Digital Modulation 1

– Lecture Notes –

Ingmar Land and Bernard H. Fleury

Navigation and Communications (NavCom)
Department of Electronic Systems
Aalborg University, DK

Version: February 5, 2007

Contents

References

[1] J. G. Proakis and M. Salehi, *Communication Systems Engineering*, 2nd ed. Prentice-Hall, 2002.

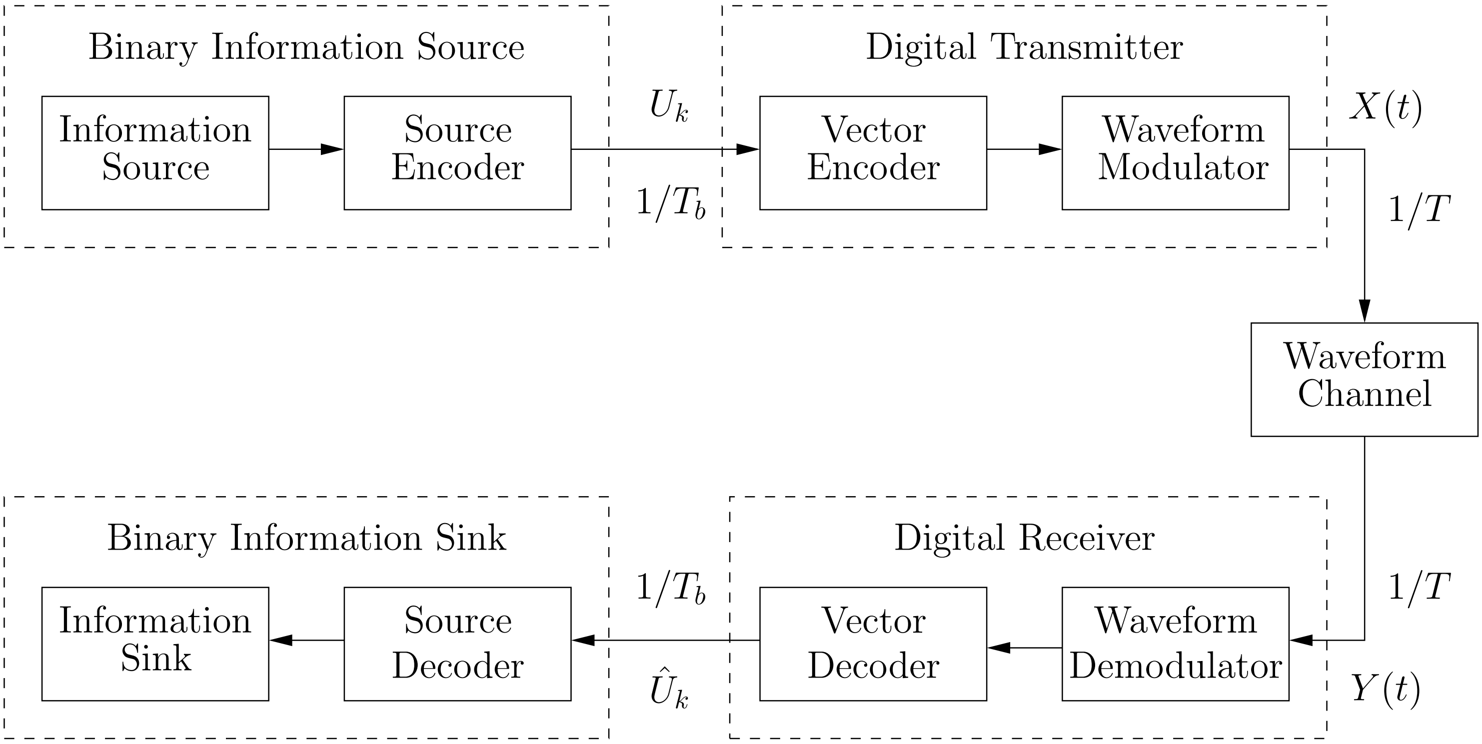
[2] B. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inform. Theory*, vol. 34, no. 2, pp. 260–270, Mar. 1988.

Part I

Basic Concepts

1 The Basic Constituents of a Digital Communication System

Block diagram of a digital communication system:



- Time duration of one binary symbol (bit) U_k: T_b
- Time duration of one waveform $X(t)$: T
- Waveform channel may introduce
  - distortion,
  - interference,
  - noise

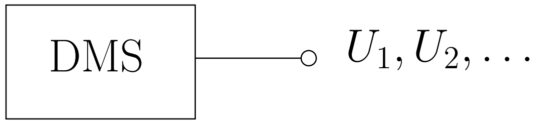
Goal: Signal of information source and signal at information sink should be as similar as possible according to some measure.

Some properties:

- Information source: may be analog or digital
- Source encoder: may perform sampling, quantization and compression; generates one binary symbol (bit) $U_k \in \{0, 1\}$ per time interval T_b
- Waveform modulator: generates one waveform $x(t)$ per time interval T
- Vector decoder: generates one binary symbol (bit) $\hat{U}_k \in \{0, 1\}$ (estimates of U_k) per time interval T_b
- Source decoder: reconstructs the original source signal

1.1 Discrete Information Sources

A discrete memoryless source (DMS) is a source that generates a sequence $U_1, U_2, \ldots$ of independent and identically distributed (i.i.d.) discrete random variables, called an i.i.d. sequence.



Properties of the output sequence:

discrete $U_1, U_2, \ldots \in \{a_1, a_2, \ldots, a_Q\}$.

memoryless $U_1, U_2, \ldots$ are statistically independent.

stationary $U_1, U_2, \ldots$ are identically distributed.

Notice: The symbol alphabet $\{a_1, a_2, \ldots, a_Q\}$ has cardinality Q. The value of Q may be infinite, the elements of the alphabet only have to be countable.

Example: Discrete sources

(i) Output of the PC keyboard, SMS
(usually not memoryless).

(ii) Compressed file
(near memoryless).

(iii) The numbers drawn on a roulette table in a casino
(ought to be memoryless, but may not...).

◇

A DMS is statistically completely described by the probabilities

$$p_U(a_q) = \Pr(U = a_q)$$

of the symbols a_q, $q = 1, 2, \ldots, Q$. Notice that

(i) $p_U(a_q) \geq 0$ for all $q = 1, 2, \ldots, Q$, and

(ii) $\sum_{q=1}^{Q} p_U(a_q) = 1$.

As the sequence is i.i.d., we have $\Pr(U_i = a_q) = \Pr(U_j = a_q)$. For convenience, we may write $p_U(a)$ shortly as $p(a)$.

Binary Symmetric Source

A binary memoryless source is a DMS with a binary symbol alphabet.

Remark:
Commonly used binary symbol alphabets are $\mathbb{F}_2 := \{0, 1\}$ and $\mathbb{B} := \{-1, +1\}$. In the following, we will use $\mathbb{F}_2$.

A binary symmetric source (BSS) is a binary memoryless source with equiprobable symbols,

$$p_U(0) = p_U(1) = \frac{1}{2}.$$

Example: Binary Symmetric Source

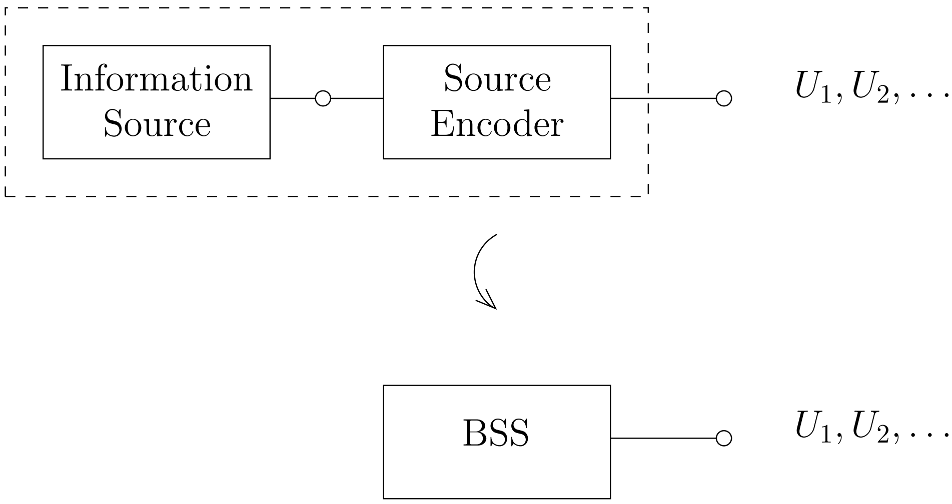
All length-K sequences of a BSS have the same probability

$$p_{U_1 U_2 \ldots U_K}(u_1 u_2 \ldots u_K) = \prod_{i=1}^{K} p_U(u_i) = \left(\frac{1}{2}\right)^K.$$

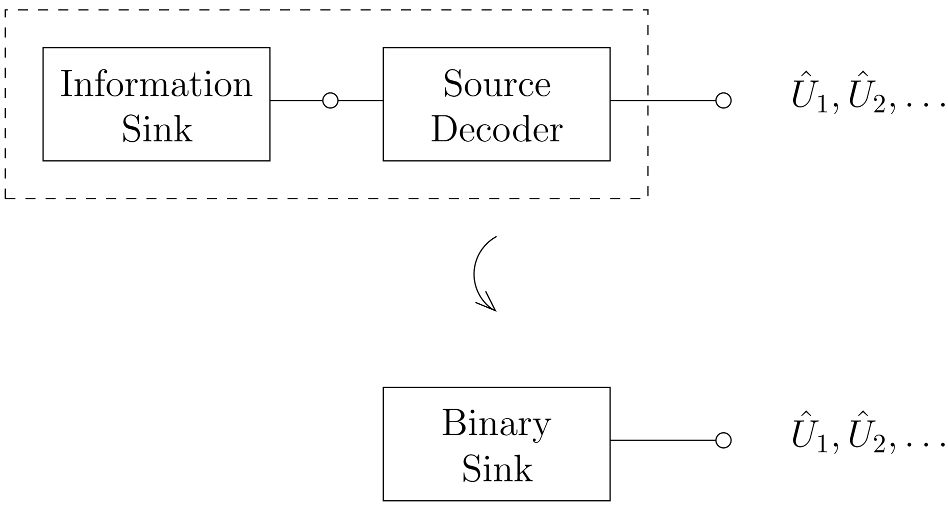
◇

1.2 Considered System Model

This course focuses on the transmitter and the receiver. Therefore, we replace the information source by a BSS



and the information sink by a binary (information) sink.

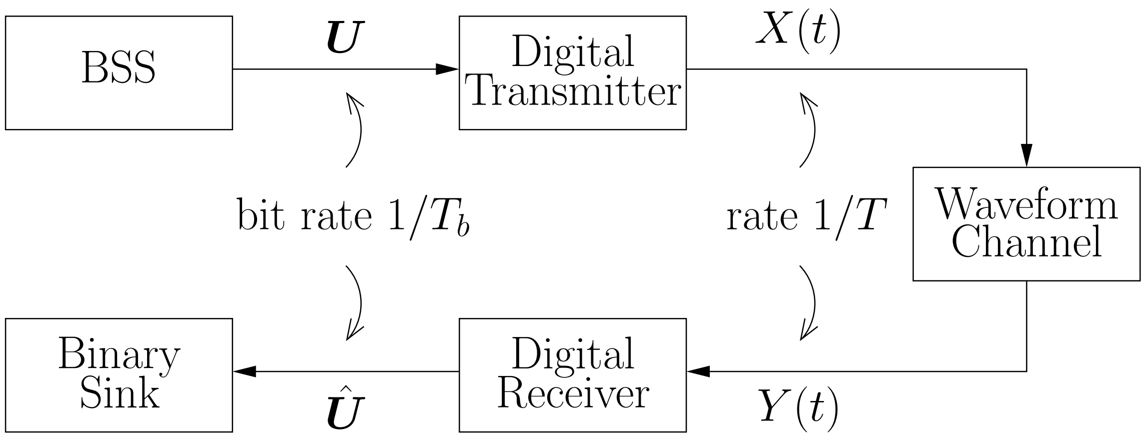


Some Remarks

(i) There is no loss of generality resulting from these substitutions. Indeed it can be demonstrated within Shannon's Information Theory that an efficient source encoder converts the output of an information source into a random binary independent and uniformly distributed (i.u.d.) sequence. Thus, the output of a perfect source encoder looks like the output of a BSS.

(ii) Our main concern is the design of communication systems for reliable transmission of the output symbols of a BSS. We will not address the various methods for efficient source encoding.

Digital communication system considered in this course:



- Source: $\boldsymbol{U} = [U_1, \ldots, U_K]$, $U_i \in \{0, 1\}$
- Transmitter: $X(t) = x(t, \boldsymbol{u})$
- Sink: $\hat{\boldsymbol{U}} = [\hat{U}_1, \ldots, \hat{U}_K]$, $\hat{U}_i \in \{0, 1\}$
- Bit rate: $1/T_b$
- Rate (of waveforms): $1/T = 1/(KT_b)$
- Bit error probability

$$P_b = \frac{1}{K} \sum_{k=1}^{K} \Pr(U_k \neq \hat{U}_k)$$

Objective: Design of efficient digital communication systems

$$\text{Efficiency means } \begin{cases} \text{small bit error probability and} \\ \text{high bit rate } 1/T_b. \end{cases}$$

Constraints and limitations:

- limited power
- limited bandwidth
- impairments (distortion, interference, noise) of the channel

Design goals:

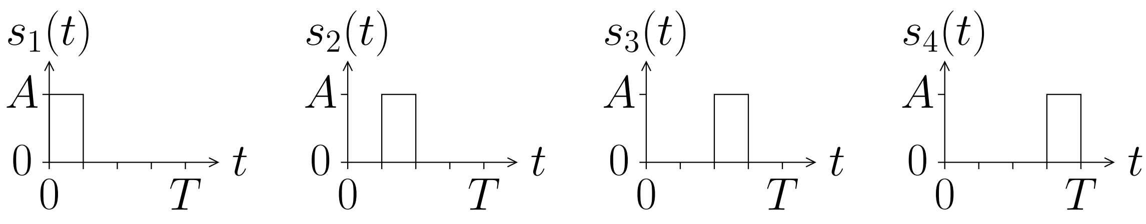
- "good" waveforms
- low-complexity transmitters
- low-complexity receivers

1.3 The Digital Transmitter

1.3.1 Waveform Look-up Table

Example: 4PPM (Pulse-Position Modulation)

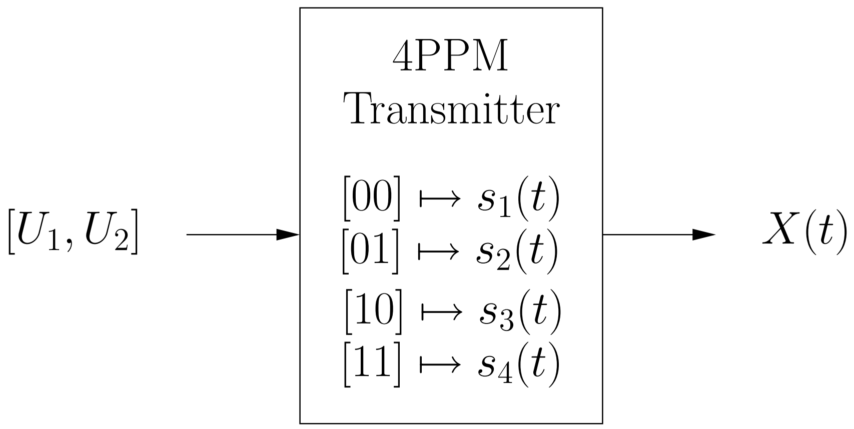
Set of four different waveforms, $\mathbb{S} = \{s_1(t), s_2(t), s_3(t), s_4(t)\}$:



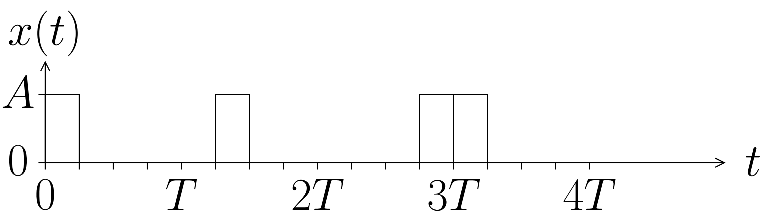
Each waveform $X(t) \in \mathbb{S}$ is addressed by vector $[U_1, U_2] \in \{0,1\}^2$:

$$[U_1, U_2] \quad \mapsto \quad X(t).$$

The mapping may be implemented by a waveform look-up table.

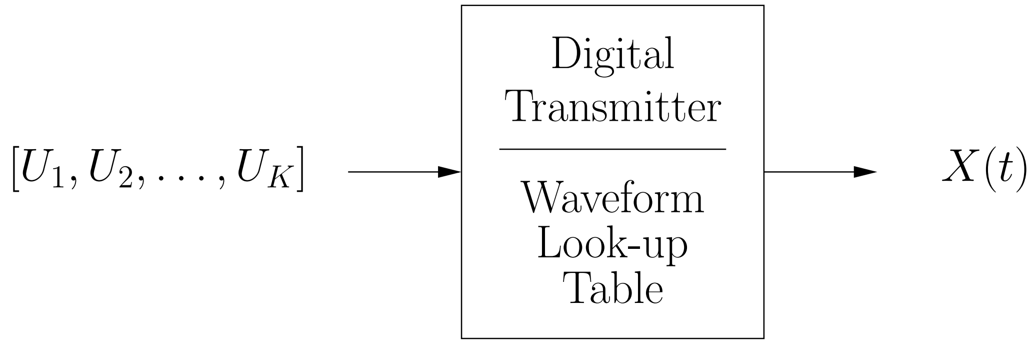


Example: [00011100] is transmitted as



◇

For an arbitrary digital transmitter, we have the following:



Input Binary vectors of length K from the input set $\mathbb{U}$:

$$[U_1, U_2, \ldots, U_K] \in \mathbb{U} := \{0, 1\}^K.$$

The "duration" of one binary symbol U_k is T_b.

Output Waveforms of duration T from the output set $\mathbb{S}$:

$$x(t) \in \mathbb{S} := \{s_1(t), s_2(t), \ldots, s_M(t)\}.$$

Waveform duration T means that for $m = 1, 2, \ldots, M$,

$$s_m(t) = 0 \text{ for } t \notin [0, T].$$

The look-up table maps each input vector $[u_1, \ldots, u_K] \in \mathbb{U}$ to one waveform $x(t) \in \mathbb{S}$. Thus the digital transmitter may defined by a mapping $\mathbb{U} \to \mathbb{S}$ with

$$[u_1, \ldots, u_K] \mapsto x(t).$$

The mapping is one-to-one and onto such that

$$M = 2^K.$$

Relation between signaling interval (waveform duration) T and bit interval ("bit duration") T_b:

$$T = K \cdot T_b.$$

1.3.2 Waveform Synthesis

The set of waveforms,

$$\mathbb{S} := \{s_1(t), s_2(t), \ldots, s_M(t)\},$$

spans a vector space. Applying the Gram-Schmidt orthogonalization procedure, we can find a set of orthonormal functions

$$\mathbb{S}_\psi := \{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\} \quad \text{with } D \leq M$$

such that the space spanned by $\mathbb{S}_\psi$ contains the space spanned by $\mathbb{S}$.

Hence, each waveform $s_m(t)$ can be represented by a linear combination of the orthonormal functions:

$$s_m(t) = \sum_{i=1}^{D} s_{m,i} \cdot \psi_i(t),$$

$m = 1, 2, \ldots, M$. Each signal $s_m(t)$ can thus be geometrically represented by the D-dimensional vector

$$\boldsymbol{s}_m = [s_{m,1}, s_{m,2}, \ldots, s_{m,D}]^\mathsf{T} \in \mathbb{R}^D$$

with respect to the set $\mathbb{S}_\psi$.

Further details are given in Appendix A.

1.3.3 Canonical Decomposition

The digital transmitter may be represented by a waveform look-up table:

$$\boldsymbol{u} = [u_1, \ldots, u_K] \quad \mapsto \quad x(t),$$

where $\boldsymbol{u} \in \mathbb{U}$ and $x(t) \in \mathbb{S}$.

From the previous section, we know how to synthesize the waveform $s_m(t)$ from $\boldsymbol{s}_m$ (see also Appendix A) with respect to a set of basis functions

$$\mathbb{S}_\psi := \{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\}.$$

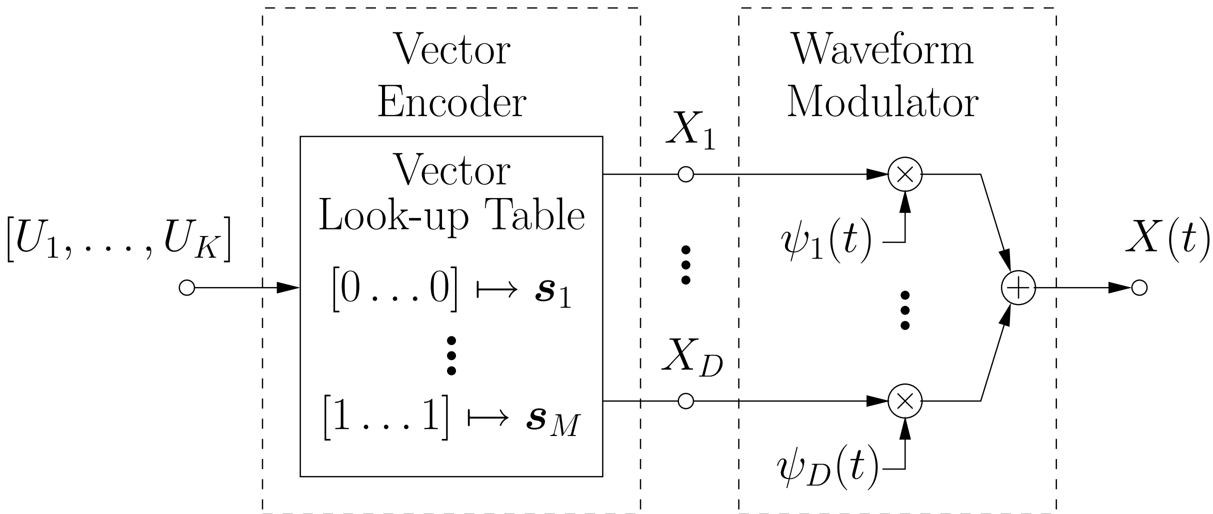
Making use of this method, we can split the waveform look-up table into a vector look-up table and a waveform synthesizer:

$$\boldsymbol{u} = [u_1, \ldots, u_K] \quad \mapsto \quad \boldsymbol{x} = [x_1, x_2, \ldots, x_D] \quad \mapsto \quad x(t),$$

where

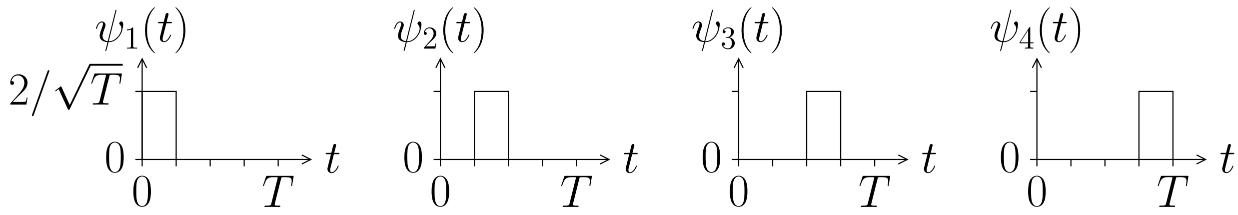
$$\begin{aligned} \boldsymbol{u} &\in \mathbb{U} = \{0, 1\}^K, \\ \boldsymbol{x} &\in \mathbb{X} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_D\} \subset \mathbb{R}^D, \\ x(t) &\in \mathbb{S} = \{s_1(t), s_2(t), \ldots, s_M(t)\}. \end{aligned}$$

This splitting procedure leads to the sought canonical decomposition of a digital transmitter:

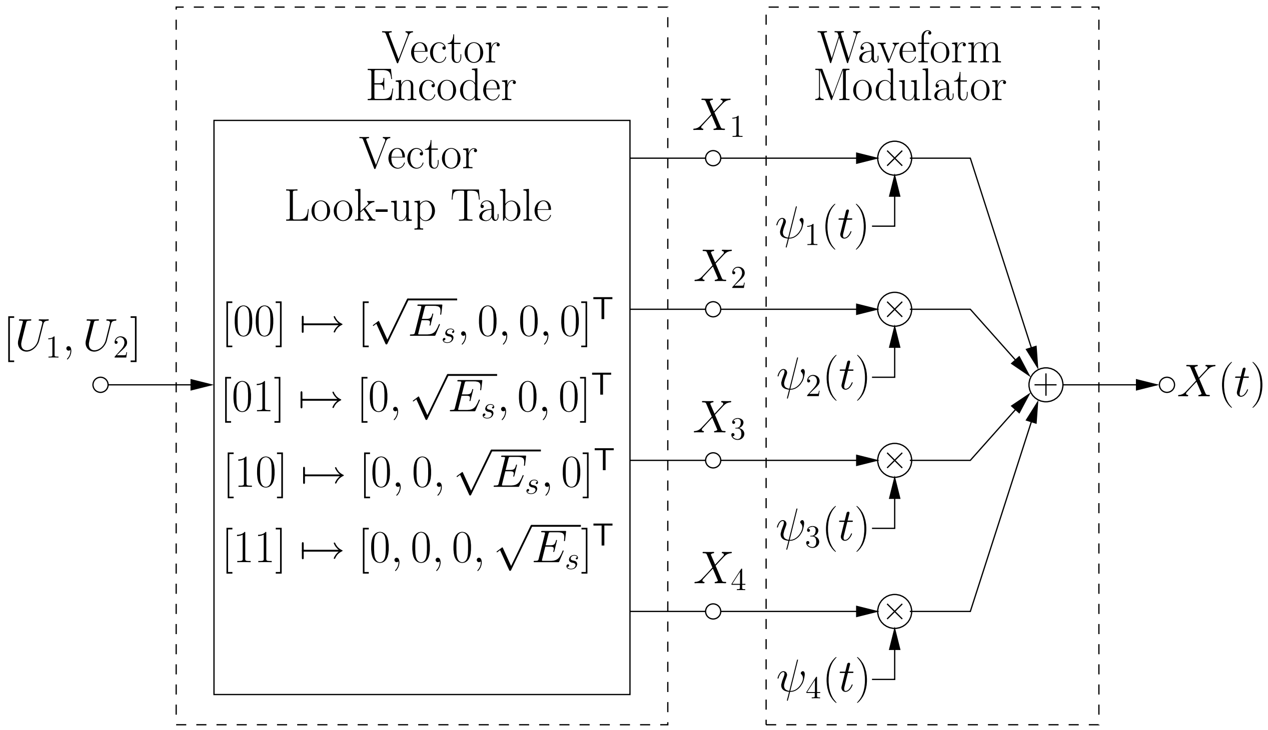


Example: 4PPM

The four orthonormal basis functions are



The canonical decomposition of the receiver is the following, where $\sqrt{E_s} = A \cdot \sqrt{T}/2$.

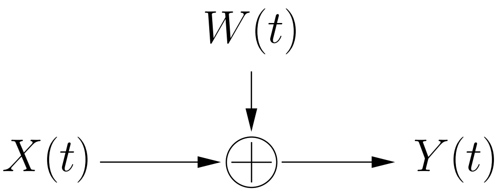


Remarks:

- The signal energy of each basis function is equal to one: $\int \psi_d(t)dt = 1$ (due to their construction).
- The basis functions are orthonormal (due to their construction).
- The value $\sqrt{E_s}$ is chosen such that the synthesized signals have the same energy as the original signals, namely E_s (compare Example 3).

◇

1.4 The Additive White Gaussian-Noise Channel



The additive white Gaussian noise (AWGN) channel-model is widely used in communications. The transmitted signal $X(t)$ is superimposed by a stochastic noise signal $W(t)$, such that the transmitted signal reads

$$Y(t) = X(t) + W(t).$$

The stochastic process $W(t)$ is a **stationary** process with the following properties:

(i) $W(t)$ is a **Gaussian process**, i.e., for each time t, the samples $v = w(t)$ are Gaussian distributed with zero mean and variance σ^2:

$$p_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{v^2}{2\sigma^2}\big).$$

(ii) $W(t)$ has a **flat power spectrum** with height $N_0/2$:

$$S_W(f) = \frac{N_0}{2}$$

(Therefore it is called “white”.)

(iii) $W(t)$ has the autocorrelation function

$$R_W(\tau) = \frac{N_0}{2}\delta(\tau).$$

Remarks

1. Autocorrelation function and power spectrum:

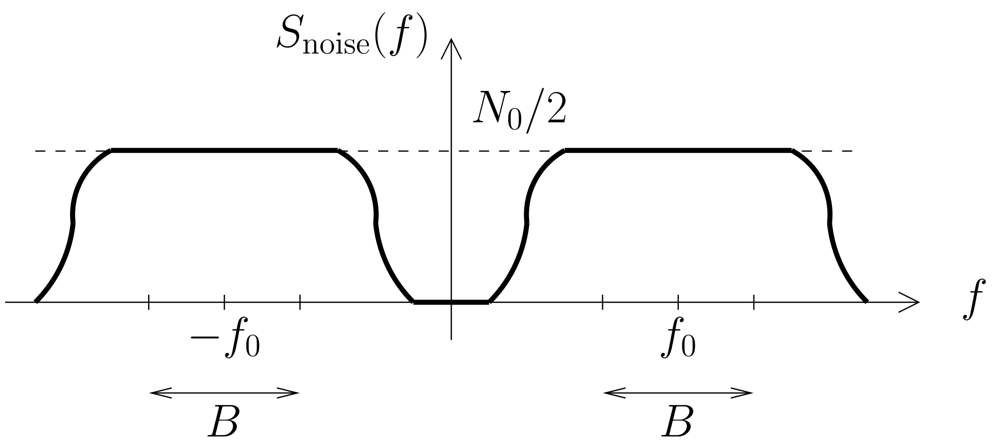
$$R_W(\tau) = \mathrm{E}\big[W(t)W(t+\tau)\big], \quad S_W(f) = \mathcal{F}\big\{R_W(\tau)\big\}.$$

2. For Gaussian processes, weak stationarity implies strong stationarity.

3. An WGN is an idealized process without physical reality:

   - The process is so "wild" that its realizations are not ordinary functions of time.
   - Its power is infinite.

   However, a WGN is a useful approximation of a noise with a flat power spectrum in the bandwidth B used by a communication system:



4. In satellite communications, $W(t)$ is the thermal noise of the receiver front-end. In this case, $N_0/2$ is proportional to the squared temperature.

1.5 The Digital Receiver

Consider a transmission system using the waveforms

$$\mathbb{S} = \Big\{ s_1(t), s_2(t), \ldots, s_M(t) \Big\}$$

with $s_m(t) = 0$ for $t \notin [0, T]$, $m = 1, 2, \ldots, M$, i.e., with duration T.

Assume transmission over an AWGN channel, such that

$$y(t) = x(t) + w(t),$$

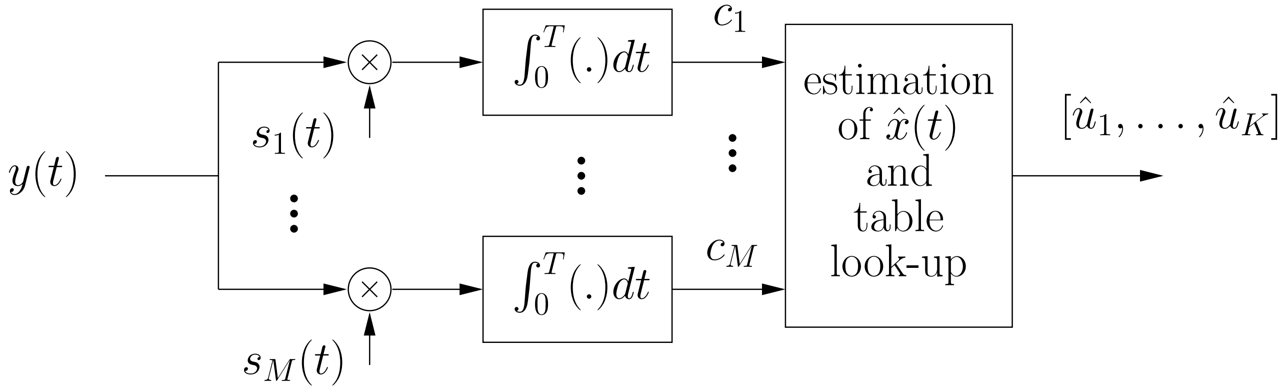
$x(t) \in \mathbb{S}$.

1.5.1 Bank of Correlators

The transmitted waveform may be recovered from the received waveform $y(t)$ by correlating $y(t)$ with all possible waveforms:

$$c_m = \langle y(t), s_m(t) \rangle,$$

$m = 1, 2, \ldots, M$. Based on the correlations $[c_1, \ldots, c_M]$, the waveform $\hat{s}(t)$ with the highest correlation is chosen and the corresponding (estimated) source vector $\hat{\boldsymbol{u}} = [\hat{u}_1, \ldots, \hat{u}_K]$ is output.



Disadvantage: High complexity.

1.5.2 Canonical Decomposition

Consider a set of orthonormal functions

$$\mathbb{S}_\psi = \Big\{\psi_1(t), \psi_2(t), \dots, \psi_D(t)\Big\}$$

obtained by applying the Gram-Schmidt procedure to $s_1(t), s_2(t), \dots, s_M(t)$. Then,

$$s_m(t) = \sum_{d=1}^{D} s_{m,d} \cdot \psi_d(t),$$

$m = 1, 2, \dots, M$. The vector

$$\boldsymbol{s}_m = \big[s_{m,1}, s_{m,2}, \dots, s_{m,D}\big]^\mathsf{T}$$

entirely determines $s_m(t)$ with respect to the orthonormal set $\mathbb{S}_\psi$.

The set $\mathbb{S}_\psi$ spans the vector space

$$\mathcal{S}_\psi := \Big\{s(t) = \sum_{i=1}^{D} s_i \psi_i(t) : [s_1, s_2, \dots, s_D] \in \mathbb{R}^D\Big\}.$$

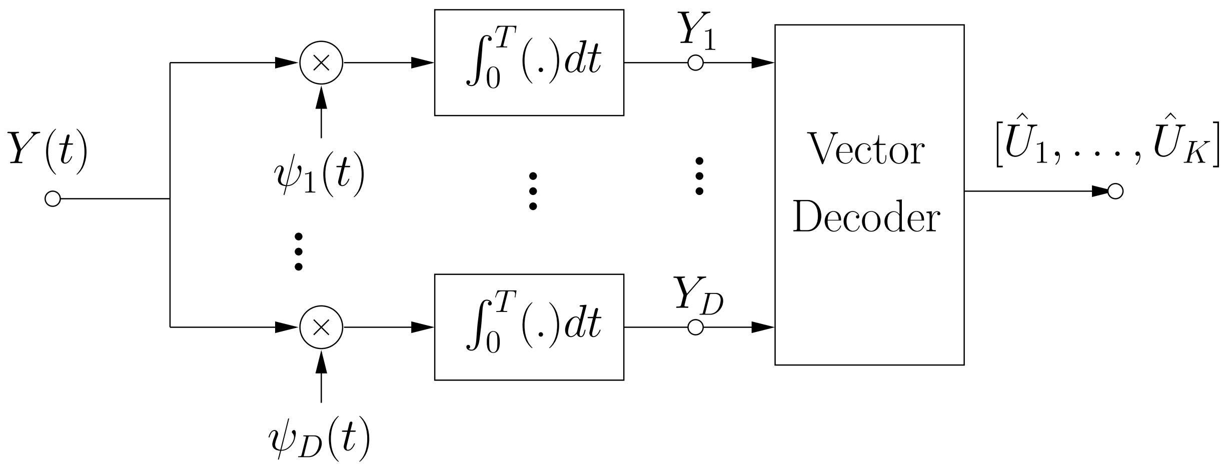
Basic Idea

- The received waveform $y(t)$ may contain components outside of $\mathcal{S}_\psi$. However, only the components inside $\mathcal{S}_\psi$ are relevant, as $x(t) \in \mathcal{S}_\psi$.
- Determine the vector representation $\boldsymbol{y}$ of the components of $y(t)$ that are inside $\mathcal{S}_\psi$.
- This vector $\boldsymbol{y}$ is sufficient for estimating the transmitted waveform.

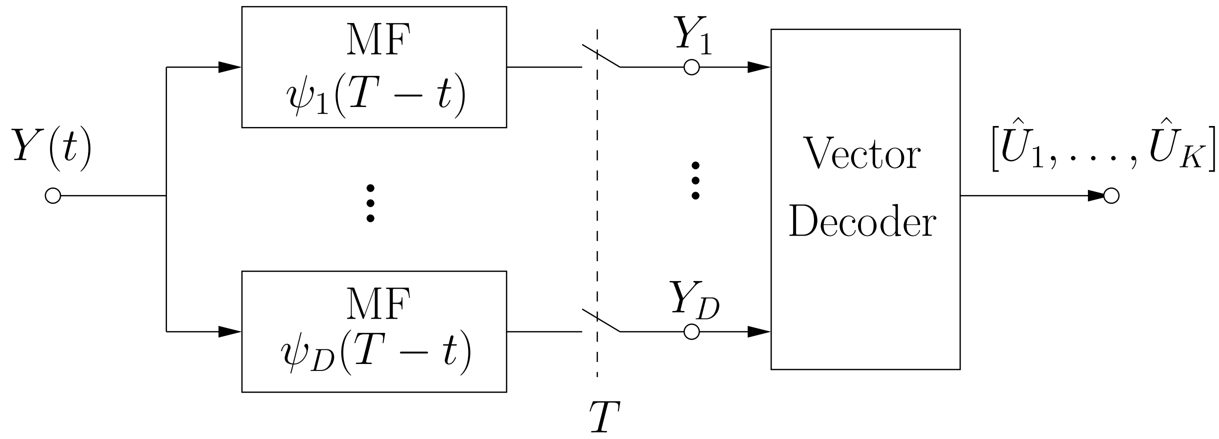
Canonical decomposition of the optimal receiver for the AWGN channel

Appendix A describes two ways to compute the vector representation of a waveform. Accordingly, the demodulator may be implemented in two ways, leading to the following two receiver structures.

Correlator-based Demodulator and Vector Decoder



Matched-filter based Demodulator and Vector Decoder



The optimality of the receiver structures is shown in the following sections.

1.5.3 Analysis of the correlator outputs

Assume that the waveform $s_m(t)$ is transmitted, i.e.,

$$x(t) = s_m(t).$$

The received waveform is thus

$$y(t) = x(t) + w(t) = s_m(t) + w(t).$$

The correlator outputs are

$$\begin{aligned}
y_d &= \int_0^T y(t)\psi_d(t)\mathsf{d}t \\
&= \int_0^T \Big[s_m(t) + w(t)\Big]\psi_d(t)\mathsf{d}t \\
&= \underbrace{\int_0^T s_m(t)\psi_d(t)\mathsf{d}t}_{s_{m,d}} + \underbrace{\int_0^T w(t)\psi_d(t)\mathsf{d}t}_{w_d} \\
&= s_{m,d} + w_d,
\end{aligned}$$

$d = 1, 2, \dots, D$.

Using the notation

$$\boldsymbol{y} = [y_1, \dots, y_D]^\mathsf{T},$$

$$\boldsymbol{w} = [w_1, \dots, w_D]^\mathsf{T}, \qquad w_d = \int_0^T w(t)\psi_d(t)\mathsf{d}t,$$

we can recast the correlator outputs in the compact form

$$\boldsymbol{y} = \boldsymbol{s}_m + \boldsymbol{w}.$$

The waveform channel between $x(t) = s_m(t)$ and $y(t)$ is thus transformed into a vector channel between $\boldsymbol{x} = \boldsymbol{s}_m$ and $\boldsymbol{y}$.

Remark:
Since the vector decoder "sees" noisy observations of the vectors $\boldsymbol{s}_m$, these vectors are also called **signal points** with respect to $\mathbb{S}_\psi$, and the set of signal points is called the **signal constellation**.

From Appendix A, we know that $s_m(t)$ entirely characterizes $s_m(t)$. The vector $\boldsymbol{y}$, however, does not represent the complete waveform $y(t)$; it represents only the waveform

$$y'(t) = \sum_{d=1}^{D} y_d \psi_d(t)$$

which is the "part" of $y(t)$ within the vector space spanned by the set $\mathcal{S}_\psi$ of orthonormal functions, i.e.,

$$y'(t) \in \mathcal{S}_\psi.$$

The "remaining part" of $y(t)$, i.e., the part outside of $\mathcal{S}_\psi$, reads

$$\begin{aligned}
y^\circ(t) &= y(t) - y'(t) \\
&= y(t) - \sum_{d=1}^{D} y_d \psi_d(t) \\
&= s_m(t) + w(t) - \sum_{d=1}^{D} [s_{m,d} + w_d] \psi_d(t) \\
&= \underbrace{s_m(t) - \sum_{d=1}^{D} s_{m,d} \psi_d(t)}_{=0} + w(t) - \sum_{d=1}^{D} w_d \psi_d(t) \\
&= w(t) - \sum_{d=1}^{D} w_d \psi_d(t).
\end{aligned}$$

Observations

- The waveform $y^{\circ}(t)$ contains no information about $s_m(t)$.
- The waveform $y'(t)$ is completely represented by $\boldsymbol{y}$.

Therefore, the receiver structures do not lead to a loss of information, and so they are optimal.

1.5.4 Statistical Properties of the Noise Vector

The components of the noise vector

$$\boldsymbol{w} = [w_1, \dots, w_D]^\mathsf{T}$$

are given by

$$w_d = \int\limits_0^T w(t)\psi_d(t)\mathsf{d}t,$$

$d = 1, 2, \dots, D$. The random variables W_d are Gaussian as they result from a linear transformation of the Gaussian process $W(t)$. Hence the probability density function of W_d is of the form

$$p_{W_d}(w_d) = \frac{1}{\sqrt{2\pi\sigma_d^2}} \exp\bigl(-\frac{(w_d - \mu_d)^2}{2\sigma_d^2}\bigr).$$

where the mean value μ_d and the variance σ_d^2 are given by

$$\mu_d = \mathrm{E}\bigl[W_d\bigr], \qquad \sigma_d^2 = \mathrm{E}\bigl[(W_d - \mu_d)^2\bigr].$$

These values are now determined.

Mean value

Using the definition of W_d and taking into account that the process $W(t)$ has zero mean, we obtain

$$\begin{aligned} \mathrm{E}\bigl[W_d\bigr] &= \mathrm{E}\Bigl[\int\limits_0^T W(t)\psi_d(t)\mathsf{d}t\Bigr] \\ &= \int\limits_0^T \mathrm{E}\bigl[W(t)\bigr]\psi_d(t)\mathsf{d}t \\ &= 0. \end{aligned}$$

Thus, all random variables W_d have zero mean.

Variance

For computing the variance, we use $\mu_d = 0$, the definition of W_d, and the autocorrelation function of $W(t)$,

$$R_W(\tau) = \frac{N_0}{2}\delta(\tau).$$

We obtain the following chain of equations:

$$\begin{aligned}
\mathrm{E}\big[(W_d - \mu_d)^2\big] &= \mathrm{E}\big[W_d^2\big] \\
&= \mathrm{E}\Big[\Big(\int_0^T W(t)\psi_d(t)\mathsf{d}t\Big)\Big(\int_0^T W(t')\psi_d(t')\mathsf{d}t'\Big)\Big] \\
&= \int_0^T \int_0^T \underbrace{\mathrm{E}\big[W(t)W(t')\big]}_{R_W(t'-t)} \psi_d(t)\psi_d(t')\mathsf{d}t\mathsf{d}t' \\
&= \frac{N_0}{2}\int_0^T \underbrace{\Big[\int_0^T \delta(t-t')\psi_d(t')\mathsf{d}t'\Big]}_{\delta(t)*\psi_d(t) = \psi_d(t)} \psi_d(t)\mathsf{d}t \\
&= \frac{N_0}{2}\int_0^T \psi_d^2(t)\mathsf{d}t = \frac{N_0}{2}.
\end{aligned}$$

Distribution of the noise vector components

Thus we have $\mu_d = 0$ and $\sigma_d^2 = N_0/2$, and the distributions read

$$p_{W_d}(w_d) = \frac{1}{\sqrt{\pi N_0}}\exp\Big(-\frac{w_d^2}{N_0}\Big).$$

for $d = 1, 2, \ldots, D$. Notice that the components W_d are all identically distributed.

As the Gaussian distribution is very important, the **shorthand notation**

$$A \sim \mathcal{N}(\mu, \sigma^2)$$

is often used. This notation means: The random variable A is Gaussian distributed with mean value μ and variance σ^2.

Using this notation, we have for the components of the noise vector

$$W_d \sim \mathcal{N}(0, N_0/2),$$

$d = 1, 2, \ldots, D$.

