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.

convert bits to waveform (analog signal)

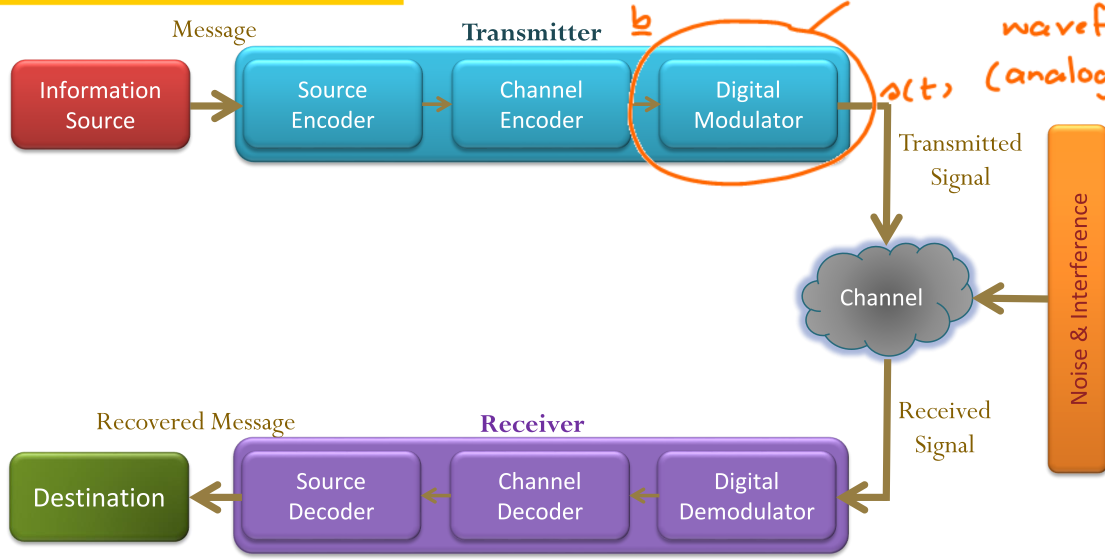


Figure 20: Basic elements of a digital communication system

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 memoryless or 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 signal set regardless of the previously transmitted signals.

- $M = 2^b$. → bits in a block of input $\Rightarrow b = \log_2 M$
- This mapping from M possible messages to M (distinct) signals is called M**-ary modulation**.

In this example, we work with blocks of 4 bits. Therefore, we need $2^4 = 16$ different waveforms.

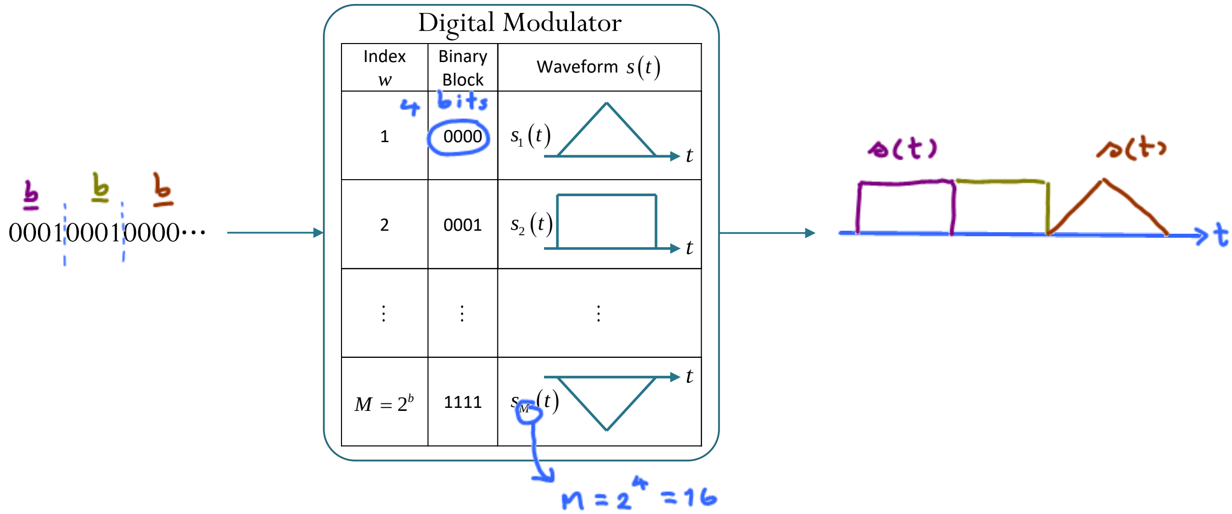


Figure 21: Digital Modulator and the mapping from binary blocks to waveforms.

- 2-ary = **binary**; 3-ary = **ternary**; 4-ary = **quarternary**.
- 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 $s_2(t)$.

$b = 1$

$M = 2^1 = 2$

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

- 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** $R = b \times R_s = (\log_2 M) R_s = \frac{\log_2 M}{T_s}$

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,

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

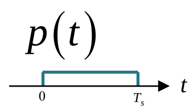
- If all signals have the same energy, then
  - $E_m \equiv E$ for some E and
  - $E_s = E$.

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

$$s_m(t) = A_m p(t), \quad 1 \le m \le M \tag{35}$$

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	$s_4(t)$

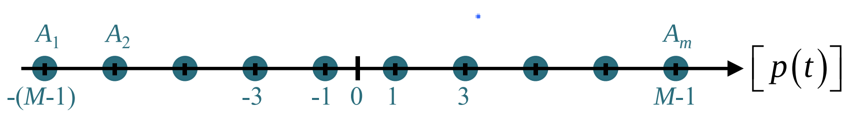
- When $M = 2$ (binary modulation) and $s_1(t) = -s_2(t)$, such signals are called antipodal. This case is sometimes called **binary antipodal signaling**.

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

$$A_m = 2m - 1 - M, \quad m = 1, 2, \ldots, M \tag{36}$$

  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

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

- For equiprobable signals,

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

When $\mathcal{A} = \{\pm 1\}$,

$M=2$ $\quad E_s = \frac{E_p}{2}\left(1^2 + (-1)^2\right) = E_p$

When $\mathcal{A} = \{\pm 1, \pm 3\}$,

$M=4$ $\quad E_s = \frac{E_p}{4}\left((-3)^2 + (-1)^2 + 1^2 + 3^2\right) = 5 E_p$

For the general $\mathcal{A}$ defined in (36),

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

Definition 6.8. In **Amplitude-Shift Keying** (ASK), the (common) pulse $p(t)$ in (35) 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.

