6 Digital Modulation

6.1 Introduction to Digital Modulation

6.1. We once again return to Figure 1 which is repeated here as Figure 20. In this chapter, digital modulator-demodulator boxes are the main focus. The **digital modulator** serves as the interface to the physical (analog) communication channel.

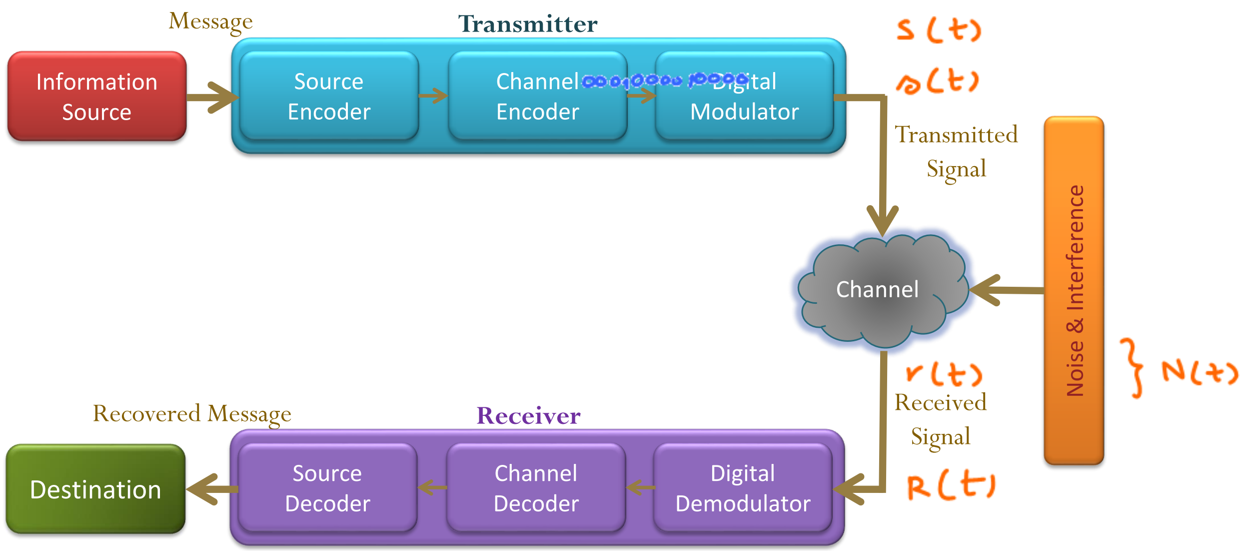


Figure 20: Basic elements of a digital communication system

The mapping between the digital sequence (which we may assume to be a binary sequence) and the (continuous-time) signal sequence to be transmitted over the channel can be either 1 memoryless or 2 with memory, resulting in memoryless modulation schemes and modulation schemes with memory.

Definition 6.2. In a **memoryless modulation** scheme, each particular digital modulation has a **signal set** which is simply a collection of M signals (or waveforms): $\{s_1(t), s_2(t), \ldots, s_M(t)\}$. The binary sequence (from the channel encoder) is parsed into blocks each of length b, and each block is mapped into one of the signals in the collection regardless of the previously transmitted signals.

- $M = 2^b$ $\Rightarrow b = \log_2 M$
- This mapping from M possible messages to M (distinct) signals is called **M-ary modulation**.