Distribution of the noise vector

Using the same reasoning as before, we can conclude that $\boldsymbol{W}$ is a Gaussian noise vector, i.e.,

$$\boldsymbol{W} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where the expectation $\boldsymbol{\mu} \in \mathbb{R}^{D\times 1}$ and the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$ are given by

$$\boldsymbol{\mu} = \mathrm{E}\big[\boldsymbol{W}\big], \qquad \boldsymbol{\Sigma} = \mathrm{E}\big[(\boldsymbol{W} - \boldsymbol{\mu})(\boldsymbol{W} - \boldsymbol{\mu})^{\mathsf{T}}\big].$$

From the previous considerations, we have immediately

$$\boldsymbol{\mu} = \mathbf{0}.$$

Using a similar approach as for computing the variances, the statistical independence of the components of $\boldsymbol{W}$ can be shown. That is,

$$\mathrm{E}\big[W_i W_j\big] = \frac{N_0}{2}\delta_{i,j},$$

where $\delta_{i,j}$ denotes the Kronecker symbol defined as

$$\delta_{i,j} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

As a consequence, the covariance matrix of $\boldsymbol{W}$ is

$$\boldsymbol{\Sigma} = \frac{N_0}{2} \boldsymbol{I}_D,$$

where $\boldsymbol{I}_D$ denotes the $D \times D$ identity matrix.

To sum up, the noise vector $\boldsymbol{W}$ is distributed as

$$\boldsymbol{W} \sim \mathcal{N}\Big(\mathbf{0}, \frac{N_0}{2} \boldsymbol{I}_D\Big).$$

As the components of $\boldsymbol{W}$ are statistically independent, the probability density function of $\boldsymbol{W}$ can be written as

$$\begin{aligned}
p_{\boldsymbol{W}}(\boldsymbol{w}) &= \prod_{d=1}^{D} p_{W_d}(w_d) \\
&= \prod_{d=1}^{D} \frac{1}{\sqrt{\pi N_0}} \exp\big(-\frac{w_d^2}{N_0}\big) \\
&= \Big(\frac{1}{\sqrt{\pi N_0}}\Big)^D \exp\big(-\frac{1}{N_0} \sum_{d=1}^{D} w_d^2\big) \\
&= \big(\pi N_0\big)^{-D/2} \exp\big(-\frac{1}{N_0} \|\boldsymbol{w}\|^2\big)
\end{aligned}$$

with $\|\boldsymbol{w}\|$ denoting the Euclidean norm

$$\|\boldsymbol{w}\| = \Big(\sum_{d=1}^{D} w_d^2\Big)^{1/2}$$

of the D-dimensional vector $\boldsymbol{w} = [w_1, \ldots, w_D]^\mathsf{T}$.

The independency of the components of $\boldsymbol{W}$ give rise to the AWGN vector channel.

1.6 The AWGN Vector Channel

The additive white Gaussian noise vector channel is defined as a channel with input vector

$$\boldsymbol{X} = [X_1, X_2, \ldots, X_D]^\mathsf{T},$$

output vector

$$\boldsymbol{Y} = [Y_1, Y_2, \ldots, Y_D]^\mathsf{T},$$

and the relation

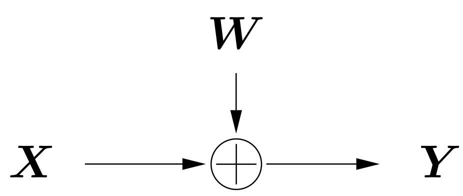
$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{W},$$

where

$$\boldsymbol{W} = [W_1, W_2, \ldots, W_D]^\mathsf{T}$$

is a vector of D independent random variables distributed as

$$W_d \sim \mathcal{N}(0, N_0/2).$$



(As nobody would have been expected differently…)

1.7 Vector Representation of a Digital Communication System

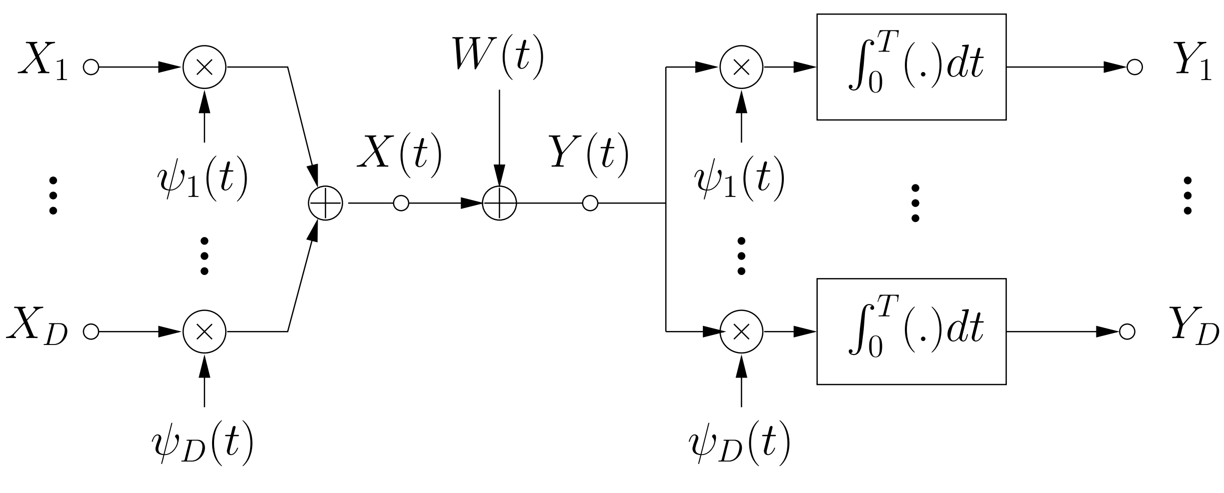
Consider a digital communication system **transmitting over an AWGN channel**. Let $s_1(t), s_2(t), \ldots, s_M(t)$, $s_m(t) = 0$ for $t \notin [0, T]$, $m = 1, 2, \ldots, M$, denote the waveforms. Let further

$$\mathbb{S}_\psi = \left\{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\right\}$$

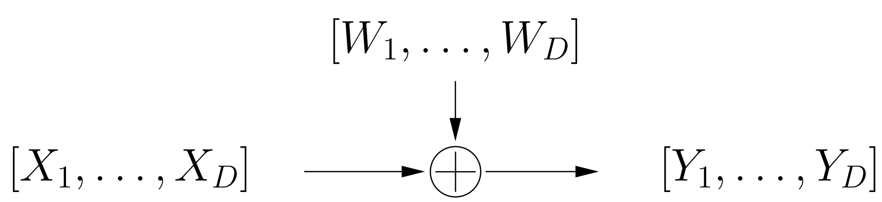
denote a set of orthonormal functions for these waveforms.

The canonical decompositions of the digital transmitter and the digital receiver convert the block formed by the waveform modulator, the AWGN channel, and the waveform demodulator into an AWGN vector channel.

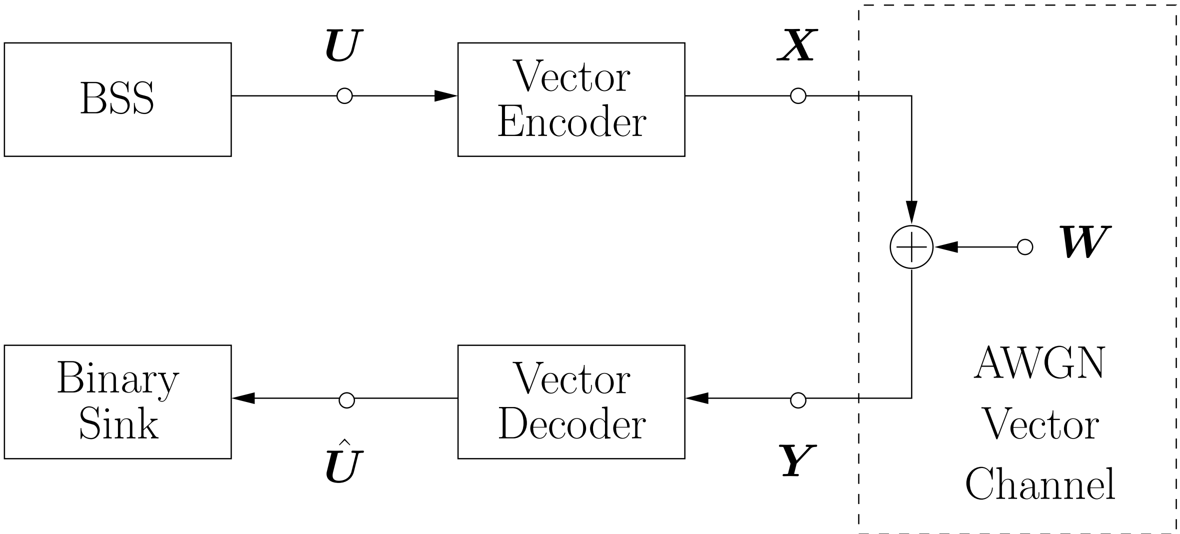
Modulator, AWGN channel, demodulator



AWGN vector channel



Using this equivalence, the communication system operating over a (waveform) AWGN channel can be represented as a communication system operating over an AWGN vector channel:



- Source vector: $\boldsymbol{U} = [U_1, \ldots, U_K]$, $U_i \in \{0, 1\}$
- Transmit vector: $\boldsymbol{X} = [X_1, \ldots, X_D]$
- Receive vector: $\boldsymbol{Y} = [Y_1, \ldots, Y_D]$
- Estimated source vector: $\hat{\boldsymbol{U}} = [\hat{U}_1, \ldots, \hat{U}_K]$, $U_i \in \{0, 1\}$
- AWGN vector: $\boldsymbol{W} = [W_1, \ldots, W_D]$,

$$\boldsymbol{W} \sim \mathcal{N}\Big(\boldsymbol{0}, \frac{N_0}{2}\boldsymbol{I}_D\Big)$$

This is the vector representation of a digital communication system transmitting over an AWGN channel.

For estimating the transmitted vector $\boldsymbol{x}$, we are interested in the **likelihood** of $\boldsymbol{x}$ for the given received vector $\boldsymbol{y}$:

$$\begin{aligned} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) &= p_{\boldsymbol{W}}(\boldsymbol{y}-\boldsymbol{x}) \\ &= (\pi N_0)^{-D/2} \exp\bigl(-\frac{1}{N_0}\|\boldsymbol{y}-\boldsymbol{x}\|^2\bigr). \end{aligned}$$

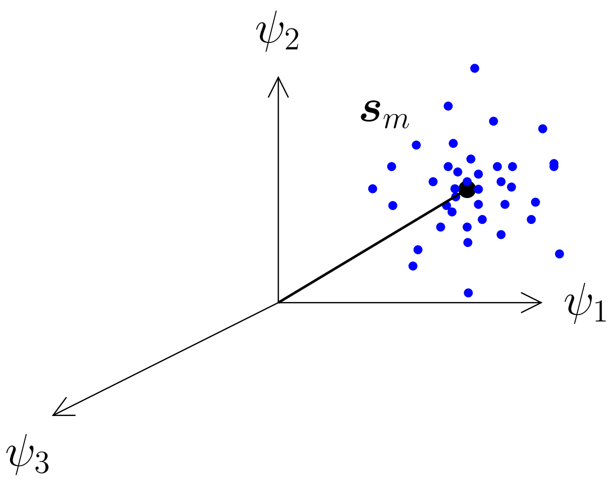
Symbolically, we may write

$$\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{s}_m \quad \sim \quad \mathcal{N}\Bigl(\boldsymbol{s}_m, \frac{N_0}{2}\boldsymbol{I}_D\Bigr).$$

Remark:
The likelihood of $\boldsymbol{x}$ is the conditional probability density function of $\boldsymbol{y}$ given $\boldsymbol{x}$, and it is regarded as a function of $\boldsymbol{x}$.

Example: Signal buried in AWGN

Consider $D = 3$ and $\boldsymbol{x} = \boldsymbol{s}_m$. Then the "Gaussian cloud" around $\boldsymbol{s}_m$ may look like this:



◇

1.8 Signal-to-Noise Ratio for the AWGN Channel

Consider a digital communication system using the set of waveforms

$$\mathbb{S} = \Big\{ s_1(t), s_2(t), \ldots, s_M(t) \Big\},$$

$s_m(t) = 0$ for $t \notin [0, T]$, $m = 1, 2, \ldots, M$, for transmission over an AWGN channel with power spectrum

$$S_W(f) = \frac{N_0}{2}.$$

Let E_{s_m} denote the energy of waveform $s_m(t)$, and let $\overline{E}_s$ denote the average waveform energy:

$$E_{s_m} = \int_0^T \big(s_m(t)\big)^2 \mathsf{d}t, \qquad \overline{E}_s = \frac{1}{M} \sum_{m=1}^{M} E_{s_m}$$

Thus, $\overline{E}_s$ is the average **energy per waveform** $x(t)$ and also the average energy per vector $\boldsymbol{x}$ (due to the one-to-one correspondence.)

The average **signal-to-noise ratio (SNR) per waveform** is then defined as

$$\gamma_s = \frac{\overline{E}_s}{N_0}.$$

On the other hand, the average energy per source bit u_k is of interest. Since K source bits u_k are transmitted per waveform, the average **energy per bit** is

$$\overline{E}_b = \frac{1}{K}\overline{E}_s.$$

Accordingly, the average **signal-to-noise ratio (SNR) per bit** is then defined as

$$\gamma_b = \frac{\overline{E}_b}{N_0}.$$

As K bits are transmitted per waveform, we have the relation

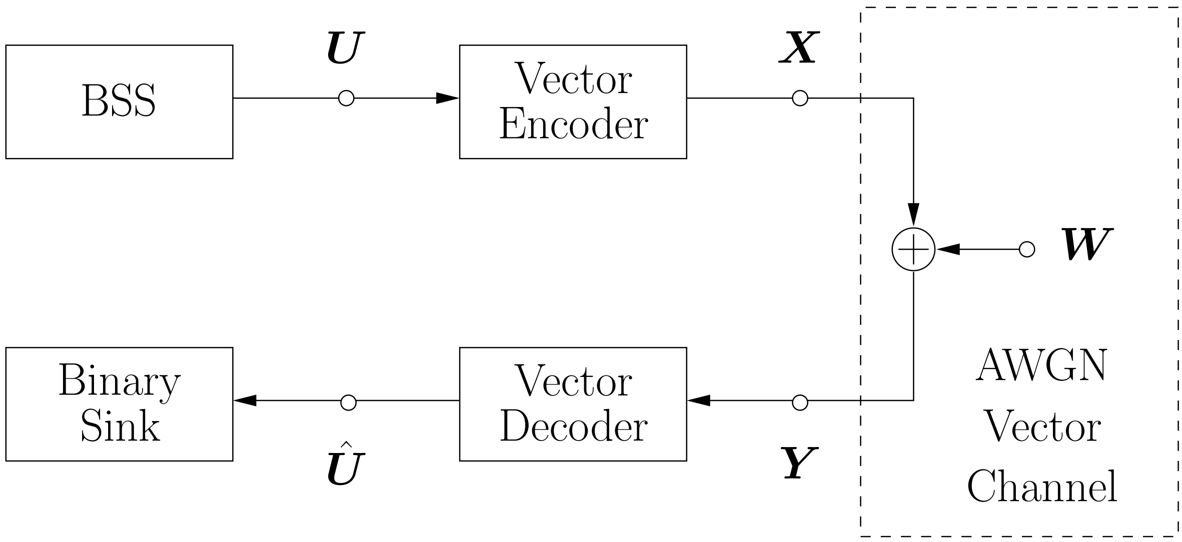
$$\gamma_b = \frac{1}{K}\gamma_s.$$

Remark:
The energies and signal-to-noise ratios usually denote *received* values.

2 Optimum Decoding for the AWGN Channel

2.1 Optimality Criterion

In the previous chapter, the vector representation of a digital transceiver transmitting over an AWGN channel was derived.



Nomenclature

$$
\begin{aligned}
\boldsymbol{u} &= [u_1, \dots, u_K] \in \mathbb{U}, & \boldsymbol{x} &= [x_1, \dots, x_D] \in \mathbb{X}, \\
\hat{\boldsymbol{u}} &= [\hat{u}_1, \dots, \hat{u}_K] \in \mathbb{U}, & \boldsymbol{y} &= [y_1, \dots, y_D] \in \mathbb{R}^D, \\
\mathbb{U} &= \{0, 1\}^K, & \mathbb{X} &= \{\boldsymbol{s}_1, \dots, \boldsymbol{s}_M\}.
\end{aligned}
$$

$\mathbb{U}$: set of all binary sequences of length K,
$\mathbb{X}$: set of vector representations of the waveforms.

The set $\mathbb{X}$ is called the **signal constellation**, and its elements are called signals, signal points, or **modulation symbols**.

For convenience, we may refer to u_k as bits and to $\boldsymbol{x}$ as symbols. (But we have to be careful not to cause ambiguity.)

As the vector representation implies no loss of information, recovering the transmitted block $\boldsymbol{u} = [u_1, \ldots, u_K]$ from the received waveform $y(t)$ is equivalent to recovering $\boldsymbol{u}$ from the vector $\boldsymbol{y} = [y_1, \ldots, y_D]$.

Functionality of the vector encoder

$$\begin{aligned} \mathrm{enc}: \quad \boldsymbol{u} &\mapsto \boldsymbol{x} = \mathrm{enc}(\boldsymbol{u}) \\ \mathbb{U} &\rightarrow \mathbb{X} \end{aligned}$$

Functionality of the vector decoder

$$\begin{aligned} \mathrm{dec}_u: \quad \boldsymbol{y} &\mapsto \hat{\boldsymbol{u}} = \mathrm{dec}_u(\boldsymbol{y}) \\ \mathbb{R}^D &\rightarrow \mathbb{U} \end{aligned}$$

Optimality criterion

A vector decoder is called an **optimal vector decoder** if it minimizes the block error probability $\Pr(\boldsymbol{U} \neq \hat{\boldsymbol{U}})$.

2.2 Decision Rules

Since the encoder mapping is one-to-one, the functionality of the vector decoder can be decomposed into the following two operations:

(i) Modulation-symbol decision:

$$\begin{aligned} \mathrm{dec}: \quad \boldsymbol{y} &\mapsto \hat{\boldsymbol{x}} = \mathrm{dec}(\boldsymbol{y}) \\ \mathbb{R}^D &\rightarrow \mathbb{X} \end{aligned}$$

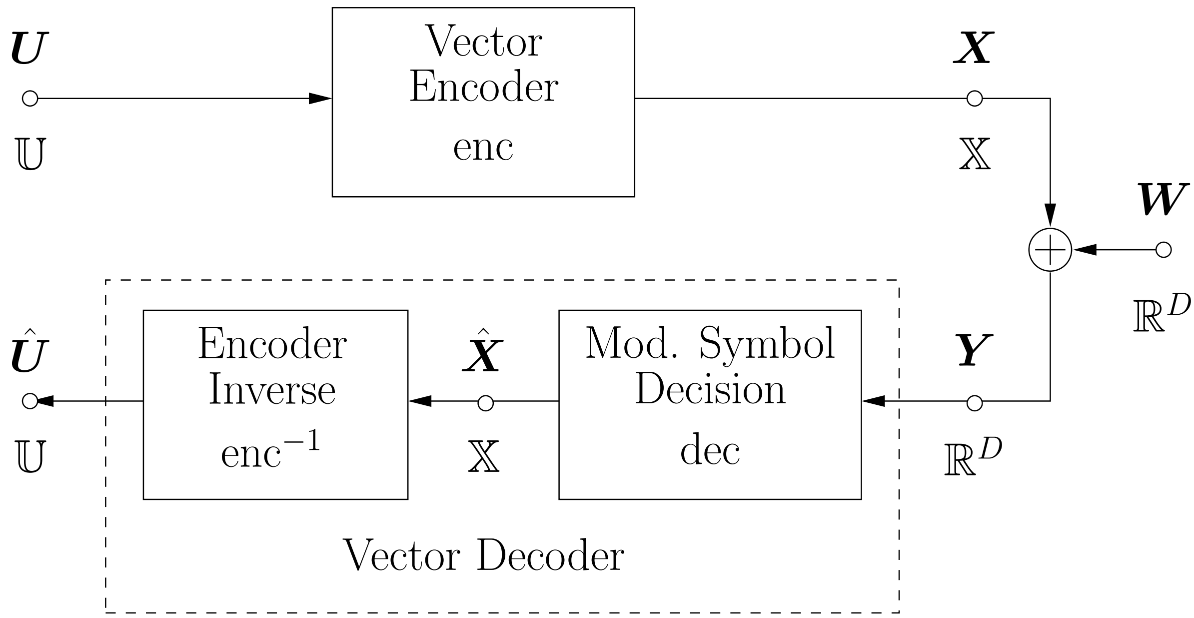
(ii) Encoder inverse:

$$\begin{aligned} \mathrm{enc}^{-1}: \quad \hat{\boldsymbol{x}} &\mapsto \hat{\boldsymbol{u}} = \mathrm{enc}^{-1}(\hat{\boldsymbol{x}}) \\ \mathbb{X} &\rightarrow \mathbb{U} \end{aligned}$$

Thus, we have the following chain of mappings:

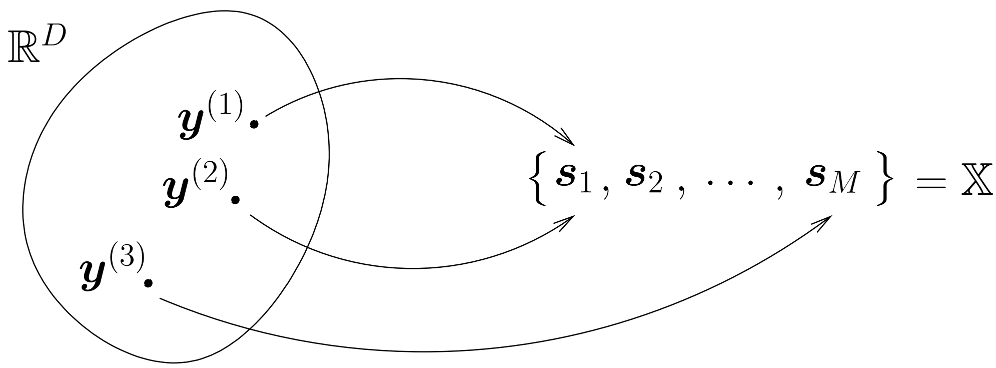
$$\boldsymbol{y} \overset{\text{dec}}{\longmapsto} \hat{\boldsymbol{x}} \overset{\text{enc}^{-1}}{\longmapsto} \hat{\boldsymbol{u}}.$$

Block Diagram



A **decision rule** is a function which maps each observation $\boldsymbol{y} \in \mathbb{R}^D$ to a modulation symbol $\boldsymbol{x} \in \mathbb{X}$, i.e., to an element of the signal constellation $\mathbb{X}$.

EXAMPLE: Decision rule



◇

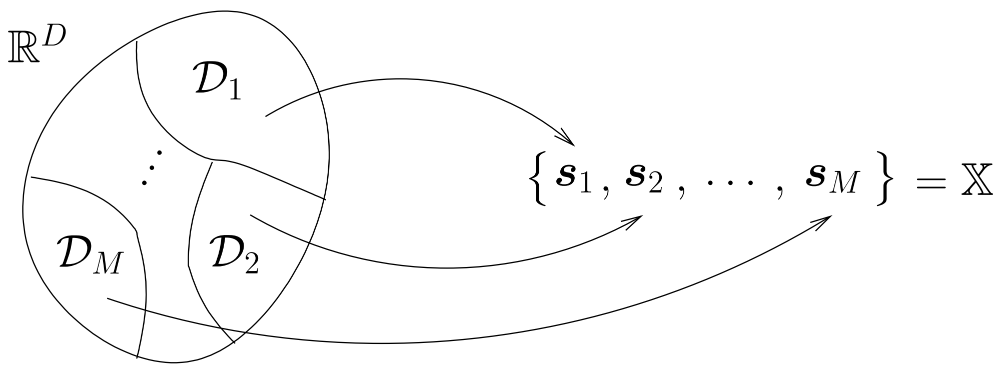
2.3 Decision Regions

A decision rule $\mathrm{dec} : \mathbb{R}^D \to \mathbb{X}$ defines a partition of $\mathbb{R}^D$ into decision regions

$$\mathcal{D}_m := \left\{ \boldsymbol{y} \in \mathbb{R}^D : \mathrm{dec}(\boldsymbol{y}) = \boldsymbol{s}_m \right\},$$

$m = 1, 2, \ldots, M$.

EXAMPLE: Decision regions



◇

Since the decision regions $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M$ form a partition, they have the following properties:

(i) They are disjoint:

$$\mathcal{D}_i \cap \mathcal{D}_j = \emptyset \quad \text{for } i \neq j.$$

(ii) They cover the whole observation space $\mathbb{R}^D$:

$$\mathcal{D}_1 \cup \mathcal{D}_2 \cup \cdots \cup \mathcal{D}_M = \mathbb{R}^D.$$

Obviously, any partition of $\mathbb{R}^D$ defines a decision rule as well.

2.4 Optimum Decision Rules

Optimality criterion for decision rules

Since the mapping enc : $\mathbb{U} \rightarrow \mathbb{X}$ of the vector encoder is one-to-one, we have

$$\hat{\boldsymbol{U}} \neq \boldsymbol{U} \quad \Leftrightarrow \quad \hat{\boldsymbol{X}} \neq \boldsymbol{X}.$$

Therefore, the decision rule of an optimal vector decoder minimizes the modulation-symbol error probability $\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X})$.

Such a decision rule is called optimal.

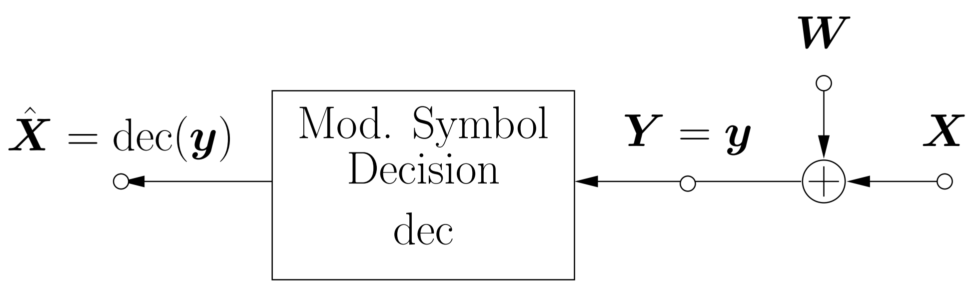
Sufficient condition for an optimal decision rule

Let $\mathrm{dec} : \mathbb{R}^D \rightarrow \mathbb{X}$ be a decision rule which minimizes the a-posteriori probability of making a decision error,

$$\Pr\big(\underbrace{\mathrm{dec}(\boldsymbol{y})}_{\hat{\boldsymbol{x}}} \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big),$$

for any observation $\boldsymbol{y}$. Then dec is optimal.

Remark:
The **a-posteriori probability of a decoding error** is the probability of the event $\{\mathrm{dec}(\boldsymbol{y}) \neq \boldsymbol{X}\}$ conditioned on the event $\{\boldsymbol{Y} = \boldsymbol{y}\}$:



Proof:

Let $\mathrm{dec}: \mathbb{R}^D \to \mathbb{X}$ denote any decision rule, and let $\mathrm{dec}_{\mathsf{opt}}: \mathbb{R}^D \to \mathbb{X}$ denote an optimal decision rule. Then we have

$$\begin{aligned}
\Pr\big(\mathrm{dec}(\boldsymbol{Y}) \neq \boldsymbol{X}\big) &= \int_{\mathbb{R}^D} \Pr\big(\mathrm{dec}(\boldsymbol{Y}) \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big) \cdot p_{\boldsymbol{Y}}(\boldsymbol{y}) \mathsf{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^D} \Pr\big(\mathrm{dec}(\boldsymbol{y}) \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big) \cdot p_{\boldsymbol{Y}}(\boldsymbol{y}) \mathsf{d}\boldsymbol{y} \\
&\overset{(\star)}{\geq} \int_{\mathbb{R}^D} \Pr\big(\mathrm{dec}_{\mathsf{opt}}(\boldsymbol{y}) \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big) \cdot p_{\boldsymbol{Y}}(\boldsymbol{y}) \mathsf{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^D} \Pr\big(\mathrm{dec}_{\mathsf{opt}}(\boldsymbol{Y}) \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big) \cdot p_{\boldsymbol{Y}}(\boldsymbol{y}) \mathsf{d}\boldsymbol{y} \\
&= \Pr\big(\mathrm{dec}_{\mathsf{opt}}(\boldsymbol{Y}) \neq \boldsymbol{X}\big).
\end{aligned}$$

$(\star)$: by definition of $\mathrm{dec}_{\mathsf{opt}}$, we have

$$\Pr\big(\mathrm{dec}(\boldsymbol{y}) \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big) \geq \Pr\big(\mathrm{dec}_{\mathsf{opt}}(\boldsymbol{y}) \neq \boldsymbol{X} \big| \boldsymbol{Y} = \boldsymbol{y}\big)$$

for all $\boldsymbol{x} \in \mathbb{X}$.

2.5 MAP Symbol Decision Rule

Starting with the necessary condition for an optimal decision rule given above, we derive now a method to make an optimal decision. The corresponding decision rule is called the maximum a-posteriori (MAP) decision rule.

Consider an optimal decision rule

$$\mathrm{dec} : \boldsymbol{y} \mapsto \hat{\boldsymbol{x}} = \mathrm{dec}(\boldsymbol{y}),$$

where the estimated modulation-symbol is denoted by $\hat{\boldsymbol{x}}$.

As the decision rule is assumed to be optimum, the probability

$$\Pr(\hat{\boldsymbol{x}} \neq \boldsymbol{X} | \boldsymbol{Y} = \boldsymbol{y})$$

is minimal.

For any $\boldsymbol{x} \in \mathbb{X}$, we have

$$\begin{aligned}\Pr(\boldsymbol{X} \neq \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}) &= 1 - \Pr(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}) \\ &= 1 - p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \\ &= 1 - \frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\, p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{Y}}(\boldsymbol{y})},\end{aligned}$$

where Bayes' rule has been used in the last line. Notice that

$$\Pr(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}) = p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$$

is the a-posteriori probability of $\{X = x\}$ given $\{Y = y\}$.

Using the above equalities, we can conclude that

$$\begin{array}{rrl} & \Pr(\boldsymbol{X} \neq \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y}) & \text{is minimal} \\ \Leftrightarrow & p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) & \text{is maximal} \\ \Leftrightarrow & p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\, p_{\boldsymbol{X}}(\boldsymbol{x}) & \text{is maximal,} \end{array}$$

since $p_{\boldsymbol{Y}}(\boldsymbol{y})$ does not depend on $\boldsymbol{x}$.

Thus, the estimated modulation-symbol $\hat{\boldsymbol{x}} = \mathrm{dec}(\boldsymbol{y})$ resulting from the optimal decision rule satisfies the two equivalent conditions

$$\begin{array}{rcll} p_{\boldsymbol{X}|\boldsymbol{Y}}(\hat{\boldsymbol{x}}|\boldsymbol{y}) & \geq & p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) & \text{for all } \boldsymbol{x} \in \mathbb{X}; \\ p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\hat{\boldsymbol{x}})\, p_{\boldsymbol{X}}(\hat{\boldsymbol{x}}) & \geq & p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\, p_{\boldsymbol{X}}(\boldsymbol{x}) & \text{for all } \boldsymbol{x} \in \mathbb{X}. \end{array}$$

These two conditions may be used to define the optimal decision rule.

Maximum a-posteriori (MAP) decision rule:

$$\begin{aligned} \hat{\boldsymbol{x}} = \mathrm{dec}_{\mathsf{MAP}}(\boldsymbol{y}) &:= \underset{\boldsymbol{x} \in \mathbb{X}}{\operatorname{argmax}} \Big\{ p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \Big\} \\ &:= \underset{\boldsymbol{x} \in \mathbb{X}}{\operatorname{argmax}} \Big\{ p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})\, p_{\boldsymbol{X}}(\boldsymbol{x}) \Big\}. \end{aligned}$$

The decision rule is called the MAP decision rule as the selected modulation-symbol $\hat{\boldsymbol{x}}$ maximizes the a-posteriori probability density function $p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$. Notice that the two formulations are equivalent.

2.6 ML Symbol Decision Rule

If the modulation-symbols at the output of the vector encoder are equiprobable, i.e., if

$$\Pr(\boldsymbol{X} = \boldsymbol{s}_m) = \frac{1}{M}$$

for $m = 1, 2, \ldots, M$, the MAP decision rule reduces to the following decision rule.

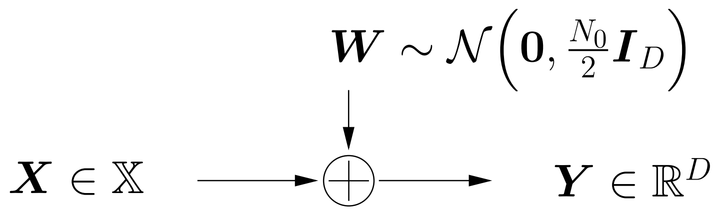
Maximum likelihood (ML) decision rule:

$$\hat{\boldsymbol{x}} = \mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) := \operatorname*{argmax}_{\boldsymbol{x} \in \mathbb{X}} \Big\{ p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \Big\}.$$

In this case, the the selected modulation-symbol $\hat{\boldsymbol{x}}$ maximizes the likelihood function $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ of $\boldsymbol{x}$.

2.7 ML Decision for the AWGN Channel

Using the canonical decomposition of the transmitter and of the receiver, we obtain the AWGN vector-channel:



The conditional probability density function of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x} \in \mathbb{X}$ is

$$p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \big(\pi N_0\big)^{-D/2} \exp\Big(-\frac{1}{N_0}\|\boldsymbol{y}-\boldsymbol{x}\|^2\Big).$$

Since exp(.) is a monotonically increasing function, we obtain the following formulation of the **ML decision rule**

$$\begin{aligned}
\hat{\boldsymbol{x}} = \mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) &= \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{X}} \Big\{ p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) \Big\} \\
&= \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{X}} \Big\{ \exp\Big(-\frac{1}{N_0}\|\boldsymbol{y}-\boldsymbol{x}\|^2\Big) \Big\} \\
&= \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{X}} \Big\{ \|\boldsymbol{y}-\boldsymbol{x}\|^2 \Big\}.
\end{aligned}$$

Thus, the ML decision rule for the AWGN vector channel selects the vector in the signal constellation $\mathbb{X}$ that is the closest one to the observation $\boldsymbol{y}$, where the distance measure is the squared Euclidean distance.

Implementation of the ML decision rule

The squared Euclidean distance can be expanded as

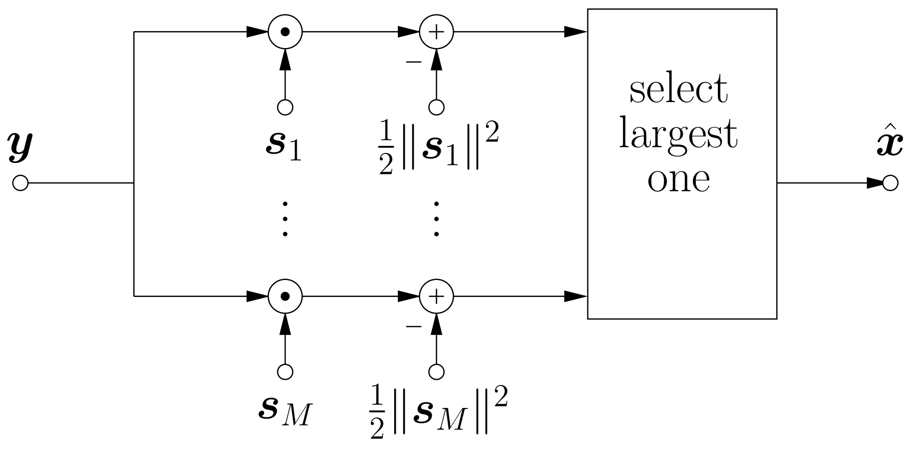
$$\|\boldsymbol{y}-\boldsymbol{x}\|^2 = \|\boldsymbol{y}\|^2 - 2\langle \boldsymbol{y},\boldsymbol{x}\rangle + \|\boldsymbol{x}\|^2.$$

The term $\|\boldsymbol{y}\|^2$ is irrelevant for the minimization with respect to $\boldsymbol{x}$.

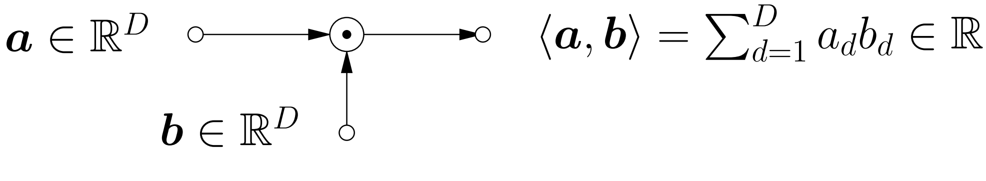
Thus we get the following equivalent formulation of the ML decision rule:

$$\begin{aligned}\hat{\boldsymbol{x}} = \mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) &= \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{X}} \Big\{ -2\langle \boldsymbol{y},\boldsymbol{x}\rangle + \|\boldsymbol{x}\|^2 \Big\} \\ &= \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{X}} \Big\{ 2\langle \boldsymbol{y},\boldsymbol{x}\rangle - \|\boldsymbol{x}\|^2 \Big\} \\ &= \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{X}} \Big\{ \langle \boldsymbol{y},\boldsymbol{x}\rangle - \tfrac{1}{2}\|\boldsymbol{x}\|^2 \Big\}.\end{aligned}$$

Block diagram of the ML decision rule for the AWGN vector-channel:



Vector correlator:



Remark:
Some signal constellations $\mathbb{X}$ have the property that $\|\boldsymbol{x}\|^2$ has the same value for all $\boldsymbol{x} \in \mathbb{X}$. In these cases (but *only* then), the ML decision rule simplifies further to

$$\hat{\boldsymbol{x}} = \text{dec}_{\mathsf{ML}}(\boldsymbol{y}) = \underset{\boldsymbol{x}\in\mathbb{X}}{\text{argmax}} \Big\{ \langle \boldsymbol{y}, \boldsymbol{x} \rangle \Big\}.$$

2.8 Symbol-Error Probability

The error probability for modulation-symbols $\boldsymbol{X}$, or for short, the symbol-error probability is defined as

$$P_s = \Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}) = \Pr(\text{dec}(\boldsymbol{Y}) \neq \boldsymbol{X}).$$

The symbol-error probability coincides with the block error probability

$$P_s = \Pr(\hat{\boldsymbol{U}} \neq \boldsymbol{U}) = \Pr\Big([\hat{U}_1, \ldots, \hat{U}_K] \neq [U_1, \ldots, U_K]\Big).$$

We compute P_s in two steps:

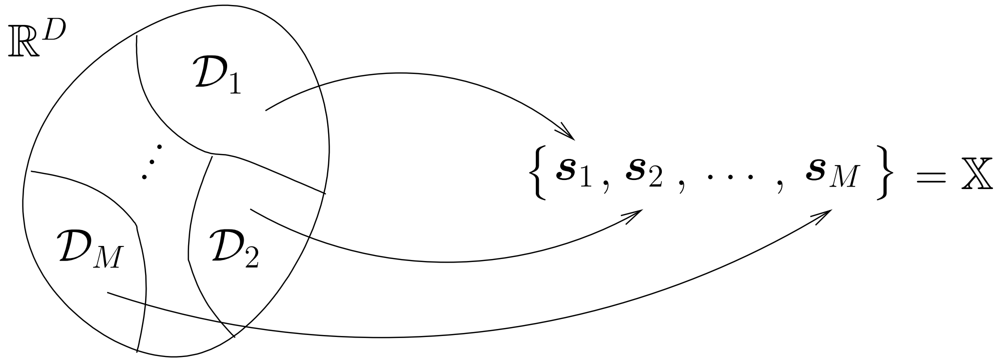
(i) Compute the conditional symbol-error probability given $\boldsymbol{X} = \boldsymbol{x}$ for any $\boldsymbol{x} \in \mathbb{X}$:

$$\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{x}).$$

(ii) Average over the signal constellation:

$$\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}) = \sum_{\boldsymbol{x}\in\mathbb{X}} \Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{x}) \cdot \Pr(\boldsymbol{X} = \boldsymbol{x}).$$

Remember: Decision regions for the symbol decision rule:



Assume that $\boldsymbol{x} = \boldsymbol{s}_m$ is transmitted. Then,

$$\begin{aligned}\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_m) &= \Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{s}_m | \boldsymbol{X} = \boldsymbol{s}_m) \\ &= \Pr(\boldsymbol{Y} \notin \mathcal{D}_m | \boldsymbol{X} = \boldsymbol{s}_m) \\ &= 1 - \Pr(\boldsymbol{Y} \in \mathcal{D}_m | \boldsymbol{X} = \boldsymbol{s}_m).\end{aligned}$$

For the AWGN vector channel, the conditional probability density function of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{s}_m$ is

$$p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_m) = \big(\pi N_0\big)^{-D/2} \exp\Big(-\frac{1}{N_0}\|\boldsymbol{y} - \boldsymbol{s}_m\|^2\Big).$$

Hence,

$$\begin{aligned}\Pr(\boldsymbol{Y} \in \mathcal{D}_m | \boldsymbol{X} = \boldsymbol{s}_m) &= \int_{\mathcal{D}_m} p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_m) \mathsf{d}\boldsymbol{y} = \\ &= \big(\pi N_0\big)^{-D/2} \int_{\mathcal{D}_m} \exp\Big(-\frac{1}{N_0}\|\boldsymbol{y} - \boldsymbol{s}_m\|^2\Big) \mathsf{d}\boldsymbol{y}.\end{aligned}$$

The **symbol-error probability** is then obtained by averaging over the signal constellation:

$$\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}) = 1 - \sum_{m=1}^{M} \Pr(\boldsymbol{Y} \in \mathcal{D}_m | \boldsymbol{X} = \boldsymbol{s}_m) \cdot \Pr(\boldsymbol{X} = \boldsymbol{s}_m).$$

If the modulation symbols are equiprobable, the formula simplifies to

$$\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}) = 1 - \frac{1}{M} \sum_{m=1}^{M} \Pr(\boldsymbol{Y} \in \mathcal{D}_m | \boldsymbol{X} = \boldsymbol{s}_m).$$

2.9 Bit-Error Probability

The error probability for the source bits U_k, or for short, the bit-error probability is defined as

$$P_b = \frac{1}{K} \sum_{k=1}^{K} \Pr(\hat{U}_k \neq U_k),$$

i.e., as the average number of erroneous bits in the block $\hat{\boldsymbol{U}} = [\hat{U}_1, \ldots, \hat{U}_K]$.

Relation between the bit-error probability and the symbol-error probability

Clearly, the relation between P_b and P_s depends on the encoding function

$$\text{enc}: \quad \boldsymbol{u} = [u_1, \ldots, u_k] \mapsto \boldsymbol{x} = \text{enc}(\boldsymbol{u}).$$

The analytical derivation of the relation between P_b and P_s is usually cumbersome. However, we can derive an upper bound and a lower bound for P_b when P_s is given (and also vice versa). These bounds are valid for any encoding function.

The lower bound provides a guideline for the design of a "good" encoding function.

Lower bound:
If a symbol is erroneous, there is at least one erroneous bit in the block of K bits. Therefore

$$P_b \geq \tfrac{1}{K} P_s.$$

Upper bound:
If a symbol is erroneous, there are at most K erroneous bits in the block of K bits. Therefore

$$P_b \leq P_s.$$

Combination:
Combining the lower bound and the upper bound yields

$$\tfrac{1}{K} P_s \leq P_b \leq P_s.$$

An efficient encoder should achieve the lower bound for the bit-error probability, i.e.,

$$P_b \approx \tfrac{1}{K} P_s.$$

Remark:
The bounds can also be obtained by considering the union bound of the events

$$\{\hat{U}_1 \neq U_1\}, \{\hat{U}_2 \neq U_2\}, \ldots, \{\hat{U}_K \neq U_K\},$$

i.e, by relating the probability

$$\Pr\Big(\{\hat{U}_1 \neq U_1\} \cup \{\hat{U}_2 \neq U_2\} \cup \ldots \cup \{\hat{U}_K \neq U_K\}\Big)$$

to the probabilities

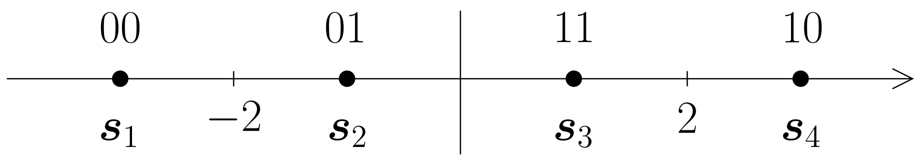
$$\Pr\big(\{\hat{U}_1 \neq U_1\}\big), \Pr\big(\{\hat{U}_2 \neq U_2\}\big), \ldots, \Pr\big(\{\hat{U}_K \neq U_K\}\big).$$

2.10 Gray Encoding

In Gray encoding, blocks of bits, $\boldsymbol{u} = [u_1, \ldots, u_K]$, which differ in only one bit u_k are assigned to neighboring vectors (modulation symbols) $\boldsymbol{x}$ in the signal constellation $\mathbb{X}$.