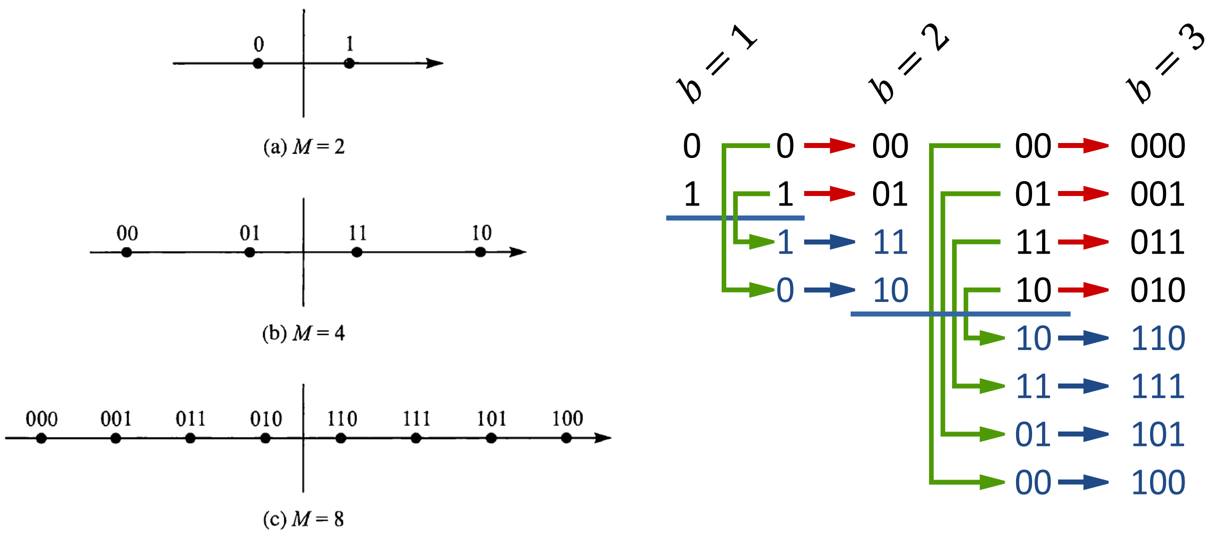


Figure 22: Gray coding and its reflect-and-prefix construction for PAM signaling

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.

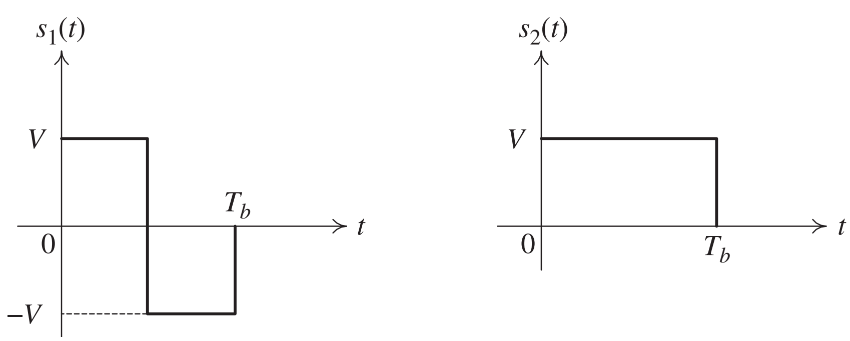


Figure 23: Signal set for Example 6.11.

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. It is possible to find such representation systematically via a process called Gram-Schmidt Orthogonalization Procedure (GSOP).

For easier analysis, we want to represent waveforms as points with respect to some "axes".

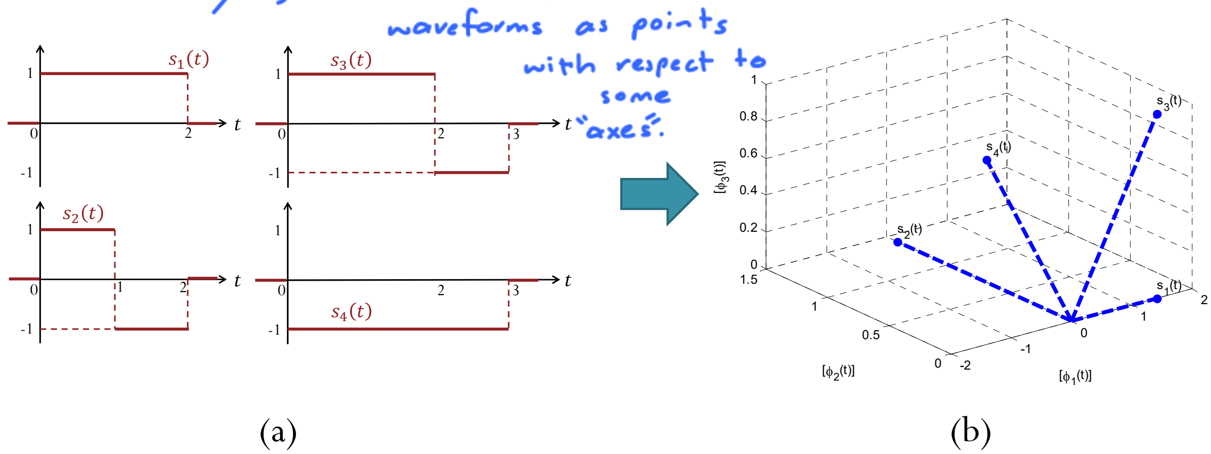


Figure 24: From Waveforms to Constellation

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

6.2 Vector Space and Inner Product Space in $\mathbb{R}^n$

In linear algebra, an inner product space is a vector space[19] with an additional structure called an inner product.

Definition 6.15. When we have a list of vectors, we use **superscripts** in parentheses as indices of vectors. **Subscripts** represent element indices inside individual vectors.

Example 6.16. Here is a list of four vectors:

$$\mathbf{v}^{(1)} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}^{(2)} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}^{(3)} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \text{ and } \mathbf{v}^{(4)} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

For the second vector, we have $v_1^{(2)} = 1$, $v_2^{(2)} = -1$, and $v_3^{(2)} = 0$.

Definition 6.17. The **inner product** of two real-valued n-dimensional (column) vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as

Transpose (replaced by $(\cdot)^H$ if vectors are complex-valued.)

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T \mathbf{u} = \sum_{k=1}^{n} u_k v_k.$$

In elementary linear algebra class, you may encounter this quantity in the form of the **dot product** between two vectors.

