} \to \mathcal{X}$ is called the symbol mapping function. For example, one can transmit binary information symbols by mapping "0" to +1V and "1" to −1V; mathematically, this is done by choosing $\mathcal{U} = \{0, 1\}$, $\mathcal{X} = \{-1, 1\}$, and $M(u) = 1 - 2u$. This particular type of PAM is called **binary phase-shift keying** (BPSK) or 2-PAM.

Suppose a BPSK signal is transmitted through our discrete-time AWGN channel model. The detector must measure the voltage and decide whether a 0 or 1 one was transmitted. A natural choice is to define a decoder function that associates positive voltages with 0 and negative voltages with 1. Let $\hat{U}_n = D(Y_n)$ be the output of the detector function

$$D(y) = \begin{cases} 0 & \text{if } y \geq 0 \\ 1 & \text{if } y < 0 \end{cases}.$$

This detector is optimal if 0's and 1's are transmitted with equal probability.

One of the main challenges in communication systems is providing reliable data transmission. In this example, the noise variance σ^2 is proportional to the power of the thermal noise in the wire. Notice that, if a 1 is transmitted, the detector described above will make an incorrect decision with probability

$$\Pr\left(Y_n \geq 0 | x_n = 1\right) = \Pr\left(Z_n \geq 1\right) = \int_1^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/(2\sigma^2)} dy = Q\left(\frac{1}{\sigma}\right).$$

This probability can be reduced by increasing the transmitted voltage, which increases the power dissipated due to resistive losses, or by reducing the ther-
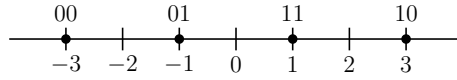
Figure 3.2: The symbol set $\mathcal{X}$ for 4-PAM with gray coded binary labels.

mal noise. For this reason, receivers used for large satellite dish receivers often use liquid nitrogen to cool the first-stage amplifier to reduce this thermal noise. On the other hand, the maximum transmitted power is typically limited by physical constraints (e.g., the wire thickness) or FCC regulations.

To send more information per channel use, one can use larger sets of PAM symbols. For example, 4-PAM uses the 4 symbols $\mathcal{X} = \{-3, -1, 1, 3\}$ while 8-PAM uses the 8 symbols $\mathcal{X} = \{-7, -5, -3, -1, 1, 3, 5, 7\}$. Notice that these two sets of symbols are centered around 0 to minimize the transmitted energy.

The mapping function $M(\cdot)$ determines the information symbol associated with each channel input value. In general, sets of input values with 2^m elements are associated with binary strings. There are still some choices to be made, however, because one can map the 4-PAM symbol set to binary strings in either the standard binary order $\{00, 01, 10, 11\}$ or with a Gray code $\{00, 01, 11, 10\}$.

For the 4-PAM symbol set with the mapping function $M(\cdot)$ that maps $\{00, 01, 10, 11\}$ (in order) to $\{-3, -1, 1, 3\}$, the natural decision function is

$$
D(y) = \begin{cases} 00 & \text{if } y < -2 \\ 01 & \text{if } -2 \leq y < 0 \\ 10 & \text{if } 0 \leq y < 2 \\ 11 & \text{if } y \geq 2 \end{cases}.
$$

3.3   Quadrature Amplitude Modulation

In most communication systems, the baseband waveform is modulated onto a high-frequency carrier to enable better propagation. This is because low-frequency signals often do not propagate well through physical media. While this process will be discussed later in more detail, the following key detail affects the discrete-time model. High frequency modulation allows two independent signals to be modulated onto the same carrier frequency; one onto the sine wave and the other onto the cosine wave. This allows one to treat the
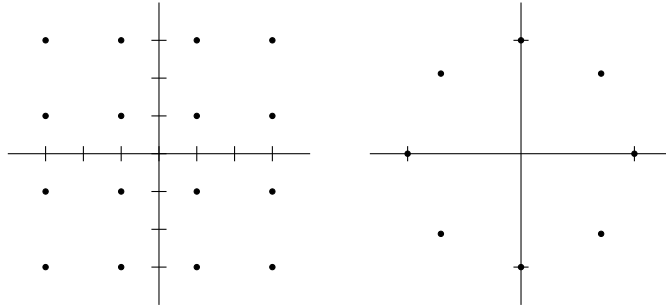
Figure 3.3: The symbol constellations $\mathcal{X}$ for 16-QAM (left) and 8-PSK (right).

transmitted value x_n and received value Y_n as points in 2-dimensional space. The set $\mathcal{X}$ of possible transmitted points in 2-dimensional space in called the **symbol constellation**.