EXAMPLE: 4PAM (Pulse Amplitude Modulation)

Signal constellation: $\mathbb{X} = \{[-3], [-1], [+1], [+3]\}$.



◇

Motivation

When a decision error occurs, it is very likely that the transmitted symbol and the detected symbol are neighbors. When the bit blocks of neighboring symbols differ only in one bit, there will be only one bit error per symbol error.

Effect

Gray encoding achieves

$$P_b \approx \tfrac{1}{K} P_s$$

for high signal-to-noise ratios (SNR).

Sketch of the proof:

We consider the case $K = 2$, i.e., $\boldsymbol{U} = [U_1, U_2]$.

$$
\begin{aligned}
P_s &= \Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}) \\
&= \Pr\Big([\hat{U}_1, \hat{U}_2] \neq [U_1, U_2]\Big) \\
&= \Pr\Big(\{\hat{U}_1 \neq U_1\} \cup \{\hat{U}_2 \neq U_2\}\Big) \\
&= \underbrace{\Pr(\hat{U}_1 \neq U_1) + \Pr(\hat{U}_2 \neq U_2)}_{2 \cdot P_b} \\
&\qquad - \underbrace{\Pr\Big(\{\hat{U}_1 \neq U_1\} \cap \{\hat{U}_2 \neq U_2\}\Big)}_{(\star)}.
\end{aligned}
$$

The term $(\star)$ is small compared to the other two terms if the SNR is high. Hence

$$P_s \approx 2 \cdot P_b.$$

Part II

Digital Communication Techniques

3 Pulse Amplitude Modulation (PAM)

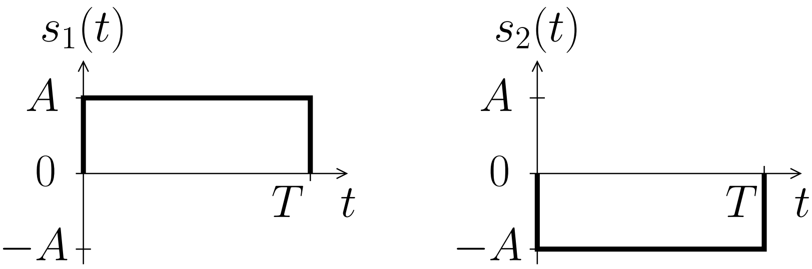
In pulse amplitude modulation (PAM), the information is conveyed by the amplitude of the transmitted waveforms.

3.1 BPAM

PAM with only two waveforms is called binary PAM (BPAM).

3.1.1 Signaling Waveforms

Set of two rectangular waveforms with amplitudes A and $-A$ over $t \in [0, T]$, $\mathbb{S} := \{s_1(t), s_2(t)\}$:



As $|\mathbb{S}| = 2$, one bit is sufficient to address the waveforms ($K = 1$). Thus, the BPAM transmitter performs the mapping

$$U \;\mapsto\; x(t)$$

with

$$U \in \mathbb{U} = \{0, 1\} \qquad \text{and} \qquad x(t) \in \mathbb{S}.$$

Remark: In the general case, the basic waveform may also have another shape.

3.1.2 Decomposition of the Transmitter

Gram-Schmidt procedure

$$\psi_1(t) = \frac{1}{\sqrt{E_1}} s_1(t)$$

with

$$E_1 = \int_0^T \big(s_1(t)\big)^2 \mathsf{d}t = \int_0^T A^2 \mathsf{d}t = A^2 \cdot T.$$

Thus

$$\psi_1(t) = \begin{cases} \frac{1}{\sqrt{T}} & \text{for } t \in [0, T]\ , \\ 0 & \text{otherwise.} \end{cases}$$

Both $s_1(t)$ and $s_2(t)$ are a linear combination of $\psi(t)$:

$$\begin{aligned} s_1(t) &= +\sqrt{E_s}\ \psi_1(t) = +A\sqrt{T}\ \psi_1(t), \\ s_2(t) &= -\sqrt{E_s}\ \psi_1(t) = -A\sqrt{T}\ \psi_1(t), \end{aligned}$$

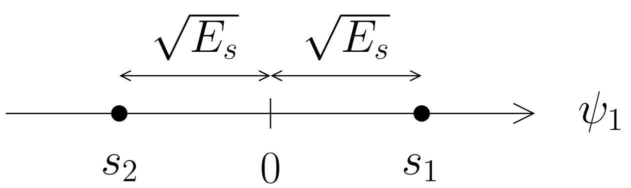
where $\sqrt{E_s} = A\sqrt{T}$ and $E_s = E_1 = E_2$. (The two waveforms have the same energy.)

Signal constellation

Vector representation of the two waveforms:

$$\begin{aligned} s_1(t) \quad &\leftrightarrow \quad \boldsymbol{s}_1 = s_1 = +\sqrt{E_s} \\ s_2(t) \quad &\leftrightarrow \quad \boldsymbol{s}_2 = s_2 = -\sqrt{E_s} \end{aligned}$$

Thus the signal constellation is $\mathbb{X} = \big\{-\sqrt{E_s}, +\sqrt{E_s}\big\}$.



Vector Encoder

The vector encoder is usually defined as

$$x = \mathrm{enc}(u) = \begin{cases} +\sqrt{E_s} & \text{for } u = 0, \\ -\sqrt{E_s} & \text{for } u = 1. \end{cases}$$

The signal constellation comprises only two waveforms, and thus per symbol and per bit, the average transmit energies and the SNRs are the same:

$$E_b = E_s, \qquad\qquad \gamma_s = \gamma_b.$$

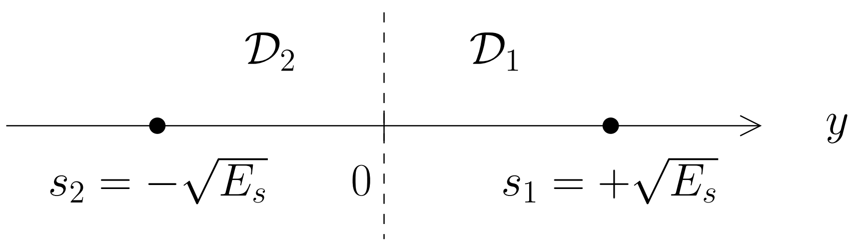
3.1.3 ML Decoding

ML symbol decision rule

$$\mathrm{dec}_{\mathsf{ML}}(y) = \operatorname*{argmin}_{x \in \{-\sqrt{E_s}, +\sqrt{E_s}\}} \|y - x\|^2 = \begin{cases} +\sqrt{E_s} & \text{for } y \geq 0, \\ -\sqrt{E_s} & \text{for } y < 0. \end{cases}$$

Decision regions

$$\mathcal{D}_2 = \{y \in \mathbb{R} : y < 0\}, \qquad \mathcal{D}_1 = \{y \in \mathbb{R} : y \geq 0\}.$$



Remark: The decision rule

$$\mathrm{dec}_{\mathsf{ML}}(y) = \begin{cases} +\sqrt{E_s} & \text{for } y > 0, \\ -\sqrt{E_s} & \text{for } y \leq 0. \end{cases}$$

is also a ML decision rule.

The mapping of the **encoder inverse** reads

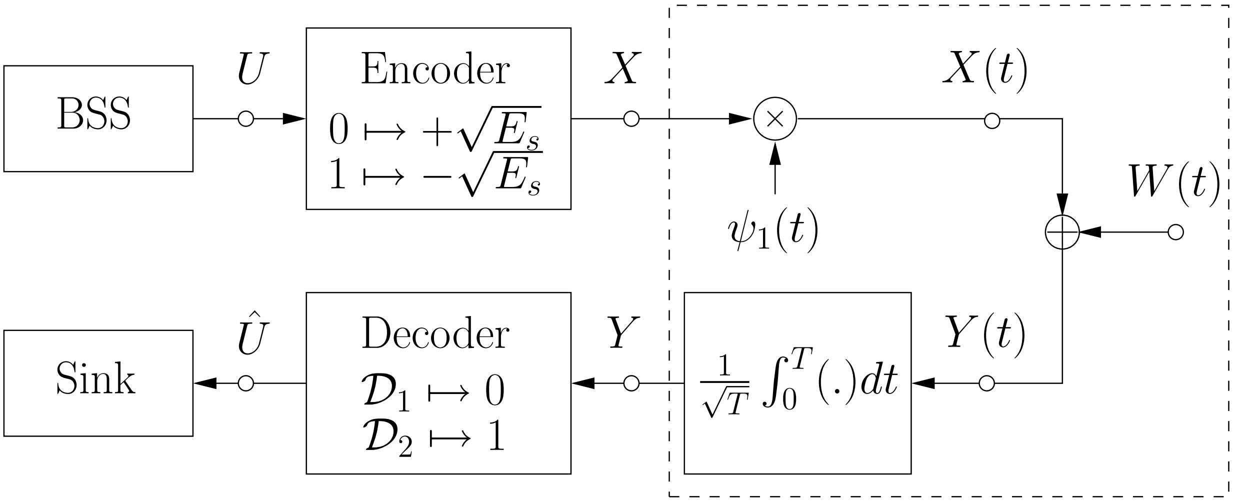
$$\hat{u} = \mathrm{enc}^{-1}(\hat{x}) = \begin{cases} 0 & \text{for } \hat{x} = +\sqrt{E_s}, \\ 1 & \text{for } \hat{x} = -\sqrt{E_s}. \end{cases}$$

Combining the (modulation) symbol decision rule and the encoder inverse yields the **ML decoder for BPAM**:

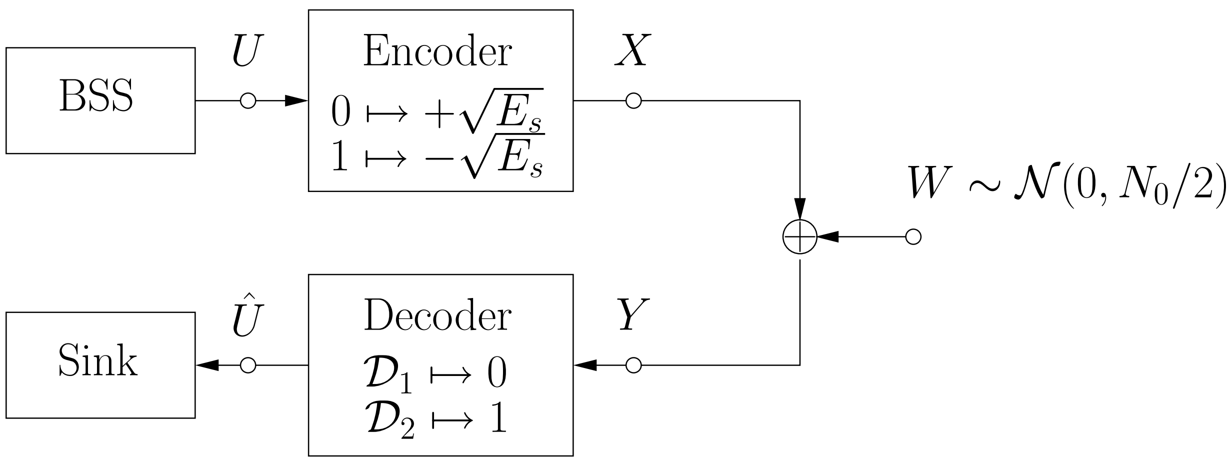
$$\hat{u} = \begin{cases} 0 & \text{for } y \geq 0, \\ 1 & \text{for } y < 0. \end{cases}$$

3.1.4 Vector Representation of the Transceiver

BPAM transceiver:



Vector representation:



3.1.5 Symbol-Error Probability

Conditional probability densities

$$p_{Y|X}(y|+\sqrt{E_s}) = \frac{1}{\sqrt{\pi N_0}} \exp\bigl(-\frac{1}{N_0}|y-\sqrt{E_s}|^2\bigr)$$
$$p_{Y|X}(y|-\sqrt{E_s}) = \frac{1}{\sqrt{\pi N_0}} \exp\bigl(-\frac{1}{N_0}|y+\sqrt{E_s}|^2\bigr)$$

Conditional symbol-error probability

Assume that $X = x = +\sqrt{E_s}$ is transmitted.

$$\begin{aligned}
\Pr(\hat{X} \neq X|X = +\sqrt{E_s}) &= \Pr(Y \in \mathcal{D}_2|X = +\sqrt{E_s}) \\
&= \Pr(Y < 0|X = +\sqrt{E_s}) \\
&= \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^{0} \exp\bigl(-\frac{1}{N_0}|y-\sqrt{E_s}|^2\bigr)\mathsf{d}y \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\sqrt{2E_s/N_0}} \exp\bigl(-\tfrac{1}{2}z^2\bigr)\mathsf{d}z,
\end{aligned}$$

where the last equality results from substituting

$$z = (y-\sqrt{E_s})/\sqrt{N_0/2}.$$

The Q-function is defined as

$$Q(v) := \int_{v}^{\infty} q(z)\mathsf{d}z, \qquad q(z) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\tfrac{1}{2}z^2\bigr).$$

The function $q(z)$ denotes a Gaussian pdf with zero mean and variance 1; i.e., it may be interpreted as the pdf of a random variable $Z \sim \mathcal{N}(0, 1)$.

Using this definition, the conditional symbol-error probabilities may be written as

$$\begin{aligned}
\Pr(\hat{X} \neq X | X = +\sqrt{E_s}) &= Q\Big(\sqrt{\frac{2E_s}{N_0}}\Big), \\
\Pr(\hat{X} \neq X | X = -\sqrt{E_s}) &= Q\Big(\sqrt{\frac{2E_s}{N_0}}\Big).
\end{aligned}$$

The second equation follows from symmetry.

Symbol-error probability

The symbol-error probability results from averaging over the conditional symbol-error probabilities:

$$\begin{aligned}
P_s &= \Pr(\hat{X} \neq X) \\
&= \Pr(X = +\sqrt{E_s}) \cdot \Pr(\hat{X} \neq X | X = +\sqrt{E_s}) \\
&\quad + \Pr(X = -\sqrt{E_s}) \cdot \Pr(\hat{X} \neq X | X = -\sqrt{E_s}) \\
&= Q\Big(\sqrt{\frac{2E_s}{N_0}}\Big) = Q(\sqrt{2\gamma_s})
\end{aligned}$$

as $\Pr(X = +\sqrt{E_s}) = \Pr(X = -\sqrt{E_s}) = 1/2$.

Alternative method to derive the conditional symbol-error probability

Assume again that $X = \sqrt{E_s}$ is transmitted. Instead of considering the conditional pdf of Y, we directly use the pdf of the white Gaussian noise W.

$$\begin{aligned}
\Pr(\hat{X} \neq X | X = +\sqrt{E_s}) &= \Pr(Y < 0 | X = +\sqrt{E_s}) \\
&= \Pr(\sqrt{E_s} + W < 0 | X = +\sqrt{E_s}) \\
&= \Pr(W < -\sqrt{E_s} | X = +\sqrt{E_s}) \\
&= \Pr(W < -\sqrt{E_s}) \\
&= \Pr(W' < -\sqrt{2E_s/N_0}) \\
&= Q(\sqrt{2E_s/N_0}),
\end{aligned}$$

where we substituted $W' = W/\sqrt{N_0/2}$; notice that $W' \sim \mathcal{N}(0,1)$. A plot of the symbol-error probability can found in [1, p. 412].

The initial event (here $\{Y < 0\}$) is transformed into an equivalent event (here $\{W' < -\sqrt{2E_s/N_0}\}$), of which the probability can easily be calculated. This method is frequently applied to determine symbol-error probabilities.

3.1.6 Bit-Error Probability

In BPAM, each waveform conveys one source bit. Accordingly, each symbol-error leads to exactly one bit-error, and thus

$$P_b = \Pr(\hat{U} \neq U) = \Pr(\hat{X} \neq X) = P_s.$$

Due to the same reason, the transmit energy per symbol is equal to the transmit energy per bit, and so also the SNR per symbol is equal to the SNR per bit:

$$E_b = E_s, \qquad\qquad \gamma_s = \gamma_b.$$

Therefore, the bit-error probability (BEP) of BPAM results as

$$P_b = Q(\sqrt{2\gamma_b}).$$

Notice: $P_b \approx 10^{-5}$ is achieved for $\gamma_b \approx 9.6$ dB.

3.2 MPAM

In the general case of M-ary pulse amplitude modulation (MPAM) with $M = 2^K$, input blocks of K bits modulate the amplitude of a unique pulse.

3.2.1 Signaling Waveforms

Waveform encoder:

$$\begin{aligned} \boldsymbol{u} = [u_1, \ldots, u_K] \ &\mapsto \ x(t) \\ \mathbb{U} = \{0,1\}^K \ &\rightarrow \ \mathbb{S} = \big\{s_1(t), \ldots, s_M(t)\big\}, \end{aligned}$$

where

$$s_m(t) = \begin{cases} (2m - M - 1)\, A\, g(t) & \text{for } t \in [0, T], \\ 0 & \text{otherwise,} \end{cases}$$

$A \in \mathbb{R}$, and $g(t)$ is a predefined pulse of duration T,

$$g(t) = 0 \quad \text{for } t \notin [0, T].$$

Values of the factors $(2m - M - 1)A$ that are weighting $g(t)$:

$$-(M-1)A, -(M-3)A, \ldots, -A, A, \ldots, (M-3)A, (M-1)A.$$

Example: 4PAM with rectangular pulse.

Using the rectangular pulse

$$g(t) = \begin{cases} 1 & \text{for } t \in [0, T], \\ 0 & \text{otherwise,} \end{cases}$$

the resulting waveforms are

$$\mathbb{S} = \big\{-3Ag(t), -Ag(t), Ag(t), 3Ag(t)\big\}.$$

◇

3.2.2 Decomposition of the Transmitter

Gram-Schmidt procedure:

Orthonormal function

$$\psi(t) = \frac{1}{\sqrt{E_1}} s_1(t) = \frac{1}{\sqrt{E_g}} g(t),$$

$$E_1 = \int_0^T \big(s_1(t)\big)^2 \mathsf{d}t, \qquad E_g = \int_0^T \big(g(t)\big)^2 \mathsf{d}t.$$

Every waveform in $\mathbb{S}$ is a linear combination of (only) $\psi(t)$:

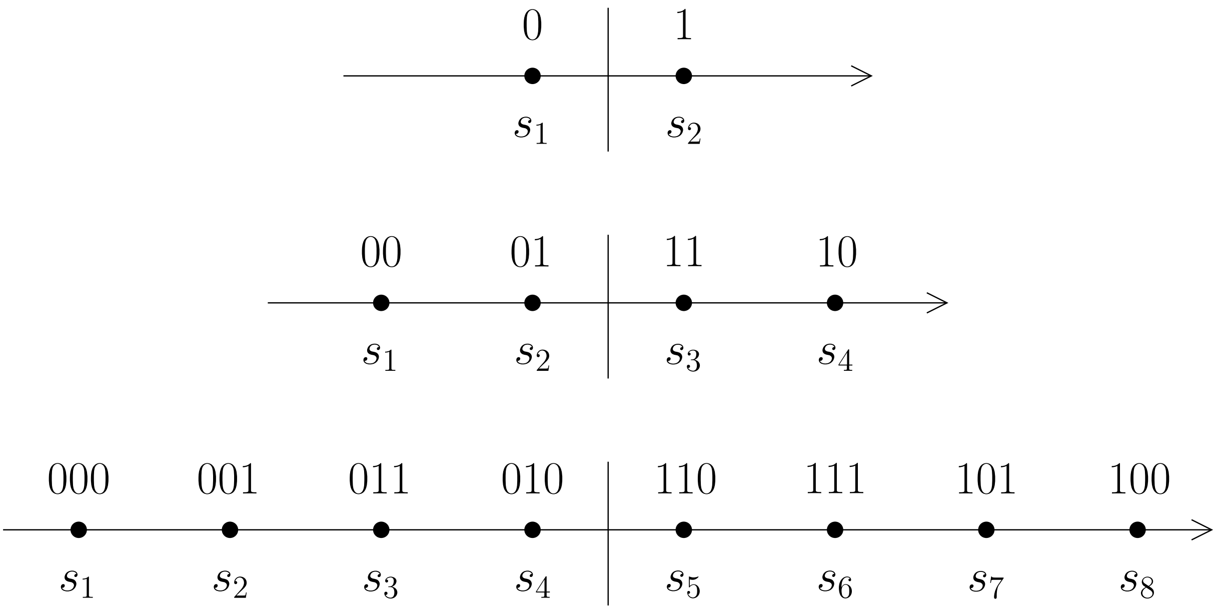
$$\begin{aligned} s_m(t) &= (2m - M - 1)\, \underbrace{A\sqrt{E_g}}_{d}\, \psi(t) \\ &= (2m - M - 1) d\, \psi(t). \end{aligned}$$

Signal constellation

$$\begin{array}{rcl} s_1(t) & \longleftrightarrow & s_1 = -(M-1)d \\ s_2(t) & \longleftrightarrow & s_2 = -(M-3)d \\ & \dots & \\ s_m(t) & \longleftrightarrow & s_m = (2m - M - 1)d \\ & \dots & \\ s_{M-1}(t) & \longleftrightarrow & s_{M-1} = (M-3)d \\ s_M(t) & \longleftrightarrow & s_M = (M-1)d \end{array}$$

$$\begin{aligned} \mathbb{X} &= \big\{s_1, s_2, \dots, s_M\big\} \\ &= \big\{-(M-1)d, -(M-3)d, \dots, -d, d, \dots \\ &\qquad \dots, (M-3)d, (M-1)d\big\}. \end{aligned}$$

Example: $M = 2, 4, 8$



◇

Vector Encoder

The vector encoder

$$\begin{array}{rrcl} \mathrm{enc}: & \boldsymbol{u} = [u_1, \dots, u_K] & \mapsto & x = \mathrm{enc}(\boldsymbol{u}) \\ & \mathbb{U} = \{0,1\}^K & \rightarrow & \mathbb{X} = \big\{-(M-1)d, \dots, (M-1)d\big\} \end{array}$$

is any Gray encoding function.

Energy and SNR

The source bits are generated by a BSS, and so the symbols of the signal constellation $\mathbb{X}$ are equiprobable:

$$\Pr(X = s_m) = \frac{1}{M}$$

for all $m = 1, 2, \dots, M$.

The average symbol energy results as

$$\begin{aligned}\overline{E}_s = \mathrm{E}\big[X^2\big] &= \frac{1}{M}\sum_{m=1}^{M}(s_m)^2 \\ &= \frac{d^2}{M}\sum_{m=1}^{M}(2m - M - 1)^2 = \frac{d^2}{M}\frac{1}{3}M(M^2 - 1)\end{aligned}$$

Thus,

$$\overline{E}_s = \frac{M^2 - 1}{3}d^2 \qquad \Longleftrightarrow \qquad d = \sqrt{\frac{3\overline{E}_s}{M^2 - 1}}$$

As $M = 2^K$, the average energy per transmitted source bit results as

$$\overline{E}_b = \frac{\overline{E}_s}{K} = \frac{M^2 - 1}{3\log_2 M}d^2 \qquad \Longleftrightarrow \qquad d = \sqrt{\frac{3\log_2 M}{M^2 - 1}\overline{E}_b}.$$

The SNR per symbol and the SNR per bit are related as

$$\gamma_b = \frac{1}{K}\gamma_s = \frac{1}{\log_2 M}\gamma_s.$$

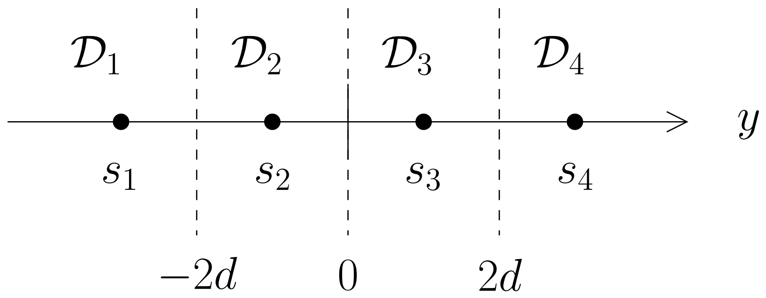
3.2.3 ML Decoding

ML symbol decision rule

$$\mathrm{dec}_{\mathsf{ML}}(y) = \underset{x \in \mathbb{X}}{\operatorname{argmin}} \|y - x\|^2$$

Decision regions

EXAMPLE: Decision regions for $M = 4$.



◇

Starting from the special cases $M = 2, 4, 8$, the decision regions can be inferred:

$$\begin{aligned}
\mathcal{D}_1 &= \{y \in \mathbb{R} : y \leq -(M-2)d\},\\
\mathcal{D}_2 &= \{y \in \mathbb{R} : -(M-2)d < y \leq -(M-4)d\},\\
&\dots\\
\mathcal{D}_m &= \{y \in \mathbb{R} : (2m-M-2)d < y \leq (2m-M)d\},\\
&\dots\\
\mathcal{D}_{M-1} &= \{y \in \mathbb{R} : (M-4)d < y \leq (M-2)d\},\\
\mathcal{D}_M &= \{y \in \mathbb{R} : (M-2)d < y\}.
\end{aligned}$$

Accordingly, the ML symbol decision rule may also be written as

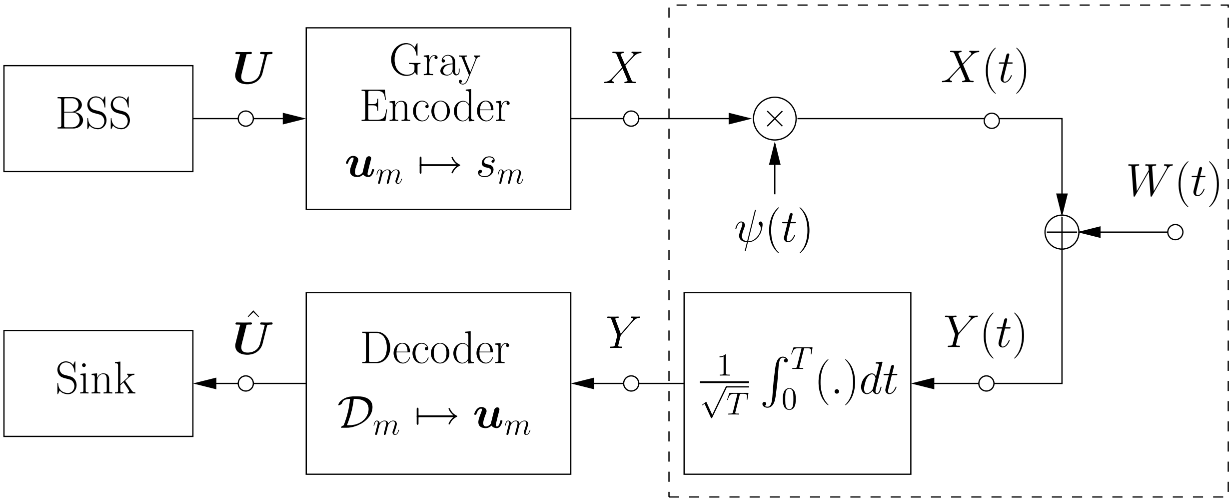
$$\mathrm{dec}_{\mathsf{ML}}(y) = s_m \quad \text{for } y \in \mathcal{D}_m.$$

ML decoder for MPAM

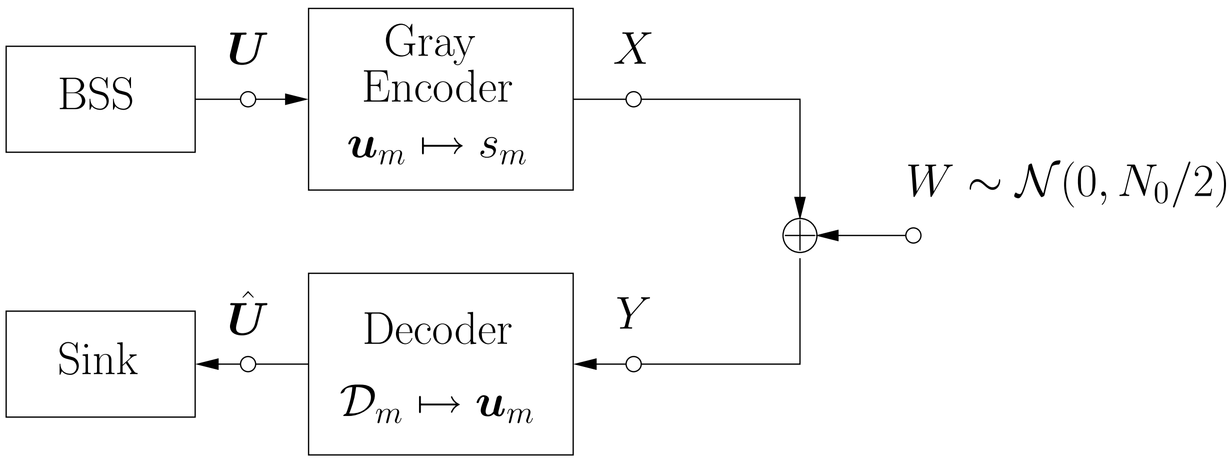
$$\hat{\boldsymbol{u}} = \mathrm{enc}^{-1}(\mathrm{dec}_{\mathsf{ML}}(y))$$

3.2.4 Vector Representation of the Transceiver

BPAM transceiver:



Vector representation:



Notice: $\boldsymbol{u}_m$ denotes the block of source bits corresponding to s_m, i.e.,

$$s_m = \mathrm{enc}(\boldsymbol{u}_m), \qquad \boldsymbol{u}_m = \mathrm{enc}^{-1}(s_m).$$

3.2.5 Symbol-Error Probability

Conditional symbol-error probability

Two cases must be considered:

1. The transmitted symbol is an edge symbol (a symbol at the edge of the signal constellation):
$$X \in \{s_1, s_M\}.$$
2. The transmitted symbol is not an edge symbol:
$$X \in \{s_2, s_3, \dots, s_{M-1}\}.$$

Case 1: $X \in \{s_1, s_M\}$

Assume first $X = s_1$. Then

$$\begin{aligned}
\Pr(\hat{X} \neq X | X = s_1) &= \Pr(Y \notin \mathcal{D}_1 | X = s_1) \\
&= \Pr(Y > s_1 + d | X = s_1) \\
&= \Pr(W > d) \\
&= \Pr\Big(\frac{W}{\sqrt{N_0/2}} > \frac{d}{\sqrt{N_0/2}}\Big) \\
&= Q\Big(\frac{d}{\sqrt{N_0/2}}\Big).
\end{aligned}$$

Notice that $W/\sqrt{N_0/2} \sim \mathcal{N}(0,1)$.

By symmetry,

$$\Pr(\hat{X} \neq X | X = s_M) = Q\Big(\frac{d}{\sqrt{N_0/2}}\Big).$$

Case 2: $X \in \{s_2, \ldots, s_{M-1}\}$
Assume $X = s_m$. Then

$$\begin{aligned}
\Pr(\hat{X} \neq X | X = s_m) &= \Pr(Y \notin \mathcal{D}_m | X = s_m) \\
&= \Pr(W < -d \text{ or } W > d) \\
&= \Pr(W < -d) + \Pr(W > d) \\
&= \Pr\Big(\frac{W}{\sqrt{N_0/2}} < -\frac{d}{\sqrt{N_0/2}}\Big) \\
&\quad + \Pr\Big(\frac{W}{\sqrt{N_0/2}} > \frac{d}{\sqrt{N_0/2}}\Big) \\
&= Q\Big(\frac{d}{\sqrt{N_0/2}}\Big) + Q\Big(\frac{d}{\sqrt{N_0/2}}\Big) \\
&= 2 \cdot Q\Big(\frac{d}{\sqrt{N_0/2}}\Big).
\end{aligned}$$

Symbol-error probability

Averaging over the symbols, one obtains

$$\begin{aligned}
P_s &= \frac{1}{M} \sum_{m=1}^{M} \Pr(\hat{X} \neq X | X = s_m) \\
&= \frac{2}{M}\, Q\Big(\frac{d}{\sqrt{N_0/2}}\Big) + \frac{2(M-2)}{M}\, Q\Big(\frac{d}{\sqrt{N_0/2}}\Big) \\
&= 2\frac{M-1}{M}\, Q\Big(\frac{d}{\sqrt{N_0/2}}\Big).
\end{aligned}$$

Expressing the symbol-error probability as a function of the SNR per symbol, γ_s, or the SNR per bit, γ_b, yields

$$\begin{aligned}
P_s &= 2\frac{M-1}{M}\, Q\Big(\sqrt{\frac{6}{M^2-1}\gamma_s}\Big) \\
&= 2\frac{M-1}{M}\, Q\Big(\sqrt{\frac{6\log_2 M}{M^2-1}\gamma_b}\Big).
\end{aligned}$$

Plots of the symbol-error probabilities can be found in [1, p. 412].

Remark: When comparing BPAM and 4PAM for the same P_s, the SNRs per bit, γ_b, differ by $10\log_{10}(5/2) \approx 4$ dB. (Proof?)

3.2.6 Bit-Error Probability

When Gray encoding is used, we have the relation

$$\begin{aligned} P_b &\approx \frac{P_s}{\log_2 M} \\ &= 2\frac{M-1}{M\log_2 M}\, Q\Big(\sqrt{\frac{6\log_2 M}{M^2-1}\gamma_b}\Big) \end{aligned}$$

for high SNR $\gamma_b = \overline{E}_b/N_0$.

4 Pulse Position Modulation (PPM)

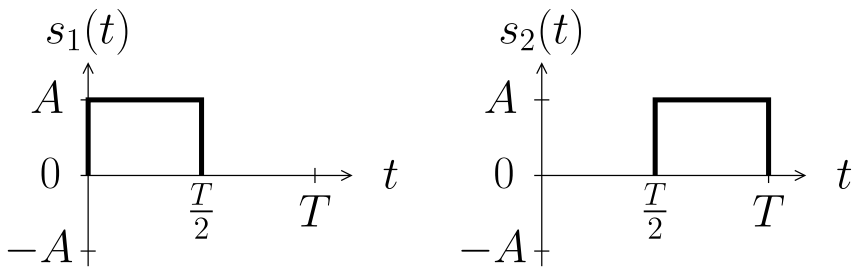
In pulse position modulation (PPM), the information is conveyed by the position of the transmitted waveforms.

4.1 BPPM

PPM with only two waveforms is called binary PPM (BPPM).

4.1.1 Signaling Waveforms

Set of two rectangular waveforms with amplitude A and duration $T/2$ over $t \in [0, T]$, $\mathbb{S} := \{s_1(t), s_2(t)\}$:



As the cardinality of the set is $|\mathbb{S}| = 2$, one bit is sufficient to address each waveforms, i.e., $K = 1$. Thus, the BPPM transmitter performs the mapping

$$U \;\mapsto\; x(t)$$

with

$$U \in \mathbb{U} = \{0, 1\} \qquad \text{and} \qquad x(t) \in \mathbb{S}.$$

Remark: In the general case, the basic waveform may also have another shape.

4.1.2 Decomposition of the Transmitter

Gram-Schmidt procedure

The waveforms $s_1(t)$ and $s_2(t)$ are orthogonal and have the same energy

$$E_s = \int_0^T s_1^2(t)\mathsf{d}t = \int_0^T s_2^2(t)\mathsf{d}t = A^2 \cdot \tfrac{T}{2}.$$

Notice that $\sqrt{E_s} = A\sqrt{T/2}$.

The two orthonormal functions and the corresponding waveform representations are

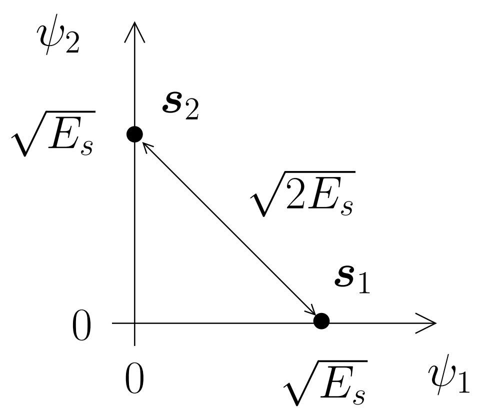
$$\psi_1(t) = \frac{1}{\sqrt{E_s}} s_1(t), \qquad s_1(t) = \sqrt{E_s}\cdot\psi_1(t) + 0\cdot\psi_2(t),$$
$$\psi_2(t) = \frac{1}{\sqrt{E_s}} s_2(t), \qquad s_2(t) = 0\cdot\psi_1(t) + \sqrt{E_s}\cdot\psi_2(t).$$

Signal constellation

Vector representation of the two waveforms:

$$s_1(t) \quad \longleftrightarrow \quad \boldsymbol{s}_1 = (\sqrt{E_s}, 0)^{\mathsf{T}},$$
$$s_2(t) \quad \longleftrightarrow \quad \boldsymbol{s}_2 = (0, \sqrt{E_s})^{\mathsf{T}}.$$

Thus the signal constellation is $\mathbb{X} = \big\{(\sqrt{E_s}, 0)^{\mathsf{T}}, (0, \sqrt{E_s})^{\mathsf{T}}\big\}$.



Vector Encoder

The vector encoder is defined as

$$\boldsymbol{x} = \mathrm{enc}(u) = \begin{cases} (\sqrt{E_s}, 0)^\mathsf{T} & \text{for } u = 0, \\ (0, \sqrt{E_s})^\mathsf{T} & \text{for } u = 1. \end{cases}$$

The signal constellation comprises only two waveforms, and thus per (modulation) symbol and per (source) bit, the average transmit energies and the SNRs are the same:

$$E_b = E_s, \qquad\qquad \gamma_s = \gamma_b.$$

4.1.3 ML Decoding

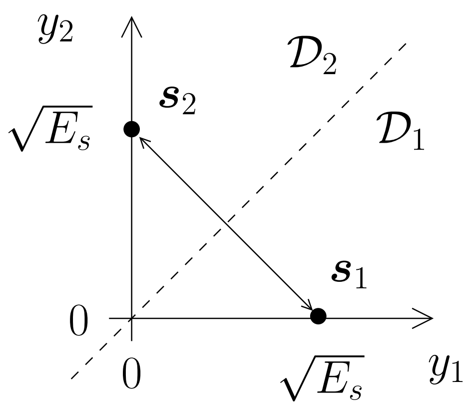
ML symbol decision rule

$$\mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) = \underset{x \in \{\boldsymbol{s}_1, \boldsymbol{s}_2\}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{x}\|^2 = \begin{cases} (\sqrt{E_s}, 0)^\mathsf{T} & \text{for } y_1 \geq y_2, \\ (0, \sqrt{E_s})^\mathsf{T} & \text{for } y_1 < y_2. \end{cases}$$

Notice: $\|\boldsymbol{y} - \boldsymbol{s}_1\|^2 \leq \|\boldsymbol{y} - \boldsymbol{s}_2\|^2 \Leftrightarrow y_1 \geq y_2$.

Decision regions

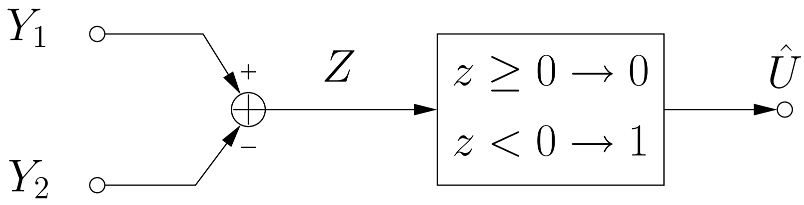
$$\mathcal{D}_1 = \{\boldsymbol{y} \in \mathbb{R}^2 : y_1 \geq y_2\}, \quad \mathcal{D}_2 = \{\boldsymbol{y} \in \mathbb{R}^2 : y_1 < y_2\}.$$



Combining the (modulation) symbol decision rule and the encoder inverse yields the **ML decoder for BPPM**:

$$\hat{u} = \mathrm{enc}^{\text{-}1}(\mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y})) = \begin{cases} 0 & \text{for } y_1 \geq y_2, \\ 1 & \text{for } y_1 < y_2. \end{cases}$$

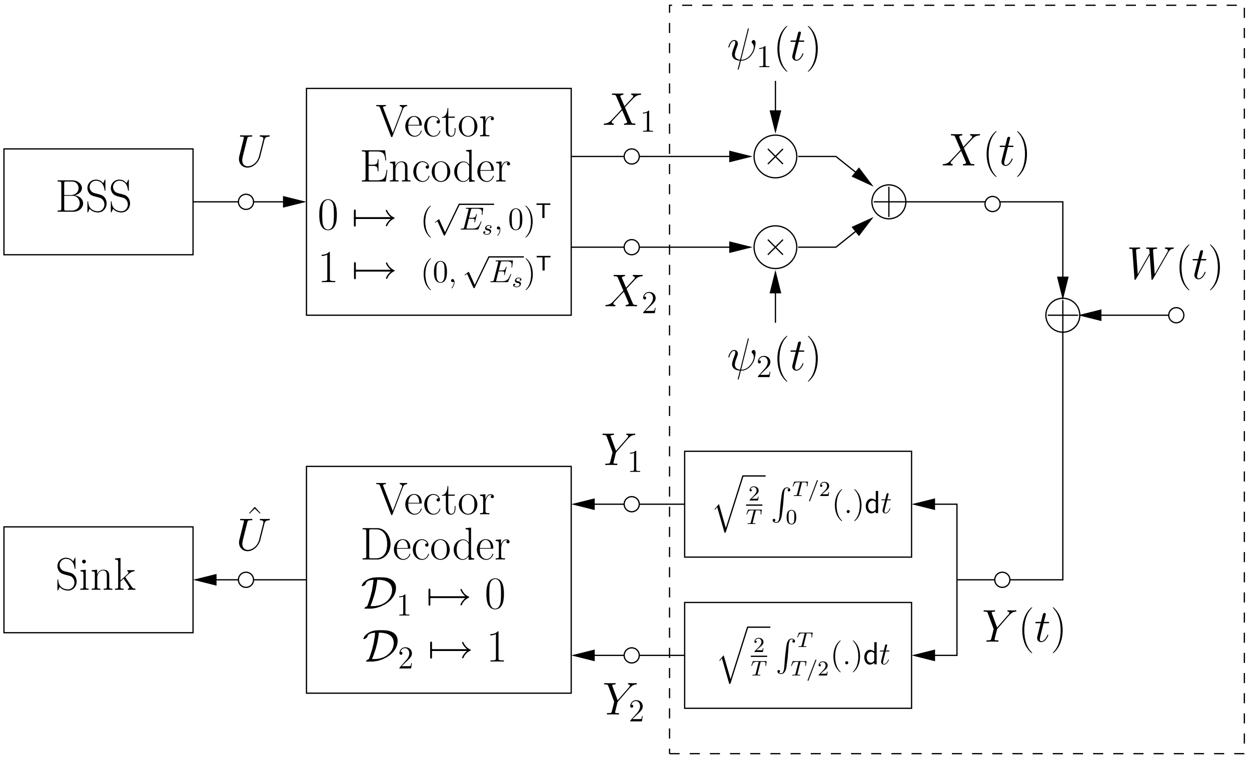
The ML decoder may be implemented by



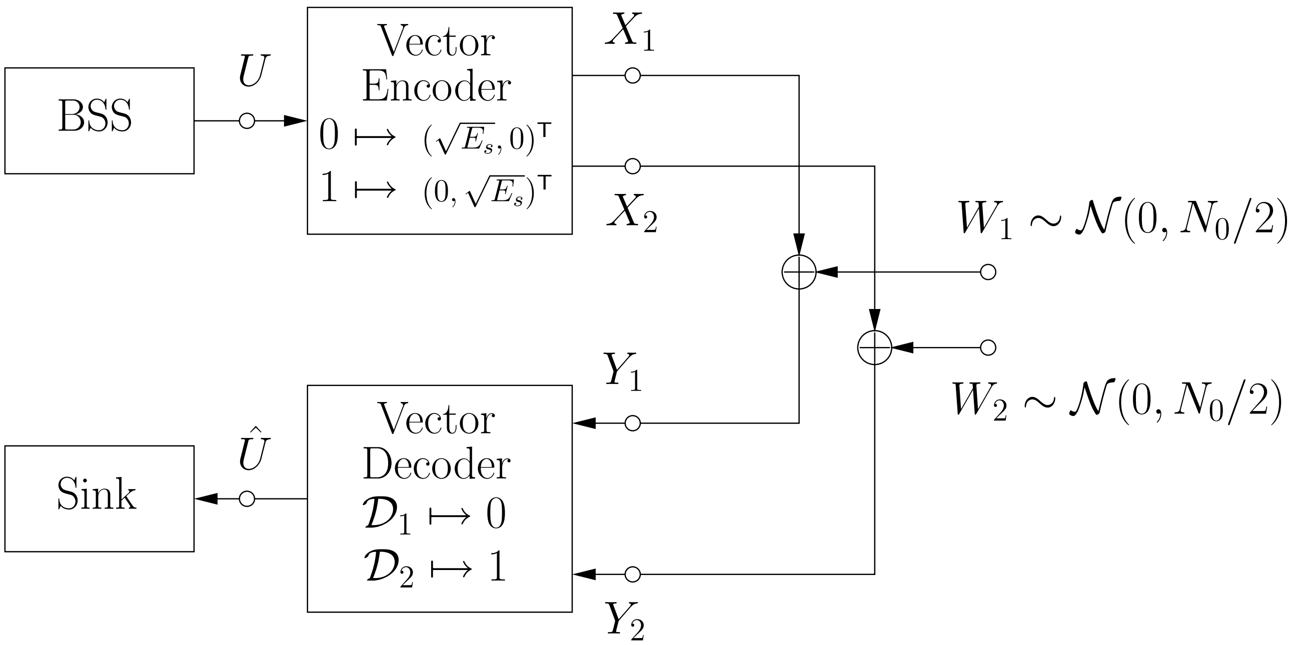
Notice that $Z = Y_1 - Y_2$ is then the decision variable.