Definition 6.18. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

More generally, a set of N vectors $\mathbf{v}^{(k)}$, $1 \le k \le N$, are **orthogonal** if $\left\langle \mathbf{v}^{(i)}, \mathbf{v}^{(j)} \right\rangle = 0$ for all $1 \le i, j \le N$, and $i \ne j$.

Definition 6.19. The **norm** of a vector $\mathbf{v}$ is denoted by $\|\mathbf{v}\|$ and is defined as

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_1^2 + v_2^2 + \cdots v_n^2}$$

which in the n-dimensional Euclidean space is simply the **length** of the vector.

Definition 6.20. A collection of vectors is said to be **orthonormal** if the vectors are orthogonal ① and each vector has a unit norm ②.

[19] Recall that a vector space is a mathematical structure formed by a collection of elements called vectors, which may be added together and multiplied ("scaled") by numbers, called scalars in this context.

Example 6.21. Let $\mathbf{v} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$ and $\mathbf{u} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$.

(a) $\langle \mathbf{v}, \mathbf{u} \rangle = v_1 u_1 + v_2 u_2 = (5 \times 0) + (5 \times 4) = 20$

(b) $\langle \mathbf{v}, \mathbf{v} \rangle = v_1^2 + v_2^2 = 5^2 + 5^2 = 50$

(c) $\langle \mathbf{u}, \mathbf{u} \rangle = u_1^2 + u_2^2 = 0^2 + 4^2 = 16$

(d) $\|\mathbf{v}\| = \sqrt{\langle v, v \rangle} = \sqrt{50} = 5\sqrt{2}$

(e) $\|\mathbf{u}\| = \sqrt{\langle u, u \rangle} = \sqrt{16} = 4$

Definition 6.22. A **unit vector**, usually denoted by $\mathbf{e}$ ($\vec{e}$), is a vector whose $\|\mathbf{e}\| = 1$.

6.23. Any vector in a vector space may also be represented as a linear combination of orthogonal unit vectors or an **orthonormal basis** $\left\{\mathbf{e}^{(i)}, 1 \le i \le N\right\}$ (for that vector space), i.e.,

$$\mathbf{v} = \sum_{i=1}^{N} c_i \mathbf{e}^{(i)}$$

where

$$c_i = \left\langle \mathbf{v}, \mathbf{e}^{(i)} \right\rangle.$$

Example 6.24. In many applications, the standard choice for the orthonormal basis of a collection of (all possible real-valued) n-dimensional vectors is

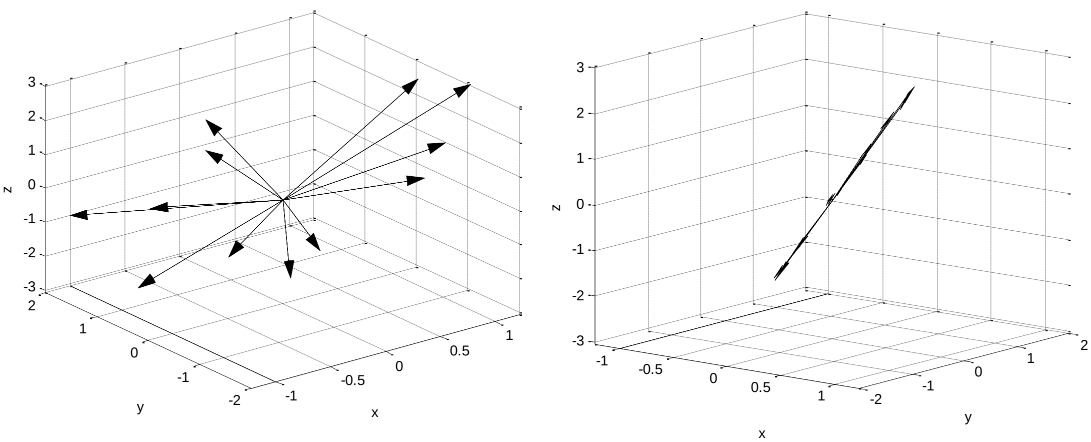
$$\mathbf{e}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}^{(n)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

In Example 6.21, all vectors are expressed via the standard basis

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

6.25. Suppose we start with a collection of M n-dimensional vectors. Do these M vectors really need to be represented in n dimensions?

Example 6.26. Figure 25a shows a particular collection of 10 vectors in 3-D. When viewed from appropriate angle (as in Figure 25b), we can see that they all reside on a 2-D plane. We only need a two-vector (orthonormal) basis. All ten vectors can be represented as linear combinations of these two vectors.



(a) (b)

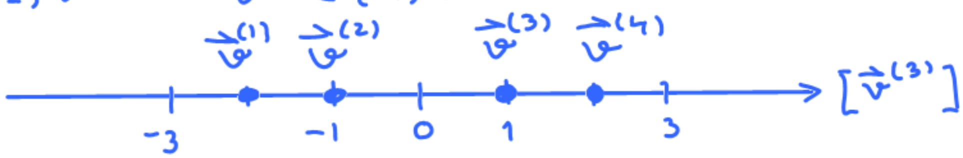
Figure 25: Ten vectors on a plane

Example 6.27. Consider the four vectors below:

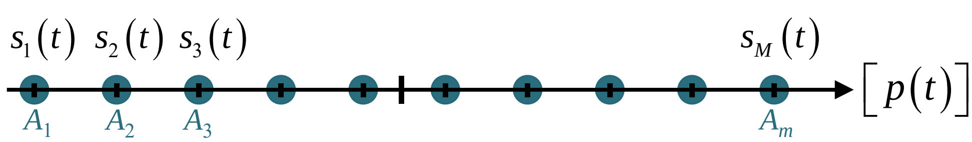
$$\mathbf{v}^{(1)} = \begin{pmatrix} -2 \\ -6 \\ 2 \end{pmatrix}, \mathbf{v}^{(2)} = \begin{pmatrix} -1 \\ -3 \\ 1 \end{pmatrix}, \mathbf{v}^{(3)} = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}, \text{ and } \mathbf{v}^{(4)} = \begin{pmatrix} 2 \\ 6 \\ -2 \end{pmatrix}.$$

They are all multiples of one another.

$\vec{v}^{(1)} = (-2)\,\vec{v}^{(3)}$ $\quad \vec{v}^{(2)} = (-1)\,\vec{v}^{(3)}$ $\quad \vec{v}^{(4)} = 2\,\vec{v}^{(3)}$



6.28. Similar idea applies to waveforms. In PAM, we have M waveforms that are simply multiples of a pulse $p(t)$. Therefore, one may represent them as points in one dimension as we had discussed in the previous section.