* possible waveforms

* bits in a block of input

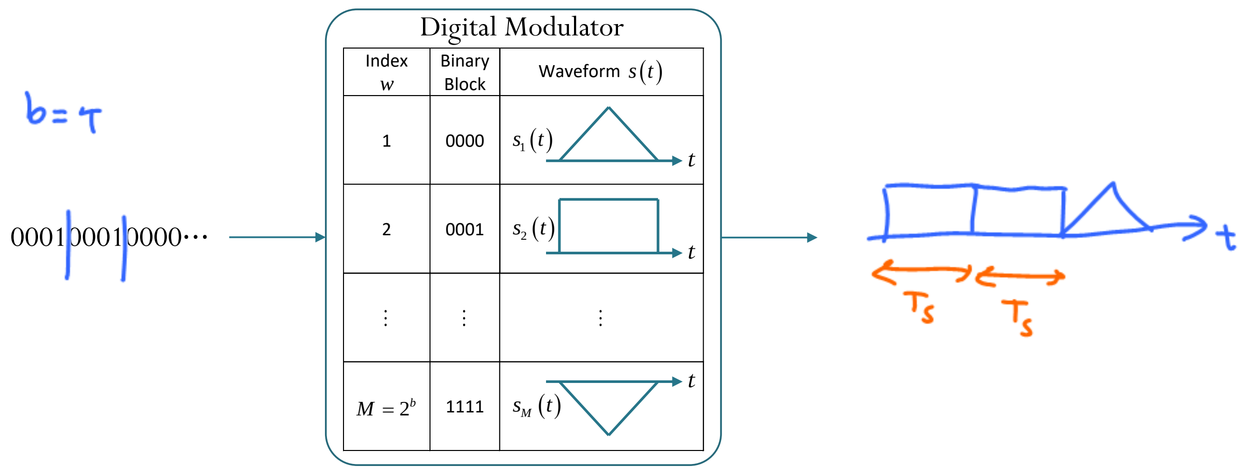


Figure 21: Digital Modulator and the mapping from binary blocks to waveforms.

- 2-ary = **binary**; 3-ary = **ternary**; 4-ary = **quarternary**.

$0 \to s_1(t)$
$1 \to s_2(t)$

- In **binary modulation**, each bit from the channel encoder is transmitted separately. The digital modulator simply map the binary digit 0 into a waveform $s_1(t)$ and the binary digit 1 into a waveform $s_2(t)$.
- The waveforms $s_m(t)$ can be, in general, of any shape. However, usually these waveforms are bandpass signals which may differ in amplitude or phase or frequency, or some combination of two or more signal parameters.

Definition 6.3. In a modulation scheme with memory, the mapping is from the set of the current b bits and the past $(L-1)b$ bits to the set of possible $M = 2^b$ messages.

- Modulation systems with memory are effectively represented by Markov chains.
- The transmitted signal depends on the current b bits as well as the most recent $L-1$ blocks of b bits.
- This defines a finite-state machine with $2^{(L-l)b}$ states.
- The mapping that defines the modulation scheme can be viewed as a mapping from the current state and the current input of the modulator to the set of output signals resulting in a new state of the modulator.
- Parameter L is called the **constraint length** of modulation.
- The case of $L = 1$ corresponds to a memoryless modulation scheme.

Definition 6.4. We assume that these signals are transmitted at every T_s seconds.

- T_s is called the **signaling interval**.
- This means that in each second

$$R_s = \frac{1}{T_s}$$

  symbols are transmitted.
  Parameter R_s is called the **signaling rate**, **symbol (transmission) rate**, or **baud rate**.

- **Bit rate** $R = b \times R_s = \frac{b}{T_s} = R_s \log_2 M = \frac{\log_2 M}{T_s}$

Definition 6.5. The energy content of a signal $s_m(t)$ is denoted by E_m. It can be calculated from

$$E_m = \int_{-\infty}^{\infty} |s_m(t)|^2 dt.$$

6.6. The **average signal energy** (per symbol) for the M-ary modulation in Definition 6.2 is given by

$$E_s = \sum_{m=1}^{M} p_m E_m$$

where p_m indicates the probability of the mth signal (message probability).

- **(Average) energy per bit**: $E_b = \frac{E_s}{b} = \frac{E_s}{\log_2 M}$
- For equiprobable signals,

$$p_m = \frac{1}{M} \qquad E_s = \sum_{m=1}^{M} \frac{1}{M} E_m = \frac{1}{M} \sum_{m=1}^{M} E_m$$

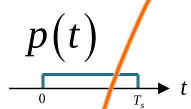
- If all signals have the same energy, then
  - $E_m \equiv E$ for some E and
  - $E_s = E$.