4.1.4 Vector Representation of the Transceiver

BPPM transceiver for the waveform AWGN channel



Vector representation including a AWGN vector channel



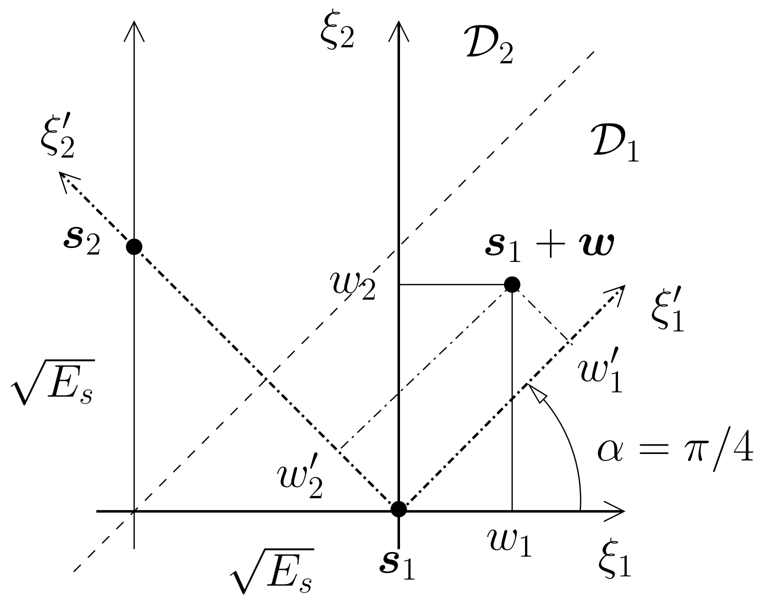
4.1.5 Symbol-Error Probability

Conditional symbol-error probability

Assume that $\boldsymbol{X} = \boldsymbol{s}_1 = (\sqrt{E_s}, 0)^\mathsf{T}$ is transmitted.

$$\begin{aligned}\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= \Pr(\boldsymbol{Y} \in \mathcal{D}_2 | \boldsymbol{X} = \boldsymbol{s}_1) \\ &= \Pr(\boldsymbol{s}_1 + \boldsymbol{W} \in \mathcal{D}_2 | \boldsymbol{X} = \boldsymbol{s}_1) \\ &= \Pr(\boldsymbol{s}_1 + \boldsymbol{W} \in \mathcal{D}_2)\end{aligned}$$

The noise vector $\boldsymbol{W}$ may be represented in the coordinate system (ξ_1, ξ_2) or in the rotated coordinate system (ξ_1', ξ_2'). The latter is more suited to computing the above probability, as only the noise component along the ξ_2'-axis is relevant.



The elements of the noise vector in the rotated coordinate system have the same statistical properties as in the non-rotated coordinate system (see Appendix B), i.e.,

$$W_1' \sim \mathcal{N}(0, N_0/2), \qquad W_2' \sim \mathcal{N}(0, N_0/2).$$

Thus, we obtain

$$\begin{aligned}\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= \Pr(W_2' > \sqrt{E_s/2}) \\ &= \Pr(\frac{W_2'}{\sqrt{N_0/2}} > \sqrt{\frac{E_s}{N_0}}) \\ &= \mathrm{Q}(\sqrt{E_s/N_0}) = Q(\sqrt{\gamma_s}).\end{aligned}$$

By symmetry,

$$\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_2) = Q(\sqrt{\gamma_s}).$$

Symbol-error probability

Averaging over the conditional symbol-error probabilities, we obtain

$$P_s = \Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X}) = Q(\sqrt{\gamma_s}).$$

4.1.6 Bit-Error Probability

For BPPM, $E_b = E_s$, $\gamma_s = \gamma_b$, $P_b = P_s$ (see above), and so the bit-error probability (BEP) results as

$$P_b = \Pr(\hat{U} \neq U) = Q(\sqrt{\gamma_b}).$$

A plot of the BEP versus γ_b may be found in [1, p. 426]

Remark

Due to their signal constellations, BPPM is called orthogonal signaling and BPAM is called antipodal signaling. Their BEPs are given by

$$P_{b,\mathrm{BPPM}} = Q(\sqrt{\gamma_b}), \qquad P_{b,\mathrm{BPAM}} = Q(\sqrt{2\gamma_b}).$$

When comparing the SNR for a certain BEP, BPPM requires 3 dB more than BPAM. For a plot, see [1, p. 409]

4.2 MPPM

In the general case of M-ary pulse position modulation (MPPM) with $M = 2^K$, input blocks of K bits determine the position of a unique pulse.

4.2.1 Signaling Waveforms

Waveform encoder

$$\begin{aligned} \boldsymbol{u} = [u_1, \ldots, u_K] \; &\mapsto \; x(t) \\ \mathbb{U} = \{0,1\}^K \; &\rightarrow \; \mathbb{S} = \big\{s_1(t), \ldots, s_M(t)\big\}, \end{aligned}$$

where

$$s_m(t) = g\Big(t - \frac{(m-1)T}{M}\Big),$$

$m = 1, 2, \ldots, M$, and $g(t)$ is a predefined pulse of duration T/M,

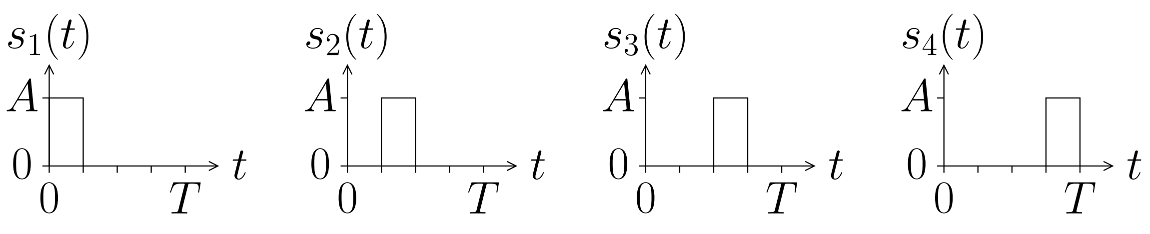
$$g(t) = 0 \quad \text{for } t \notin [0, T/M].$$

The waveforms are orthogonal and have the same energy as $g(t)$,

$$E_s = E_g = \int_0^T g^2(t)\mathrm{d}t.$$

Example: (M=4)

Consider a rectangular pulse $g(t)$ with amplitude A and duration $T/4$. Then, we have the set of four waveforms, $\mathbb{S} = \{s_1(t), s_2(t), s_3(t), s_4(t)\}$:



◇

4.2.2 Decomposition of the Transmitter

Gram-Schmidt procedure

As all waveforms are orthogonal, we have M orthonormal functions:

$$\psi_m(t) = \frac{1}{\sqrt{E_s}} s_m(t),$$

$m = 1, 2, \ldots, M$.

The waveforms and the orthonormal functions are related as

$$s_m(t) = \sqrt{E_s} \cdot \psi_m(t),$$

$m = 1, 2, \ldots, M$.

Signal constellation

Waveforms and their vector representations:

$$\begin{array}{ccc}
s_1(t) & \longleftrightarrow & \boldsymbol{s}_1 = (\overbrace{\sqrt{E_s}, 0, 0, \ldots, 0}^{M})^{\mathsf{T}} \\
s_2(t) & \longleftrightarrow & \boldsymbol{s}_2 = (0, \sqrt{E_s}, 0, \ldots, 0)^{\mathsf{T}} \\
 & \vdots & \\
s_M(t) & \longleftrightarrow & \boldsymbol{s}_M = (0, 0, 0, \ldots, \sqrt{E_s})^{\mathsf{T}}
\end{array}$$

Signal constellation:

$$\mathbb{X} = \left\{ \boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M \right\} \in \mathbb{R}^M.$$

The modulation symbols $\boldsymbol{x} \in \mathbb{X}$ are orthogonal, and they are placed on corners of an M-dimensional hypercube.

EXAMPLE: $M = 3$

The signal constellation

$$\mathbb{X} = \left\{(\sqrt{E_s}, 0, 0)^\mathsf{T}, (0, \sqrt{E_s}, 0)^\mathsf{T}, (0, \sqrt{E_s}, 0)^\mathsf{T}\right\} \in \mathbb{R}^3$$

may be visualized using a cube with side-length $\sqrt{E_s}$. ◇

Vector Encoder

$$\begin{aligned} \text{enc}: \quad \boldsymbol{u} = [u_1, \ldots, u_K] &\mapsto \boldsymbol{x} = \text{enc}(\boldsymbol{u}) \\ \mathbb{U} = \{0,1\}^K &\to \mathbb{X} = \left\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M\right\} \end{aligned}$$

Energy and SNR

The symbol energy is E_s, and as $M = 2^K$, the average energy per transmitted source bit results as

$$E_b = \frac{E_s}{K}$$

The SNR per symbol and the SNR per bit are related as

$$\gamma_b = \frac{1}{K}\gamma_s = \frac{1}{\log_2 M}\gamma_s.$$

4.2.3 ML Decoding

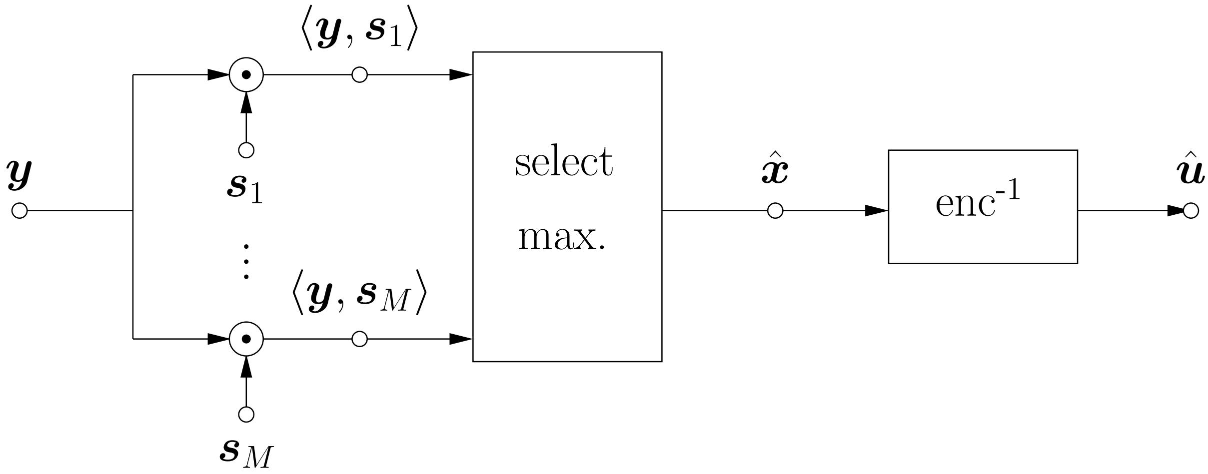
ML symbol decision rule

$$
\begin{aligned}
\mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) &= \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{X}} \|\boldsymbol{y}-\boldsymbol{x}\|^2 \\
&= \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{X}} \|\boldsymbol{y}\|^2 - 2\langle\boldsymbol{y},\boldsymbol{x}\rangle + \underbrace{\|\boldsymbol{x}\|^2}_{E_s} \\
&= \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{X}} \langle\boldsymbol{y},\boldsymbol{x}\rangle
\end{aligned}
$$

ML decoder for MPAM

$$\hat{\boldsymbol{u}} = \mathrm{enc}^{-1}(\mathrm{dec}_{\mathsf{ML}}(y))$$

Implementation of the ML detector using vector correlators:



Notice that $\hat{\boldsymbol{u}} = [\hat{u}_1, \ldots, \hat{u}_K]$.

4.2.4 Symbol-Error Probability

Conditional symbol-error probability

Assume first that $\boldsymbol{X} = \boldsymbol{s}_1$ is transmitted. Then

$$\begin{aligned}\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= 1 - \Pr(\hat{\boldsymbol{X}} = \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) \\ &= 1 - \Pr\Big(\langle \boldsymbol{y}, \boldsymbol{s}_1 \rangle \geq \langle \boldsymbol{y}, \boldsymbol{s}_m \rangle \text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1\Big).\end{aligned}$$

For $\boldsymbol{X} = \boldsymbol{s}_1 = (\sqrt{E_s}, 0, \ldots, 0)$, the received vector is

$$\boldsymbol{Y} = \boldsymbol{s}_1 + \boldsymbol{W} = (\sqrt{E_s}, 0, \ldots, 0) + (W_1, \ldots, W_M),$$

so that

$$\langle \boldsymbol{y}, \boldsymbol{s}_m \rangle = \begin{cases} E_s + \sqrt{E_s} \cdot W_1 & \text{for } m = 1, \\ \sqrt{E_s} \cdot W_m & \text{for } m = 2, \ldots, M. \end{cases}$$

Therefore, the symbol-error probability may be written as follows:

$$\begin{aligned}&\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) = \\ &\quad = 1 - \Pr\big(E_s + \sqrt{E_s} W_1 \geq \sqrt{E_s} W_m \quad \text{for all } m \neq 1\big) \\ &\quad = 1 - \Pr\Big(\sqrt{\frac{E_s}{N_0/2}} + \underbrace{\frac{W_1}{\sqrt{N_0/2}}}_{W_1'} \geq \underbrace{\frac{W_m}{\sqrt{N_0/2}}}_{W_m'} \quad \text{for all } m \neq 1\Big) \\ &\quad = 1 - \Pr\Big(W_m' \leq \sqrt{\frac{2E_s}{N_0}} + W_1' \quad \text{for all } m \neq 1\Big)\end{aligned}$$

with

$$\boldsymbol{W}' = (W_1', \ldots, W_M')^{\mathsf{T}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}),$$

i.e., the components W_m' are independent and Gaussian distributed with zero mean and variance 1.

The probability is now written using the pdf of W_1':

$$\Pr\Big(W_m' \leq \sqrt{\tfrac{2E_s}{N_0}} + W_1' \quad \text{for all } m \neq 1\Big) =$$

$$= \int_{-\infty}^{+\infty} \underbrace{\Pr\Big(W_m' \leq \sqrt{\tfrac{2E_s}{N_0}} + W_1' \quad \text{for all } m \neq 1 \Big| W_1' = v\Big)}_{(\star)} p_{W_1'}(v) \mathsf{d}v.$$

The events corresponding to different values of m are independent, and thus

$$(\star) = \prod_{m=2}^{M} \underbrace{\Pr\Big(W_m' \leq \sqrt{\tfrac{2E_s}{N_0}} + W_1' \Big| W_1' = v\Big)}_{\Pr\big(W_m' \leq \sqrt{\frac{2E_s}{N_0}} + v\big)}$$

$$= \prod_{m=2}^{M} \mathrm{Q}\Big(\sqrt{\tfrac{2E_s}{N_0}} + v\Big) = \mathrm{Q}^{M-1}(\sqrt{2\gamma_s} + v).$$

Using this result and inserting $p_{W_1'}(v)$, we obtain the conditional symbol-error probability

$$\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) =$$

$$= \int_{-\infty}^{+\infty} \Big[1 - \mathrm{Q}^{M-1}(\sqrt{2\gamma_s} + v)\Big] \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{v^2}{2}\Big) \mathsf{d}v$$

Symbol-error probability

The above expression is independent of $\boldsymbol{s}_1$ and thus valid for all $\boldsymbol{x} \in \mathbb{X}$. Therefore, the average symbol-error probability results as

$$P_s = \int_{-\infty}^{+\infty} \Big[1 - \mathrm{Q}^{M-1}(\sqrt{2\gamma_s} + v)\Big] \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{v^2}{2}\Big) \mathsf{d}v.$$

4.2.5 Bit-Error Probability

Hamming distance

The Hamming distance $\mathrm{d_H}(\boldsymbol{u}, \boldsymbol{u}')$ between two length-K vectors $\boldsymbol{u}$ and $\boldsymbol{u}'$ is the number of positions in which they differ:

$$\mathrm{d_H}(\boldsymbol{u}, \boldsymbol{u}') = \sum_{k=1}^{K} \delta(u_k, u'_k),$$

where the indicator function is defined as

$$\delta(u, u') = \begin{cases} 1 & \text{for } u = u', \\ 0 & \text{for } u \neq u'. \end{cases}$$

EXAMPLE: 4PPM

Let the blocks of source bits and the modulation symbols be associated as

$$\begin{aligned} \boldsymbol{u}_1 &= \mathrm{enc}^{-1}(\boldsymbol{s}_1) = [0, 0] \\ \boldsymbol{u}_2 &= \mathrm{enc}^{-1}(\boldsymbol{s}_2) = [0, 1] \\ \boldsymbol{u}_3 &= \mathrm{enc}^{-1}(\boldsymbol{s}_3) = [1, 0] \\ \boldsymbol{u}_4 &= \mathrm{enc}^{-1}(\boldsymbol{s}_4) = [1, 1]. \end{aligned}$$

Then we have for example

$$\begin{aligned} \mathrm{d_H}(\boldsymbol{u}_1, \boldsymbol{u}_1) &= 0, & \mathrm{d_H}(\boldsymbol{u}_1, \boldsymbol{u}_2) &= 1, \\ \mathrm{d_H}(\boldsymbol{u}_2, \boldsymbol{u}_3) &= 2, & \mathrm{d_H}(\boldsymbol{u}_2, \boldsymbol{u}_4) &= 1. \end{aligned}$$

◇

Conditional bit-error probability

Assume that $\boldsymbol{X} = \boldsymbol{s}_1$, and let

$$\boldsymbol{u}_m = [u_{m,1}, \ldots, u_{m,K}] = \mathrm{enc}^{-1}(\boldsymbol{s}_m).$$

The conditional bit-error probability is defined as

$$\Pr(U \neq \hat{U} | \boldsymbol{X} = \boldsymbol{s}_1) = \frac{1}{K} \sum_{k=1}^{K} \Pr(U_k \neq \hat{U}_k | \boldsymbol{X} = \boldsymbol{s}_1).$$

The argument of the sum may be expanded as

$$\begin{aligned}
&\Pr(U_k \neq \hat{U}_k | \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\quad = \sum_{m=1}^{M} \Pr(U_k \neq \hat{U}_k | \hat{\boldsymbol{X}} = \boldsymbol{s}_m, \boldsymbol{X} = \boldsymbol{s}_1) \Pr(\hat{\boldsymbol{X}} = \boldsymbol{s}_m | \boldsymbol{X} = \boldsymbol{s}_1).
\end{aligned}$$

Since

$$\begin{array}{ccccc}
\boldsymbol{X} = \boldsymbol{s}_1 & \Leftrightarrow & \boldsymbol{U} = \boldsymbol{u}_1 & \Rightarrow & U_k = u_{1,k}, \\
\hat{\boldsymbol{X}} = \boldsymbol{s}_m & \Leftrightarrow & \hat{\boldsymbol{U}} = \boldsymbol{u}_m & \Rightarrow & \hat{U}_k = u_{m,k},
\end{array}$$

we have

$$\begin{aligned}
&\Pr(U_k \neq \hat{U}_k | \hat{\boldsymbol{X}} = \boldsymbol{s}_m, \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\quad = \Pr(u_{1,k} \neq u_{m,k} | \hat{\boldsymbol{X}} = \boldsymbol{s}_m, \boldsymbol{X} = \boldsymbol{s}_1) \\
&\quad = \delta(u_{1,k}, u_{m,k}).
\end{aligned}$$

Using this result in the expression for the conditional bit-error probability, we obtain

$$\begin{aligned}
&\Pr(U \neq \hat{U} | \boldsymbol{X} = \boldsymbol{s}_1) \\
&\quad = \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} \delta(u_{1,k}, u_{m,k}) \;\Pr(\hat{\boldsymbol{X}} = \boldsymbol{s}_m | \boldsymbol{X} = \boldsymbol{s}_1) \\
&\quad = \frac{1}{K} \sum_{m=1}^{M} \underbrace{\sum_{k=1}^{K} \delta(u_{1,k}, u_{m,k})}_{\mathrm{d_H}(\boldsymbol{u}_1, \boldsymbol{u}_m)} \;\Pr(\hat{\boldsymbol{X}} = \boldsymbol{s}_m | \boldsymbol{X} = \boldsymbol{s}_1)
\end{aligned}$$

Due to the symmetry of the signal constellation, the probabilities of making a specific decision error are identical, i.e.,

$$\Pr(\hat{\boldsymbol{X}} = \boldsymbol{s}_m | \boldsymbol{X} = \boldsymbol{s}_{m'}) = \frac{P_s}{M-1} = \frac{P_s}{2^K - 1}$$

for all $m \neq m'$. Using this in the above sum yields

$$\Pr(U \neq \hat{U} | \boldsymbol{X} = \boldsymbol{s}_1) = \frac{1}{K} \cdot \frac{P_s}{2^K - 1} \sum_{m=1}^{M} \mathrm{d_H}(\boldsymbol{u}_1, \boldsymbol{u}_m).$$

Without loss of generality, we assume $\boldsymbol{u}_1 = \boldsymbol{0}$. The sum over the Hamming distances becomes then the overall number of ones in all bit blocks of length K.

Example:

Consider $K = 3$ and write the binary vectors of length 3 in the columns of a matrix:

$$\left[\boldsymbol{u}_1^\mathsf{T}, \ldots, \boldsymbol{u}_8^\mathsf{T}\right] = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

In every row, half of the entries are ones. The overall number of ones is thus $3 \cdot 2^{3-1}$. ◇

Generalizing the example, the above sum can shown to be

$$\sum_{m=1}^{M} \mathrm{d_H}(\boldsymbol{0}, \boldsymbol{u}_m) = K \cdot 2^{K-1}.$$

Applying this result, we obtain the conditional bit-error probability

$$\Pr(U \neq \hat{U} | \boldsymbol{X} = \boldsymbol{s}_1) = \frac{2^{K-1}}{2^K - 1} \cdot P_s,$$

which is independent of $\boldsymbol{s}_1$ and thus the same for all transmitted symbols $\boldsymbol{X} = \boldsymbol{s}_m$, $m = 1, \ldots, M$.

Average bit-error probability

From above, we obtain

$$\Pr(U \neq \hat{U}) = \frac{2^{K-1}}{2^K - 1} \cdot P_s.$$

Plots of the BEPs may be found in [1, p. 426]

Notice that for large K,

$$\Pr(U \neq \hat{U}) \approx \frac{1}{2} \cdot P_s.$$

Due to the signal constellation, Gray mapping is not possible.

Remarks

(i) Provided that the SNR $\gamma_b = E_b/N_0$ is larger than $\ln 2$ (corresponding to -1.6 dB), **the BEP of MPPM can be made arbitrarily small** by selecting M large enough. The price to pay is an increasing decoding delay and an increasing required bandwidth.

Let the transmission rate be denoted by

$$R = \frac{K}{T} = \frac{\log_2 M}{T} \quad \text{[bit/sec]}.$$

Decoding delay:

$$T = \frac{1}{R} \cdot \log_2 M.$$

Hence, for a fixed rate R, $T \to \infty$ for $M \to \infty$.

Required bandwidth:

$$B = B_g \approx \frac{1}{T/M} = \frac{M}{T} = R \cdot \frac{M}{\log_2 M},$$

where B_g denotes the bandwidth of the pulse $g(t)$.

Hence, for a fixed rate R, $B \to \infty$ for $M \to \infty$.

(ii) The power of the MPPM signal is

$$P = \frac{E_b \cdot K}{T} = E_b \cdot R.$$

Hence, the condition $\gamma_b > \ln 2$ is equivalent to the condition

$$R < \frac{1}{\ln 2} \cdot \frac{P}{N_0} = C_\infty.$$

The value C_∞ denotes the capacity of the (infinite bandwidth) AWGN channel (see Appendix C).

5 Phase-Shift Keying (PSK)

In phase-shift keying (PSK), the waveforms are sines with the same frequency, and the information is conveyed by the phase of the waveforms.

5.1 Signaling Waveforms

Set of sines $s_m(t)$ with the same frequency f_0, duration T, and different phases ϕ_m:

$$s_m(t) = \begin{cases} \sqrt{2P}\cos\bigl(2\pi f_0 t + \underbrace{2\pi\tfrac{m-1}{M}}_{\phi_m}\bigr) & \text{for } t \in [0,T], \\ 0 & \text{elsewhere,} \end{cases}$$

$m = 1, 2, \ldots, M$. The phase can take the values

$$\phi_m \in \bigl\{0, 2\pi\tfrac{1}{M}, 2\pi\tfrac{2}{M}, \ldots, 2\pi\tfrac{M-1}{M}\bigr\}$$

Hence, the digital information is embedded in the phase of a carrier. The frequency f_0 is called the carrier frequency.

For $M = 2$, this modulation scheme is called **binary phase-shift keying** (BPSK), and for $M = 4$, it is called **quadrature phase-shift keying**. (QPSK).

All waveforms have the same energy

$$E_s = \int_{-\infty}^{\infty} s_m^2(t)\,\mathsf{d}t = PT.$$

(Notice: $\cos^2 x = 1 + \cos 2x$.)

Usually, $f_0 \gg 1/T$. For simplifying the theoretical analysis, we assume that $f_0 = L/T$ for some large $L \in \mathbb{N}$.

BPSK

For $M = 2$, we obtain $\phi_m \in \{0, \pi\}$ and thus the two waveforms

$$
\begin{aligned}
s_1(t) &= \sqrt{2P}\cos\bigl(2\pi f_0 t\bigr), \\
s_2(t) &= \sqrt{2P}\cos\bigl(2\pi f_0 t + \pi\bigr) = -s_1(t)
\end{aligned}
$$

for $t \in [0, T]$.

QPSK

For $M = 4$, we obtain $\phi_m \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$ and thus the waveforms are

$$
\begin{aligned}
s_1(t) &= +\sqrt{2P}\cos\bigl(2\pi f_0 t\bigr), \\
s_2(t) &= -\sqrt{2P}\sin\bigl(2\pi f_0 t\bigr), \\
s_3(t) &= -\sqrt{2P}\cos\bigl(2\pi f_0 t\bigr), \\
s_4(t) &= +\sqrt{2P}\sin\bigl(2\pi f_0 t\bigr)
\end{aligned}
$$

for $t \in [0, T]$.

5.2 Signal Constellation

BPSK

Orthonormal functions:

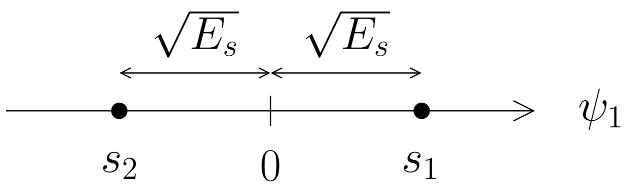
$$\psi_1(t) = \frac{1}{\sqrt{E_s}} s_1(t) = \sqrt{\frac{2}{T}} \cos\bigl(2\pi f_0 t\bigr)$$

for $t \in [0, T]$.

Geometrical representation of the waveforms:

$$\begin{aligned} s_1(t) &= +\sqrt{PT}\psi_1(t) \quad &\longleftrightarrow \quad s_1 &= +\sqrt{PT} = +\sqrt{E_s} \\ s_2(t) &= -\sqrt{PT}\psi_1(t) \quad &\longleftrightarrow \quad s_2 &= -\sqrt{PT} = -\sqrt{E_s} \end{aligned}$$

Signal constellation:



Vector encoder:

$$x = \mathrm{enc}(u) = \begin{cases} s_1 & \text{for } u = 1, \\ s_2 & \text{for } u = 0. \end{cases}$$

QPSK

Orthonormal functions:

$$\psi_1(t) = \tfrac{1}{\sqrt{E_s}} s_1(t) = +\sqrt{\tfrac{2}{T}} \cos\bigl(2\pi f_0 t\bigr)$$
$$\psi_2(t) = \tfrac{1}{\sqrt{E_s}} s_2(t) = -\sqrt{\tfrac{2}{T}} \sin\bigl(2\pi f_0 t\bigr)$$

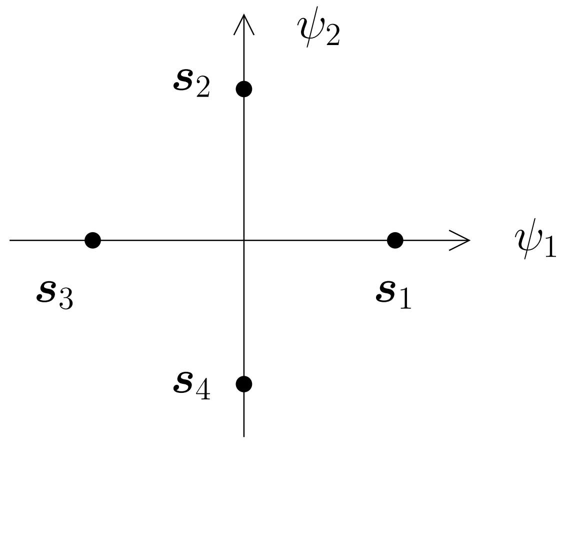
for $t \in [0, T]$.

Geometrical representation of the waveforms:

$$\begin{aligned}
s_1(t) &= +\sqrt{PT}\psi_1(t) &\longleftrightarrow\quad \boldsymbol{s}_1 &= (\sqrt{E_s}, 0)^{\mathsf{T}} \\
s_2(t) &= \sqrt{E_s}\psi_2(t) &\longleftrightarrow\quad \boldsymbol{s}_2 &= (0, \sqrt{E_s})^{\mathsf{T}} \\
s_3(t) &= -\sqrt{PT}\psi_1(t) &\longleftrightarrow\quad \boldsymbol{s}_3 &= (-\sqrt{E_s}, 0)^{\mathsf{T}} \\
s_4(t) &= -\sqrt{E_s}\psi_2(t) &\longleftrightarrow\quad \boldsymbol{s}_4 &= (0, -\sqrt{E_s})^{\mathsf{T}}
\end{aligned}$$

with $E_s = PT$.

Signal constellation:



Vector encoder: any Gray encoder

General Case

The expression for $s_m(t)$ can be expanded as follows:

$$\begin{aligned} s_m(t) &= \sqrt{E_s}\ \cos\big(2\pi f_0 t + \phi_m\big) \\ &= \sqrt{E_s}\ \cos\phi_m\ \cos(2\pi f_0 t) - \sqrt{E_s}\ \sin\phi_m\ \sin(2\pi f_0 t), \end{aligned}$$

where $\phi_m = 2\pi\frac{m-1}{M}$, $m = 1, 2, \ldots, M$.

(Notice: $\cos(x+y) = \cos x\ \cos y - \sin x\ \sin y$.)

Orthonormal functions:

$$\begin{aligned} \psi_1(t) &= \sqrt{\tfrac{2}{T}}\ \cos(2\pi f_0 t) \\ \psi_2(t) &= -\sqrt{\tfrac{2}{T}}\ \sin\big(2\pi f_0 t\big) \end{aligned}$$

for $t \in [0, T]$.

Geometrical representation of the waveforms:

$$\begin{aligned} s_m(t) &= \sqrt{E_s}\cos\phi_m \cdot \psi_1(t) + \sqrt{E_s}\sin\phi_m \cdot \psi_2(t) \\ \longleftrightarrow \quad \boldsymbol{s}_m &= \big(\sqrt{E_s}\cos\phi_m, \sqrt{E_s}\sin\phi_m\big)^\mathsf{T}, \end{aligned}$$

$m = 1, 2, \ldots, M$. Remember: $E_s = PT$.

Signal constellation:

$$\mathbb{X} = \Big\{\big(\sqrt{E_s}\ \cos 2\pi\tfrac{m-1}{M}, \sqrt{E_s}\ \sin 2\pi\tfrac{m-1}{M}\big)^\mathsf{T} : m = 1, 2, \ldots, M\Big\}$$

Example: 8PSK

For $M = 8$, we obtain $\phi_m = \pi\frac{m-1}{4}$, $m = 1, 2, \ldots, 8$. (Figure?) $\diamond$

5.3 ML Decoding

ML symbol decision rule

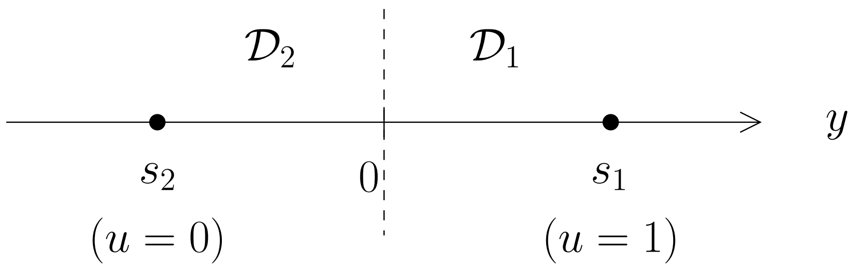
$$\mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) = \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{X}} \|\boldsymbol{y}-\boldsymbol{x}\|^2$$

ML decoder for MPSK

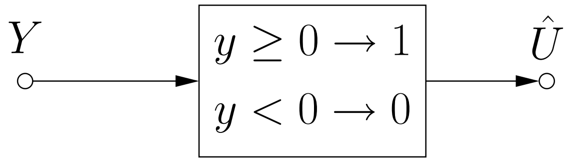
$$\hat{\boldsymbol{u}} = [\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_K] = \mathrm{enc}^{-1}(\mathrm{dec}_{\mathsf{ML}}(y))$$

BPSK

BPSK has the same signal constellation as BPAM. Thus, the ML detectors for these two modulation schemes are the same.

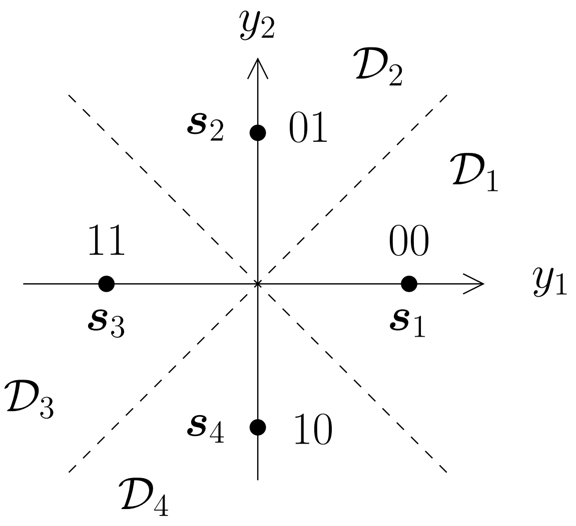


Implementation of the ML detector:



QPSK

Decision regions for the symbols $\boldsymbol{s}_m$ and corresponding bit blocks $\boldsymbol{u} = [u_1, u_2]$ (Gray encoding):

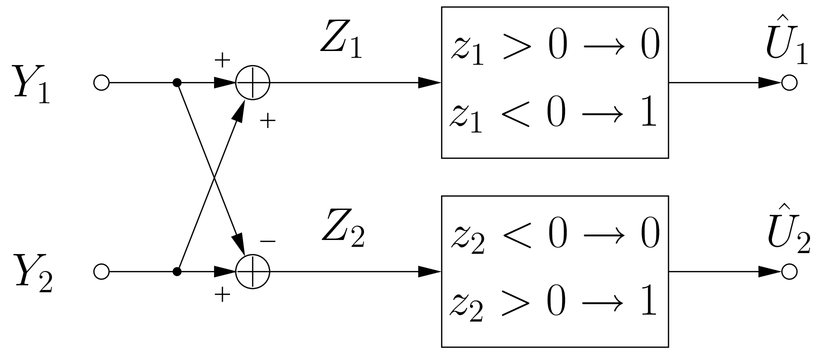


Decision rule for the data bits:

$$\hat{u}_1 = \begin{cases} 0 & \text{for } \boldsymbol{y} \in \mathcal{D}_1 \cup \mathcal{D}_2 \quad \Leftrightarrow \quad y_1 + y_2 > 0 \\ 1 & \text{for } \boldsymbol{y} \in \mathcal{D}_3 \cup \mathcal{D}_4 \quad \Leftrightarrow \quad y_1 + y_2 < 0 \end{cases}$$

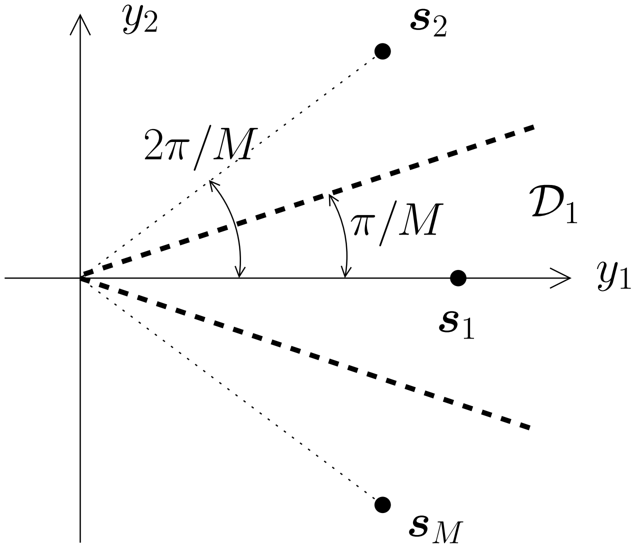
$$\hat{u}_2 = \begin{cases} 0 & \text{for } \boldsymbol{y} \in \mathcal{D}_1 \cup \mathcal{D}_4 \quad \Leftrightarrow \quad y_2 - y_1 < 0 \\ 1 & \text{for } \boldsymbol{y} \in \mathcal{D}_2 \cup \mathcal{D}_3 \quad \Leftrightarrow \quad y_2 - y_1 > 0 \end{cases}$$

Implementation of the ML decoder:

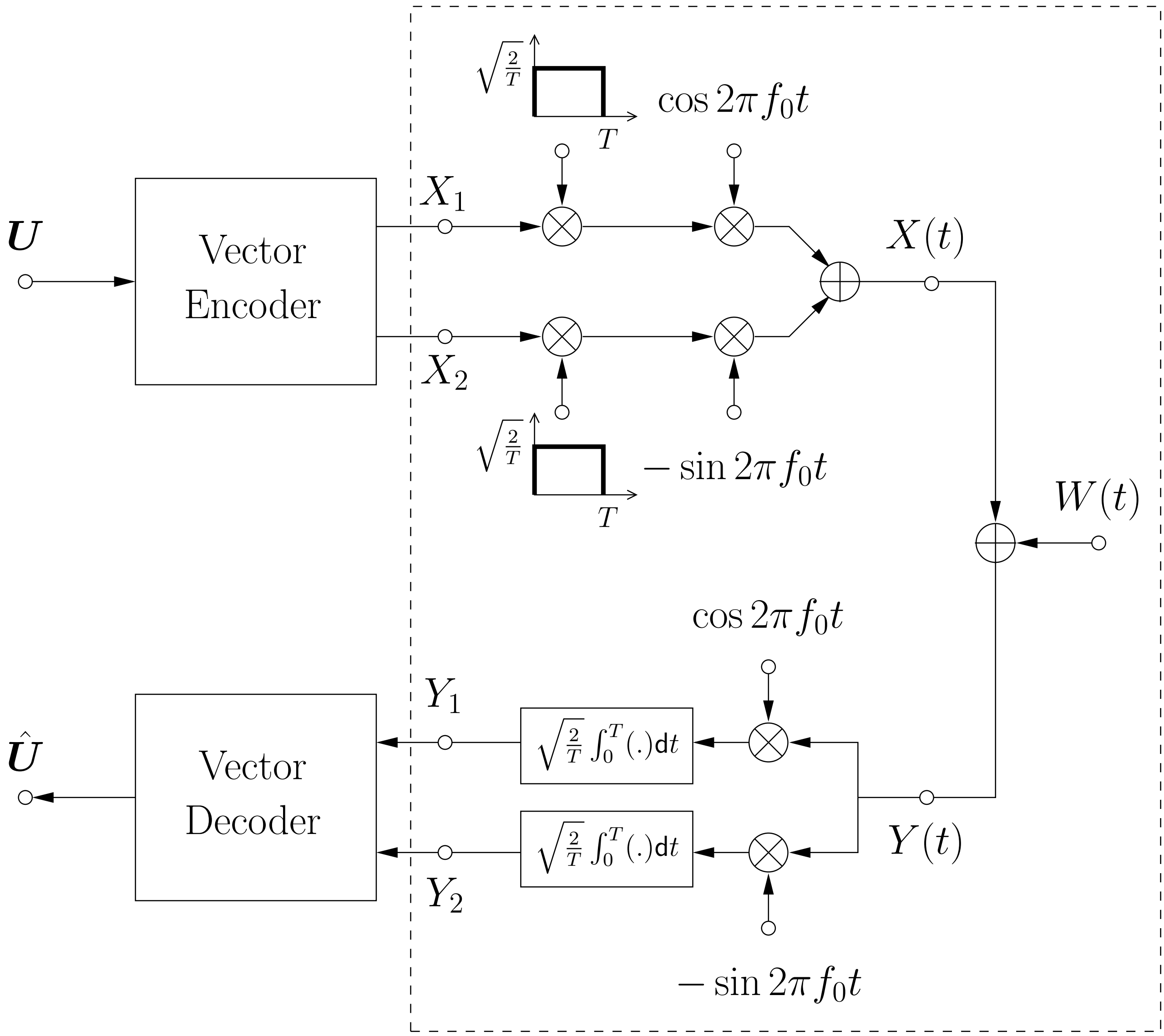


General Case

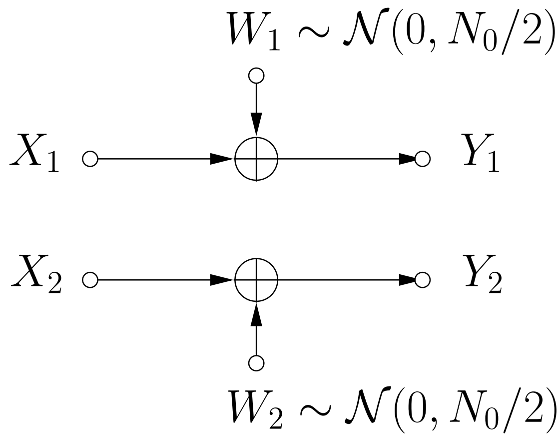
The decision regions for the general case can be determined in a straight-forward way:



5.4 Vector Representation of the MPSK Transceiver



AWGN Vector Channel:



5.5 Symbol-Error Probability

BPSK

Since BPSK has the same signal constellation as BPAM, the two modulation schemes have the same symbol-error probability:

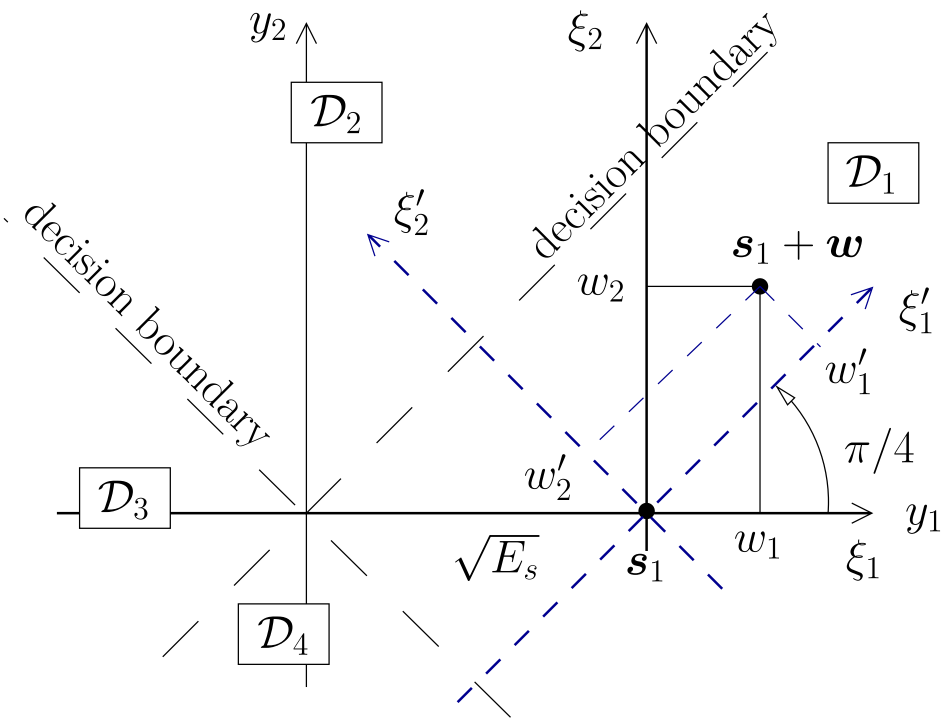
$$P_s = \mathrm{Q}\left(\sqrt{\tfrac{2E_s}{N_0}}\right) = \mathrm{Q}(\sqrt{2\gamma_s}).$$

QPSK

Assume that $\boldsymbol{X} = \boldsymbol{s}_1$ is transmitted. Then, the conditional symbol-error probability is given by

$$\begin{aligned}
\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= \Pr(\boldsymbol{y} \notin \mathcal{D}_1 | \boldsymbol{X} = \boldsymbol{s}_1) \\
&= 1 - \Pr(\boldsymbol{y} \in \mathcal{D}_1 | \boldsymbol{X} = \boldsymbol{s}_1) \\
&= 1 - \Pr(\boldsymbol{s}_1 + \boldsymbol{W} \in \mathcal{D}_1).
\end{aligned}$$

For computing this probability, the noise vector is projected on the rotated coordinate system (ξ_1', ξ_2') (compare BPPM):



The probability may then be written as

$$\begin{aligned}
\Pr(\boldsymbol{s}_1 + \boldsymbol{W} \in \mathcal{D}_1) &= \Pr\Big(W_1' > -\sqrt{\tfrac{E_s}{2}} \quad \text{and} \quad W_2' < \sqrt{\tfrac{E_s}{2}}\Big) \\
&= \Pr\Big(W_1' > -\sqrt{\tfrac{E_s}{2}}\Big) \cdot \Pr\Big(W_2' < \sqrt{\tfrac{E_s}{2}}\Big) \\
&= \Pr\Big(\tfrac{W_1'}{\sqrt{N_0/2}} > -\sqrt{\tfrac{E_s}{N_0}}\Big) \cdot \Pr\Big(\tfrac{W_2'}{\sqrt{N_0/2}} < \sqrt{\tfrac{E_s}{N_0}}\Big) \\
&= \Big(1 - \mathrm{Q}(\sqrt{\gamma_s})\Big)\Big(1 - \mathrm{Q}(\sqrt{\gamma_s})\Big) \\
&= \Big(1 - \mathrm{Q}(\sqrt{\gamma_s})\Big)^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= 1 - \Big(1 - \mathrm{Q}(\sqrt{\gamma_s})\Big)^2 \\
&= 2\,\mathrm{Q}(\sqrt{\gamma_s}) - \mathrm{Q}^2(\sqrt{\gamma_s}).
\end{aligned}$$

By symmetry, the four conditional symbol-error probabilities are identical, and thus the average symbol-error probability results as

$$P_s = 2\,\mathrm{Q}(\sqrt{\gamma_s}) - \mathrm{Q}^2(\sqrt{\gamma_s}).$$

General Case

A closed-form expression for P_s does not exist. It can be upper-bounded using an union-bound technique, or it can be numerically computed using an integral expression. (See literature.) A plot of P_s versus the SNR can be found in [1, p. 416].