6.29. Given a collection of M vecctors $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$, $\dots$, $\mathbf{v}^{(M)}$, we can use a technique called the **Gram-Schmidt Orthogonalization Procedure (GSOP)** to find a collection[20] of N orthonormal vectors $\mathbf{e}^{(1)}, \mathbf{e}^{(2)}, \dots, \mathbf{e}^{(N)}$ such that any $\mathbf{v}^{(j)}$ can be expressed as a linear combination:

$$\mathbf{v}^{(j)} = \sum_{i=1}^{N} c_{i,j}\mathbf{e}^{(i)},$$

where the constants (weights) $c_{i,j} = \left\langle \mathbf{v}^{(j)}, \mathbf{e}^{(i)} \right\rangle$. We can then think of the vector $\mathbf{c}^{(j)} = (c_{1,j}, c_{2,j}, \dots, c_{N,j})^T$ as the new coordinates of $\mathbf{v}^{(j)}$ based on the new "axes" $\mathbf{e}^{(1)}$, $\mathbf{e}^{(2)}$, $\dots$, $\mathbf{e}^{(N)}$.

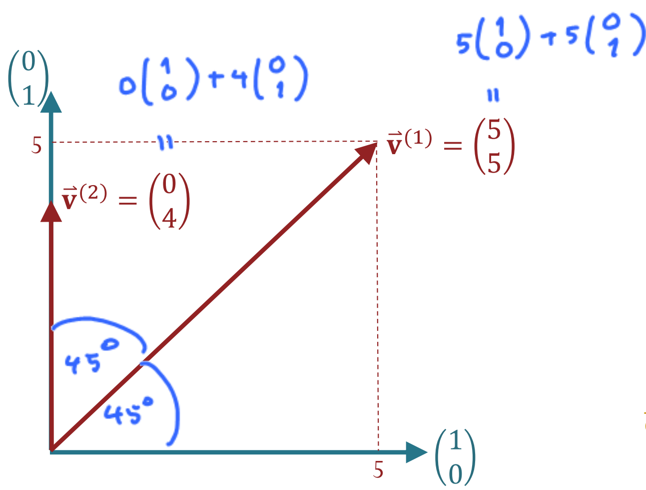
Example 6.30. Consider a collection of two vectors:

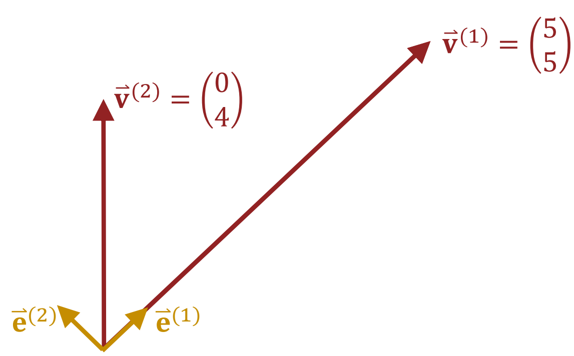
$$\mathbf{v}^{(1)} = \begin{pmatrix} 5 \\ 5 \end{pmatrix} \text{ and } \mathbf{v}^{(2)} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.$$

- In their original (default) coordinate systems, the basis contains two vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
- Alternatively, consider two orthonormal vectors

① $\langle \vec{e}^{(1)}, \vec{e}^{(2)} \rangle = 0$
② $\|\vec{e}^{(1)}\| = 1$
$\|\vec{e}^{(2)}\| = 1$

$$\mathbf{e}^{(1)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \mathbf{e}^{(2)} = \frac{1}{2\sqrt{2}}\begin{pmatrix} -2 \\ 2 \end{pmatrix}.$$





[20]In linear algebra, this collection is an orthonormal basis for the vector space spanned by $\mathbf{v}^{(1)}$, $\mathbf{v}^{(2)}$, $\dots$, $\mathbf{v}^{(M)}$.

Using $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ as the new axes, we can express $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$ in the new coordinate system as

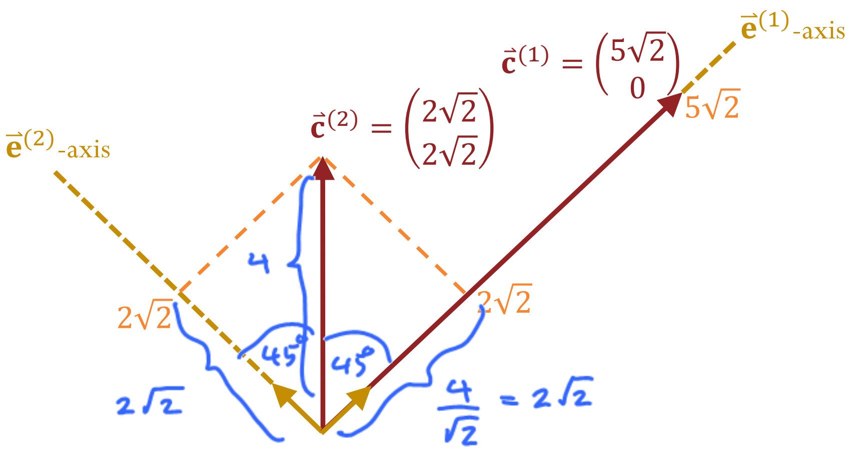
$\langle \vec{v}^{(2)}, \vec{e}^{(1)} \rangle = 2\sqrt{2}$

$\langle \vec{v}^{(1)}, \vec{e}^{(1)} \rangle = 5\sqrt{2}$

$$\mathbf{c}^{(1)} = \begin{pmatrix} 5\sqrt{2} \\ 0 \end{pmatrix} \text{ and } \mathbf{c}^{(2)} = \begin{pmatrix} 2\sqrt{2} \\ 2\sqrt{2} \end{pmatrix}.$$

$\langle \vec{v}^{(2)}, \vec{e}^{(2)} \rangle = 2\sqrt{2}$

$\langle \vec{v}^{(1)}, \vec{e}^{(2)} \rangle = 0$



6.31. Important properties: the transformation from $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(M)}$ to $\mathbf{c}^{(1)}, \mathbf{c}^{(2)}, \ldots, \mathbf{c}^{(M)}$ preserve many geometric quantities.

(a) Parseval's Identity: Same inner product.

$\langle \vec{v}^{(i)}, \vec{v}^{(j)} \rangle = \langle \vec{c}^{(i)}, \vec{c}^{(j)} \rangle$

(b) Same norm.

