Digital Communication Systems

By:
Behnaam Aazhang

Digital Communication Systems

By:
Behnaam Aazhang

Online:
$<$ http://cnx.org/content/col10134/1.3/ $>$

C O N N E X I O N S

Rice University, Houston, Texas

Table of Contents

Chapter 1

Chapter 1

Chapter 2

Chapter 2

2.1 Review of Probability Theory[1]

The focus of this course is on digital communication, which involves transmission of information, in its most general sense, from source to destination using digital technology. Engineering such a system requires modeling both the information and the transmission media. Interestingly, modeling both digital or analog information and many physical media requires a probabilistic setting. In this chapter and in the next one we will review the theory of probability, model random signals, and characterize their behavior as they traverse through deterministic systems disturbed by noise and interference. In order to develop practical models for random phenomena we start with carrying out a random experiment. We then introduce definitions, rules, and axioms for modeling within the context of the experiment. The outcome of a random experiment is denoted by ω. The sample space Ω is the set of all possible outcomes of a random experiment. Such outcomes could be an abstract description in words. A scientific experiment should indeed be repeatable where each outcome could naturally have an associated probability of occurrence. This is defined formally as the ratio of the number of times the outcome occurs to the total number of times the experiment is repeated.

2.1.1 Random Variables

A random variable is the assignment of a real number to each outcome of a random experiment.
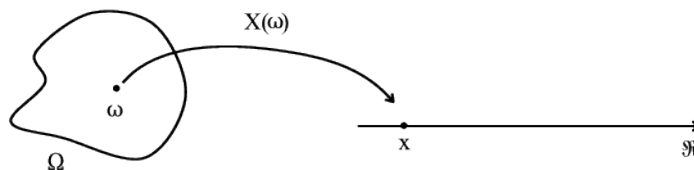
Figure 2.1

Example 2.1
Roll a dice. Outcomes $\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$

[1]This content is available online at <http://cnx.org/content/m10224/2.16/>.

$\omega_i = i$ dots on the face of the dice.

$X(\omega_i) = i$

2.1.2 Distributions

Probability assignments on intervals $a < X \le b$

Definition 2.1: Cumulative distribution

The cumulative distribution function of a random variable X is a function $F_X\ (\mathbb{R} \mapsto \mathbb{R})$ such that

$$
\begin{aligned}
F_X\ (b) &= Pr\,[X \le b] \\
&= Pr\,[\{\,\omega \in \Omega \mid X(\omega) \le b\,\}]
\end{aligned}
\tag{2.1}
$$

$$A = \{\omega \in \Omega \mid X(\omega) \le b\ \}$$
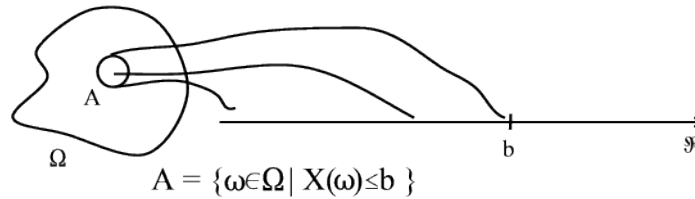
Figure 2.2

Definition 2.2: Continuous Random Variable

A random variable X is continuous if the cumulative distribution function can be written in an integral form, or

$$
F_X\ (b) = \int_{-\infty}^{b} f_X\ (x)\,dx
\tag{2.2}
$$

and $f_X\ (x)$ is the probability density function (pdf) (e.g., $F_X\ (x)$ is differentiable and $f_X\ (x) = \frac{d}{dx}(F_X\ (x))$)

Definition 2.3: Discrete Random Variable

A random variable X is discrete if it only takes at most countably many points (i.e., $F_X\ (\cdot)$ is piecewise constant). The probability mass function (pmf) is defined as

$$
\begin{aligned}
p_X\ (x_k) &= Pr\,[X = x_k] \\
&= F_X\ (x_k) - \lim_{x(x \to x_k)\ \cdot\ (x < x_k)} F_X\ (x)
\end{aligned}
\tag{2.3}
$$