Remark: Union Bound

For two random events A and B, we have

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \underbrace{\Pr(A \cap B)}_{\leq 0} \leq \Pr(A) + \Pr(B).$$

The expression on the right-hand side is called the union bound for the probability on the left-hand side. In a similar way, we obtain the union bound for multiple events:

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_M) \leq \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_M).$$

Each of these events may denote the case that the received vector is within a certain decision region.

5.6 Bit-Error Probability

BPSK

Since $E_b = E_s$, $\gamma_s = \gamma_b$, $P_b = P_s$, the bit-error probability is given by

$$P_b = \mathrm{Q}(\sqrt{2\gamma_b}).$$

QPSK

For Gray encoding, P_b can directly be computed.

Assume first that $\boldsymbol{X} = \boldsymbol{s}_1$ is transmitted, which corresponds to $[U_1, U_2] = [0, 0]$. Then, the conditional bit-error probability is given by

$$\begin{aligned}
&\Pr(\hat{U} \neq U | \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\qquad = \tfrac{1}{2}\Big[\Pr(\hat{U}_1 \neq U_1 | \boldsymbol{X} = \boldsymbol{s}_1) + \Pr(\hat{U}_2 \neq U_2 | \boldsymbol{X} = \boldsymbol{s}_1)\Big] \\
&\qquad = \tfrac{1}{2}\Big[\Pr(\hat{U}_1 \neq 0 | \boldsymbol{X} = \boldsymbol{s}_1) + \Pr(\hat{U}_2 \neq 0 | \boldsymbol{X} = \boldsymbol{s}_1)\Big] \\
&\qquad = \tfrac{1}{2}\Big[\Pr(\hat{U}_1 = 1 | \boldsymbol{X} = \boldsymbol{s}_1) + \Pr(\hat{U}_2 = 1 | \boldsymbol{X} = \boldsymbol{s}_1)\Big].
\end{aligned}$$

From the direct decision rule for data bits, we have

$$\begin{aligned}
\hat{U}_1 = 1 \qquad &\Leftrightarrow \qquad \boldsymbol{Y} \in \mathcal{D}_3 \cup \mathcal{D}_4, \\
\hat{U}_2 = 1 \qquad &\Leftrightarrow \qquad \boldsymbol{Y} \in \mathcal{D}_2 \cup \mathcal{D}_3.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\Pr(\hat{U} \neq U | \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\qquad = \tfrac{1}{2}\Big[\Pr(\boldsymbol{Y} \in \mathcal{D}_3 \cup \mathcal{D}_4) + \Pr(\boldsymbol{Y} \in \mathcal{D}_2 \cup \mathcal{D}_3)\Big] \\
&\qquad = \tfrac{1}{2}\Big[\Pr\Big(W_1' < -\sqrt{\tfrac{E_s}{2}}\Big) + \Pr\Big(W_2' > \sqrt{\tfrac{E_s}{2}}\Big)\Big] \\
&\qquad = \tfrac{1}{2}\Big[\Pr\Big(\tfrac{W_1'}{\sqrt{N_0/2}} < -\sqrt{\tfrac{E_s}{N_0}}\Big) + \Pr\Big(\tfrac{W_2'}{\sqrt{N_0/2}} > \sqrt{\tfrac{E_s}{N_0}}\Big)\Big] \\
&\qquad = \mathrm{Q}(\sqrt{\gamma_s}).
\end{aligned}
$$

By symmetry, all four conditional bit-error probabilities are identical. Since two data bits are transmitted per modulation symbol (waveform) in QPSK, we have the relations $E_s = 2E_b$ and $\gamma_s = 2\gamma_b$. Therefore,

$$P_b = \mathrm{Q}(\sqrt{2\gamma_b}).$$

(Remember: $P_b \approx 10^{-5}$ for $\gamma_b \approx 9.6$ dB.)

Comparison of QPSK and BPSK

(i) QPSK has exactly the same bit-error probability as BPSK.

(ii) Relation between symbol duration T and (average) bit duration $T_b = T/K$:

$$\begin{aligned} \text{BPSK:} \quad & T = T_b, \\ \text{QPSK:} \quad & T = 2T_b. \end{aligned}$$

Hence, for a fixed symbol duration T, the bit rate of QPSK is twice as fast as that of BPSK.

General Case

Since Gray encoding is used, we have

$$P_b \approx \frac{1}{\log_2 M} P_s$$

for high SNR.

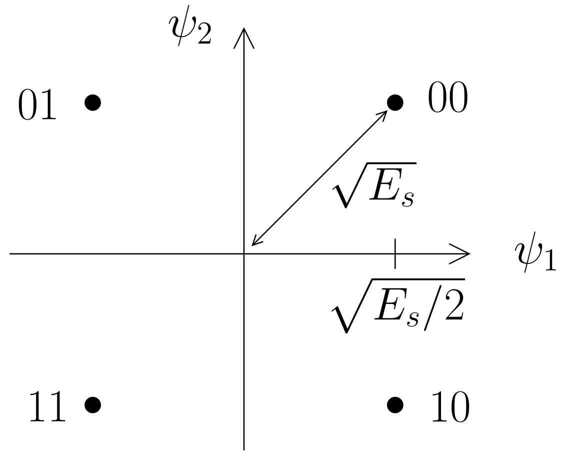
The symbol-error probability may be determined using one of the methods mentioned above.

6 Offset Quadrature Phase-Shift Keying (OQPSK)

Offset QPSK is a variant of QPSK. It has the same performance as QPSK, but the requirements on the transmitter and receiver hardware are lower.

6.1 Signaling Scheme

Consider QPSK with the following **“tilted” signal constellation** and Gray encoding:



The phase of QPSK can change by maximal π. This is the case for instance, when two consecutive 2-bit blocks are $[00], [11]$ or $[01], [10]$. These phase changes put high requirements on the dynamic behavior of the RF-part of the QPSK transmitter and receiver.

Phase changes of π can be avoided by staggering the quadrature branch of QPSK by $T/2 = T_b$. The resulting modulation scheme is called **staggered QPSK** or **offset QPSK** (OQPSK).

Without significant loss of generality, we assume

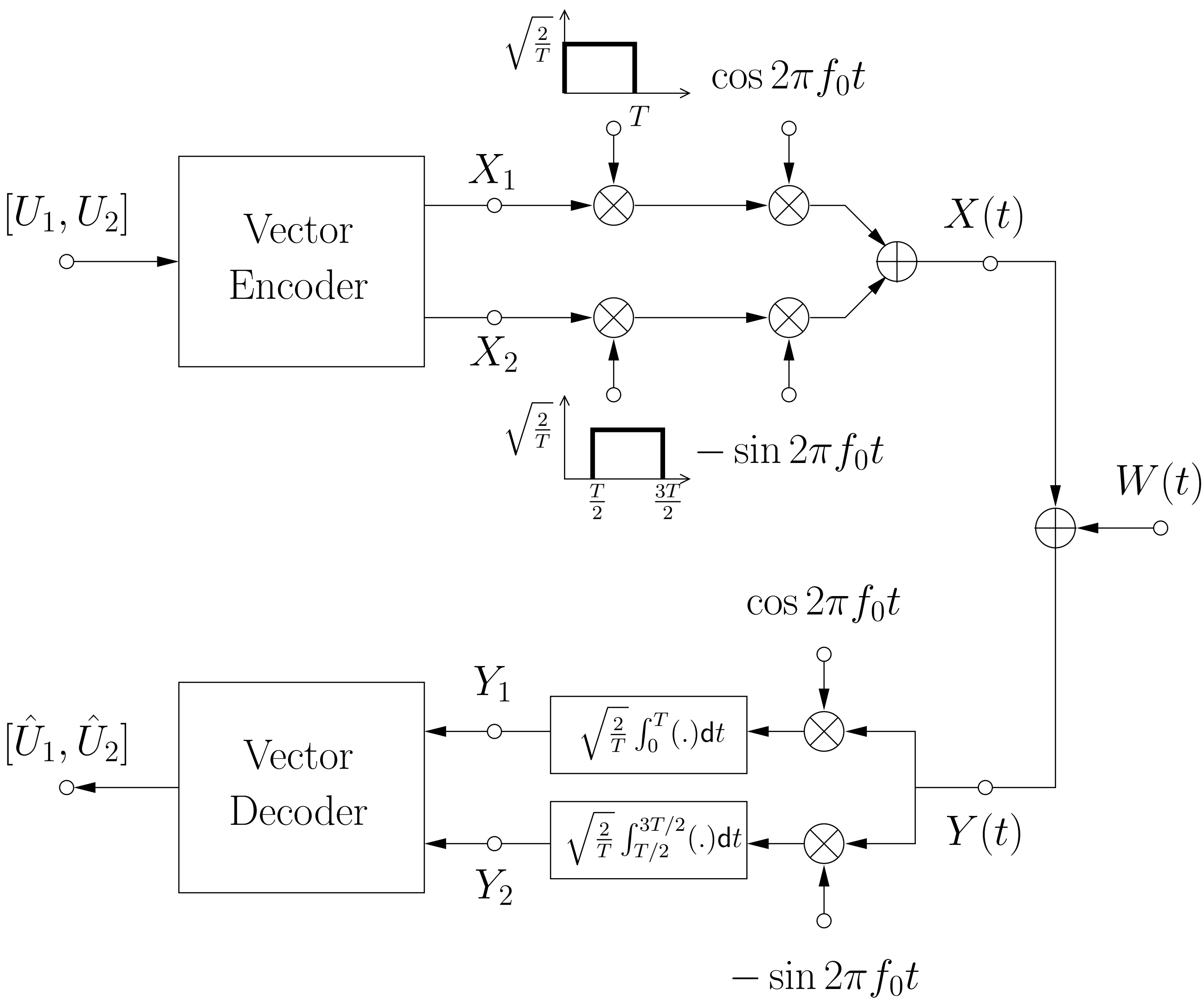
$$f_0 = \frac{2L}{T} \quad \text{for some } L \gg 1,\ L \in \mathbb{N}.$$

Remark

The ψ_1 component of the modulation symbol is called the in-phase component (I), and the ψ_2 component is called the quadrature component (Q).

6.2 Transceiver

We consider communication over an AWGN channel.



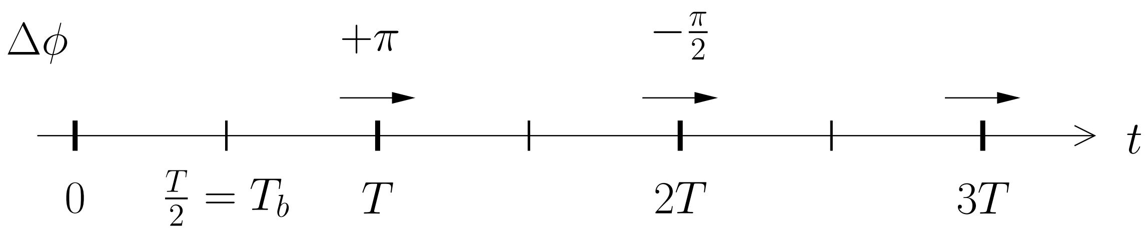
6.3 Phase Transitions for QPSK and OQPSK

EXAMPLE: Transmission of three symbols

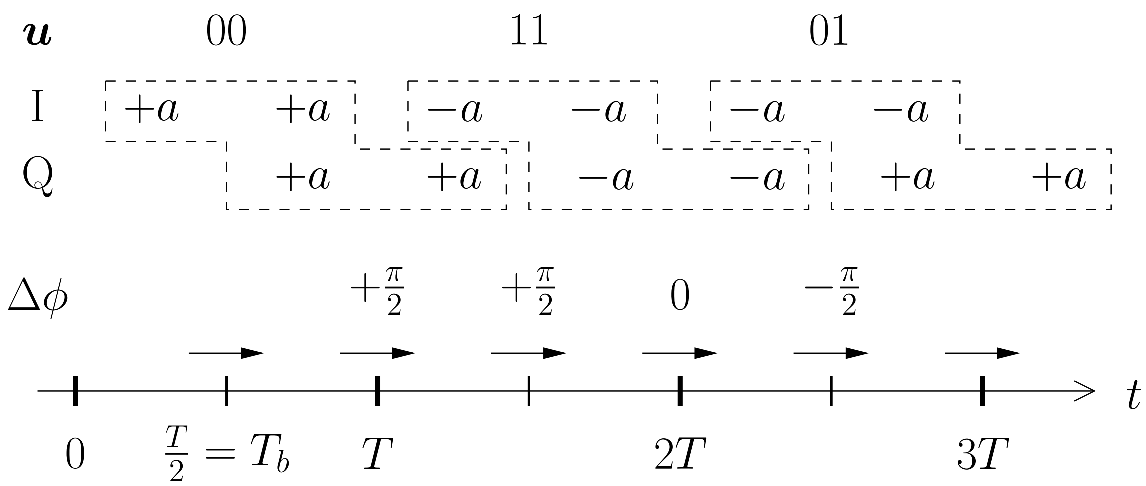
For the transmission of three symbols, we consider the data bits $\boldsymbol{u} = [u_1, u_2]$, the in-phase component (I) and the quadrature component (Q) of the modulation symbol, and the phase change $\Delta\phi$. For convenience, let $a := \sqrt{E_s/2}$.

QPSK

$\boldsymbol{u}$	00	11	01
I	$+a$	$-a$	$-a$
Q	$+a$	$-a$	$+a$



OQPSK



◇

QPSK exhibits phase changes $\{\pm\frac{\pi}{2}, \pm\pi\}$ at times that are multiples of the signaling interval T. In contrast to this, OQPSK exhibits phase changes $\{\pm\frac{\pi}{2}\}$ at times that are multiples of the bit duration $T_b = T/2$.

Remarks

(i) OQPSK has the same bit-error probability as QPSK.

(ii) Staggering the quadrature component of QPSK by $T/2$ does not prevent phase changes of π if the signaling constellation of QPSK is not tilted. (Why?)

7 Frequency-Shift Keying (FSK)

In FSK, the waveforms are sines, and the information is conveyed by the frequencies of the waveforms. FSK with two waveforms is called binary FSK (BFSK), and FSK with M waveforms is called M-ary FSK (MFSK).

7.1 BFSK with Coherent Demodulation

7.1.1 Signal Waveforms

Set $\mathbb{S}$ of two sinus waveforms with different frequencies f_1, f_2 and (possibly) different but known phases $\phi_1, \phi_2 \in [0, 2\pi)$:

$$s_1(t) = \sqrt{2P}\cos\big(2\pi f_1 t + \phi_1\big),$$
$$s_2(t) = \sqrt{2P}\cos\big(2\pi f_2 t + \phi_2\big).$$

Both values are limited to the time interval $t \in [0, T]$. Thus we have

$$\mathbb{S} = \big\{s_1(t), s_2(t)\big\}.$$

Usually, $f_1, f_2 \gg 1/T$. To simplify the theoretical analysis, we assume that

$$f_1 = \frac{k_1}{T}, \qquad f_2 = \frac{k_2}{T},$$

where $k_1, k_2 \in \mathbb{N}$ and $k_1, k_2 \gg 1$.

Both waveforms have the same energy

$$E_s = \int_{-\infty}^{\infty} s_1^2(t)\mathsf{d}t = \int_{-\infty}^{\infty} s_2^2(t)\mathsf{d}t = PT.$$

We assume the following **BFSK transmitter**:

$$\begin{aligned} \{0,1\} &\rightarrow \mathbb{S} \\ u &\mapsto x(t) \end{aligned}$$

with the mapping

$$\begin{aligned} 0 &\mapsto s_1(t) \\ 1 &\mapsto s_2(t). \end{aligned}$$

The **frequency separation**

$$\Delta f = |f_1 - f_2|$$

is selected such that $s_1(t)$ and $s_2(t)$ are orthogonal, i.e.,

$$\langle s_1(t), s_2(t) \rangle = \int_0^T s_1(t) s_2(t) \mathsf{d}t = 0.$$

The necessary condition for orthogonality is

$$\Delta f = \frac{k}{T}$$

with $k \in \mathbb{N}$. Notice that the minimum frequency separation is $1/T$.

The phases are assumed to be known to the receiver. Demodulation *with* phase information (known phases or estimated phases) is called **coherent demodulation**.

7.1.2 Signal Constellation

Orthonormal functions:

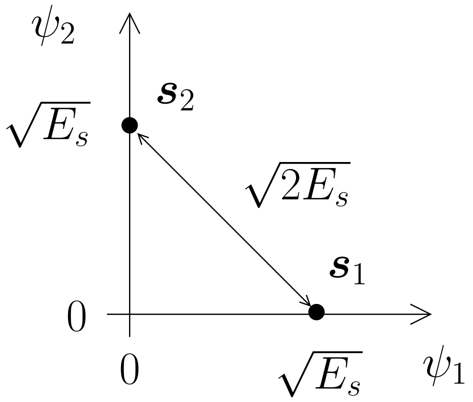
$$\psi_1(t) = \tfrac{1}{\sqrt{E_s}} s_1(t) = \sqrt{\tfrac{2}{T}} \cos\bigl(2\pi f_1 t + \phi_1\bigr)$$
$$\psi_2(t) = \tfrac{1}{\sqrt{E_s}} s_2(t) = \sqrt{\tfrac{2}{T}} \cos\bigl(2\pi f_2 t + \phi_2\bigr)$$

for $t \in [0, T]$.

Geometrical representation of the waveforms:

$$s_1(t) = \sqrt{E_s} \cdot \psi_1(t) + 0 \cdot \psi_2(t) \quad \longleftrightarrow \quad \boldsymbol{s}_1 = (\sqrt{E_s}, 0)^{\mathsf{T}}$$
$$s_2(t) = 0 \cdot \psi_1(t) + \sqrt{E_s} \cdot \psi_2(t) \quad \longleftrightarrow \quad \boldsymbol{s}_1 = (0, \sqrt{E_s})^{\mathsf{T}}.$$

Signal constellation:



BFSK with coherent demodulation has the same signal constellation as BPPM.

Accordingly, these two modulation methods are equivalent.

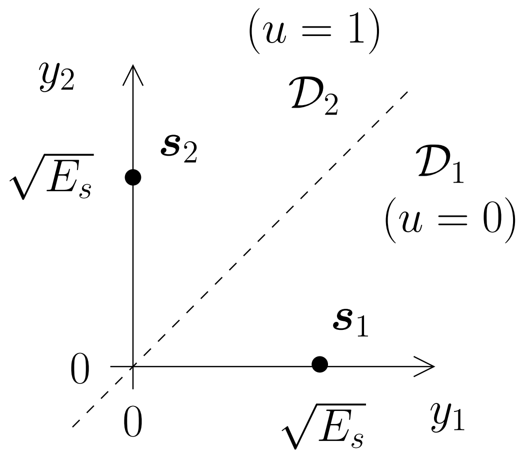
7.1.3 ML Decoding

ML symbol decision rule

$$\mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) = \operatorname*{argmin}_{x\in\{\boldsymbol{s}_1,\boldsymbol{s}_2\}} \|\boldsymbol{y}-\boldsymbol{x}\|^2 = \begin{cases} (\sqrt{E_s},0)^{\mathsf{T}} & \text{for } y_1 \geq y_2, \\ (0,\sqrt{E_s})^{\mathsf{T}} & \text{for } y_1 < y_2. \end{cases}$$

Decision regions

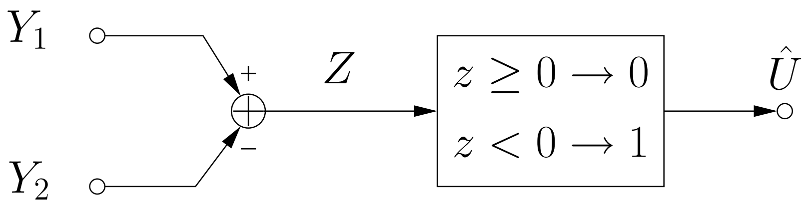
$$\mathcal{D}_1 = \{\boldsymbol{y}\in\mathbb{R}^2 : y_1 \geq y_2\}, \qquad \mathcal{D}_2 = \{\boldsymbol{y}\in\mathbb{R}^2 : y_1 < y_2\}.$$



ML decoder for BFSK

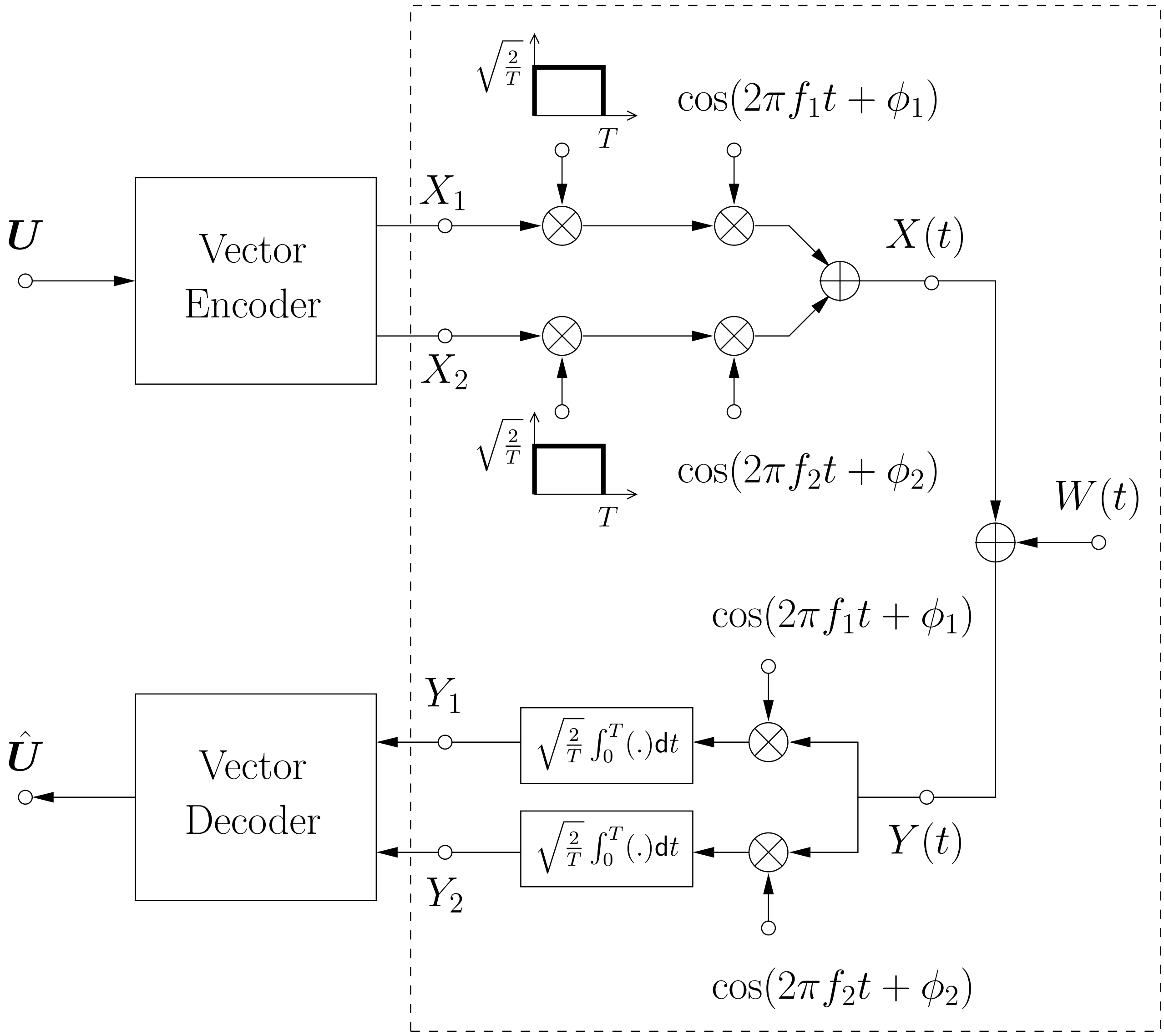
$$\hat{u} = \mathrm{enc}^{-1}(\mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y})) = \begin{cases} 0 & \text{for } y_1 \geq y_2, \\ 1 & \text{for } y_1 < y_2. \end{cases}$$

Implementation of the ML decoder



7.1.4 Vector Representation of the Transceiver

We assume BFSK over the AWGN channel and coherent demodulation.



The dashed box forms a 2-dimensional AWGN vector channel with independent noise components

$$W_1, W_2 \sim \mathcal{N}(0, N_0/2).$$

Remark: Notice the similarities to BPSK and BPPM.

7.1.5 Error Probabilities

As the signal constellation of coherently demodulated BFSK is the same as that of BPPM, the error probabilities are the same:

$$
\begin{aligned}
P_s &= \mathrm{Q}(\sqrt{\gamma_s}), \\
P_b &= \mathrm{Q}(\sqrt{\gamma_b}).
\end{aligned}
$$

For the derivation, see analysis of BPPM.

7.2 MFSK with Coherent Demodulation

7.2.1 Signal Waveforms

The signal waveforms are

$$s_m(t) = \begin{cases} \sqrt{2P}\cos\bigl(2\pi f_m t + \phi_m\bigr) & \text{for } t \in [0,T], \\ 0 & \text{elsewhere,} \end{cases}$$

$m = 1, 2, \ldots, M$.

The parameters of the waveforms are as follows:

(a) frequencies $f_m = k_m/T$ for $k_m \in \mathbb{N}$ and $k_m \gg 1$;

(b) frequency separations $\Delta f_{m,n} = |f_m - f_n| = k_{m,n}/T$, $k_{m,n} \in \mathbb{N}$;

(c) phases $\phi_m \in [0, 2\pi)$;

$m, n = 1, 2, \ldots, M$.

The phases are assumed to be known to the receiver. Thus, we consider coherent demodulation.

As in the binary case, all waveforms have the same energy

$$E_s = \int_{-\infty}^{\infty} s_m^2(t)\,\mathsf{d}t = PT.$$

7.2.2 Signal Constellation

Orthonormal functions:

$$\psi_m(t) = \tfrac{1}{\sqrt{E_s}} s_m(t) = \sqrt{\tfrac{2}{T}} \cos\big(2\pi f_m t + \phi_m\big)$$

for $t \in [0, T]$, $m = 1, 2, \ldots, M$.

Geometrical representation of the waveforms:

$$\begin{aligned} s_1(t) = \sqrt{E_s} \cdot \psi_1(t) \quad &\longleftrightarrow \quad \boldsymbol{s}_1 = (\overbrace{\sqrt{E_s}, 0, \ldots, 0}^{M \text{ elements}})^\mathsf{T} \\ &\qquad\qquad \vdots \\ s_M(t) = \sqrt{E_s} \cdot \psi_M(t) \quad &\longleftrightarrow \quad \boldsymbol{s}_1 = (0, \ldots, 0, \sqrt{E_s})^\mathsf{T}. \end{aligned}$$

MFSK with coherent demodulation has the same signal constellation as MPPM.

Accordingly, these two modulation methods are equivalent and have the same performance.

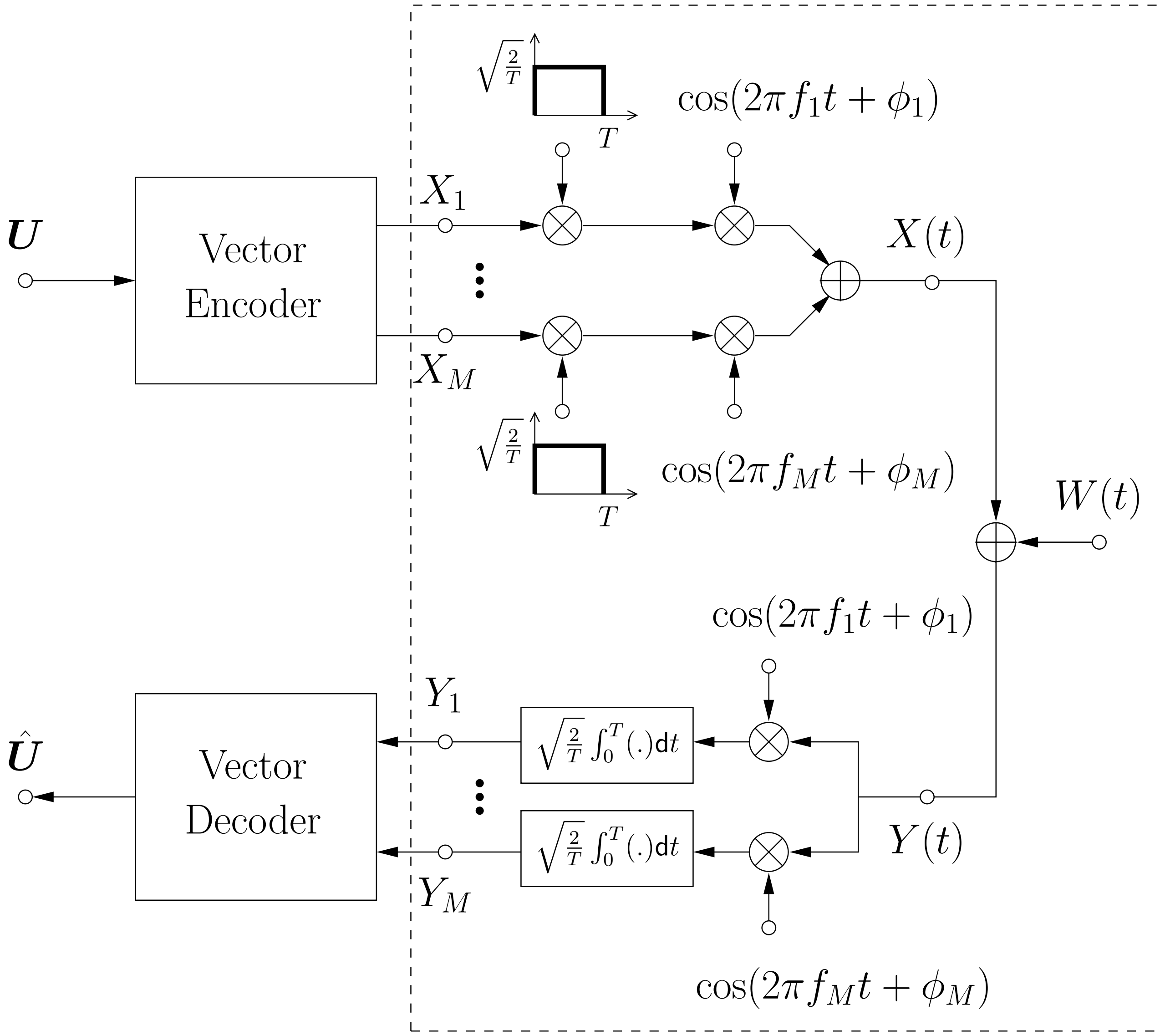
Vector Encoder

$$\begin{aligned} \text{enc}: \quad \boldsymbol{u} = [u_1, \ldots, u_K] \ &\mapsto \ \boldsymbol{x} = \text{enc}(\boldsymbol{u}) \\ \mathbb{U} = \{0, 1\}^K \ &\to \ \mathbb{X} = \big\{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M\big\} \end{aligned}$$

with $K = \log_2 M$.

7.2.3 Vector Representation of the Transceiver

We assume MFSK over the AWGN channel and coherent demodulation.



The dashed box forms an M-dimensional AWGN vector channel with mutually independent components

$$W_m \sim \mathcal{N}(0, N_0/2),$$

$m = 1, 2, \ldots, M$.

Remarks

The M waveforms $s_m(t) \in \mathbb{S}$ are orthogonal. The receiver simply correlates the received waveform $y(t)$ to each of the possibly transmitted waveforms $s_m(t) \in \mathbb{S}$ and selects the one with the highest correlation as the estimate.

7.2.4 Error Probabilities

The signal constellation of coherently demodulated MFSK is the same as that of MPPM. Therefor, the symbol-error probability and the bit-error probability are the same. For the analysis, see MPPM.

7.3 BFSK with Noncoherent Demodulation

7.3.1 Signal Waveforms

The signal waveforms are

$$s_1(t) = \sqrt{2P}\cos\bigl(2\pi f_1 t + \phi_1\bigr),$$
$$s_2(t) = \sqrt{2P}\cos\bigl(2\pi f_2 t + \phi_2\bigr).$$

for $t \in [0, T)$ with

(i) frequencies $f_m = k_m/T$ for $k_m \in \mathbb{N}$ and $k_m \gg 1$, $m = 1, 2$, and

(ii) frequency separation $\Delta f = |f_1 - f_2| = k/T$ for some $k \in \mathbb{N}$.

As opposed to the case of coherent demodulation, the phases ϕ_1 and ϕ_2 are not known to the receiver. Instead, they are assumed as random variables that are uniformly distributed, i.e.,

$$p_{\Phi_1}(\phi_1) = \frac{1}{2\pi}, \qquad\qquad p_{\Phi_2}(\phi_2) = \frac{1}{2\pi}$$

for $\phi_1, \phi_2 \in [0, 2\pi)$.

Demodulation *without* use of phase information is called **noncoherent demodulation**. As the phases need not be estimated, the noncoherent demodulator has a lower complexity than the coherent demodulator.

7.3.2 Signal Constellation

Using the same method as for the canonical decomposition of PSK, the **orthonormal functions** are obtained as

$$
\begin{aligned}
\psi_1(t) &= \sqrt{\tfrac{2}{T}}\cos\bigl(2\pi f_1 t\bigr)\\
\psi_2(t) &= -\sqrt{\tfrac{2}{T}}\sin\bigl(2\pi f_1 t\bigr)\\
\psi_3(t) &= \sqrt{\tfrac{2}{T}}\cos\bigl(2\pi f_2 t\bigr)\\
\psi_4(t) &= -\sqrt{\tfrac{2}{T}}\sin\bigl(2\pi f_2 t\bigr)
\end{aligned}
$$

for $t \in [0, T]$.

(Remember: $\cos(x+y) = \cos x\ \cos y - \sin x\ \sin y$.)

Geometrical representations of the waveforms:

$$
\begin{aligned}
s_1(t) &= \sqrt{E_s}\cos\phi_1 \cdot \psi_1(t) + \sqrt{E_s}\sin\phi_1 \cdot \psi_2(t)\\
&\longleftrightarrow \quad \boldsymbol{s}_1 = \bigl(\sqrt{E_s}\cos\phi_1, \sqrt{E_s}\sin\phi_1, 0, 0\bigr)^{\mathsf{T}},
\end{aligned}
$$

$$
\begin{aligned}
s_2(t) &= \sqrt{E_s}\cos\phi_2 \cdot \psi_3(t) + \sqrt{E_s}\sin\phi_2 \cdot \psi_4(t)\\
&\longleftrightarrow \quad \boldsymbol{s}_2 = \bigl(0, 0, \sqrt{E_s}\cos\phi_2, \sqrt{E_s}\sin\phi_2\bigr)^{\mathsf{T}}.
\end{aligned}
$$

The **signal constellation** is thus

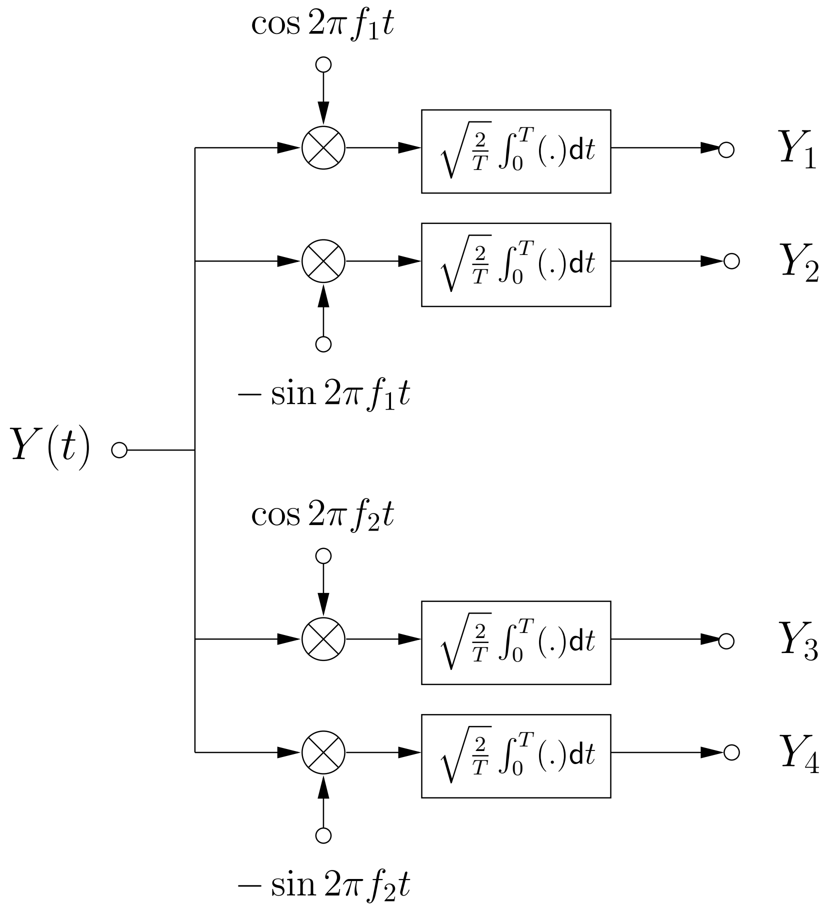
$$
\mathbb{X} = \bigl\{\boldsymbol{s}_1, \boldsymbol{s}_2\bigr\}.
$$

The **transmitted (modulation) symbols** are the four-dimensional(!) vectors

$$
\boldsymbol{X} = \bigl(\underbrace{X_1, X_2}_{\text{“for } \boldsymbol{s}_1\text{”}}, \underbrace{X_3, X_4}_{\text{“for } \boldsymbol{s}_2\text{”}}\bigr)^{\mathsf{T}} \in \mathbb{X}.
$$

7.3.3 Noncoherent Demodulator

As there are four orthonormal basis functions, the demodulator has four branches:



Notice the similarity to PSK.

We introduce the vectors

$$\boldsymbol{Y}_1 = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \qquad \boldsymbol{Y}_2 = \begin{pmatrix} Y_3 \\ Y_4 \end{pmatrix}$$

and

$$\boldsymbol{Y} = \left(\boldsymbol{Y}_1^{\mathsf{T}}, \boldsymbol{Y}_2^{\mathsf{T}}\right)^{\mathsf{T}} = \left(Y_1, Y_2, Y_3, Y_4\right)^{\mathsf{T}}.$$

(Loosely speaking, $\boldsymbol{Y}_1$ corresponds to $\boldsymbol{s}_1$ and $\boldsymbol{Y}_2$ corresponds to $\boldsymbol{s}_2$.)

7.3.4 Conditional Probability of Received Vector

Assume that the transmitted vector is

$$\boldsymbol{X} = \boldsymbol{s}_1 = \big(\sqrt{E_s}\cos\phi_1, \sqrt{E_s}\sin\phi_1, 0, 0\big)^{\mathsf{T}}.$$

Then, the received vector is

$$\begin{aligned}\boldsymbol{Y} &= \big(\underbrace{\sqrt{E_s}\cos\phi_1 + W_1, \sqrt{E_s}\sin\phi_1 + W_2}_{\boldsymbol{Y}_1^{\mathsf{T}}}, \underbrace{W_3, W_4}_{\boldsymbol{Y}_2^{\mathsf{T}}}\big)^{\mathsf{T}} \\ &= \big(\boldsymbol{Y}_1^{\mathsf{T}}, \boldsymbol{Y}_2^{\mathsf{T}}\big)^{\mathsf{T}}.\end{aligned}$$

The conditional probability density of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{s}_1$ and ϕ_1 is

$$\begin{aligned}&p_{\boldsymbol{Y}|\boldsymbol{X},\Phi_1}(\boldsymbol{y}|\boldsymbol{s}_1, \phi_1) = \\ &\quad= \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\big(-\tfrac{1}{N_0}\|\boldsymbol{y} - \boldsymbol{s}_1\|^2\big) \\ &\quad= \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\Big(-\tfrac{1}{N_0}\Big[\big(y_1 - \sqrt{E_s}\cos\phi_1\big)^2 + \\ &\qquad\qquad + \big(y_2 - \sqrt{E_s}\sin\phi_1\big)^2 + y_3^2 + y_4^2\Big]\Big) \\ &\quad= \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\big(-\tfrac{E_s}{N_0}\big) \cdot \exp\Big(-\tfrac{1}{N_0}\Big[\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2\Big]\Big) \cdot \\ &\qquad \cdot \exp\Big(\tfrac{2\sqrt{E_s}}{N_0}\big(y_1\cos\phi_1 + y_2\sin\phi_1\big)\Big)\end{aligned}$$

With $\alpha_1 = \tan^{-1}(y_2/y_1)$, we may write

$$\begin{aligned}y_1\cos\phi_1 + y_2\sin\phi_1 &= \|\boldsymbol{y}_1\|\ \big(\cos\alpha_1\cos\phi_1 + \sin\alpha_1\sin\phi_1\big) \\ &= \|\boldsymbol{y}_1\|\ \cos(\phi_1 - \alpha_1).\end{aligned}$$

The phase can be removed from the condition by averaging over it. The phase is uniformly distributed in $[0, 2\pi)$, and thus

$$
\begin{aligned}
p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_1) &= \\
&= \int_0^{2\pi} p_{\boldsymbol{Y}|\boldsymbol{X},\Phi_1}(\boldsymbol{y}|\boldsymbol{s}_1,\phi_1)\; p_{\phi_1}(\phi_1)\,\mathsf{d}\phi_1 \\
&= \tfrac{1}{2\pi}\int_0^{2\pi} p_{\boldsymbol{Y}|\boldsymbol{X},\Phi_1}(\boldsymbol{y}|\boldsymbol{s}_1,\phi_1)\,\mathsf{d}\phi_1 \\
&= \tfrac{1}{2\pi}\int_0^{2\pi} \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\big(-\tfrac{E_s}{N_0}\big)\cdot \exp\Big(-\tfrac{1}{N_0}\Big[\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2\Big]\Big)\cdot \\
&\qquad \cdot \exp\Big(\tfrac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_1\|\cos(\phi_1-\alpha_1)\Big)\,\mathsf{d}\phi_1 \\
&= \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\big(-\tfrac{E_s}{N_0}\big)\cdot \exp\Big(-\tfrac{1}{N_0}\Big[\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2\Big]\Big)\cdot \\
&\qquad \cdot \underbrace{\tfrac{1}{2\pi}\int_0^{2\pi} \exp\Big(\tfrac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_1\|\cos(\phi_1-\alpha_1)\Big)\,\mathsf{d}\phi_1}_{I_0\left(\frac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_1\|\right)}.
\end{aligned}
$$

The above function is the modified Bessel function of order 0, and it is defined as

$$
I_0(v) := \tfrac{1}{2\pi}\int_0^{2\pi} \exp\big(v\,\cos(\phi-\alpha)\big)\,\mathsf{d}\phi
$$

for $\alpha \in [0, 2\pi)$. It is monotonically increasing. A plot can be found in [1, p. 403].