$\|\vec{v}^{(i)}\| = \|\vec{c}^{(i)}\|$

(c) Same distance.

$d(\vec{v}^{(i)}, \vec{v}^{(j)}) = \|\vec{v}^{(j)} - \vec{v}^{(i)}\| = \|\vec{c}^{(j)} - \vec{c}^{(i)}\| = d(\vec{c}^{(i)}, \vec{c}^{(j)})$

6.3 Signal Space Concepts

As in the case of vectors, we now discuss a parallel treatment for a set of signals (waveforms).

Definition 6.32.

(a) The **inner product** of two real-valued signals $x_1(t)$ and $x_2(t)$ is denoted by $\langle x_1(t), x_2(t)\rangle$ and defined by

$$\langle x_1(t), x_2(t)\rangle = \int_{-\infty}^{\infty} x_1(t)x_2(t)dt.$$

(b) The signals are **orthogonal** if their inner product is zero.

(c) The **norm** of a signal is defined as

$$\|x(t)\| = \sqrt{\langle x(t), x(t)\rangle} = \sqrt{E_x}$$

where E_x is the energy in $x(t)$:

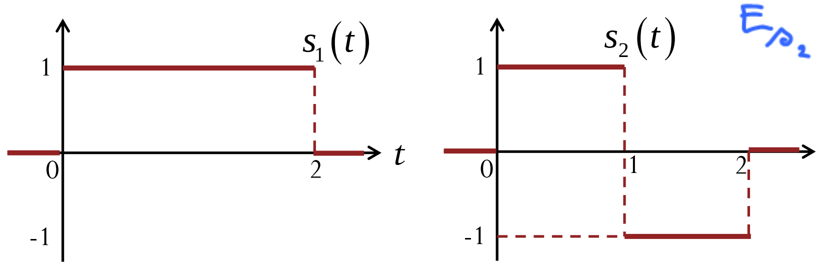
$$\langle x(t), x(t)\rangle = \int_{-\infty}^{\infty} x(t)\,x(t)\,dt = \int_{-\infty}^{\infty} |x(t)|^2 dt \equiv E_x$$

(d) A collection of N signals is **orthonormal** if the signals are orthogonal and their norms are all unity.

Example 6.33. Consider the two waveforms shown in Figure 26.

$$E_{s_1} = \int_{-\infty}^{\infty} s_1^2(t)\,dt = 1^2 \times 2 = 2$$

$$\|s_1\| = \sqrt{E_{s_1}} = \sqrt{2}$$



$$E_{s_2} = \int_{-\infty}^{\infty} s_2^2(t)\,dt = 1^2 \times 2 = 2$$

$$\|s_2\| = \sqrt{E_{s_2}} \cdot \sqrt{2}$$

Figure 26: Two Waveforms in Example 6.33

$$\langle s_1, s_2\rangle = \int_{-\infty}^{\infty} s_1(t)\,s_2(t)\,dt = 0$$

$s_2(t)$

6.34. For a signal, note that if its norm is unity, then its energy is also unity as well. We will use $\phi(t)$ to denote a unit-energy signal.

6.35. Similar to 6.29, given a collection of M signals $s_1(t)$, $s_2(t)$, $\ldots$, $s_M(t)$, we can (use a technique called GSOP) to find a collection of N orthonormal signals $\phi_1(t)$, $\phi_2(t)$, $\ldots$, $\phi_N(t)$ such that any signal $s_j(t)$ can be expressed as a linear combination:

$$s_j(t) = \sum_{i=1}^{N} s_i^{(j)} \phi_i(t) \tag{37}$$

where the constants (weights)

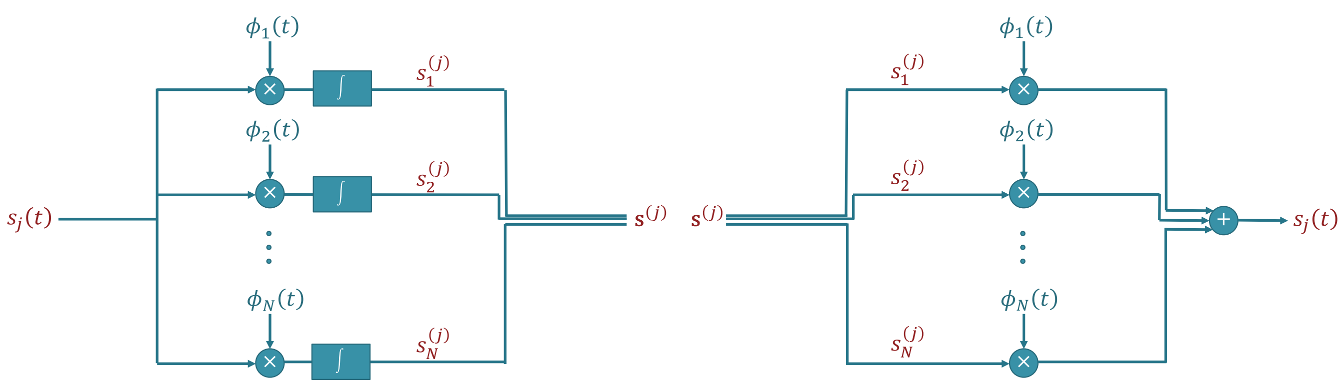
$$s_i^{(j)} = \langle s_j(t), \phi_i(t) \rangle . \tag{38}$$

Each signal can then be represented by a vector (or sequence)

$$\mathbf{s}^{(j)} = (s_1^{(j)}, s_2^{(j)}, \ldots, s_N^{(j)})^T, \tag{39}$$

or, equivalently, as a point in the N-dimensional (in general, complex) signal space.

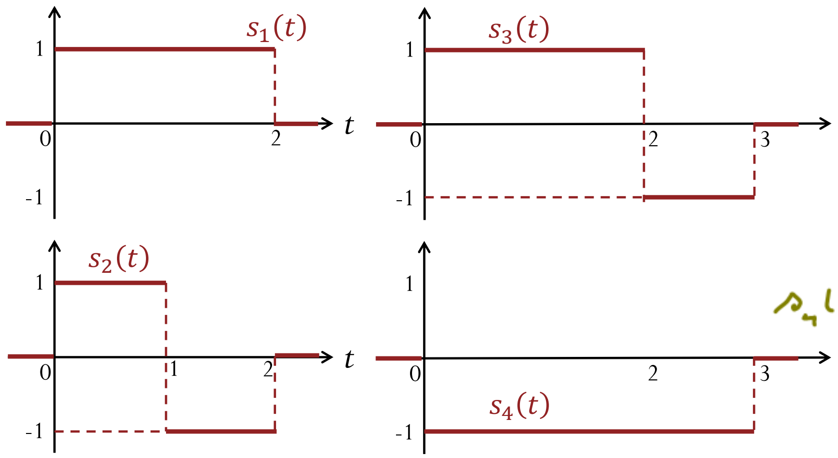
The (mathematical/conceptual) conversion/mapping from waveform to it corresponding vector in (39) and (38) is shown in Figure 27a. The inverse mapping from vector to waveform in (37) is shown in Figure 27b.