For mathematical convenience, points in these two-dimensional symbol constellations are represented by complex numbers. The set of complex numbers is $\mathbb{C}$ and the constellation is a subset $\mathcal{X} \subset \mathbb{C}$. Likewise, the transmitted symbol is $x_n \in \mathbb{C}$ and the received value is $Y_n \in \mathbb{C}$. The noise term Z_n now consists of two i.i.d. Gaussian random variables (one in each direction). The probabilistic observation model is formed by treating the real and imaginary parts separately, and is given by

$$f_{Y_n^{(r)}, Y_n^{(i)}}(y_n^{(r)}, y_n^{(i)}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(y_n^{(r)} - x_n^{(r)}\right)^2 / (2\sigma^2)} \right) \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(y_n^{(i)} - x_n^{(i)}\right)^2 / (2\sigma^2)} \right)$$

$$= \frac{1}{2\pi\sigma^2} e^{-|y_n - x_n|^2 / (2\sigma^2)}.$$

Therefore, the probability of receiving a y_n value is simply a function of its Euclidean distance $|y_n - x_n|^2$ from the actual transmitted symbol. This leads to a nice geometric characterization of the optimal decision regions for the detector.

The **signal-to-noise ratio** (SNR) of a communication system is typically denoted by E_s / N_0 where E_s is the average **energy per channel input symbol** and N_0 is the **noise spectral density**. For equiprobable signaling, one finds that

$$E_s = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} |x|^2.$$

The noise spectral density measures how much the AWGN is affects the channel and later we will see that $N_0 = 2\sigma^2$ for our discrete-time model.

Constellations are typically defined by first choosing the set of channel input values $\mathcal{X}$, and then choosing the mapping function $M : \mathcal{U} \to \mathcal{X}$. This second step is called **labeling** the constellation. While the labeling does not affect the symbol error rate of the system, it generally does affect the bit error rate. Therefore, one can optimize the mapping function for a particular application.

Constellations can be chosen and optimized for a variety of reasons. Still, there are a few very common choices:

- M-ary PAM (M-PAM) is M points equally spaced along a line or

$$\mathcal{X} = \big\{ 2a - (M - 1) \,\big|\, a \in \{0, 1, \dots, M - 1\} \big\} \subset \mathbb{C}.$$

- M^2-ary QAM (M^2-QAM) is an M by M square grid of points or

$$\mathcal{X} = \big\{ (2a - (M - 1)) + (2b - (M - 1))\, i \,\big|\, a, b \in \{0, 1, \dots, M - 1\} \big\} \subset \mathbb{C}.$$

- M-ary PSK (M-PSK) is M points equally spaced around a circle or

$$\mathcal{X} = \big\{ e^{2\pi i k/M} \,\big|\, k \in \{0, 1, \dots, M - 1\} \big\} \subset \mathbb{C}.$$

Example 3.3.1. *The standard QAM constellation with 16 points (known as 16-QAM) is given by*

$$\mathcal{X} = \big\{ a + bi \,\big|\, a, b \in \{-3, -1, 1, 3\} \big\} \subset \mathbb{C}.$$

The average energy of this constellation is given by

$$E_s = \frac{1}{16} \sum_{a,b \in \{-3,-1,1,3\}} (a^2 + b^2) = \frac{8}{16} \sum_{a \in \{-3,-1,1,3\}} a^2 = 10.$$

3.4   Optimal Symbol Detection

In this section, we consider the problem of designing a symbol detector that minimizes the probability of error. This is known as an **optimal detection** problem and has an elegant solution that is related to the classical problem of **hypothesis testing** in statistics.

3.4.1 Hypothesis Testing

Sir Ronald Fisher, one of the founders of statistical decision theory, was at a tea party when Ms. Bristol mentioned that she preferred tea poured into milk over milk poured into tea. Fisher commented that surely she could not tell the difference, but his colleague William Roach suggested that they design an experiment. At that point, they prepared eight cups of tea: four milk-into-tea and four tea-into-milk. The cups were presented in a random order and she correctly identified enough (all eight cups by some accounts) to prove her point.

The mathematics behind this type of hypothesis test is based on defining a **null hypothesis**, which states that the difference in preparation has no effect on the outcome. In this example, the null hypothesis is $H_0 =$ "the order of pouring the tea and milk does not affect Ms. Bristol's answer". Notice that there are $\binom{8}{4} = 70$ ways that Ms. Bristol can divide the cups of tea into two categories, but that only one identifies all of them correctly. Since the cups are presented in a random order, we can compute

$$\Pr(\text{she identifies all cups correctly}|H_0) = \frac{1}{70}.$$

This number is small enough so that it is reasonable to conclude that H_0 is false and the order does influence Ms. Bristol's decisions.