Definition 6.7. In (the digital version of) **Pulse Amplitude Modulation (PAM)**, the signal waveforms are of the form

$$s_m(t) = A_m p(t), \quad 1 \le m \le M \tag{33}$$

where $p(t)$ is a (common) pulse and $\mathcal{A} = \{A_m, 1 \le m \le M\}$ denotes the set of M possible "amplitudes".

- For $M = 2$, we may have $\mathcal{A} = \{\pm 1\}$
  For $M = 4$, we may have $\mathcal{A} = \{\pm 1, \pm 3\}$



PAM Example: $M = 2$

Index m	Binary Block $\underline{\mathbf{b}}$	Amplitude A_m	Waveform $s_m(t)$
1	0	-1	$s_1(t)$
$M = 2$	1	1	$s_2(t)$

PAM Example: $M = 4$

Index m	Binary Block $\underline{\mathbf{b}}$	Amplitude A_m	Waveform $s_m(t)$
1	00	-3	$s_1(t)$
2	01	-1	$s_2(t)$
3	10	1	$s_3(t)$
$M = 4$	11	3	$s_4(t)$

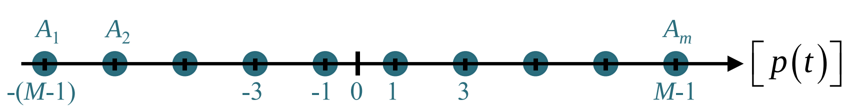
- When $M = 2$ (binary modulation) and $s_1(t) = -s_2(t)$, such signals are called antipodal. This case is sometimes called **binary antipodal signaling**.

- More generally, the signal "amplitudes" A_m may take the discrete values

  $$A_m = 2m - 1 - M, \quad m = 1, 2, \ldots, M \tag{34}$$

  i.e., the "amplitudes" are $\pm 1, \pm 3, \pm 5, \ldots, \pm(M-1)$.

- These M waveforms can be visualized as M points on an axis as shown below. Note how the axis is scaled by the common pulse $p(t)$.



- The shape of $p(t)$ influences the spectrum of the transmitted signal.
- The energy in signal $s_m(t)$ is given by

$$E_m \equiv \int_{-\infty}^{\infty} |s_m(t)|^2 dt = \int_{-\infty}^{\infty} |A_m p(t)|^2 dt = A_m^2 \underbrace{\int_{-\infty}^{\infty} |p(t)|^2 dt}_{E_p} = A_m^2 E_p$$

- For equiprobable signals,

$$E_s = \sum_{m=1}^{M} p_m E_m = \sum_{m=1}^{M} \frac{1}{M} E_m = \frac{1}{M} E_p \sum_{m=1}^{M} A_m^2$$

When $\mathcal{A} = \{\pm 1\}$, $M=2$
$$E_s = \frac{1}{2} E_p \left((-1)^2 + (1)^2 \right) = E_p$$

When $\mathcal{A} = \{\pm 1, \pm 3\}$, $M=4$
$$E_s = \frac{1}{4} E_p \left((-3)^2 + (-1)^2 + (1)^2 + (3)^2 \right) = 5 E_p$$

For the general $\mathcal{A}$ defined in (34),

HWc

$$E_s = \frac{1}{M} E_p \left((-(M-1))^2 + \cdots + (-3)^2 + (-1)^2 + (1)^2 + (3)^2 + \cdots + (M-1)^2 \right) \stackrel{?}{=} ?$$

Definition 6.8. In **Amplitude-Shift Keying** (ASK), the (common) pulse $p(t)$ in (33) for PAM is replaced by

$$p(t) = g(t)\cos(2\pi f_c t).$$

where f_c is the carrier frequency.

- Note that $E_p = \frac{E_g}{2}$.

6.9. The mapping or assignment of b (encoded) bits to the $M = 2^b$ possible signals may be done in a number of ways. The preferred assignment is one in which the adjacent signal amplitudes differ by one binary digit. This mapping is called **Gray coding**.