(a) (b)

Figure 27: Waveform to vector (a), and vector to waveform (b) mappings.

Example 6.36. Consider the four waveforms illustrated in Figure 28.

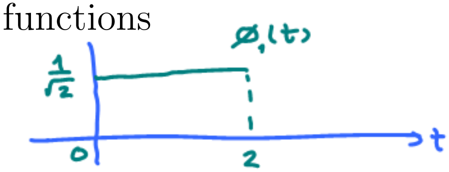


$s_4(t) = -s_1(t) + \phi_3(t)$
$= -\sqrt{2}\,\phi_1(t) + \phi_3(t)$

Figure 28: Four signals for orthogonalization in Example 6.36

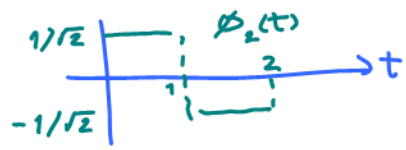
Consider the following ~~orthogonal~~ **orthonormal** functions

(a) $\phi_1(t) = \frac{s_1(t)}{\sqrt{E_{s_1}}} = \frac{1}{\sqrt{2}} s_1(t)$



$s_1(t) = \sqrt{2}\,\phi_1(t)$

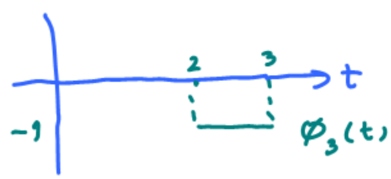
(b) $\phi_2(t) = \frac{s_2(t)}{\sqrt{E_{s_2}}} = \frac{1}{\sqrt{2}} s_2(t)$



$s_2(t) = \sqrt{2}\,\phi_2(t)$

(c) $\phi_3(t) = s_3(t) - s_1(t)$

$s_3(t) = s_1(t) + \phi_3(t)$
$= \sqrt{2}\,\phi_1(t) + \phi_3(t)$



Then,

$$s_1(t) = \underline{\sqrt{2}}\,\phi_1(t) + \underline{\ 0\ }\,\phi_2(t) + \underline{\ 0\ }\,\phi_3(t) \Rightarrow \mathbf{s}^{(1)} = \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}$$

$$s_2(t) = \underline{\ 0\ }\,\phi_1(t) + \underline{\sqrt{2}}\,\phi_2(t) + \underline{\ 0\ }\,\phi_3(t) \Rightarrow \mathbf{s}^{(2)} = \begin{pmatrix} 0 \\ \sqrt{2} \\ 0 \end{pmatrix}$$

$$s_3(t) = \underline{\sqrt{2}}\,\phi_1(t) + \underline{\ 0\ }\,\phi_2(t) + \underline{\ 1\ }\,\phi_3(t) \Rightarrow \mathbf{s}^{(3)} = \begin{pmatrix} \sqrt{2} \\ 0 \\ 1 \end{pmatrix}$$

$$s_4(t) = \underline{-\sqrt{2}}\,\phi_1(t) + \underline{\ 0\ }\,\phi_2(t) + \underline{\ 1\ }\,\phi_3(t) \Rightarrow \mathbf{s}^{(4)} = \begin{pmatrix} -\sqrt{2} \\ 0 \\ 1 \end{pmatrix}$$

Definition 6.37. From 6.35, a set of M signals $\{s_j(t), j = 1, 2, \dots, M\}$ can be represented by a set of M vectors $\{\mathbf{s}^{(j)}\}$ in the N-dimensional space. The corresponding set of vectors is called the **signal space representation**, or **constellation**, of $\{s_j(t), j = 1, 2, \dots, M\}$.

6.38. From the orthonormality of the basis, we have

(a) the inner product of two signals is equal to the inner product of the corresponding vectors:

waveform versions

$$\langle s_i(t), s_j(t)\rangle = \left\langle \mathbf{s}^{(i)}, \mathbf{s}^{(j)} \right\rangle.$$

vector versions

(b) $E_j \equiv E_{s^{(j)}} = \|s_j(t)\|^2 = \left\|\mathbf{s}^{(j)}\right\|^2.$

$E_{s_1} = (\sqrt{2})^2 + 0^2 + 0^2 = 2$

So, we can find the energy of any waveform represented in the constellation from its corresponding vector simply by the sum of its squared elements

6.39. The vector representation of the signals $\{s_j(t)\}$ will depend on the orthonormal functions $\{\phi_i(t)\}$, which are not unique. Nevertheless, the dimensionality of the signal space (N) will not change, and the vectors $\mathbf{s}^{(j)}$ will retain their geometric configuration; i.e., their lengths and their inner products will be invariant to the choice of the orthonormal functions $\{\phi_i(t)\}$.

6.4 Constellations for Digital Modulation Schemes

6.4.1 PAM

Definition 6.40. Recall, from 6.7, that **PAM signal waveforms** are represented as

$$s_m(t) = A_m p(t), \quad 1 \le m \le M$$

where $p(t)$ is a pulse and $A_m \in \mathcal{A}$.

6.41. Clearly, PAM signals are one-dimensional since all are multiples of the same basic signals. We define

$$\phi(t) = \frac{p(t)}{\sqrt{E_p}} \quad \Rightarrow p(t) = \sqrt{E_p}\,\phi(t)$$

as the basis for the PAM signals above. In which case,

$$s_m(t) = A_m \sqrt{E_p}\phi(t), \quad 1 \le m \le M$$

and the corresponding one-dimensional vector representation is

$$\mathbf{s}^{(m)} = A_m \sqrt{E_p}.$$

The corresponding signal space diagrams for $M = 2$, $M = 4$, and $M = 8$ are shown in Figure 29.