Combining all the above results yields

$$
\begin{aligned}
p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_1) = \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\big(-\tfrac{E_s}{N_0}\big)\cdot \\
\cdot \exp\Big(-\tfrac{1}{N_0}\Big[\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2\Big]\Big) \cdot I_0\big(\tfrac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_1\|\big).
\end{aligned}
$$

Using the same reasoning, we obtain

$$
\begin{aligned}
p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_2) = \big(\tfrac{1}{\pi N_0}\big)^2 \cdot \exp\big(-\tfrac{E_s}{N_0}\big)\cdot \\
\cdot \exp\Big(-\tfrac{1}{N_0}\Big[\|\boldsymbol{y}_1\|^2 + \|\boldsymbol{y}_2\|^2\Big]\Big) \cdot I_0\big(\tfrac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_2\|\big).
\end{aligned}
$$

7.3.5 ML Decoding

Remember the definitions

$$
\boldsymbol{Y} = \big(\underbrace{Y_1, Y_2}_{\boldsymbol{Y}_1}, \underbrace{Y_3, Y_4}_{\boldsymbol{Y}_2}\big)^{\mathsf{T}} = \big(\boldsymbol{Y}_1^{\mathsf{T}}, \boldsymbol{Y}_2^{\mathsf{T}}\big)^{\mathsf{T}}.
$$

ML symbol decision rule

$$
\hat{\boldsymbol{x}} = \mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) = \operatorname*{argmax}_{\boldsymbol{x}\in\mathbb{X}} \; p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})
$$

Let $\hat{\boldsymbol{x}} = \boldsymbol{s}_{\hat{m}}$. Then, the index of the ML vector is

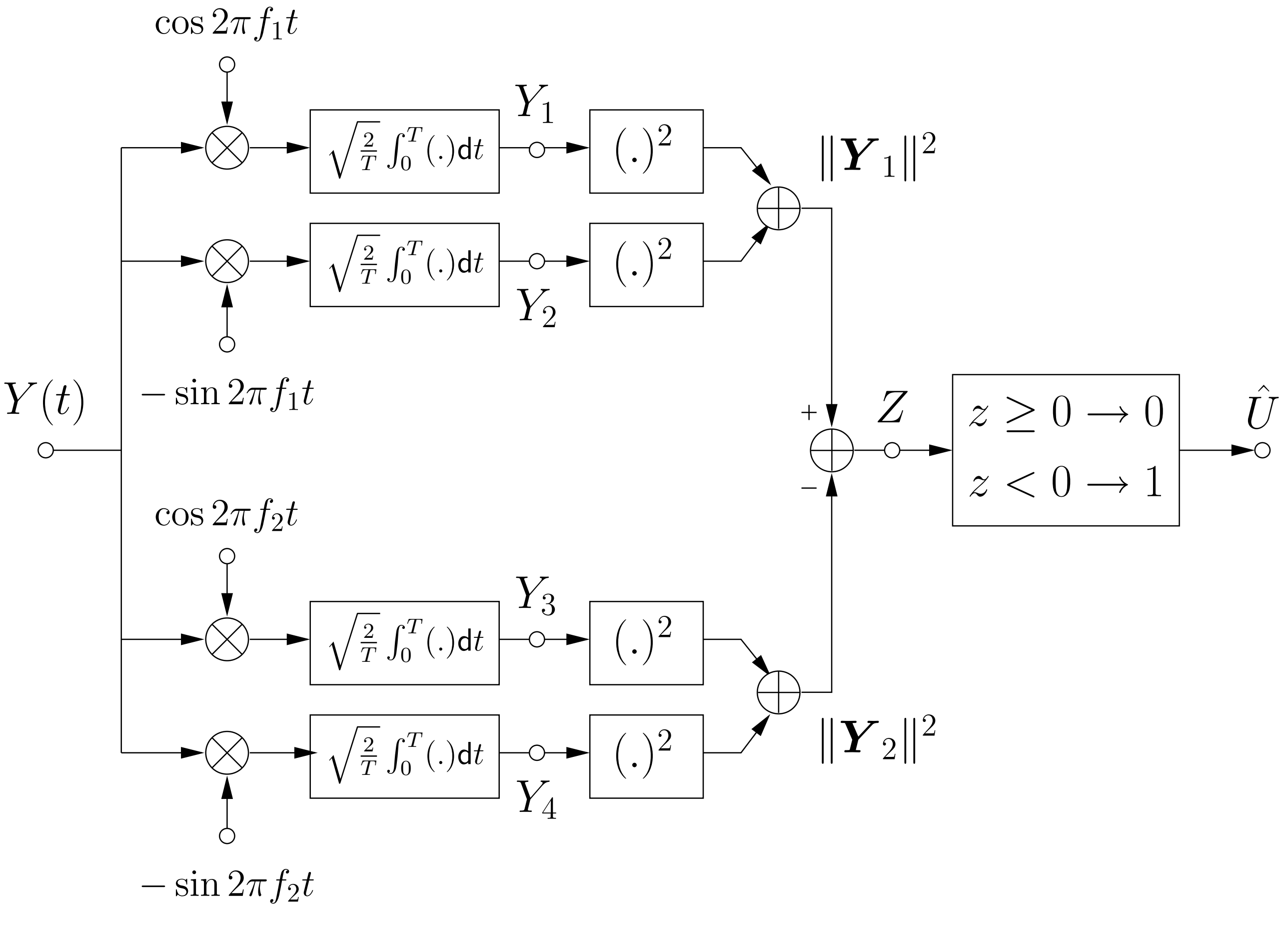
$$
\begin{aligned}
\hat{m} &= \operatorname*{argmax}_{m\in\{1,2\}} \; I_0\big(\tfrac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_m\|\big) \\
&= \operatorname*{argmax}_{m\in\{1,2\}} \|\boldsymbol{y}_m\| \\
&= \operatorname*{argmax}_{m\in\{1,2\}} \|\boldsymbol{y}_m\|^2.
\end{aligned}
$$

Notice that selecting $\hat{\boldsymbol{x}}$ and selecting $\hat{m}$ are equivalent.

ML decoder

$$\hat{u} = \begin{cases} 0 & \text{for } \hat{\boldsymbol{x}} = \boldsymbol{s}_1 \Leftrightarrow \|\boldsymbol{y}_1\| \geq \|\boldsymbol{y}_2\|, \\ 1 & \text{for } \hat{\boldsymbol{x}} = \boldsymbol{s}_2 \Leftrightarrow \|\boldsymbol{y}_1\| < \|\boldsymbol{y}_2\|. \end{cases}$$

Optimal noncoherent receiver for BFSK



Notice:

$$\|\boldsymbol{Y}_1\|^2 = \sqrt{Y_1^2 + Y_2^2}, \qquad Z = \|\boldsymbol{Y}_1\|^2 - \|\boldsymbol{Y}_2\|^2,$$

$$\|\boldsymbol{Y}_2\|^2 = \sqrt{Y_3^2 + Y_4^2}.$$

7.4 MFSK with Noncoherent Demodulation

7.4.1 Signal Waveforms

The signal waveforms are

$$s_m(t) = \sqrt{2P}\ \cos\bigl(2\pi f_m t + \phi_m\bigr)$$

for $t \in [0, T)$, $m = 1, 2, \ldots, M$, with

(i) $f_m = k_m/T$ for $k_m \in \mathbb{N}$ and $k_m \gg 1$, $m = 1, 2, \ldots, M$,

(ii) $\Delta f m, n = |f_m - f_n| = k/T$ for some $k_{m,n} \in \mathbb{N}$, $m, n = 1, 2, \ldots, M$.

The phases $\phi_1, \phi_2, \ldots, \phi_M$ are not known to the receiver. They are assumed to be uniformly distributed in $[0, 2\pi)$.

7.4.2 Signal Constellation

Orthonormal functions:

$$\begin{aligned}
\psi_1(t) &= \sqrt{\tfrac{2}{T}}\cos\bigl(2\pi f_1 t\bigr)\\
\psi_2(t) &= -\sqrt{\tfrac{2}{T}}\sin\bigl(2\pi f_1 t\bigr)\\
&\vdots\\
\psi_{2m-1}(t) &= \sqrt{\tfrac{2}{T}}\cos\bigl(2\pi f_m t\bigr)\\
\psi_{2m}(t) &= -\sqrt{\tfrac{2}{T}}\sin\bigl(2\pi f_m t\bigr)\\
&\vdots\\
\psi_{2M-1}(t) &= \sqrt{\tfrac{2}{T}}\cos\bigl(2\pi f_M t\bigr)\\
\psi_{2M}(t) &= -\sqrt{\tfrac{2}{T}}\sin\bigl(2\pi f_M t\bigr)
\end{aligned}$$

for $t \in [0, T]$.

Signal vectors (of length $2M$):

$$\begin{aligned}
\boldsymbol{s}_1 &= \bigl(\sqrt{E_s}\cos\phi_1, \sqrt{E_s}\sin\phi_1, 0, \dots, 0\bigr)^{\mathsf{T}}\\
\boldsymbol{s}_2 &= \bigl(0, 0, \sqrt{E_s}\cos\phi_2, \sqrt{E_s}\sin\phi_2, 0, \dots, 0\bigr)^{\mathsf{T}}\\
&\vdots\\
\boldsymbol{s}_m &= \bigl(0, \dots, 0, \underbrace{\sqrt{E_s}\cos\phi_m, \sqrt{E_s}\sin\phi_m}_{\text{positions } 2m-1 \text{ and } 2m}, 0, \dots, 0\bigr)^{\mathsf{T}}\\
&\vdots\\
\boldsymbol{s}_M &= \bigl(0, \dots, 0, \sqrt{E_s}\cos\phi_M, \sqrt{E_s}\sin\phi_M\bigr)^{\mathsf{T}}
\end{aligned}$$

Signal constellation:

$$\mathbb{X} = \bigl\{\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_M\bigr\}.$$

7.4.3 Conditional Probability of Received Vector

Using the same reasoning as for BFSK, we can show that

$$
\begin{aligned}
p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{s}_m) = \left(\tfrac{1}{\pi N_0}\right)^M \cdot \exp\left(-\tfrac{E_s}{N_0}\right)\cdot \\
\cdot \exp\Bigl(-\tfrac{1}{N_0}\sum_{m=1}^{M}\|\boldsymbol{y}_m\|^2\Bigr) \cdot I_0\left(\tfrac{2\sqrt{E_s}}{N_0}\|\boldsymbol{y}_m\|\right)
\end{aligned}
$$

with the definition

$$
\boldsymbol{Y} = \bigl(\underbrace{Y_1, Y_2}_{\boldsymbol{Y}_1}, \underbrace{Y_3, Y_4}_{\boldsymbol{Y}_2}, \ldots, \underbrace{Y_{2M-1}, Y_{2M}}_{\boldsymbol{Y}_M},\bigr)^{\mathsf{T}} = \bigl(\boldsymbol{Y}_1^{\mathsf{T}}, \boldsymbol{Y}_2^{\mathsf{T}}, \ldots, \boldsymbol{Y}_M^{\mathsf{T}}\bigr)^{\mathsf{T}}.
$$

7.4.4 ML Decision Rule

Original ML decision rule:

$$
\hat{\boldsymbol{x}} = \mathrm{dec}_{\mathsf{ML}}(\boldsymbol{y}) = \underset{\boldsymbol{x}\in\mathbb{X}}{\operatorname{argmax}}\; p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})
$$

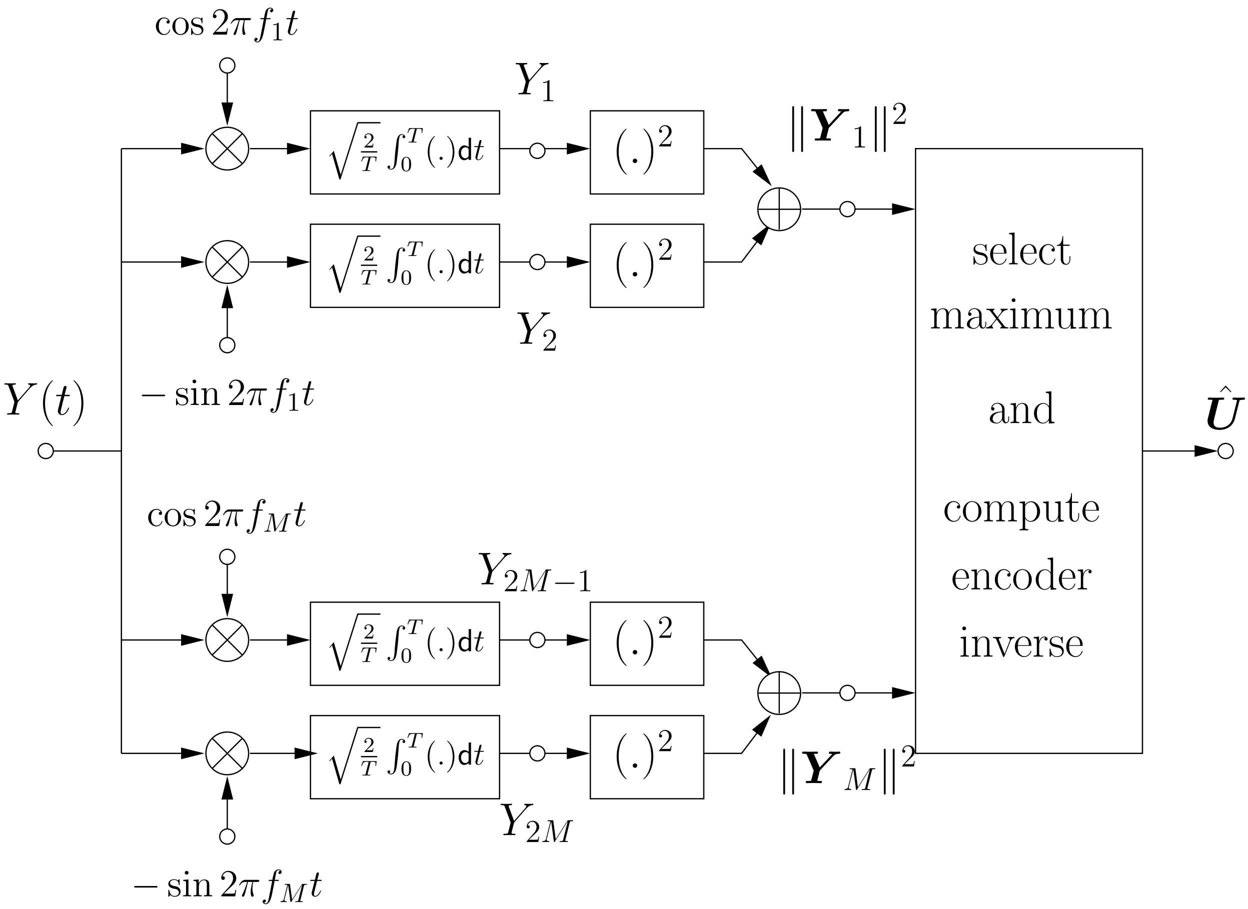
Let $\hat{\boldsymbol{x}} = \boldsymbol{s}_{\hat{m}}$. Then, the original decision rule may equivalently be expressed as

$$
\hat{m} = \underset{m\in\{1,\ldots,M\}}{\operatorname{argmax}}\; \|\boldsymbol{y}_m\|^2.
$$

7.4.5 Optimal Noncoherent Receiver for MFSK

Remember: "noncoherent demodulation" means that no phase information is used for demodulation.

The optimal noncoherent receiver for MFSK over the AWGN channel is as follows:



Notice that $\hat{\boldsymbol{U}} = [\hat{U}_1, \ldots, \hat{U}_K]$ with $K = \log_2 M$.

7.4.6 Symbol-Error Probability

Assume that $\boldsymbol{X} = \boldsymbol{s}_1$ is transmitted. Then, the probability of a correct decision is

$$\begin{aligned}
&\Pr(\hat{\boldsymbol{X}} = \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\qquad = \Pr\big(\|\boldsymbol{Y}_m\|^2 \leq \|\boldsymbol{Y}_1\|^2 \text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1\big).
\end{aligned}$$

Provided that $\boldsymbol{X} = \boldsymbol{s}_1$ is transmitted, we have

$$\begin{aligned}
\boldsymbol{Y}_1 &= \big(\sqrt{E_s}\cos\phi_1 + W_1, \sqrt{E_s}\sin\phi_1 + W_2\big)^\mathsf{T} \\
\boldsymbol{Y}_2 &= \big(W_3, W_4\big)^\mathsf{T} = \boldsymbol{W}_2 \\
&\vdots \\
\boldsymbol{Y}_m &= \big(W_{2m-1}, W_{2m}\big)^\mathsf{T} = \boldsymbol{W}_m \\
&\vdots \\
\boldsymbol{Y}_M &= \big(W_{2M-1}, W_{2M}\big)^\mathsf{T} = \boldsymbol{W}_M.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\Pr(\hat{\boldsymbol{X}} = \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\qquad = \Pr\big(\|\boldsymbol{W}_m\|^2 \leq \|\boldsymbol{Y}_1\|^2 \text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1\big) \\
&\qquad = \Pr\Big(\underbrace{\tfrac{1}{\sqrt{N_0/2}}\|\boldsymbol{W}_m\|^2}_{R_m} \leq \underbrace{\tfrac{1}{\sqrt{N_0/2}}\|\boldsymbol{Y}_1\|^2}_{R_1} \text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1\Big) \\
&\qquad = \Pr\big(R_m \leq R_1 \text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1\big) \\
&\qquad = \int_0^\infty \Pr\big(R_m \leq R_1 \text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1, R_1 = r_1\big) \cdot \\
&\qquad\qquad \cdot\, p_{R_1}(r_1)\, \mathsf{d}r_1.
\end{aligned}$$

The probability density function of R_1 is denoted by $p_{R_1}(r_1)$.

The probability of a correct decision can further be written as

$$\begin{aligned}
&\Pr(\hat{\boldsymbol{X}} = \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) = \\
&\quad = \int_0^\infty \Pr\big(R_m \le r_1 \;\text{ for all } m \neq 1 | \boldsymbol{X} = \boldsymbol{s}_1, R_1 = r_1\big)\, p_{R_1}(r_1)\, \mathsf{d}r_1 \\
&\quad = \int_0^\infty \Pr\big(R_m \le r_1 \;\text{ for all } m \neq 1\big)\, p_{R_1}(r_1)\, \mathsf{d}r_1
\end{aligned}$$

The noise vectors $W_2, \ldots, W_M$ are mutually statistically independent, and thus also the values $R_2, \ldots, R_M$ are mutually statistically independent. Therefore,

$$\begin{aligned}
\Pr(\hat{\boldsymbol{X}} = \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= \int_0^\infty \prod_{m=2}^{M} \Pr(R_m \le r_1)\, p_{R_1}(r_1)\, \mathsf{d}r_1 \\
&= \int_0^\infty \Big(\Pr(R_2 \le r_1)\Big)^{M-1} p_{R_1}(r_1)\, \mathsf{d}r_1.
\end{aligned}$$

Notice that

$$\Pr(R_m \le r_1) = \Pr(R_2 \le r_1)$$

for $m = 2, 3, \ldots, M$, because $R_2, \ldots, R_M$ are identically distributed.

The conditional probability of a symbol-error can thus be written as

$$\begin{aligned}\Pr(\hat{\boldsymbol{X}} \neq \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) &= 1 - \Pr(\hat{\boldsymbol{X}} = \boldsymbol{X} | \boldsymbol{X} = \boldsymbol{s}_1) \\ &= 1 - \int_0^\infty \Big(\Pr(R_2 \leq r_1)\Big)^{M-1} p_{R_1}(r_1)\, \mathsf{d}r_1.\end{aligned}$$

By symmetry, all conditional symbol-error probabilities are the same. Thus,

$$P_s = 1 - \int_0^\infty \Big(\Pr(R_2 \leq r_1)\Big)^{M-1} p_{R_1}(r_1)\, \mathsf{d}r_1.$$

In the above expressions, we have the following distributions (see literature):

(i) R_2 (and also $R_3, \ldots, R_M$) is **Rayleigh distributed**:

$$p_{R_2}(r_2) = r_2\, e^{-r_2^2/2}.$$

Accordingly, its cumulative distribution function results as

$$\Pr(R_2 \leq r_1) = 1 - e^{-r_1^2/2}$$

(ii) R_1 is **Rice distributed**:

$$p_{R_1}(r_1) = r_1 \cdot \exp\big(-\tfrac{1}{2}(r_1 + 2\gamma_s)^2\big) \cdot I_0\big(\sqrt{2\gamma_s r_1}\big)$$

Remark

The Rayleigh distribution and the Ricd distribution are often used to model mobile radio channels.

Using the first relation, we can expand the first term as

$$\Big(\Pr(R_2 \le r_1)\Big)^{M-1} = \Big(1 - e^{-r_1^2/2}\Big)^{M-1}$$
$$= \sum_{m=0}^{M-1} (-1)^m \binom{M-1}{m} e^{-mr_1^2/2}$$

After integrating over r_1, we obtain the following closed-form solution for the **symbol-error probability of MFSK with noncoherent demodulation**:

$$P_s = \sum_{m=1}^{M-1} (-1)^{m+1} \binom{M-1}{m} \frac{1}{m+1} \exp\Big(-\frac{m}{m+1}\,\gamma_s\Big).$$

For $M = 2$, this sum simplifies to

$$P_s = \tfrac{1}{2}\, e^{-\gamma_s/2}.$$

7.4.7 Bit-Error Probability

BFSK

We have $\gamma_b = \gamma_s$ and $P_b = P_s$.

Therefore, the bit-error probability is given by

$$P_b = \tfrac{1}{2}\, e^{-\gamma_b/2}.$$

MFSK

For $M > 2$, we have

(i) $\gamma_b = \frac{1}{K}\gamma_s$,

(ii) $P_b = \dfrac{2^{K-1}}{2^K - 1}\, P_s$ with $K = \log_2 M$ (see MPPM).

Therefore the bit-error probability results as

$$P_b = \frac{2^{K-1}}{2^K - 1} \sum_{m=1}^{M-1} (-1)^{m+1} \binom{M-1}{m} \frac{1}{m+1} \exp\Big(-\frac{mK}{m+1}\,\gamma_s\Big).$$

Plots of the error probability can be found in [1, p. 433].

8 Minimum-Shift Keying (MSK)

MSK is a binary modulation scheme with coherent detection, and it is similar to BFSK with coherent detection. The waveforms are sines, the information is conveyed by the frequencies of the waveforms, and the phases of the waveforms are chosen such that there are no phase discontinuities. Thus, the power spectrum of MSK is more compact than that of BFSK. The description of MSK follows [2].

8.1 Motivation

Consider BFSK. The signal waveforms of BFSK are

$$\begin{aligned} U = 0 \quad &\mapsto \quad s_1(t) = \sqrt{2P}\cos\bigl(2\pi f_1 t\bigr), \\ U = 1 \quad &\mapsto \quad s_2(t) = \sqrt{2P}\cos\bigl(2\pi f_2 t\bigr), \end{aligned}$$

for $t \in [0, T]$. The minimum frequency separation such that $s_1(t)$ and $s_2(t)$ are orthogonal (assuming coherent detection) is

$$\Delta f = |f_1 - f_2| = \frac{1}{2T}.$$

In the following, we assume minimum frequency separation.

Without loss of generality, we assume that

$$f_1 = \frac{k}{T}, \qquad\qquad f_2 = \frac{k}{T} + \frac{1}{2T} = \frac{k + 1/2}{T},$$

for $k \in \mathbb{N}$ and $k \gg 1$, i.e., k is a large integer. Thus, we have the following properties:

waveform	number of periods in $t \in [0, T)$
$s_1(t)$	k
$s_2(t)$	$k + 1/2$

Example: Output of BFSK Transmitter

Consider BFSK with $f_1 = 1/T$ and $f_2 = 3/2T$ such that the frequency separation $\Delta = 1/T$ is minimum. Assume the bit sequence

$$\boldsymbol{U}_{[0,4]} = [U_0, U_1, U_2, U_3, U_4] = [0, 1, 1, 0, 0].$$

The resulting output signal $x^{\bullet}(t)$ of the BFSK transmitter shows discontinuities. ◇

The discontinuities produce high side-lobes in the power spectrum of $X(t)$, which is either inefficient or problematic. This can be avoided by slightly modifying the modulation scheme.

We rewrite the signal waveforms as follows:

$$\begin{aligned} s_1(t) &= \sqrt{2P}\cos\bigl(2\pi f_1 t\bigr), \\ s_2(t) &= \sqrt{2P}\cos\bigl(2\pi f_1 t + 2\pi \frac{t}{2T}\bigr), \end{aligned}$$

for $t \in [0, T]$. (Remember that $f_2 = f_1 + 1/2T$.) Thus, we can express the output signal as

$$x(t) = \sqrt{2P}\cos\bigl(2\pi f_1 t + \phi^{\bullet}(t)\bigr),$$

$t \in [0, T)$, with the phase function

$$\phi^{\bullet}(t) = \begin{cases} 0 & \text{for } s_1(t), \\ 2\pi\frac{t}{2T} & \text{for } s_2(t). \end{cases}$$

Notice that $\phi^{\bullet}(t)$ depends directly on the transmitted source bit U.

EXAMPLE: Phase function of BFSK

The phase function $\phi^\bullet(t)$ of the previous example shows discontinuities. ◇

The discontinuities of the phase and thus of $x(t)$ can be avoided by making the phase continuous. This can be achieved by disabling the "reset to zero" of $\phi^\bullet(t)$, yielding the new phase function $\phi(t)$. Notice that this introduces memory in the transmitter.

EXAMPLE: Modified BFSK Transmitter

(We continue the previous example.) The phase function $\phi(t)$ of the modified transmitter is continuous. Accordingly, also the signal $x(t)$ generated by the modified transmitter is continuous. ◇

FSK with phase made continuous as described above is called continuous-phase FSK (CPFSK). Binary CPFSK with minimum frequency separation is termed minimum-shift keying (MSK).

8.2 Transmit Signal

The continuous-phase output signal of an MSK transmitter consists of the following waveforms (the expressions on the right-hand side are only used as abbreviations in the following):

$$
\begin{aligned}
s_1(t) &= +\sqrt{2P}\cos\bigl(2\pi f_1 t\bigr) && \text{“}f_1\text{”}\\
s_2(t) &= -\sqrt{2P}\cos\bigl(2\pi f_1 t\bigr) && \text{“}-f_1\text{”}\\
s_3(t) &= +\sqrt{2P}\cos\bigl(2\pi f_2 t\bigr) && \text{“}f_2\text{”}\\
s_4(t) &= -\sqrt{2P}\cos\bigl(2\pi f_2 t\bigr) && \text{“}-f_2\text{”}
\end{aligned}
$$

$t \in [0, T]$.

The input bit $U_j = 0$ generates either $s_1(t)$ or $s_2(t)$; the input bit $U_j = 1$ generates either $s_3(t)$ or $s_4(t)$, which means a phase difference of π between the beginning and the end of the waveform.

Consider a bit sequence of length L,

$$\boldsymbol{u}_{[0,L-1]} = [u_0, u_1, \ldots, u_{L-1}].$$

According to the above considerations (see also examples), the **output signal of the MSK transmitter** can be written as follows:

$$x(t) = \sqrt{2P}\cos\bigl(2\pi f_1 t + \underbrace{\sum_{l=0}^{L-1} u_l \cdot 2\pi\, b(t - lT)\bigr)}_{\phi(t)},$$

$t \in [0, LT)$, with the **phase pulse**

$$b(\tau) = \begin{cases} 0 & \text{for } \tau < 0, \\ \tau/2T & \text{for } 0 \leq \tau < T, \\ 1/2 & \text{for } T \leq \tau. \end{cases}$$

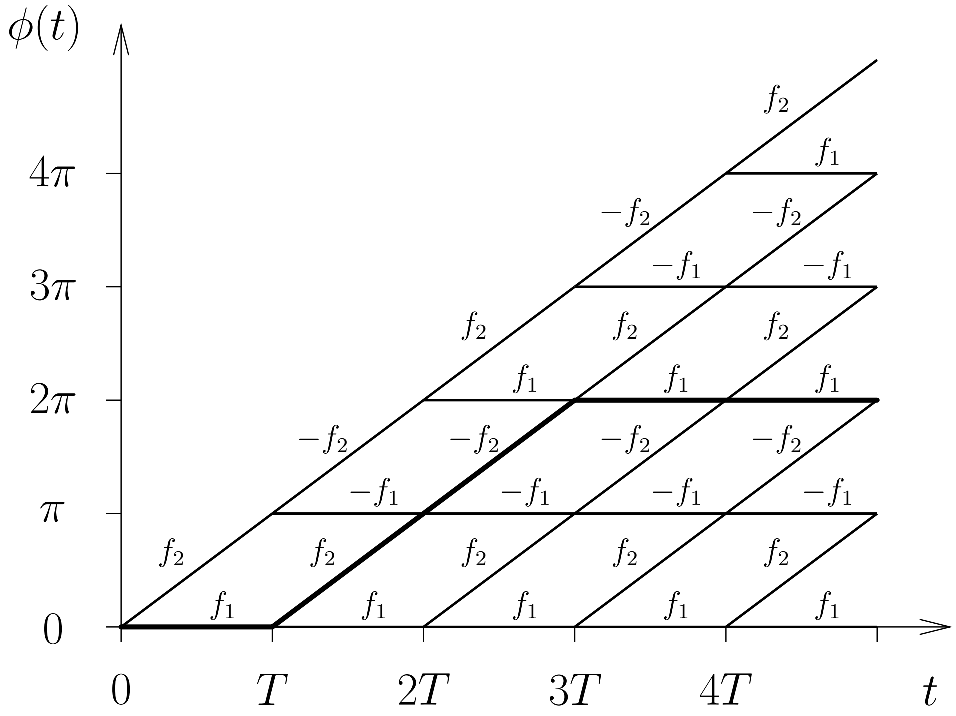
Remark

The BFSK signal $x(t)$ can be generated with the same expression as above, when the phase pulse

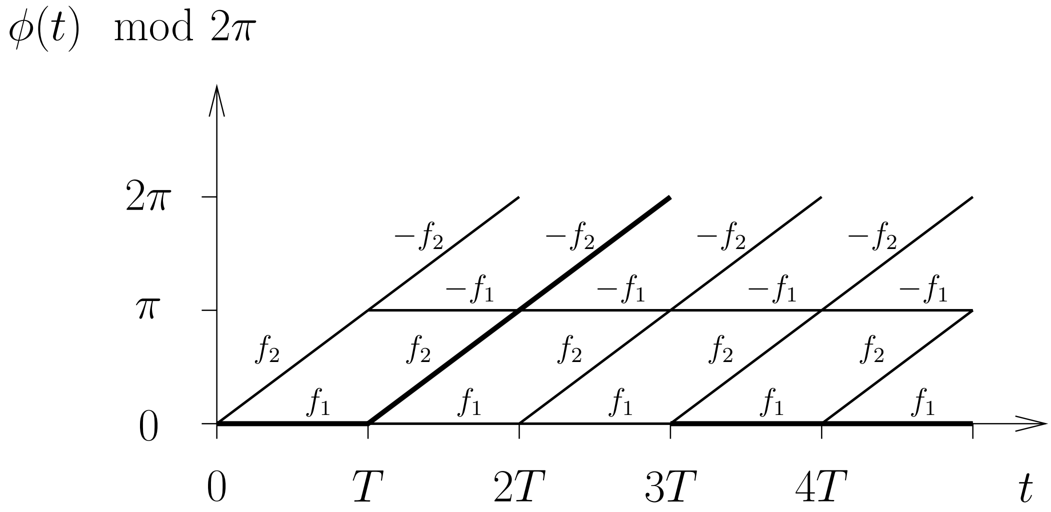
$$b^{\bullet}(\tau) = \begin{cases} 0 & \text{for } \tau < 0, \\ \tau/2T & \text{for } 0 \leq \tau < T, \\ 0 & \text{for } T \leq \tau \end{cases}$$

is used.

Plotting all possible trajectories of $\phi(t)$, we obtain the (tilted) **phase tree of MSK**.



The physical phase is obtained by taking $\phi(t)$ modulo 2π. Plotting all possible trajectories, we obtain the (tilted) **physical-phase trellis of MSK**.



8.3 Signal Constellation

Signal waveforms:

$$
\begin{aligned}
s_1(t) &= +\sqrt{2P}\cos\bigl(2\pi f_1 t\bigr),\\
s_2(t) &= -\sqrt{2P}\cos\bigl(2\pi f_1 t\bigr),\\
s_3(t) &= +\sqrt{2P}\cos\bigl(2\pi f_2 t\bigr),\\
s_4(t) &= -\sqrt{2P}\cos\bigl(2\pi f_2 t\bigr),
\end{aligned}
$$

$t \in [0, T]$.

Orthonormal functions:

$$
\begin{aligned}
\psi_1(t) &= \sqrt{2/T}\cos\bigl(2\pi f_1 t\bigr),\\
\psi_2(t) &= \sqrt{2/T}\cos\bigl(2\pi f_2 t\bigr),
\end{aligned}
$$

$t \in [0, T]$.

Vector representation of the signal waveforms:

$$
\begin{aligned}
s_1(t) &= +\sqrt{E_s}\;\psi_1(t) &&\longleftrightarrow&& \boldsymbol{s}_1 = (+\sqrt{E_s}, 0)^\mathsf{T}\\
s_2(t) &= -\sqrt{E_s}\;\psi_1(t) &&\longleftrightarrow&& \boldsymbol{s}_2 = (-\sqrt{E_s}, 0)^\mathsf{T} = -\boldsymbol{s}_1\\
s_3(t) &= +\sqrt{E_s}\;\psi_2(t) &&\longleftrightarrow&& \boldsymbol{s}_3 = (0, +\sqrt{E_s})^\mathsf{T}\\
s_4(t) &= -\sqrt{E_s}\;\psi_2(t) &&\longleftrightarrow&& \boldsymbol{s}_4 = (0, -\sqrt{E_s})^\mathsf{T} = -\boldsymbol{s}_3
\end{aligned}
$$

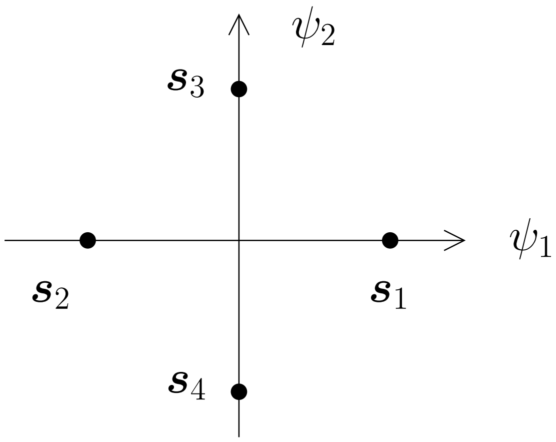
with $E_s = PT$.

Remark:
The modulation symbol at discrete time j is written as

$$\boldsymbol{X}_j = (X_{j,1}, X_{j,2})^\mathsf{T}.$$

For example, $\boldsymbol{X}_j = \boldsymbol{s}_1$ means $X_{j,1} = \sqrt{E_s}$ and $X_{j,2} = 0$.

Signal constellation:



Remark:

MSK and QPSK have the same signal constellation. However, the two modulation schemes are not equivalent because their vector encoders are different. This is discussed in the following.

Modulator:

The symbol $\boldsymbol{X}_j = (X_{j,1}, X_{j,2})^\mathsf{T}$ is transmitted in the time interval

$$t \in [jT, (j+1)T).$$

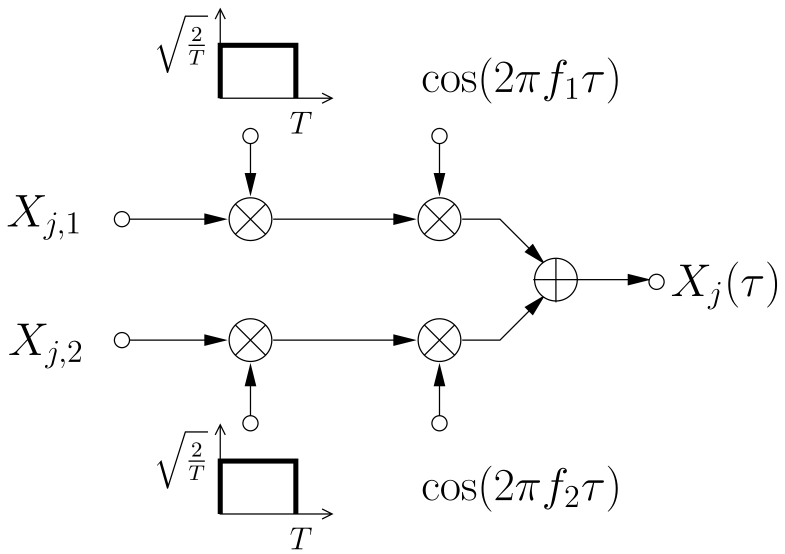
This time interval can also be written as

$$t = jT + \tau, \qquad\qquad \tau \in [0, T).$$

The modulation for $\tau = t - jT \in [0, T)$, $j = 0, 1, 2, \ldots$, is then given by the mapping

$$\boldsymbol{X}_j \mapsto X_j(\tau),$$

and it can be implemented by

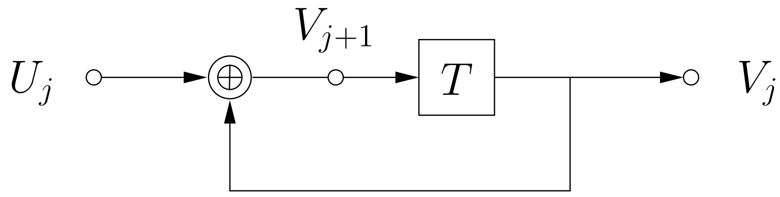


Remark:

The index j is not necessary for describing the modulator. It is only introduced to have the same notation as being used for the vector encoder.

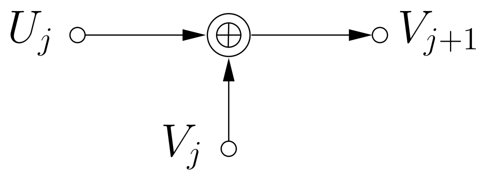
8.4 A Finite-State Machine

Consider the following device (which is a simple finite-state machine)



It consists of the following components:

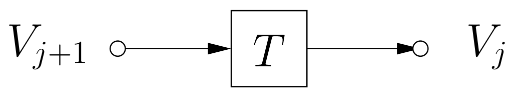
(i) A modulo-2 adder:



It has the functionality

$$V_{j+1} = U_j \oplus V_j = U_j + V_j \quad \text{mod } 2.$$

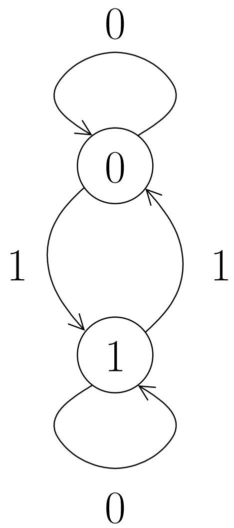
(ii) A shift register clocked at times T ("every T seconds"):



(iii) The sequence $\{U_j\}$ is a binary sequence with elements in $\{0, 1\}$, and so is the sequence $\{V_j\}$.

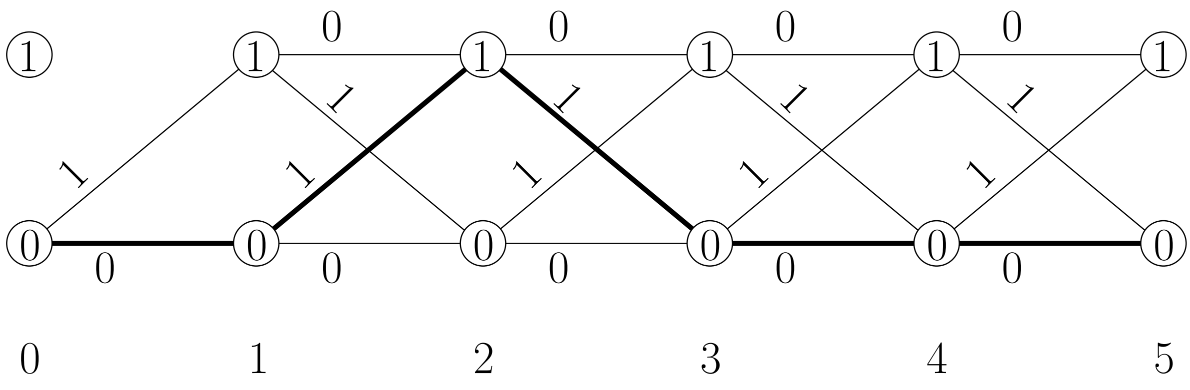
The output value V_j describes the **state** of this device.

As $V_j \in \{0, 1\}$, this device can be in two different states. All possible transitions can be depicted in a **state transition diagram**:



The values in the circles denote the states v, and the labels of the arrows denote the inputs u.

The state transitions and the associated input values can also be depicted in a **trellis diagram**. The initial state is assumed to be zero.



The time indices are written below the states.

The emphasized path corresponds to the input sequence

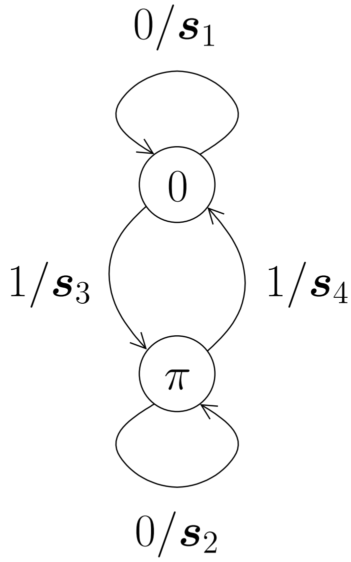
$$[U_0, U_1, U_2, U_3, U_4] = [0, 1, 1, 0, 0].$$

8.5 Vector Encoder

According to the trellis of the physical phase, the initial phase at the beginning of each symbol interval can take the two values 0 and π. These values determine the state of the encoder. We use the following association:

$$\begin{array}{lcl} \text{phase } 0 & \leftrightarrow & \text{state } V = 0, \\ \text{phase } \pi & \leftrightarrow & \text{state } V = 1. \end{array}$$

The functionality of the encoder can be described by the following state transition diagram:



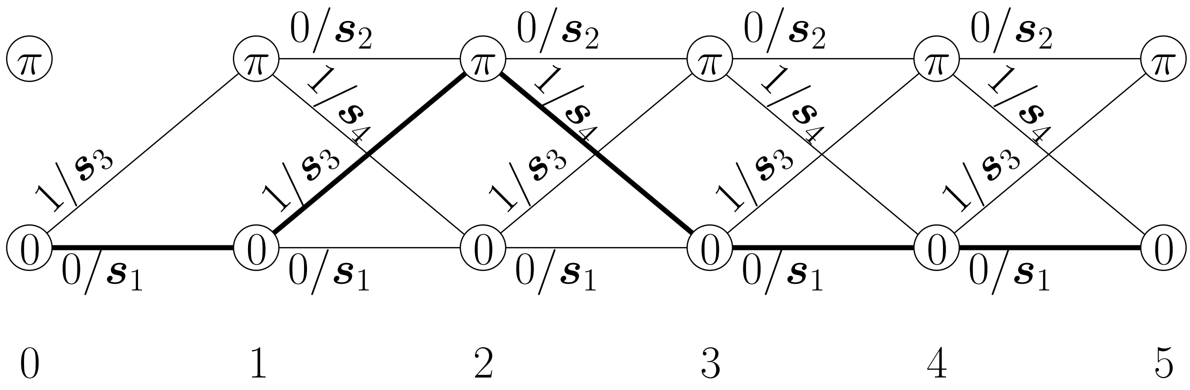
The transitions are labeled by $u/\boldsymbol{x}$. The value u denotes the binary source symbol at the encoder input, and $\boldsymbol{x}$ denotes the modulation symbol at the encoder output.

Remember: $\boldsymbol{s}_2 = -\boldsymbol{s}_1$ and $\boldsymbol{s}_4 = -\boldsymbol{s}_3$.

Remark

The vector encoder is a finite-state machine as considered in the previous section.