- It is important in the demodulation of the signal because the most likely errors caused by (additive white gaussian) noise involve the erroneous selection of an adjacent amplitude to the transmitted signal amplitude. In such a case, only a single bit error occurs in the b-bit sequence.
- Gray code list for n bits can be generated recursively from the list for $n-1$ bits by

i reflecting the list (i.e. listing the entries in reverse order),

ii concatenating the original list with the reversed list,

iii prefixing the entries in the original list with a binary 0, and then prefixing the entries in the reflected list with a binary 1.

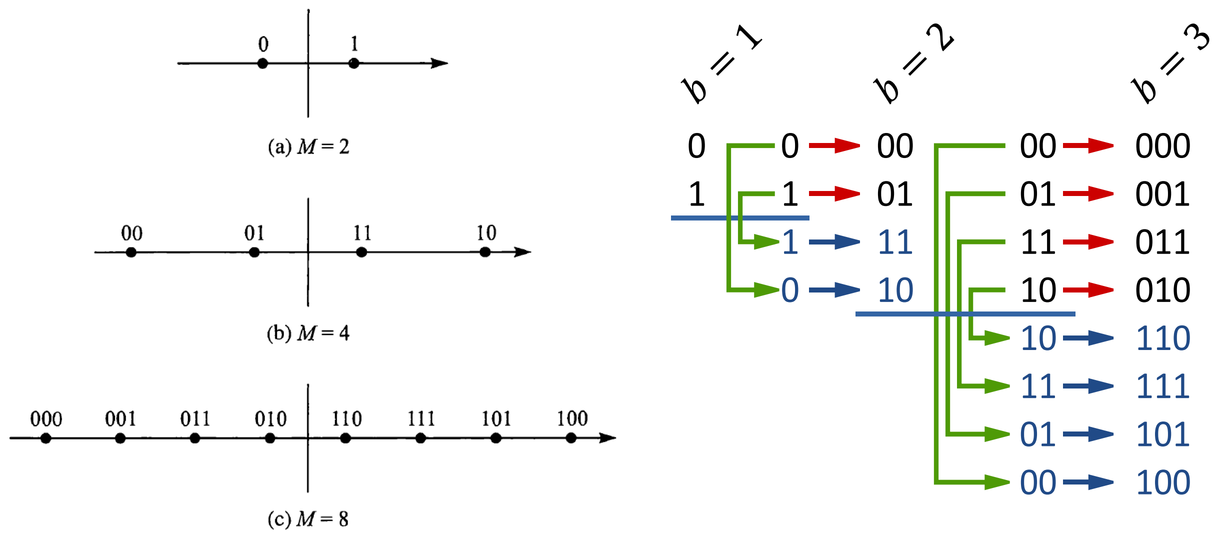


Figure 22: Gray coding and its reflect-and-prefix construction for PAM signaling

6.10. In PAM (and ASK), we use just one pulse (sinusoidal pulse in the case of ASK) and modify the amplitude of the pulse to create many waveforms $s_1(t), s_2(t), \ldots s_M(t)$ that we can use to transmit different block of bits. Next, we would like to study the case where multiple shapes are used.

Example 6.11. For (baseband) binary (digital) modulation, we may use the two waveforms $s_1(t)$ and $s_2(t)$ shown in Figure 23.