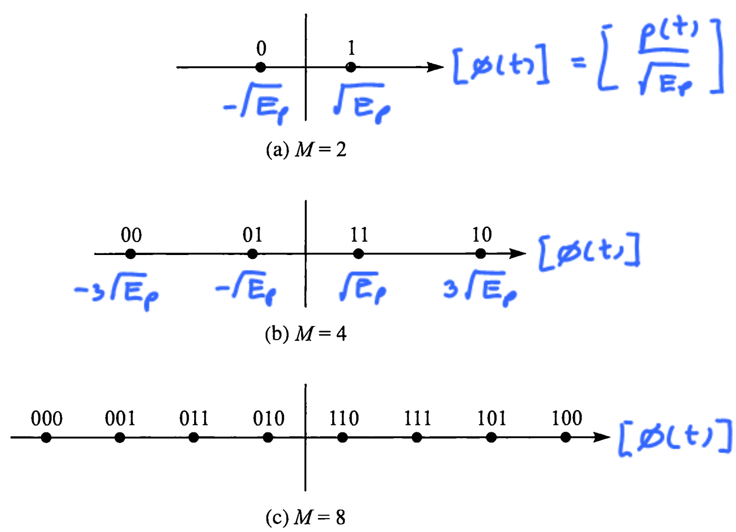


Figure 29: Constellation for PAM signaling

6.42. In **Amplitude-Shift Keying** (ASK), $p(t) = g(t)\cos(2\pi f_c t)$ where f_c is the carrier frequency.

6.4.2 Phase-Shift Keying (PSK)

Definition 6.43. In **digital phase modulation**, the M signal waveforms are represented as

$$s_m(t) = g(t)\cos\left(2\pi f_c t + \frac{2\pi}{M}(m-1)\right), \quad m = 1, 2, \ldots, M \qquad (40)$$

where

- $g(t)$ is the signal pulse shape and
- $\theta_m = \frac{2\pi}{M}(m-1)$, $m = 1, 2, \ldots, M$ is the M possible phases of the carrier that convey the transmitted information.

Digital phase modulation is usually called **phase-shift keying (PSK)**.

6.44. The PSK signal waveforms defined in (40) have equal energy:

6.45. Note that

(a) From the cos identity

$$\cos(\underbrace{\alpha}_{2\pi f_c t} \pm \overbrace{\beta}^{\theta_m}) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta,$$

we have

$$s_m(t) = g(t)\cos(\theta_m)\cos(2\pi f_c t) - g(t)\sin(\theta_m)\sin(2\pi f_c t).$$

(b) $g(t)\cos(2\pi f_c t)$ and $-g(t)\sin(2\pi f_c t)$ are orthogonal.

Suppose we define

$$\phi_1(t) = \sqrt{\frac{2}{E_g}} g(t) \cos(2\pi f_c t), \tag{41}$$

$$\phi_2(t) = -\sqrt{\frac{2}{E_g}} g(t) \sin(2\pi f_c t). \tag{42}$$

In which case,

$$s_m(t) = \sqrt{\frac{E_g}{2}} \cos(\theta_m) \phi_1(t) + \sqrt{\frac{E_g}{2}} \sin(\theta_m) \phi_2(t).$$

Therefore the signal space dimensionality is $N = 2$ and the resulting vector representations are

$$\mathbf{s}^{(m)} = \left(\sqrt{\frac{E_g}{2}} \cos(\theta_m), \sqrt{\frac{E_g}{2}} \sin(\theta_m) \right)^T.$$

6.46. Signal space diagrams for BPSK (binary PSK, $M = 2$), QPSK (quaternary PSK, $M = 4$), and 8-PSK are shown in Figure 30.

$\theta_m = \frac{2\pi}{M}(m-1)$

$M = 2$

$M = 4$

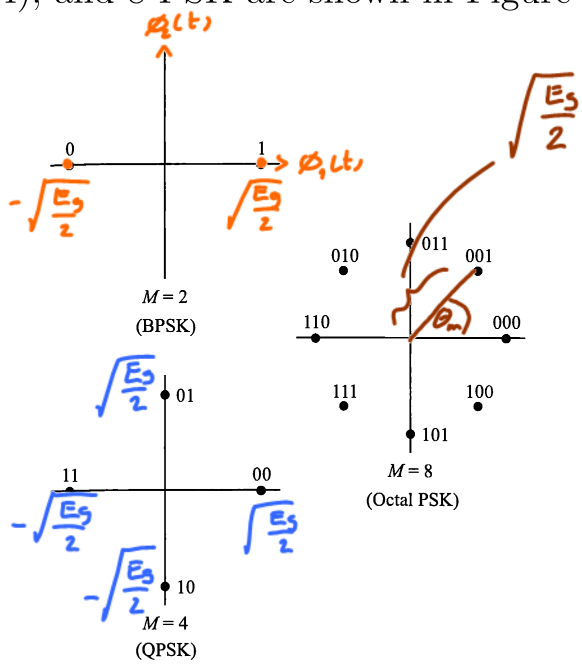


Figure 30: Signal space diagrams for BPSK, QPSK, and 8-PSK.

Note that BPSK corresponds to one-dimensional signals, which are identical to binary PAM signals.

6.4.3 Quadrature Amplitude Modulation (QAM)

Definition 6.47. In **Quadrature Amplitude Modulation (QAM)**, two separate b-bit symbols from the information sequence on two quadrature carriers $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are transmitted simultaneously. The corresponding signal waveforms may be expressed as

$$s_m(t) = A_m^{(I)} g(t) \cos(2\pi f_c t) - A_m^{(Q)} g(t) \sin(2\pi f_c t), \quad m = 1, 2, \ldots, M \quad (43)$$

where

- $A_m^{(I)}$ and $A_m^{(Q)}$ are the information-bearing signal amplitudes of the quadrature carriers and
- $g(t)$ is the signal pulse.

Equivalently,

$$\begin{aligned} s_m(t) &= \mathrm{Re}\left\{ \left(A_m^{(I)} + jA_m^{(Q)} \right) g(t)\, e^{j2\pi f_c t} \right\} && (44) \\ &= \mathrm{Re}\left\{ r_m e^{j\theta_m} g(t)\, e^{j2\pi f_c t} \right\} && (45) \\ &= r_m g(t) \cos(2\pi f_c t + \theta_m) && (46) \end{aligned}$$

where

- $r_m = \sqrt{\left(A_m^{(I)}\right)^2 + \left(A_m^{(Q)}\right)^2}$ is the magnitude and
- θ_m is the argument or phase

of the complex number $A_m^{(I)} + jA_m^{(Q)}$.

6.48. From (46), it is apparent that the QAM signal waveforms may be viewed as combined amplitude (r_m) and phase (θ_m) modulation. In fact, we may select any combination of M_1-level PAM and M_2-phase PSK to construct an $M = M_1 M_2$ combined **PAM-PSK signal constellation**.