The temporal behavior of the encoder can be described by the trellis diagram:



The emphasized path corresponds to the input sequence

$$[U_0, U_1, U_2, U_3, U_4] = [0, 1, 1, 0, 0].$$

General Conventions

(i) The input sequence of length L is written as

$$\boldsymbol{u} = [u_0, u_1, \ldots, u_{L-1}].$$

(ii) The initial state is $V_0 = 0$.

(iii) The final state is $V_L = 0$.

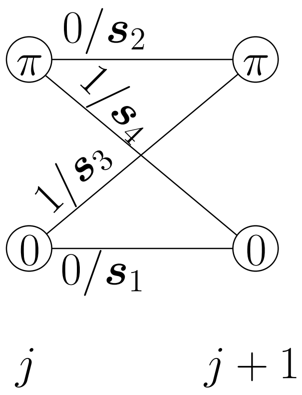
The last bit U_{L-1} is selected such that the encoder is driven back to state 0:

$$\begin{array}{lcl} V_{L-1} = 0 & \longrightarrow & U_{L-1} = 0, \\ V_{L-1} = 1 & \longrightarrow & U_{L-1} = 1. \end{array}$$

Therefore, U_{L-1} does not carry information.

Generation of the Encoder Outputs

Consider one trellis section:



Remember: $\boldsymbol{s}_2 = -\boldsymbol{s}_1$ and $\boldsymbol{s}_4 = -\boldsymbol{s}_3$.

Observations:

(i) Input symbols $\rightarrow$ output vectors

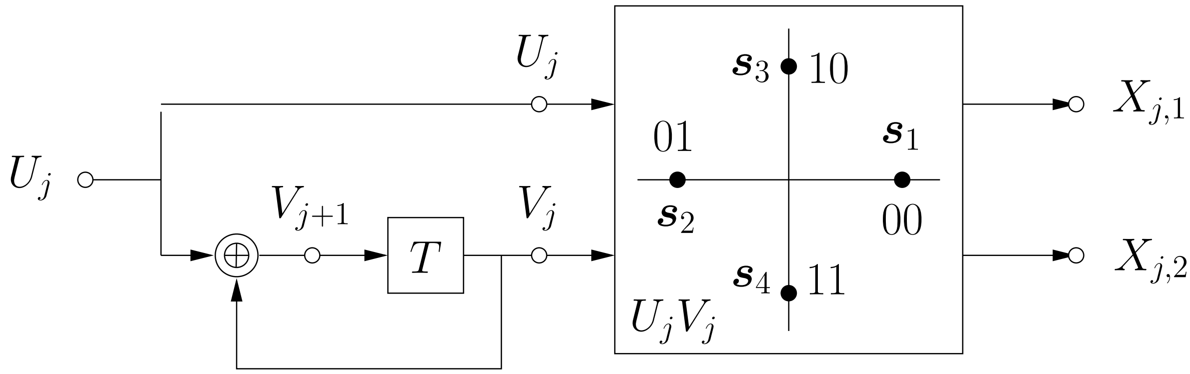
$$\begin{aligned} U_j = 0 \qquad &\longrightarrow \qquad \boldsymbol{X}_j \in \{\boldsymbol{s}_1, \boldsymbol{s}_2\} = \{\boldsymbol{s}_1, -\boldsymbol{s}_1\}, \\ U_j = 1 \qquad &\longrightarrow \qquad \boldsymbol{X}_j \in \{\boldsymbol{s}_3, \boldsymbol{s}_4\} = \{\boldsymbol{s}_3, -\boldsymbol{s}_3\}. \end{aligned}$$

(ii) Encoder state $\rightarrow$ output vectors

$$\begin{aligned} V_j = 0 \text{ (“0”)} \qquad &\longrightarrow \qquad \boldsymbol{X}_j \in \{\boldsymbol{s}_1, \boldsymbol{s}_3\} \\ V_j = 1 \text{ (“}\pi\text{”)} \qquad &\longrightarrow \qquad \boldsymbol{X}_j \in \{\boldsymbol{s}_2, \boldsymbol{s}_4\} = \{-\boldsymbol{s}_1, -\boldsymbol{s}_3\}. \end{aligned}$$

Using these observations, the above device can be supplemented to obtain an implementation of the vector encoder for MSK.

Block Diagram of the Vector Encoder for MSK



At time index j:

- input symbol U_j,
- state V_j,
- output symbol $\boldsymbol{X}_j = (X_{j,1}, X_{j,2})^\mathsf{T}$,
- subsequent state V_{j+1}.

Notice:

(i) U_j controls the frequency of the transmitted waveform.

(ii) V_j controls the polarity of the transmitted waveform.

8.6 Demodulator

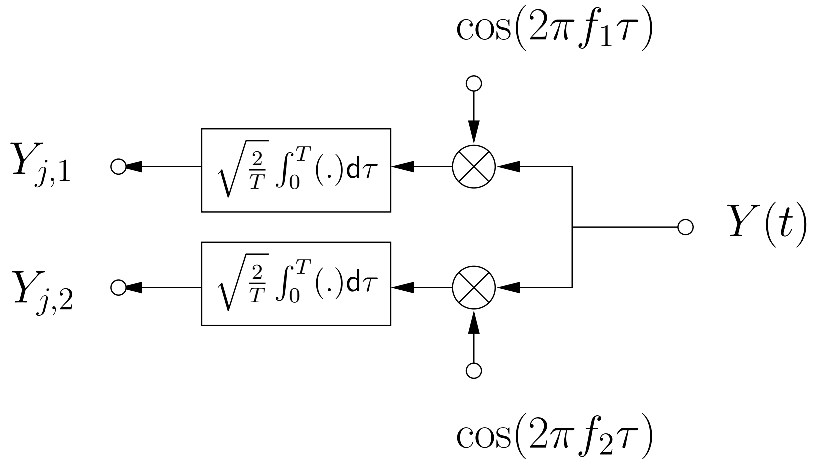
The received symbol $\boldsymbol{Y}_j = (Y_{j,1}, Y_{j,2})^\mathsf{T}$ is determined in the time interval

$$t \in [jT, (j+1)T).$$

This time interval can also be written as

$$t = jT + \tau, \qquad\qquad \tau \in [0, T).$$

The demodulation for $\tau = t - jT \in [0, T)$, $j = 0, 1, 2, \ldots$, can then be implemented as follows:



8.7 Conditional Probability Density of Received Sequence

Sequence of (two-dimensional) received symbols

$$[\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_{L-1}] = [\boldsymbol{X}_0 + \boldsymbol{W}_0, \ldots, \boldsymbol{X}_{L-1} + \boldsymbol{W}_{L-1}].$$

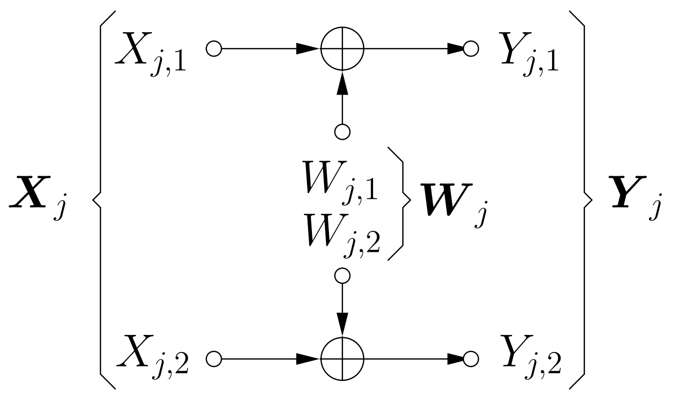
The vector $[\boldsymbol{W}_0, \ldots, \boldsymbol{W}_{L-1}]$ is a white Gaussian noise sequence with the properties:

(i) $\boldsymbol{W}_0, \ldots, \boldsymbol{W}_{L-1}$ are independent ;

(ii) each (two-dimensional) noise vector is distributed as

$$\boldsymbol{W}_j = \begin{bmatrix} W_{j,1} \\ W_{j,2} \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} N_0/2 & 0 \\ 0 & N_0/2 \end{bmatrix} \Big)$$

Illustration of the AWGN vector channel:



Conditional probability of $\boldsymbol{y}_{[0,L-1]} = [\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{L-1}]$ given that $\boldsymbol{x}_{[0,L-1]} = [\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}]$ was transmitted:

$$
\begin{aligned}
p_{\boldsymbol{Y}_{[0,L-1]}|\boldsymbol{X}_{[0,L-1]}}\big(\boldsymbol{y}_0, \ldots, \boldsymbol{y}_{L-1}|\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}\big) = \\
&= \prod_{j=0}^{L-1} p_{\boldsymbol{Y}_j|\boldsymbol{X}_j}\big(\boldsymbol{y}_j|\boldsymbol{x}_j\big) \\
&= \prod_{j=0}^{L-1} \Big[\frac{1}{\pi N_0} \cdot \exp\big(-\tfrac{1}{N_0}\|\boldsymbol{y}_j - \boldsymbol{x}_j\|^2\big)\Big] \\
&= \Big(\frac{1}{\pi N_0}\Big)^L \cdot \exp\big(-\tfrac{1}{N_0} \sum_{j=0}^{L-1}\|\boldsymbol{y}_j - \boldsymbol{x}_j\|^2\big).
\end{aligned}
$$

Notice that this is the likelihood function of $\boldsymbol{x}_{[0,L-1]}$ for a given received sequence $\boldsymbol{y}_{[0,L-1]}$.

8.8 ML Sequence Estimation

The trellis of the vector encoder has the following two properties:

(i) Each input sequence $\boldsymbol{u}_{[0,L-1]} = [u_0, \ldots, u_{L-1}]$ uniquely determines a path of length L in the trellis, and vice versa.

(ii) Each path in the trellis uniquely determines an output sequence $\boldsymbol{x}_{[0,L-1]} = [\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}]$

A sequence $\boldsymbol{x}_{[0,L-1]}$ of length L generated by the vector encoder is called an **admissible sequence** if the initial and the final state of the encoder are zero, i.e., if $V_0 = V_L = 0$. The **set of admissible sequences** is denoted by $\mathbb{A}$.

Using these definitions, the **ML decoding rule** is the following:

> Search the trellis for an admissible sequence that maximizes the likelihood function.

Or in mathematical terms:

$$\hat{\boldsymbol{x}}_{[0,L-1]} = \underset{\boldsymbol{x}_{[0,L-1]}\in\mathbb{A}}{\operatorname{argmax}}\; p(\boldsymbol{y}_{[0,L-1]}|\boldsymbol{x}_{[0,L-1]}).$$

Using the relation

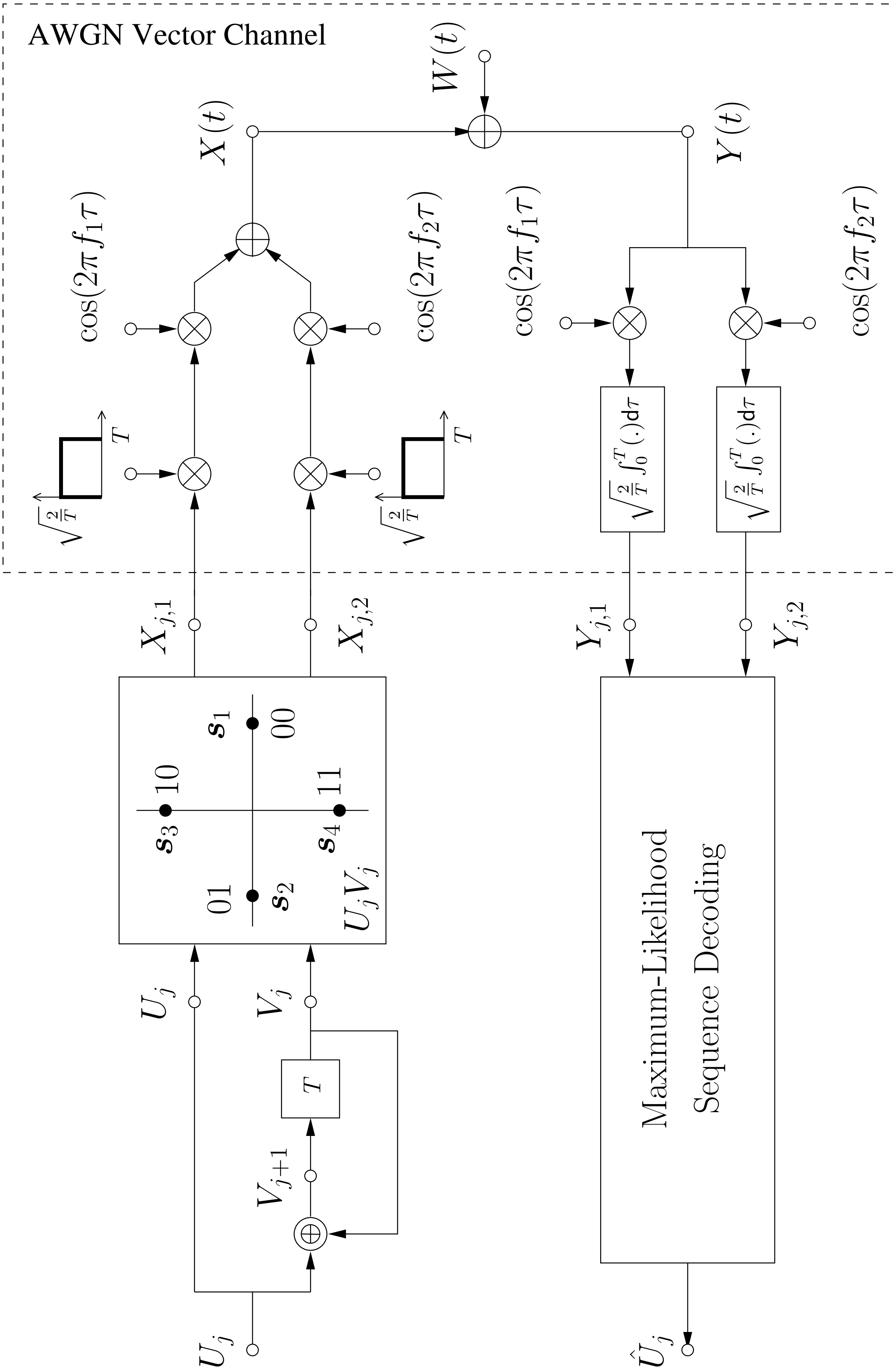
$$\|\boldsymbol{y}_j - \boldsymbol{x}_j\|^2 = \underbrace{\|\boldsymbol{y}_j\|^2}_{\text{indep. of } \boldsymbol{x}_j} - 2\langle \boldsymbol{y}_j, \boldsymbol{x}_j\rangle + \underbrace{\|\boldsymbol{x}_j\|^2}_{= E_s}$$

and the expression for the likelihood function, the **rule for ML sequence estimation (MLSE)** can equivalently be formulated as follows:

$$\begin{aligned}\hat{\boldsymbol{x}}_{[0,L-1]} &= \underset{\boldsymbol{x}_{[0,L-1]}\in\mathbb{A}}{\operatorname{argmax}}\; \exp\Big(-\frac{1}{N_0}\sum_{j=0}^{L-1}\|\boldsymbol{y}_j - \boldsymbol{x}_j\|^2\Big)\\ &= \underset{\boldsymbol{x}_{[0,L-1]}\in\mathbb{A}}{\operatorname{argmin}}\; \sum_{j=0}^{L-1}\|\boldsymbol{y}_j - \boldsymbol{x}_j\|^2\\ &= \underset{\boldsymbol{x}_{[0,L-1]}\in\mathbb{A}}{\operatorname{argmax}}\; \sum_{j=0}^{L-1}\langle \boldsymbol{y}_j, \boldsymbol{x}_j\rangle.\end{aligned}$$

Hence, the ML estimate of the transmitted sequence is a sequence in $\mathbb{A}$ that maximizes the sum $\sum_{j=0}^{L-1}\langle \boldsymbol{y}_j, \boldsymbol{x}_j\rangle$.

8.9 Canonical Decomposition of MSK



8.10 The Viterbi Algorithm

Given Problem:

(i) The admissible sequences

$$\boldsymbol{x}_{[0,L-1]} = [\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}] \in \mathbb{A}$$

can be represented in a trellis.

(ii) The value to be maximized (optimization criterion) can be written as the sum

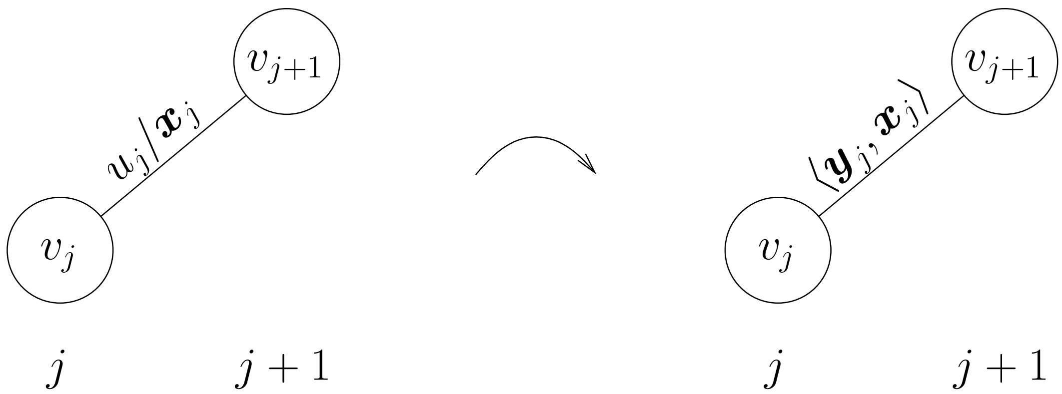
$$\sum_{j=0}^{L-1} \langle \boldsymbol{y}_j, \boldsymbol{x}_j \rangle.$$

(Remember: $\langle \boldsymbol{y}_j, \boldsymbol{x}_j \rangle = y_{j,1} \cdot x_{j,1} + y_{j,2} \cdot x_{j,2}$.)

The **Viterbi Algorithm** solves such an optimization problem in a very effective way.

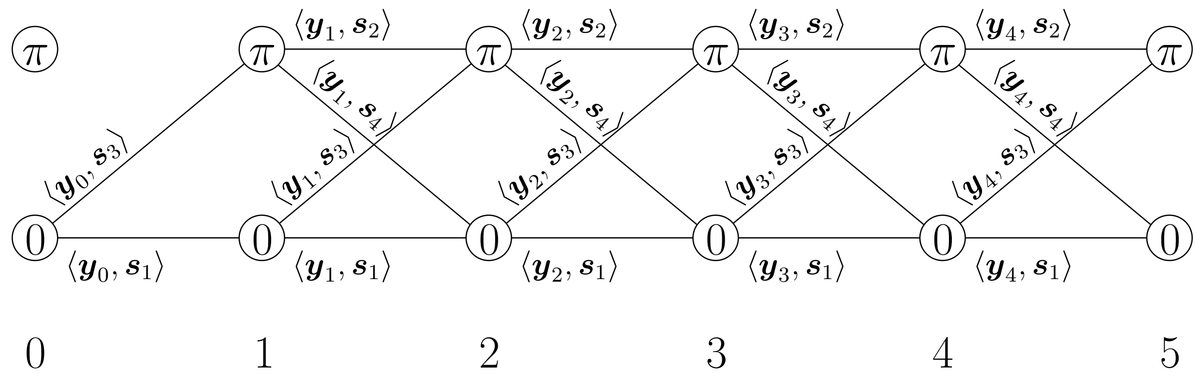
Branch Metrics and Path Metrics

The branches of the trellis are re-labeled as follows:



The new labels are called **branch metrics**. They may be interpreted as a gain or a profit made when going along this branch.

When labeling all branches with their branch metrics, we obtain the following trellis diagram:



The total gain or profit achieved by going along a path in the trellis is called the **path metric**. The path corresponding to the sequence

$$\boldsymbol{x}_{[0,L-1]} = [\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{L-1}]$$

has the path metric

$$\sum_{j=0}^{L-1} \langle \boldsymbol{y}_j, \boldsymbol{x}_j \rangle.$$

Notice: The branch metric and the path metric are not necessarily a mathematical metric.

Hence, the ML estimate $\hat{\boldsymbol{x}}_{[0,L-1]}$ of $\boldsymbol{X}_{[0,L-1]}$ is an admissible sequence ($\in \mathbb{A}$) that maximizes the path metric. The corresponding path is called an **optimum path**.

There are $2 \cdot 4^{L-1}$ paths of length L. The Viterbi algorithm is a sequential method that determines an optimum path without computing the path metrics of all $2 \cdot 4^{L-1}$ paths.

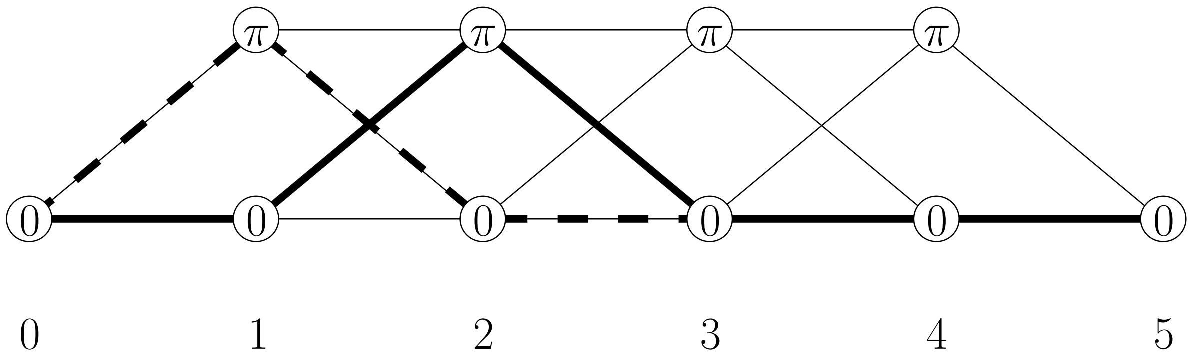
Survivors

The Viterbi algorithm relies on the following property of optimum paths:

> A path of length L is optimum if and only if its subpaths at any depth $n = 1, 2, \ldots, L-1$ are also optimum.

EXAMPLE: Optimality of Subpaths

Consider the following trellis. Assume that the thick path (solid line) is optimum.



The dashed path ending at depth 3 cannot have a metric that is larger than that of the subpath obtained by pruning the optimum path at depth 3. ◇

For each state in each depth n, there are several subpaths ending in this state. The one with the largest metric ("the best path") is called the **survivor** at this state in this depth.

From the above considerations, we have the following important results:

- All subpaths of an optimum path are survivors.
- Therefore, only survivors need to be considered for determining an optimum path.

Accordingly, non-survivor paths can be discarded without loss of optimality. (Therefore the name “survivor”.)

Viterbi Algorithm

1. Start at the beginning of the trellis (depth 0 and state 0 by convention).
2. Increase the depth by 1. Extend all previous survivors to the new depth and determine the new survivor for each state.
3. If the end of the trellis is reached (depth L and state 0 by convention), the survivor is an optimum path. The associated sequence is the ML sequence.

Remarks

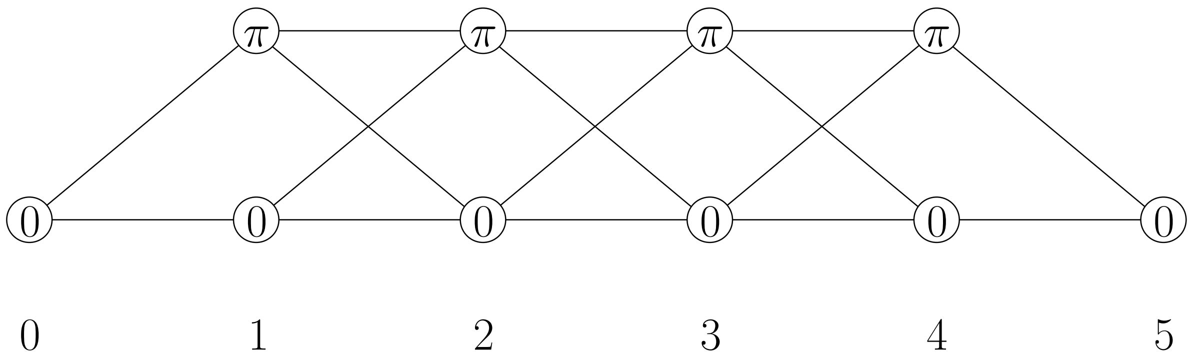
(i) The complexity of computing the path metrics of all paths is exponential in the sequence length. (Exhaustive search.)

(ii) The complexity of the Viterbi algorithm is linear in the sequence length.

(iii) The complexity of the Viterbi algorithm is exponential in the number of shift registers in the vector encoder.

Example: Viterbi Algorithm

Consider the case $L = 5$. This corresponds to the following trellis.



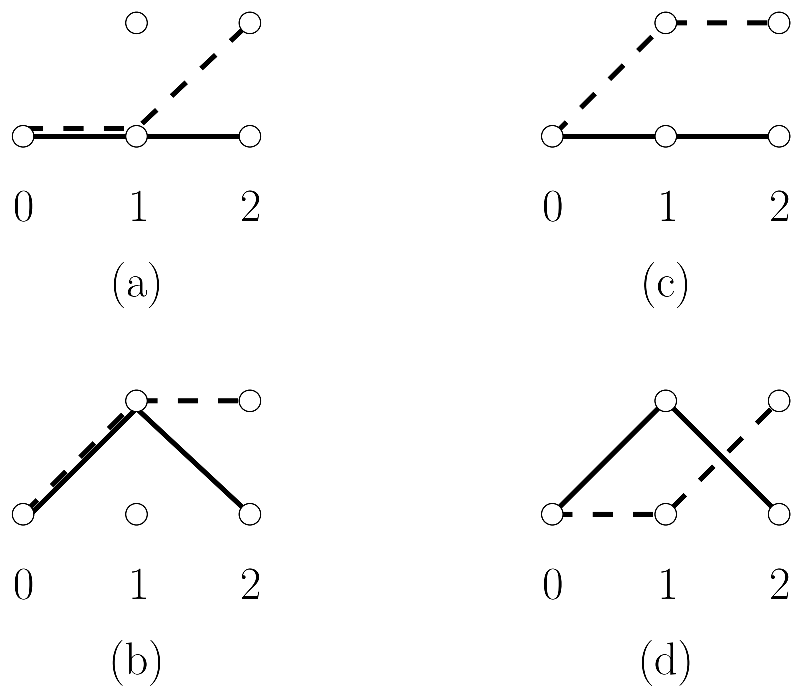
The Viterbi algorithm may be illustrated for a randomly drawn received sequence $\boldsymbol{y}_{[0,L-1]}$. (The receiver must be capable of handling every received sequence.) ◇

8.11 Simplified ML Decoder

The MSK trellis has a special structure, and this leads to a special behavior of the Viterbi algorithm. This can be exploited to derive a simpler optimal decoder.

Observation

At depth $n = 2$, the two survivors (solid line and dashed line) will look as depicted in (a) or (b), but never as in (c) or (d).



Therefore the final estimate of the state $\hat{V}_1$ is already decided at time $j = 1$. This is proved shortly afterwards.

The general result for the given MSK transmitter follows by induction:

> The two survivors at any depth n are identical for $j < n$ and differ only in the current state.

Therefore, final decisions on any state $\hat{V}_n$ (value of the vector-encoder state at time n) can already be made at time n.

We proof now that the Viterbi algorithm makes the final decision about state $\hat{V}_1$ (state at time $j = 1$) already at time 1.

Proof

The inner product is defined as

$$\langle \boldsymbol{y}_j, \boldsymbol{s}_m \rangle = y_{j,1} \cdot s_{m,1} + y_{j,2} \cdot s_{m,2}.$$

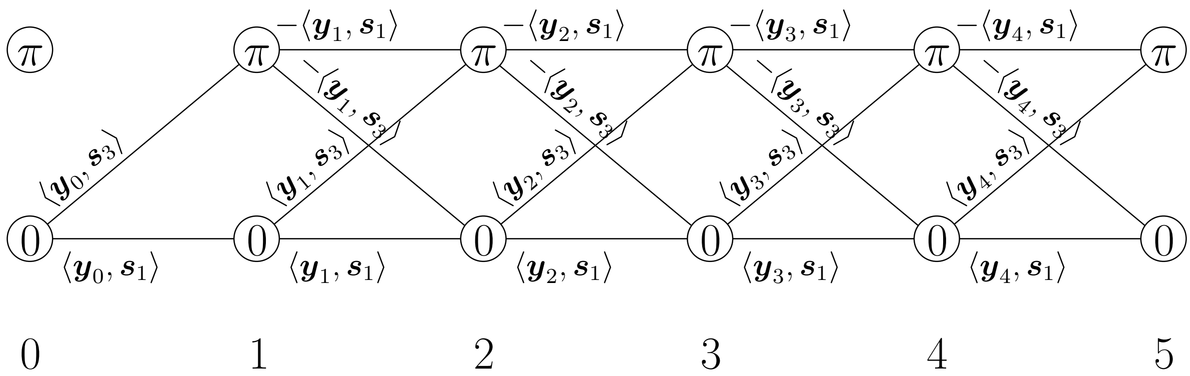
Due to the signal constellation,

$$\boldsymbol{s}_2 = -\boldsymbol{s}_1, \qquad \boldsymbol{s}_4 = -\boldsymbol{s}_3.$$

Therefore,

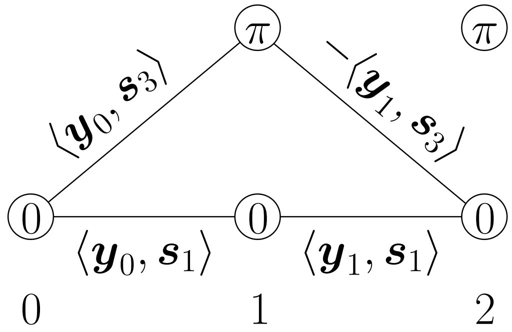
$$\begin{aligned} \langle \boldsymbol{y}_j, \boldsymbol{s}_2 \rangle &= \langle \boldsymbol{y}_j, -\boldsymbol{s}_1 \rangle = -\langle \boldsymbol{y}_j, \boldsymbol{s}_1 \rangle, \\ \langle \boldsymbol{y}_j, \boldsymbol{s}_4 \rangle &= \langle \boldsymbol{y}_j, -\boldsymbol{s}_3 \rangle = -\langle \boldsymbol{y}_j, \boldsymbol{s}_3 \rangle. \end{aligned}$$

Using these relations, the branches can equivalently be labeled as



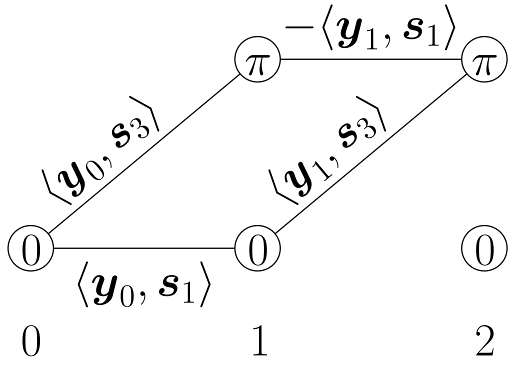
Consider the conditions for selecting the survivors, while making use of these relations:

- Survivor at state $v_2 = 0$ (time $j = 1$):



$$
\begin{aligned}
\langle \boldsymbol{y}_0, \boldsymbol{s}_3 \rangle - \langle \boldsymbol{y}_1, \boldsymbol{s}_3 \rangle &> \langle \boldsymbol{y}_0, \boldsymbol{s}_1 \rangle + \langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle \quad \Rightarrow \quad \text{upper path} \\
\langle \boldsymbol{y}_0, \boldsymbol{s}_3 \rangle - \langle \boldsymbol{y}_1, \boldsymbol{s}_3 \rangle &< \langle \boldsymbol{y}_0, \boldsymbol{s}_1 \rangle + \langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle \quad \Rightarrow \quad \text{lower path}
\end{aligned}
$$

- Survivor at state $v_2 = 1$ (“π”) (time $j = 1$):



$$
\begin{aligned}
\langle \boldsymbol{y}_0, \boldsymbol{s}_3 \rangle - \langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle &> \langle \boldsymbol{y}_0, \boldsymbol{s}_1 \rangle + \langle \boldsymbol{y}_1, \boldsymbol{s}_3 \rangle \quad \Rightarrow \quad \text{upper path} \\
\langle \boldsymbol{y}_0, \boldsymbol{s}_3 \rangle - \langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle &< \langle \boldsymbol{y}_0, \boldsymbol{s}_1 \rangle + \langle \boldsymbol{y}_1, \boldsymbol{s}_3 \rangle \quad \Rightarrow \quad \text{lower path}
\end{aligned}
$$

Obviously, either the two upper paths or the two lower paths are the survivors. Therefore, the two survivors go through the same state $\hat{V}_1$.

ML State Sequence

By induction, we obtain the following optimal decision rule for state V_j based on $\boldsymbol{y}_{j-1}$ and $\boldsymbol{y}_j$:

$$\hat{V}_j = \begin{cases} 0 & \text{if} \quad \langle \boldsymbol{y}_{j-1}, \boldsymbol{s}_3 \rangle - \langle \boldsymbol{y}_j, \boldsymbol{s}_3 \rangle < \langle \boldsymbol{y}_{j-1}, \boldsymbol{s}_1 \rangle + \langle \boldsymbol{y}_j, \boldsymbol{s}_1 \rangle, \\ 1 & \text{if} \quad \langle \boldsymbol{y}_{j-1}, \boldsymbol{s}_3 \rangle - \langle \boldsymbol{y}_j, \boldsymbol{s}_3 \rangle > \langle \boldsymbol{y}_{j-1}, \boldsymbol{s}_1 \rangle + \langle \boldsymbol{y}_j, \boldsymbol{s}_1 \rangle, \end{cases}$$

$j = 1, 2, \ldots, L-1$. (Remember: $V_L = 0$ by convention.)

This decision rule can be simplified. The modulation symbols are

$$\boldsymbol{s}_1 = (+\sqrt{E_s}, 0)^{\mathsf{T}}, \qquad \boldsymbol{s}_3 = (0, +\sqrt{E_s})^{\mathsf{T}}.$$

Thus, the inner products can be written as

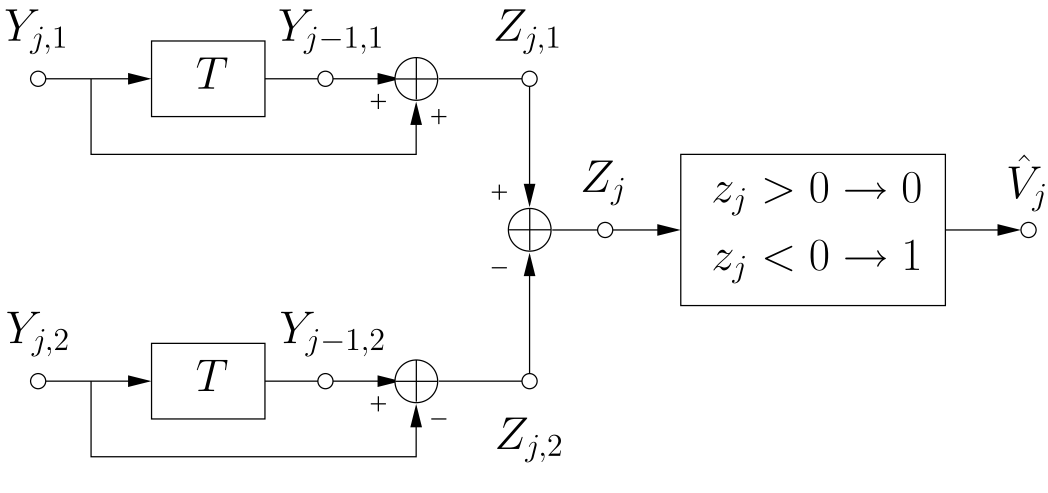
$$\langle \boldsymbol{y}_j, \boldsymbol{s}_1 \rangle = y_{j,1} \cdot \sqrt{E_s}, \qquad \langle \boldsymbol{y}_j, \boldsymbol{s}_3 \rangle = y_{j,2} \cdot \sqrt{E_s}.$$

Substituting this in the decision rule and dividing by $\sqrt{E_s}$, we obtain the following simplified (and still optimal) decision rule:

$$\hat{V}_j = \begin{cases} 0 & \text{if} \quad y_{j-1,2} - y_{j,2} < y_{j-1,1} + y_{j,1}, \\ 1 & \text{if} \quad y_{j-1,2} - y_{j,2} > y_{j-1,1} + y_{j,1}, \end{cases}$$

$j = 1, 2, \ldots, L-1$.

The following device implements this **state decision rule**:



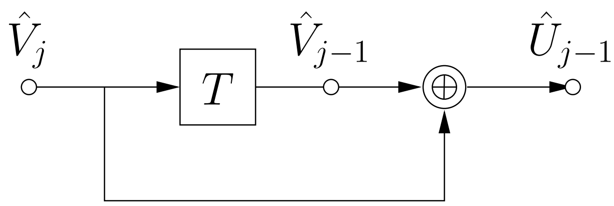
Vector-Encoder Inverse

Given the sequence of state estimates $\{\hat{V}_j\}$, the sequence of bit estimates $\{\hat{U}_j\}$ can computed according to

$$\hat{U}_j = \hat{V}_j \oplus \hat{V}_{j+1},$$

$j = 0, 1, \ldots, L-1$. (Compare vector encoder.)

The following device implements this operation:



The time indices are $j = 0, 1, \ldots, L-1$, and $\hat{V}_L = 0$ as $V_L = 0$ by convention. Remember that U_{L-1} does not convey information.

8.12 Bit-Error Probability

The BER performance of MSK is about 2 dB worse than that of BPSK and QPSK. This means, when comparing the SNR in E_b/N_0 required to achieve a certain bit-error probability (BEP), MSK needs about 2 dB more than BPSK and QPSK. This can be improved by slightly modifying the vector encoder.

The trellis of the vector encoder has the following properties:

(i) Two different paths differ in at least one state.

(ii) Two different paths differ in at least two bits.

(See trellis of vector encoder in Section 8.5.)

The most likely error-event for high SNR is that the transmitted path and the estimated path differ in one state. This error-event leads to two bit errors.

Idea for improvement:

- The state sequence should be equal to the bit sequence, such that the most likely error-event leads only to one bit error.
- This can be achieved by using a pre-coder before the vector encoder.

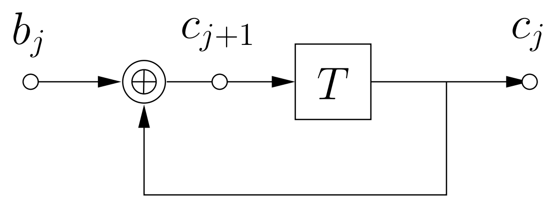
The resulting modulation scheme is called **precoded MSK**

8.13 Differential Encoding and Differential Decoding

The two terms "differential encoder" and "differential decoder" are mainly used due to historical reasons. Both may be used as encoders.

In the following, $b_j \in \{0, 1\}$ denote the bits before encoding, and $c_j \in \{0, 1\}$ denote the bits after encoding.

Differential Encoder (DE)

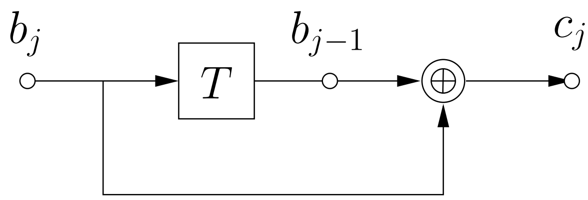


Operation:

$$c_{j+1} = b_j \oplus c_j,$$

$j = 0, 1, 2, \ldots$. Initialization: $c_0 = 0$.

Differential Decoder (DD)



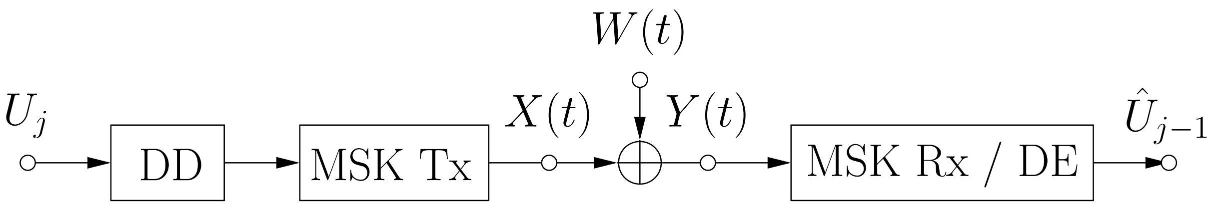
Operation:

$$c_j = b_j \oplus b_{j-1},$$

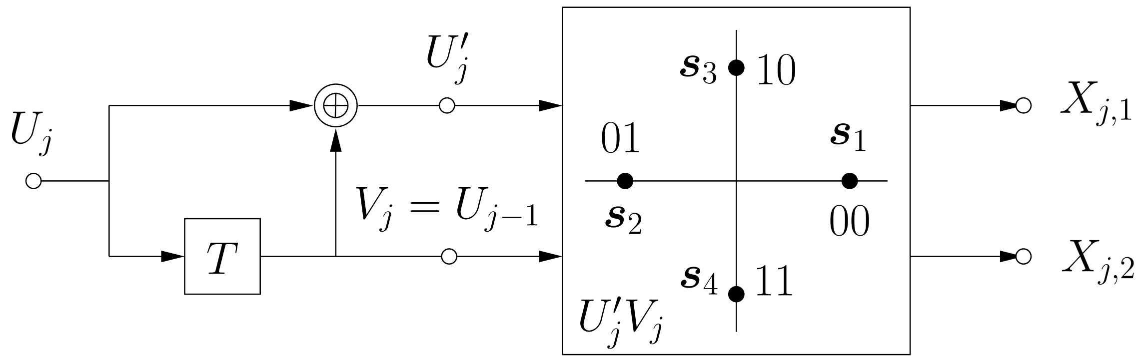
$j = 0, 1, 2, \ldots$. Initialization: $b_{-1} = 0$.

8.14 Precoded MSK

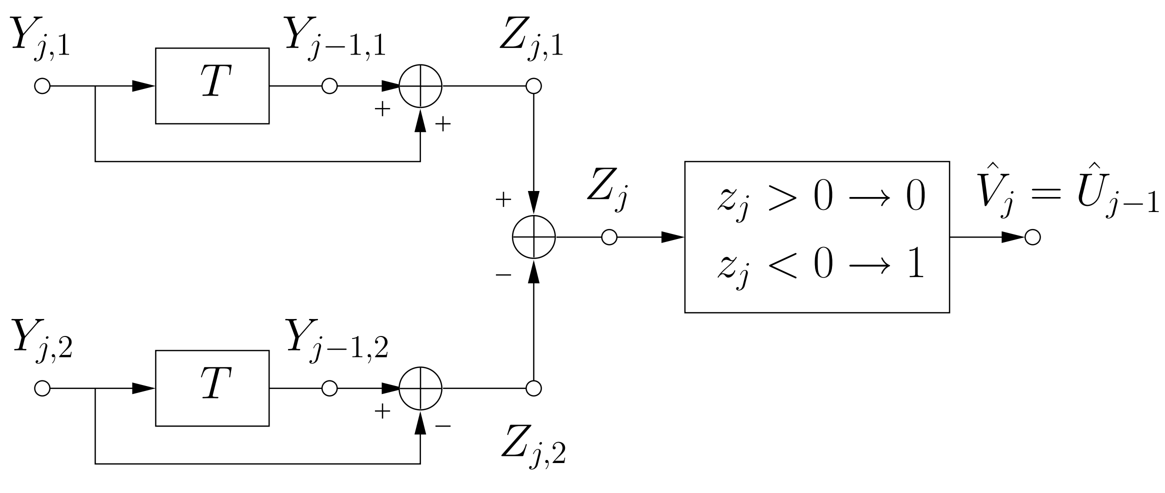
The DD(!) is used as a pre-coder and is placed before the vector encoder. The DE(!) is placed behind the vector-encoder inverse.



Vector Representation of Precoded MSK

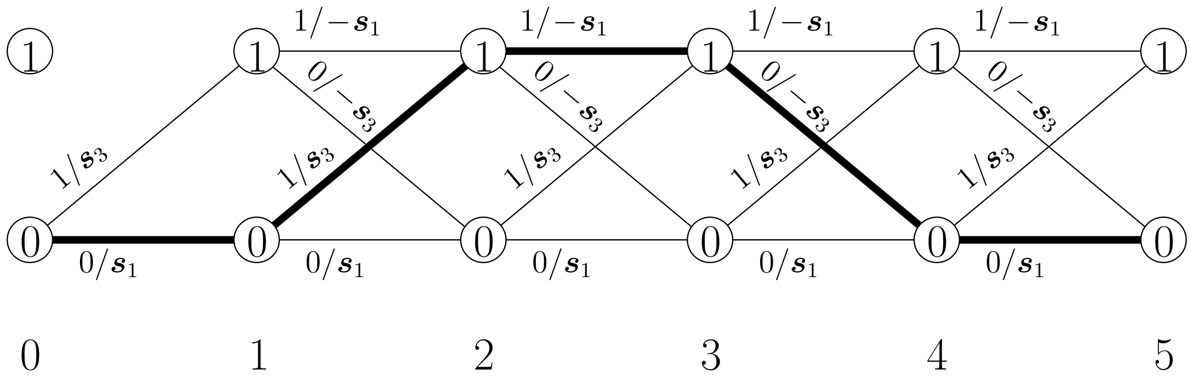


Initialization: $V_0 = 0$. Notice: $U'_j = U_j \oplus V_j = U_j \oplus U_{j-1}$.