There are more subtle issues, however, than the case where Ms. Bristol succeeds due to luck. What is more problematic is listing all other hypotheses that can lead to the same observation. For example, variations in the temperature or composition of the tea could help Ms. Bristol guess correctly. Fisher, in his essay on this experiment, argues that proper use of randomization can eliminate the effect of these variations. On the other hand, Ms. Bristol may pass by cheating but still be unable to distinguish between the two drinks. If all possible hypotheses are not explicitly considered, then one can come to an incorrect conclusion. For this reason, it commonly understood that a scientific test of an hypothesis can only *disprove* that hypothesis.

3.4.2   Multiple Hypothesis Testing

Hypothesis testing in communication theory often has the luxury that one of the hypotheses must be true. This leads to the more well-defined problem of **multiple hypothesis testing**. Instead of testing a single hypothesis to see if it is false, one can compare multiple hypothesis to see which is most supported by the observation.

Let $H_0, H_1, \ldots, H_{m-1}$ be m different hypotheses that affect a random observation Y. The probability of a hypothesis before the observation, $\Pr(H_i)$, is called the **a priori probability**. For each hypothesis, the connection with Y is defined by the **observation probability** $\Pr(Y = y \,|\, H_i)$.

The goal is to choose a decision function $D(y)$ which, for any observation, minimizes the decision error probability. Of course, this is equivalent to maximizing the probability that the decision is correct. Notice that, if $Y = y$, then the probability that hypothesis H_i is correct is given by its **a posteriori probability** $\Pr(H_i \,|\, Y = y)$. Therefore, one finds that the optimal choice is the **maximum a posteriori probability** (MAP) decision rule

$$D(y) = \underset{i \in \{0, \ldots, m-1\}}{\arg \max} \ \Pr(H_i \,|\, Y = y).$$

In practice, these probabilities can be computed with Bayes' rule using only the a priori probabilities and observation probabilities. This gives

$$\Pr\left(H_i | Y = y\right) = \frac{\Pr(H_i) \Pr(Y = y | H_i)}{\sum_{j=0}^{m-1} \Pr(H_j) \Pr(Y = y | H_j)}.$$

Since the denominator of this expression is the same for all i, the MAP rule can be simplified to

$$D(y) = \underset{i \in \{0, \ldots, m-1\}}{\arg \max} \ \Pr(H_i) \Pr(Y = y | H_i).$$

Example 3.4.1. *Consider a system which transmits BPSK over an AWGN channel. Let H_0 be the hypothesis that a zero (i.e., $+1$) was sent and H_1 be the hypothesis that a one (i.e., -1) was sent. For binary hypothesis problems, the MAP decision rule can be written as*

$$\Pr(H_0) \Pr(Y = y | H_0) \underset{H_1}{\overset{H_0}{\gtrless}} \Pr(H_0) \Pr(Y = y | H_0,$$

where this notation implies that one should pick H_0 if the LHS is greater than the RHS and H_1 otherwise. If $\Pr(H_0) = 1 - p$ and $\Pr(H_1) = p$, then one can substitute formulae to rewrite this as

$$(1-p)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-1)^2/(2\sigma^2)} \overset{H_0}{\underset{H_1}{\gtrless}} p\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y+1)^2/(2\sigma^2)}.$$

After a little algebra, taking the logarithm of both sides simplifies this to

$$y \overset{H_0}{\underset{H_1}{\gtrless}} \frac{\sigma^2}{2}\ln\frac{p}{1-p}.$$

Another popular rule is the **maximum likelihood** (ML) decision rule

$$D(y) = \underset{i\in\{0,\dots,m-1\}}{\arg\max} \ \Pr(Y = y|H_i),$$

which ignores the a priori probability. When all the hypotheses have the same a priori probability, these two rules are identical. In communication systems, this is often the case.

Example 3.4.2. *Consider a system which transmits 4-PAM (i.e, $\mathcal{X} = \{-3, -1, 1, 3\}$) over an AWGN channel. If all channel inputs are equiprobable, then the optimum detector is*

$$\begin{aligned} D(y) &= \underset{x\in\{-3,-1,1,3\}}{\arg\max} \left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-x)^2/(2\sigma^2)}\right) \\ &= \underset{x\in\{-3,-1,1,3\}}{\arg\max} \left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y-x)^2\right) \\ &= \underset{x\in\{-3,-1,1,3\}}{\arg\min} \ (y-x)^2. \end{aligned}$$

Therefore, the optimum detector chooses the constellation point closest to the channel observation. Moreover, this statement remains true for any signal constellation with equiprobable signalling and AWGN.