- If $M_1 = 2^{b_1}$ and $M_2 = 2^{b_2}$, the combined PAM-PSK signal constellation results in the simultaneous transmission of $b_1 + b_2 = \log_2 M_1 M_2$ binary digits occurring at a symbol rate $R/(b_1 + b_2)$.

6.49. From (43), it can be seen that, similar to the PSK case, $\phi_1(t)$ and $\phi_2(t)$ given in (41) and (42) can be used as an orthonormal basis for QAM signals. The dimensionality of the signal space for QAM is $N = 2$. Using this basis, we have

$$s_m(t) = A_m^{(I)}\sqrt{\frac{E_g}{2}}\phi_1(t) + A_m^{(Q)}\sqrt{\frac{E_g}{2}}\phi_2(t)$$

which results in vector representations of the form

$$\mathbf{s}^{(m)} = \left(A_m^{(I)}\sqrt{\frac{E_g}{2}}, A_m^{(Q)}\sqrt{\frac{E_g}{2}}\right)^T.$$

Example 6.50. Examples of signal space diagrams for combined PAM-PSK are shown in Figure 31, for $M = 8$ and $M = 16$.

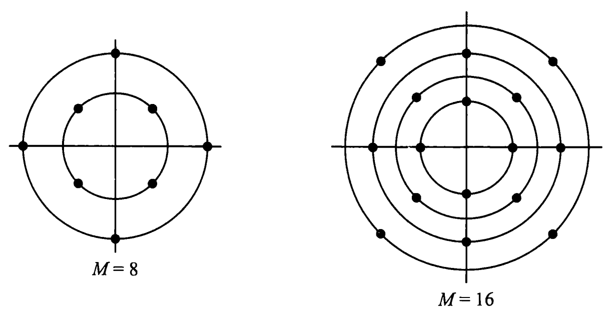


Figure 31: Examples of combined PAM-PSK constellations.

Example 6.51. In the special case where the signal amplitudes are taken from the set of discrete values $\mathcal{A} = \{(2m - 1 - M), m = 1, 2, \ldots, M\}$, the signal space diagram is rectangular, as shown in Figure 32.

6.52. PAM and PSK can be considered as special cases of QAM. In QAM signaling, both amplitude and phase carry information, whereas in PAM and PSK only amplitude or phase carries the information.

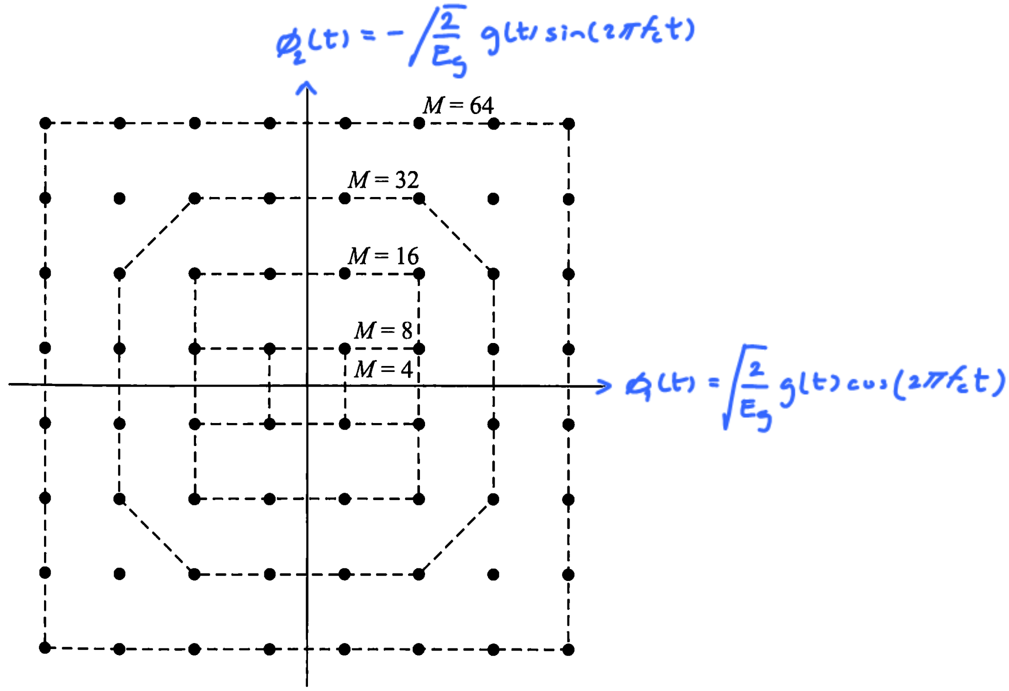


Figure 32: Several signal space diagrams for rectangular QAM.

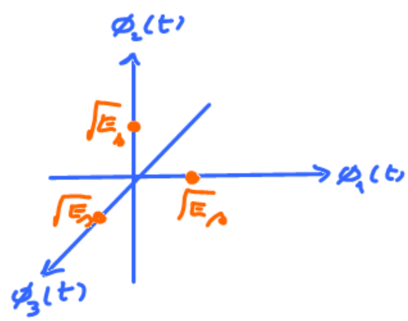
6.4.4 Orthogonal Signaling

Definition 6.53. In **orthogonal signaling**, the waveforms $s_m(t)$ are orthogonal and of equal energy E_s. In which case, the orthonormal set $\{\phi_m(t), 1 \le m \le N\}$ defined by

$$\phi_m(t) = \frac{s_m(t)}{\sqrt{E_s}}, \quad 1 \le m \le M$$

can be used as an orthonormal basis for representation of $\{s_m(t), 1 \le m \le M\}$. The resulting vector representation of the signals will be

$$\begin{aligned} \mathbf{s}^{(1)} &= \left(\sqrt{E_s}, 0, 0, \dots, 0\right), \\ \mathbf{s}^{(2)} &= \left(0, \sqrt{E_s}, 0, \dots, 0\right), \\ \vdots &= \vdots \\ \mathbf{s}^{(M)} &= \left(0, 0, 0, \dots, \sqrt{E_s}\right). \end{aligned}$$



Definition 6.54. In **Frequency-Shift Keying (FSK)**, messages are transmitted by waveforms that differ in frequency:

$$s_m(t) = A\cos(2\pi f_m t), \quad 0 \le t \le T_s, \ 1 \le m \le M.$$

Note that when $f_m = \frac{m}{T_s}$, we have orthogonal signaling.