Notice: The vector-encoder inverse and the DE compensate each other and can be removed.

Trellis of Precoded MSK



The thick path corresponds to the bit sequence

$$\boldsymbol{U}_{[0,4]} = [0, 1, 1, 0, 0].$$

Bit-Error Probability of Precoded MSK

The bit-error probability of precoded MSK is

$$P_b = \mathrm{Q}(\sqrt{2\gamma_b}),$$

$\gamma_b = E_b/N_0$.

Sketch of the Proof:

The bit-error probability is given as

$$P_b = \lim_{L\to\infty} \frac{1}{L} \sum_{j=0}^{L-1} \Pr(\hat{U}_j \neq U_j).$$

All terms in this sum can shown to be identical, and thus

$$P_b = \Pr(\hat{U}_0 \neq U_0).$$

Due to the precoding, we have $U_0 = V_1$, and therefore

$$P_b = \Pr(\hat{V}_1 \neq V_1).$$

This expression can be written using conditional probabilities:

$$P_b = \sum_{v=0}^{1} \sum_{u=0}^{1} \Pr(\hat{V}_1 \neq V_1 | V_1 = v, U_1 = u) \cdot \Pr(V_1 = v, U_1 = u).$$

(Usual trick.) These conditional probabilities can shown to be all identical, and thus

$$P_b = \Pr(\hat{V}_1 \neq V_1 | V_1 = 0, U_1 = 0).$$

As $V_1 = U_0$ (property of precoded MSK), we obtain

$$P_b = \Pr(\hat{V}_1 = 1 | U_0 = 0, U_1 = 0).$$

The ML decision rule is

$$\hat{V}_1 = \begin{cases} 0 & \text{if} \quad y_{0,2} - y_{1,2} < y_{0,1} + y_{1,1}, \\ 1 & \text{if} \quad y_{0,2} - y_{1,2} > y_{0,1} + y_{1,1}. \end{cases}$$

Given the hypothesis $U_0 = U_1 = 0$, we have

$$\boldsymbol{X}_0 = \boldsymbol{X}_1 = \boldsymbol{s}_1 = (\sqrt{E_s}, 0)^\mathsf{T},$$

and so

$$\begin{aligned} Y_{0,1} &= \sqrt{E_s} + W_{0,1}, & Y_{0,2} &= W_{0,2}, \\ Y_{1,1} &= \sqrt{E_s} + W_{1,1}, & Y_{1,2} &= W_{1,2}. \end{aligned}$$

Remember:

$$\boldsymbol{Y}_0 = (Y_{0,1}, Y_{0,2})^\mathsf{T}, \qquad \boldsymbol{Y}_1 = (Y_{1,1}, Y_{1,2})^\mathsf{T}.$$

Therefore, the bit-error probability can be written as

$$\begin{aligned}
P_b &= \Pr(Y_{0,2} - Y_{1,2} > Y_{0,1} + Y_{1,1} | U_0 = 0, U_1 = 0) \\
&= \Pr(W_{0,2} - W_{1,2} > 2\sqrt{E_s} + W_{0,1} + W_{1,1}) \\
&= \Pr(\underbrace{W_{0,2} - W_{1,2} - W_{0,1} - W_{1,1}}_{W' \sim \mathcal{N}(0, 2N_0)} > 2\sqrt{E_s}) \\
&= \Pr\big(W'/\sqrt{2N_0} > \sqrt{2E_s/N_0}\big) \\
&= \mathrm{Q}\big(\sqrt{2E_s/N_0}\big) = \mathrm{Q}\big(\sqrt{2\gamma_b}\big).
\end{aligned}$$

As (precoded) MSK is a binary modulation scheme, we have $E_s = E_b$ and $\gamma_s = \gamma_b$.

Thus, the **bit-error probability of precoded MSK** is

$$P_b = \mathrm{Q}(\sqrt{2\gamma_b}),$$

which is the same as that of BPSK and QPSK.

8.15 Gaussian MSK

Gaussian MSK is a generalization of MSK, and it is used in the GSM mobile phone system.

Motivation

The phase function of MSK is continuous, but not its dervative. These "edges" in the transmit signal lead to some high-frequency components. By avoiding these "edges", i.e., by smoothing the phase function, the spectrum of the transmit signal can be made more compact.

Concept

The output signal of the transmitter is generated similar to that in MSK:

$$x(t) = \sqrt{2P}\cos\Big(2\pi f_1 t + \underbrace{\sum_{l=0}^{L-1} u_l \cdot 2\pi\, b(t-lT)}_{\phi(t)}\Big),$$

$t \in [0, LT)$, with the **phase pulse**

$$b(\tau) = \begin{cases} 0 & \text{for } \tau < 0, \\ \text{monotonically increasing} & \text{for } 0 \le \tau < JT, \\ 1/2 & \text{for } JT \le \tau, \end{cases}$$

for some integer $J \geq 1$.

Notice that there are two new degrees of freedom:

(i) The phase pulse may stretch over more than one symbol duration T.

(ii) The phase pulse may be be selected such that its derivative is zero for $t =$ and $t = JT$.

Phase Pulse and Frequency Pulse

In GMSK, the phase pulse $b(\tau)$ is constructed via its derivative $c(\tau)$, called the frequency pulse:

(i) The **frequency pulse** $c(\tau)$ is the convolution of a rectangular pulse (as in MSK!) with a Gaussian pulse:

$$c(\tau) = \text{rectangular pulse} * \text{Gaussian pulse},$$

where $c(\tau) = 0$ for $\tau \notin [0, JT)$, and $\int c(\tau)\mathsf{d}\tau = 1/2$. The resulting pulse is

$$c(\tau) = \mathrm{Q}\Big(\frac{2\pi B}{\sqrt{\ln 2}}\big(\tau - T/2\big) - \mathrm{Q}\Big(\frac{2\pi B}{\sqrt{\ln 2}}\big(\tau + T/2\big)\Big).$$

The value B corresponds to a bandwidth and is a design parameter.

(ii) The **phase pulse** is the integral over the frequency pulse:

$$b(\tau) = \int_0^\tau c(\tau')\mathsf{d}\tau'.$$

Due to the definition of the frequency pulse, we have $b(\tau) = 1/2$ for $\tau \geq JT$.

The resulting phase pulse fulfills the conditions given above.

In **GSM**, the parameter B is chosen such that $BT \approx 0.3$. The length of the phase pulse is then $L \approx 4$.

9 Quadrature Amplitude Modulation

10 BPAM with Bandlimited Waveforms

Part III

Appendix

A Geometrical Representation of Signals

A.1 Sets of Orthonormal Functions

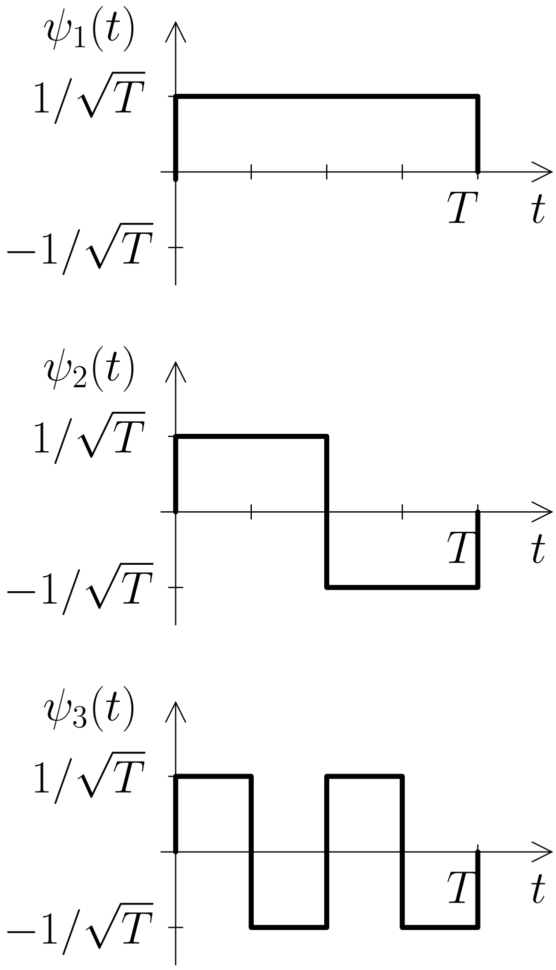
The D real-valued functions

$$\psi_1(t), \psi_2(t), \dots, \psi_D(t)$$

form an orthonormal set if

$$\int\limits_{-\infty}^{+\infty} \psi_k(t)\, \psi_l(t)\; dt = \begin{cases} 0 & \text{for } k \neq l \text{ (orthogonality)}, \\ 1 & \text{for } k = l \text{ (normalization)}. \end{cases}$$

EXAMPLE: $D = 3$



◇

A.2 Signal Space Spanned by an Orthonormal Set of Functions

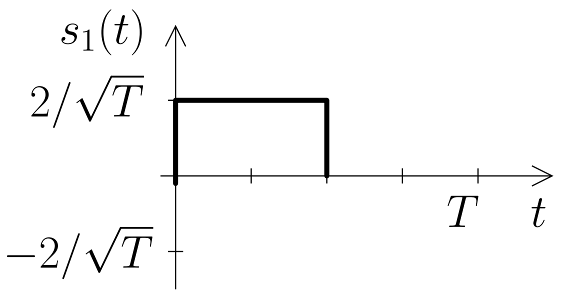
Let $\{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\}$ be a set of orthonormal functions. We denote by $\mathcal{S}_\psi$ the set containing all linear combinations of these functions:

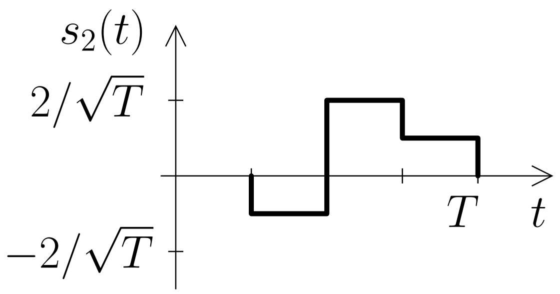
$$\mathcal{S}_\psi := \Big\{ s(t) : s(t) = \sum_{i=1}^{D} s_i \psi_i(t);\, s_1, s_2, \ldots, s_D \in \mathbb{R} \Big\}.$$

EXAMPLE: (continued)

The following functions are elements of $\mathcal{S}_\psi$:

$$s_1(t) = \psi_1(t) + \psi_2(t), \qquad s_2(t) = \tfrac{1}{2}\psi_1(t) - \psi_2(t) + \tfrac{1}{2}\psi_3(t)$$





◇

The set $\mathcal{S}_\psi$ is a vector space with respect to the following operations:

- Addition
$$+ : \big(s_1(t), s_2(t)\big) \mapsto s_1(t) + s_2(t)$$
- Multiplication with a scalar
$$\cdot : \big(a, s(t)\big) \mapsto a \cdot s(t)$$

A.3 Vector Representation of Elements in the Vector Space

With each element $s(t) \in \mathcal{S}_\psi$,

$$s(t) = \sum_{i=1}^{D} s_i \psi_i(t),$$

we associate the D-dimensional real vector

$$\boldsymbol{s} = [s_1, s_2, \dots, s_D]^\mathsf{T} \in \mathbb{R}^D.$$

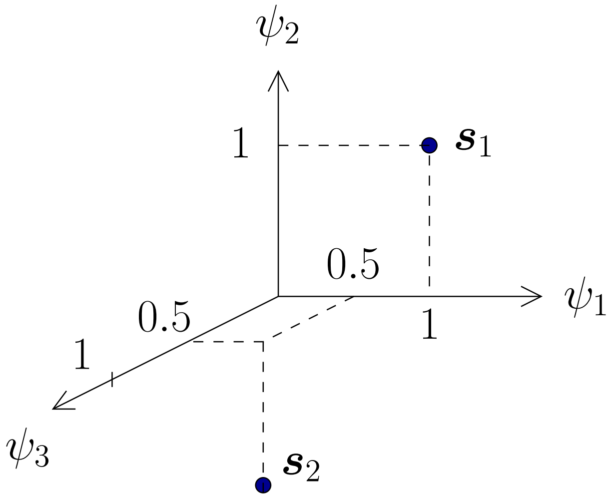
EXAMPLE: (continued)

$$\begin{aligned} s_1(t) &= \psi_1(t) + \psi_2(t) && \mapsto && \boldsymbol{s}_1 = [1, 1, 0]^T, \\ s_2(t) &= \tfrac{1}{2}\psi_1(t) - \psi_2(t) + \tfrac{1}{2}\psi_3(t) && \mapsto && \boldsymbol{s}_2 = [\tfrac{1}{2}, -1, \tfrac{1}{2}]^T. \end{aligned}$$

◇

Vectors in $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$, can be represented geometrically by points on the line, in the plane, an in the 3-dimensional Euclidean space, respectively.

EXAMPLE: (continued)



◇

A.4 Vector Isomorphism

The mapping $\mathcal{S}_\psi \to \mathbb{R}^D$,

$$s(t) = \sum_{i=1}^{D} s_i \psi_i(t) \quad \mapsto \quad \boldsymbol{s} = [s_1, s_2, \dots, s_D]^\mathsf{T},$$

is a (vector) isomorphism, i.e.:

(i) The mapping is bijective.

(ii) The mapping preserves the addition and the multiplication operators:

$$\begin{aligned} s_1(t) + s_2(t) \quad &\mapsto \quad \boldsymbol{s}_1 + \boldsymbol{s}_2, \\ a s(t) \quad &\mapsto \quad a\boldsymbol{s}. \end{aligned}$$

Since the mapping $\mathcal{S}_\psi \to \mathbb{R}^D$ is an isomorphism, $\mathcal{S}_\psi$ and $\mathbb{R}^D$ are "the same" or "equivalent" from an algebraic point of view.

A.5 Scalar Product and Isometry

Scalar product in $\mathcal{S}_\psi$:

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

for $s_1(t), s_2(t) \in \mathcal{S}_\psi$.

Scalar product in $\mathbb{R}^D$:

$$\langle \boldsymbol{s}_1, \boldsymbol{s}_2\rangle := \sum_{i=1}^{D} s_{1,i}s_{2,i}$$

for $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathbb{R}^D$.

Relation of scalar products:

Each element in $\mathcal{S}_\psi$ is a linear combination of $\psi_1(t), \psi_2(t), \ldots, \psi_D(t)$:

$$s_1(t) = \sum_{k=1}^{D} s_{1,k}\psi_k(t), \qquad s_2(t) = \sum_{l=1}^{D} s_{2,l}\psi_l(t).$$

Inserting the two sums into the definition of the scalar product, we obtain

$$\langle s_1(t), s_2(t)\rangle = \int \Big(\sum_{k=1}^{D} s_{1,k}\psi_k(t)\Big)\Big(\sum_{l=1}^{D} s_{2,l}\psi_l(t)\Big)\, dt =$$

$$= \sum_{k=1}^{D}\sum_{l=1}^{D} s_{1,k}s_{2,l} \underbrace{\int \psi_k(t)\psi_l(t)dt}_{\text{“orthonormality”}} = \sum_{k=1}^{D} s_{1,k}s_{2,k} = \langle \boldsymbol{s}_1, \boldsymbol{s}_2\rangle$$

Hence the mapping $\mathcal{S}_\psi \to \mathbb{R}^D$ preserves the scalar product, i.e., it is an **isometry**.

Norms in $\mathcal{S}_\psi$

$$\|s(t)\|^2 := \langle s(t), s(t) \rangle = \int s^2(t)\, dt = E_s$$

The value E_s is the energy of $s(t)$.

Norms in $\mathbb{R}^D$:

$$\|\boldsymbol{s}\|^2 := \langle \boldsymbol{s}, \boldsymbol{s} \rangle = \sum_{i=1}^{D} s_i^2$$

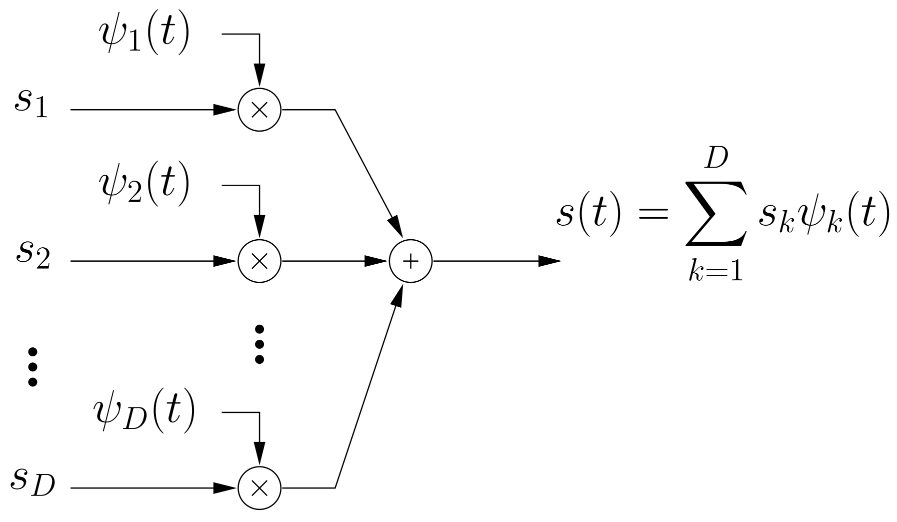
From the isometry and the definitions of the norms, we obtain the **Parseval relation**

$$\|s(t)\|^2 = \|\boldsymbol{s}\|^2.$$

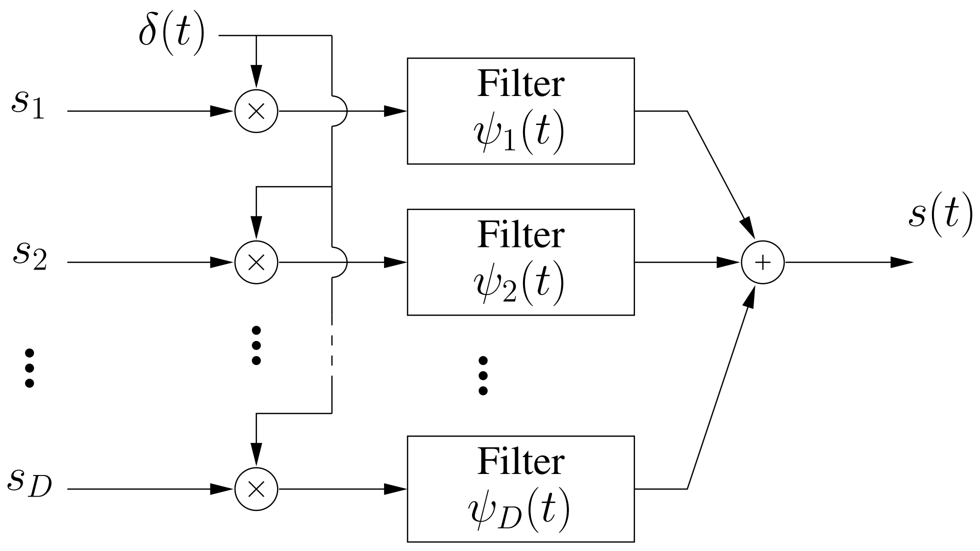
A.6 Waveform Synthesis

Let $\{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\}$ be a set of orthonormal functions spanning the vector space $\mathcal{S}_\psi$. Every element $s(t) \in \mathcal{S}_\psi$ can be synthesized from its vector representation $\boldsymbol{s} = [s_1, s_2, \ldots, s_D]^\mathsf{T}$ by one of the following devices.

Multiplicative-based Synthesizer



Filter-bank based Synthesizer



$$s(t) = \sum_{k=1}^{D} s_k \delta(t) * \psi_k(t) \quad \text{with Dirac function } \delta(t)$$

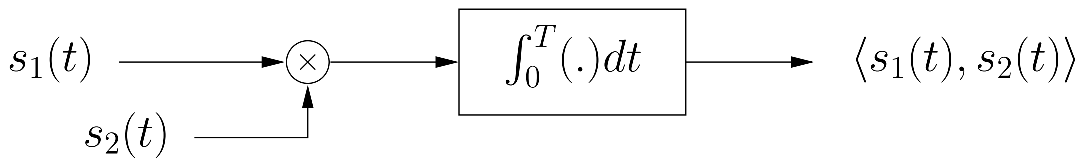
A.7 Implementation of the Scalar Product

Let $s_1(t)$ and $s_2(t)$ be time-limited to the interval $[0, T]$. Then

$$\langle s_1(t), s_2(t)\rangle := \int_0^T s_1(t)s_2(t)\, dt.$$

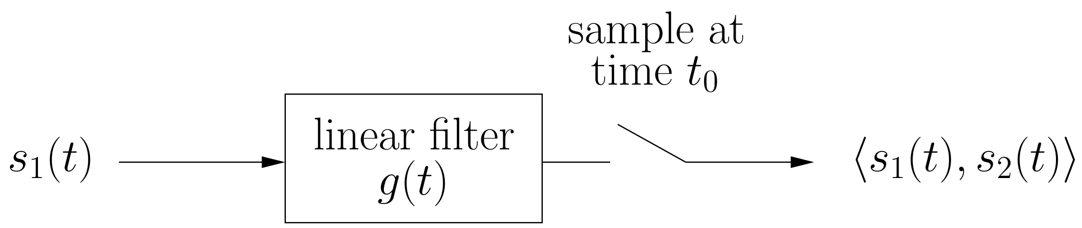
The scalar product can be computed by the following two devices.

Correlator-based Implementation



Matched-filter based Implementation

We seek a linear filter that outputs the value $\langle s_1(t), s_2(t)\rangle$ at a certain time, say t_0.



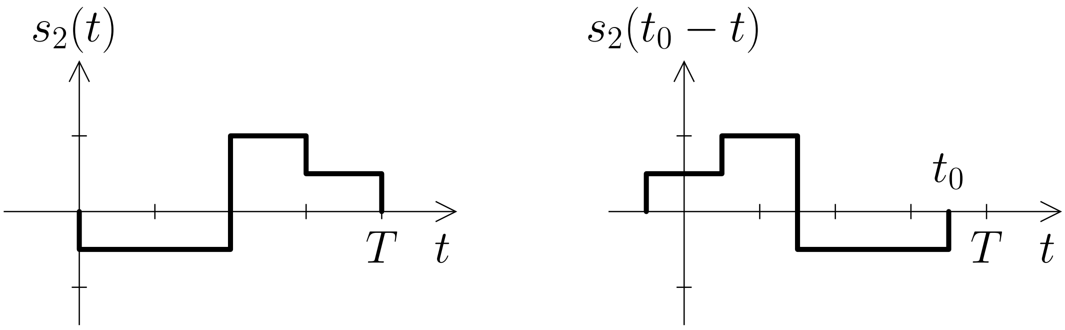
We want

$$\langle s_1(t), s_2(t)\rangle = \int_0^T s_1(\tau)s_2(\tau)\, d\tau \stackrel{!}{=} \int_{-\infty}^{+\infty} s_1(\tau)g(t'-\tau)\, d\tau\Big|_{t'=t_0}$$

The second equality holds for

$$s_2(\tau) = g(t_0 - \tau) \qquad \Leftrightarrow \qquad g(t) = s_2(t_0 - t)$$

EXAMPLE:

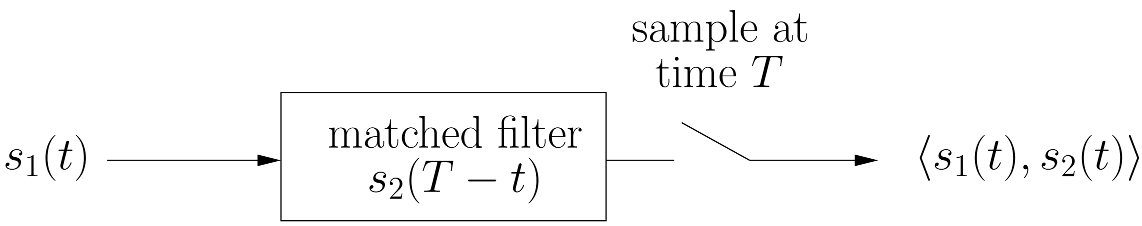


◇

The filter is causal and thus realizable for a function time-limited to the interval $[0, T]$ if

$$t_0 \geq T.$$

The value t_0 is the time at which the value $\langle s_1(t), s_2(t) \rangle$ appears at the filter output. Choosing the minimum value, i.e., $t_0 = T$, we obtain the following (realizable) implementation.



The filter is matched to $s_2(t)$.

A.8 Computation of the Vector Representation

Let $\{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\}$ be a set of orthonormal functions time-limited to $[0, T]$ and spanning the vector space $\mathcal{S}_\psi$.

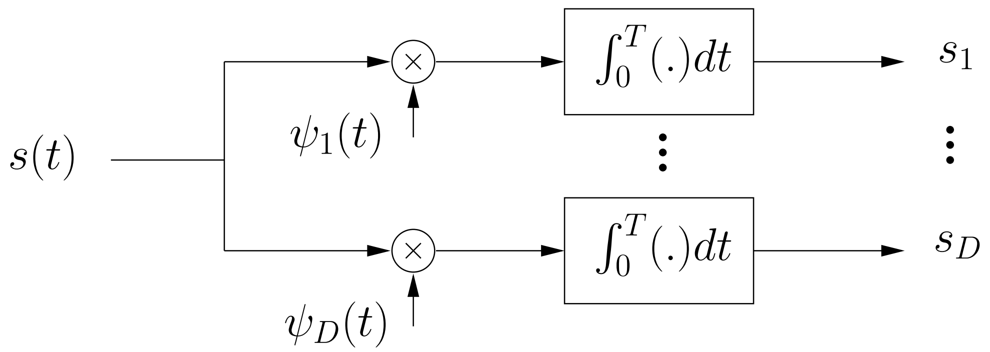
We seek a device that computes the vector representation $\boldsymbol{s} \in \mathbb{R}^D$ for any function $s(t) \in \mathcal{S}_\psi$, i.e.,

$$\boldsymbol{s} = [s_1, s_2, \ldots, s_D]^\mathsf{T} \qquad \text{such that} \qquad s(t) = \sum_{k=1}^{D} s_k \psi_k(t).$$

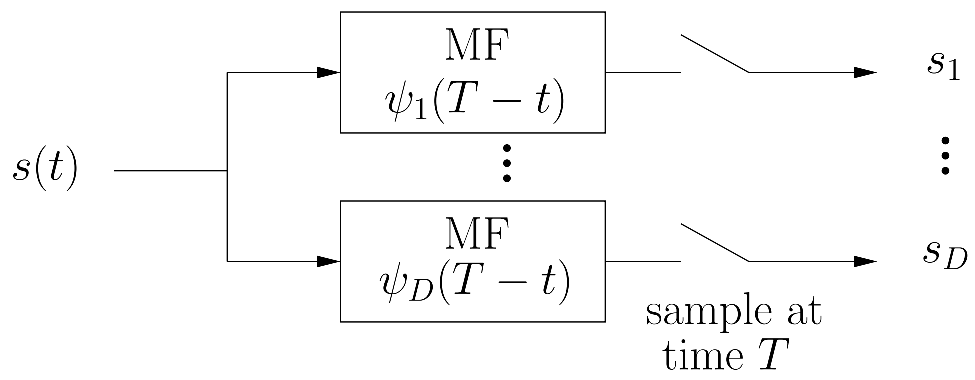
The components of $\boldsymbol{s}$ can be computed as

$$s_k = \langle s(t), \psi_k(t) \rangle, \quad k = 1, 2, \ldots, D.$$

Correlator-based Implementation



Matched-filter based Implementation



A.9 Gram-Schmidt Orthonormalization

Assume a set of M signal waveforms

$$\Big\{s_1(t), s_2(t), \ldots, s_M(t)\Big\}$$

spanning a D-dimensional vector space. From these waveforms, we can construct a set of $D \leq M$ orthonormal waveforms

$$\Big\{\psi_1(t), \psi_2(t), \ldots, \psi_D(t)\Big\}$$

by applying the following procedure.

Remember:

$$\langle s_1(t), s_2(t)\rangle = \int s_1(t)s_2(t)\, dt$$
$$\|s(t)\|^2 := \langle s(t), s(t)\rangle = E_s$$

Gram-Schmidt orthonormalization procedure

1. Take the first waveform $s_1(t)$ and normalize it to unit energy:

$$\psi_1(t) := \frac{s_1(t)}{\sqrt{\|s_1(t)\|^2}}.$$

2. Take the second waveform $s_2(t)$. Determine the projection of $s_2(t)$ onto $\psi_1(t)$, and subtract this projection from $s_2(t)$:

$$c_{2,1} = \langle s_2(t), \psi_1(t)\rangle,$$
$$s_2'(t) = s_2(t) - c_{2,1}\psi_1(t).$$

   Normalize $s_2'(t)$ to unit energy:

$$\psi_2(t) := \frac{s_2'(t)}{\sqrt{\|s_2'(t)\|^2}}.$$

3. Take the third waveform $s_3(t)$. Determine the projections of $s_3(t)$ onto $\psi_1(t)$ and onto $\psi_2(t)$, and subtract these projections from $s_3(t)$:

$$\begin{aligned} c_{3,1} &= \langle s_3(t), \psi_1(t) \rangle, \\ c_{3,2} &= \langle s_3(t), \psi_2(t) \rangle, \\ s'_3(t) &= s_3(t) - c_{3,1}\psi_1(t) - c_{3,2}\psi_2(t). \end{aligned}$$

   Normalize $s'_3(t)$ to unit energy:

$$\psi_3(t) := \frac{s'_3(t)}{\sqrt{\|s'_3(t)\|^2}}.$$

4. Proceed in this way for all other waveforms $s_k(t)$ until $k = D$. Notice that $s'_k(t) = 0$ for $k > D$. (Why?)

The **general rule** is the following. For $k = 1, 2, \ldots, D$,

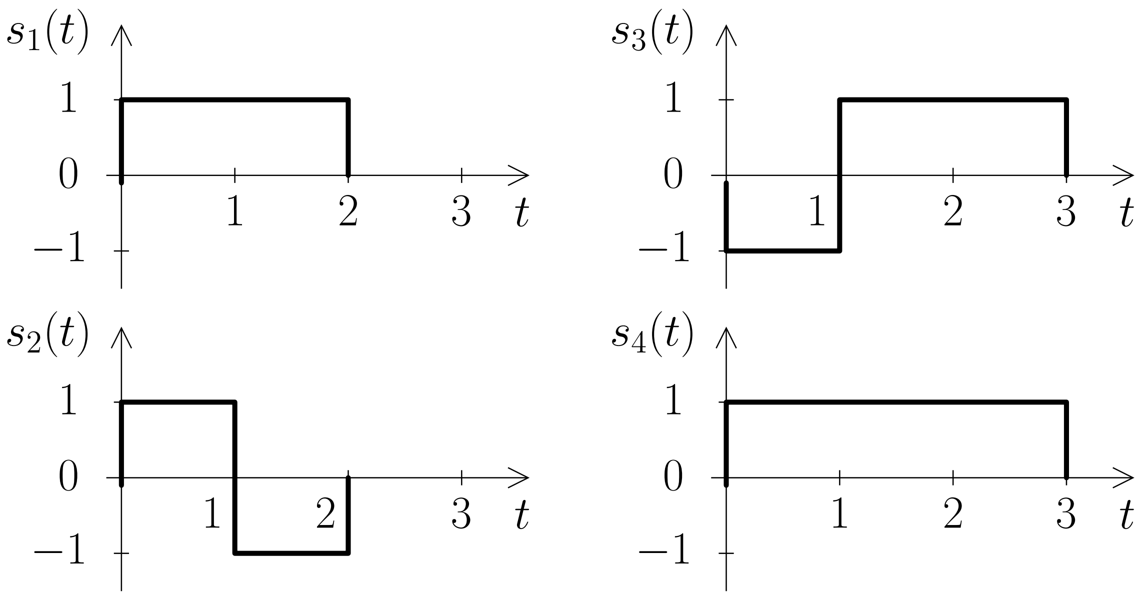
$$\begin{aligned} c_{k,i} &= \langle s_k(t), \psi_i(t) \rangle, \qquad\qquad i = 1, 2, \ldots, k-1 \\ s'_k(t) &= s_k(t) - \sum_{i=1}^{k-1} c_{k,i}\psi_i(t), \\ \psi_k(t) &:= \frac{s'_k(t)}{\sqrt{\|s'_k(t)\|^2}}. \end{aligned}$$

Due to this procedure, the resulting waveforms $\psi_k(t)$ are orthogonal and normalized, as desired, and they form an orthonormal basis of the D-dimensional vector space of the waveforms $s_i(t)$.

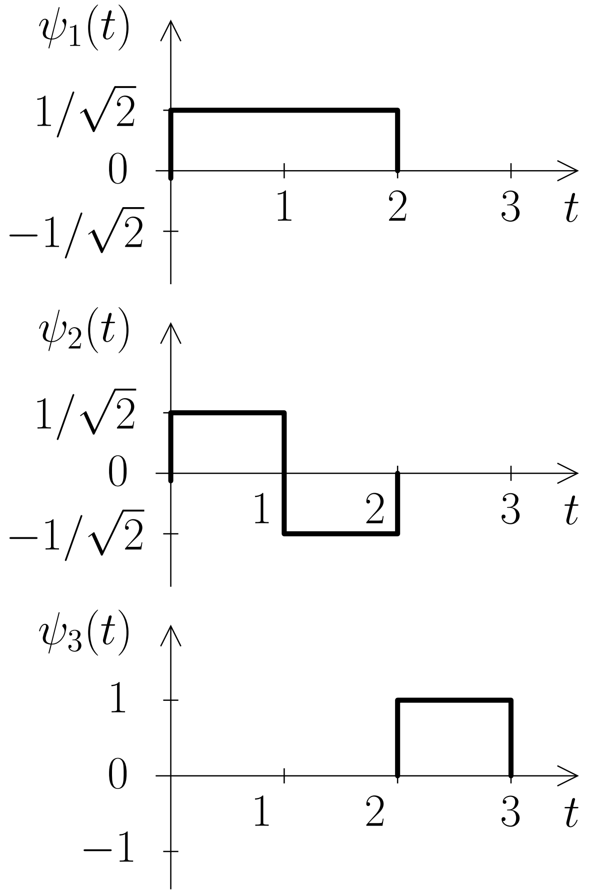
Remark: The set of orthonormal waveforms is not unique.

Example: Orthonormalization

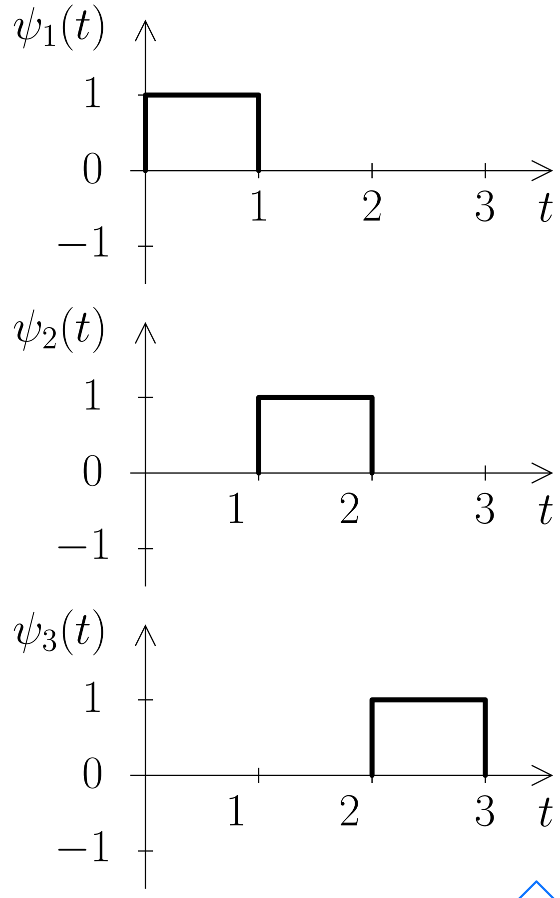
Original signal waveforms:



Set 1 of orthonormal waveforms:



Set 2 of orthonormal waveforms:



B Orthogonal Transformation of a White Gaussian Noise Vector

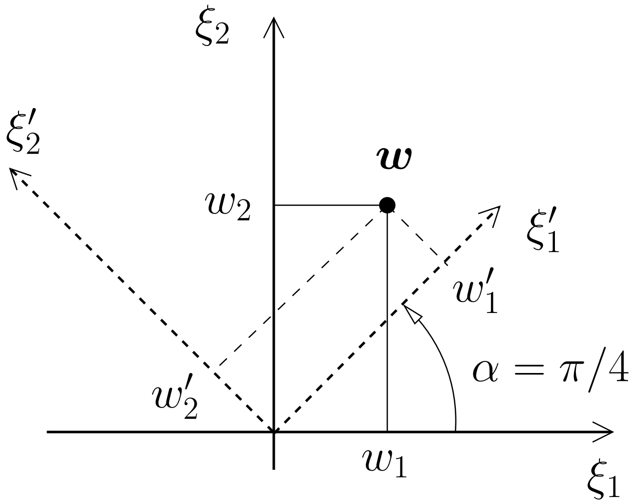
The statistical properties of a white Gaussian noise vector (WGNV) are invariant with respect to orthonormal transformations. This is first shown for a vector with two dimensions, and then for a vector with an arbitrary number of dimensions.

B.1 The Two-Dimensional Case

Let $\boldsymbol{W}$ denote a 2-dimensional WGNV, i.e.,

$$\boldsymbol{W} = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right).$$

Consider the original coordinate system (ξ_1, ξ_2) and the rotated coordinate system (ξ_1', ξ_2').



Let $[W_1', W_2']^\mathsf{T}$ denote the components of $\boldsymbol{W}$ expressed with respect to the rotated coordinate system (ξ_1', ξ_2'):

$$\begin{aligned} W_1' &= +W_1 \cos\alpha + W_2 \sin\alpha \\ W_2' &= -W_1 \sin\alpha + W_2 \cos\alpha. \end{aligned}$$

Writing this in matrix notation as

$$\begin{bmatrix} W_1' \\ W_2' \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix},$$

the vector $\boldsymbol{W}'$ can seen to be the orthonormal transformation of the vector $\boldsymbol{W}$.

The resulting vector $\boldsymbol{W}' = [W_1', W_2']^\mathsf{T}$ is also a two-dimensional WGNV with the same covariance matrix as $\boldsymbol{W} = [W_1, W_2]^\mathsf{T}$:

$$\boldsymbol{W}' = \begin{bmatrix} W_1' \\ W_2' \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \right).$$

Proof: See problem.

B.2 The General Case

Let $\boldsymbol{W} = [W_1, \ldots, W_D]^\mathsf{T}$ denote a D-dimensional WGNV with zero mean and covariance matrix $\sigma^2\boldsymbol{I}$, i.e.,

$$\boldsymbol{W} \sim \mathcal{N}\big(\boldsymbol{0}, \sigma^2\boldsymbol{I}\big).$$

($\boldsymbol{0}$ is a column vector of length D, and $\boldsymbol{I}$ is the $D \times D$ identity matrix.) Let further $\boldsymbol{V}$ denote an orthonormal transformation $\mathbb{R}^D \to \mathbb{R}^D$, i.e., a linear transformation with the property

$$\boldsymbol{V}\boldsymbol{V}^\mathsf{T} = \boldsymbol{V}^\mathsf{T}\boldsymbol{V} = \boldsymbol{I}.$$

Then the transformed vector

$$\boldsymbol{W}' = \boldsymbol{V}\boldsymbol{W}$$

is also a D-dimensional WGNV with the same covariance matrix as $\boldsymbol{W}$:

$$\boldsymbol{W}' \sim \mathcal{N}\big(\boldsymbol{0}, \sigma^2\boldsymbol{I}\big).$$

Proof

(i) A linear transformation of a Gaussian random vector is again a Gaussian random vector; thus $\boldsymbol{W}'$ is Gaussian.

(ii) The mean value is

$$\mathrm{E}\big[\boldsymbol{W}'\big] = \mathrm{E}\big[\boldsymbol{V}\boldsymbol{W}\big] = \boldsymbol{V}\,\mathrm{E}\big[\boldsymbol{W}\big] = \boldsymbol{0}.$$

(iii) The covariance matrix is

$$\begin{aligned}\mathrm{E}\big[\boldsymbol{W}'\boldsymbol{W}'^{\mathsf{T}}\big] &= \mathrm{E}\big[\boldsymbol{V}\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}}\boldsymbol{V}^{\mathsf{T}}\big] \\ &= \boldsymbol{V}\underbrace{\mathrm{E}\big[\boldsymbol{W}\boldsymbol{W}^{\mathsf{T}}\big]}_{\sigma^2\boldsymbol{I}}\boldsymbol{V}^{\mathsf{T}} = \sigma^2\underbrace{\boldsymbol{V}\boldsymbol{V}^{\mathsf{T}}}_{\boldsymbol{I}} = \sigma^2\boldsymbol{I}.\end{aligned}$$

Comment

In the case $D = 2$, $\boldsymbol{V}$ takes the form

$$\boldsymbol{V} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$$

for some $\alpha \in [0, 2\pi)$.

C The Shannon Limit

C.1 Capacity of the Band-limited AWGN Channel

Consider an AWGN channel with bandwidth W [1/sec]. The capacity of this channel is given by

$$C_W = W \ \log_2\bigl(1 + \frac{S}{WN_0}\bigr) \quad \text{[bit/sec]}.$$

Consider further a digital transmission scheme operating at transmission rate R [bit/sec] with limited transmit power S.

Channel Coding Theorem (C.E. Shannon, 1948):

> Consider a binary random sequence generated by a BSS that is transmitted over an AWGN channel with bandwidth W using some transmission scheme.
>
> Transmission with an arbitrarily small probability of error (using an appropriate coding/modulation scheme) is possible provided that the transmission rate is less than the channel capacity, i.e.,
>
> $$P_b \to 0 \qquad \text{is possible if} \qquad R < C_W.$$
>
> Conversely, error-free transmission is impossible if the rate is larger than the channel capacity, i.e.,
>
> $$P_b \to 0 \qquad \text{is impossible if} \qquad R > C_W.$$

A coding/modulation scheme can be characterized by its spectral efficiency and its power efficiency. The **spectral efficiency** is the ratio R/W, and it has the unit $\frac{\text{bit}}{\text{sec Hz}}$. The **power efficiency** is the SNR $\gamma_b = E_b/N_0$ necessary to achieve a certain error probability (e.g., 10^{-5}).

Example: MPAM

The rate is given by $R = \frac{1}{T}\log_2 M$. The bandwidth required to transmit pulses of duration T is about $W = 1/T$. The spectral efficiency is thus

$$\frac{R}{W} = \log_2 M \frac{\text{bit}}{\text{sec Hz}}.$$

Error probabilities can be found in [1, p. 412, 435]. ◇

Example: MPPM

The rate is given by $R = \frac{1}{T}\log_2 M$. The bandwidth required to transmit pulses of duration T/M is about $W = M/T$. The spectral efficiency is thus

$$\frac{R}{W} = \frac{\log_2 M}{M} \frac{\text{bit}}{\text{sec Hz}}.$$

Error probabilities can be found in [1, p. 426, 435]. ◇

A theoretical limit for the relation between the spectral efficiency and the power efficiency (for error-free transmission) can be established using the channel coding theorem.

The transmit power is given by $S = RE_b$. Combining this relation, the channel capacity, and the condition $R < C_W$, we obtain

$$\frac{R}{W} < \log_2\Big(1 + \frac{RE_b}{WN_0}\Big) = \log_2\Big(1 + \frac{R}{W}\gamma_b\Big).$$

This can be rewritten as

$$\gamma_b > \frac{2^{R/W} - 1}{R/W}.$$

Using the maximal rate for error-free transmission, $R = C$, we obtain the theoretical bound

$$\gamma_b = \frac{2^{C/W} - 1}{C/W}.$$

The plot can be found in [1, pp. 587].

The lower limit of γ_b for $R/W \to 0$ is called the **Shannon limit** for the AWGN channel:

$$\gamma_b > \lim_{R/W \to 0} \frac{2^{R/W} - 1}{R/W} = \ln 2.$$

Notice that $10 \log_{10}(\ln 2) \approx -1.6$ dB. Thus, error-free transmission over the AWGN channel is impossible if the SNR per bit is lower than -1.6 dB.

C.2 Capacity of the Band-Unlimited AWGN Channel

For the AWGN channel with **unlimited bandwidth** (but still limited transmit power), the capacity is given by

$$C_\infty = \lim_{W \to \infty} W \ \log_2\Big(1 + \frac{S}{W N_0}\Big) = \frac{1}{\ln 2} \frac{S}{N_0}.$$

The condition $R < C_\infty$ becomes then equivalent with the condition $\gamma_b > \ln 2$. (Remember: $S = RE_b$.)