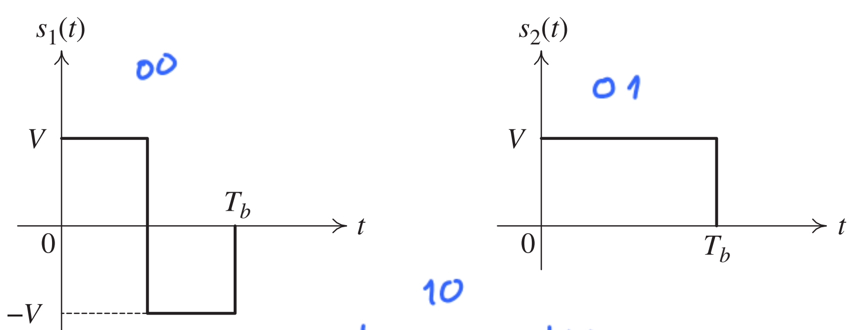
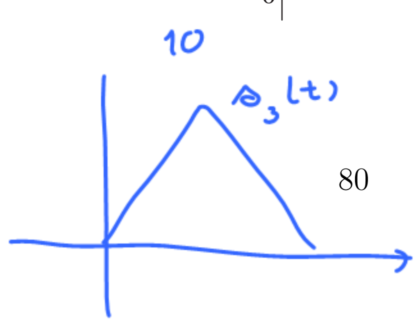
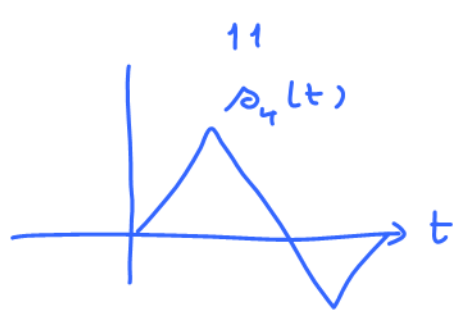


Figure 23: Signal set for Example 6.11.

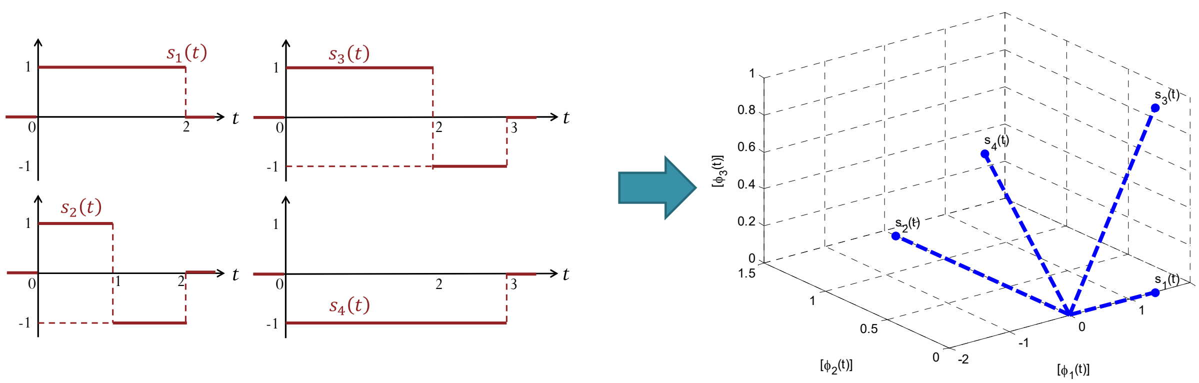




6.12. It is difficult to visualize, find relationship between, work with, or perform analysis directly on waveforms. For example, when we have many waveforms in the signal set, it is difficult to tell (by looking at their plots) how easy it is for them to get corrupted by the noise process; that is, how easy it is for one waveform to be interpreted as being another waveform at the demodulator.

In the next sections, we will study how to represent waveforms in the signal set as "equivalent" vectors (or points) in a **signal space** similar to what we saw in Figure 22. Representing waveforms as points allows us to look at them as a collection effectively.

Example 6.13. Consider a signal set containing four waveforms in Figure 24a. Note that a waveform contains infinitely many points. To represent all possible waveforms, we would need to work in infinite-dimensional space. However, we only have to consider four possible waveforms here. It turns out that we can represent these four waveforms by four vectors in a three-dimensional space as shown in Figure 24b. We will learn how to do this in the remaining parts of this chapter.



(a) (b)

Figure 24: From Waveforms to Constellation

6.14. A signal space is a vector space. So, we will first provide a review of some concepts related to vector spaces.