Two random variables defined on an experiment have joint distribution

$$
\begin{aligned}
F_{X,,,Y}\ (a,b) &= Pr\,[X \le a, Y \le b] \\
&= Pr\,[\{\,\omega \in \Omega \mid (X(\omega) \le a)\ \cdot\ (Y(\omega) \le b)\,\}]
\end{aligned}
\tag{2.4}
$$
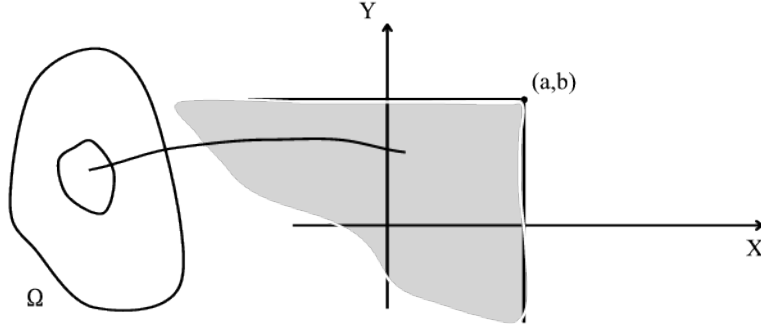
Figure 2.3

Joint pdf can be obtained if they are jointly continuous

$$F_{X,,,Y}(a,b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f_{X,Y}(x,y)\,dxdy \tag{2.5}$$

(e.g., $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,,,Y}(x,y)}{\partial x \partial y}$)
Joint pmf if they are jointly discrete

$$p_{X,Y}(x_k, y_l) = Pr[X = x_k, Y = y_l] \tag{2.6}$$

Conditional density function

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \tag{2.7}$$

for all x with $f_X(x) > 0$ otherwise conditional density is not defined for those values of x with $f_X(x) = 0$
Two random variables are **independent** if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \tag{2.8}$$

for all $x \in \mathbb{R}$ and $y \in \mathbb{R}$. For discrete random variables,

$$p_{X,Y}(x_k, y_l) = p_X(x_k)p_Y(y_l) \tag{2.9}$$

for all k and l.

2.1.3 Moments
Statistical quantities to represent some of the characteristics of a random variable.

$$\begin{aligned} \overline{g(X)} &= E[g(X)] \\ &= \begin{cases} \int_{-\infty}^{\infty} g(x)f_X(x)\,dx & \text{if continuous} \\ \sum_k g(x_k)p_X(x_k) & \text{if discrete} \end{cases} \end{aligned} \tag{2.10}$$

- Mean

$$\mu_X = \bar{X} \tag{2.11}$$

- Second moment

$$E\left[X^2\right] = \bar{X^2} \tag{2.12}$$

- Variance

$$\begin{aligned}
\mathrm{Var}\left(X\right) &= \sigma\left(X\right)^2 \\
&= \overline{\left(X - \mu_X\right)^2} \\
&= \bar{X^2} - \mu_X{}^2
\end{aligned} \tag{2.13}$$

- Characteristic function

$$\Phi_X\left(u\right) = \overline{e^{juX}} \tag{2.14}$$

for $u \in \mathbb{R}$, where $j = \sqrt{-1}$

- Correlation between two random variables

$$\begin{aligned}
R_{XY} &= \overline{XY^*} \\
&= \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy^* f_{X,Y}\left(x,y\right) dx dy & \text{if } X \text{ and } Y \text{ are jointly continuous} \\ \sum_k \sum_l x_k y_l^* p_{X,Y}\left(x_k, y_l\right) & \text{if } X \text{ and } Y \text{ are jointly discrete} \end{cases}
\end{aligned} \tag{2.15}$$

- Covariance

$$\begin{aligned}
C_{XY} &= \mathrm{Cov}\left(X,Y\right) \\
&= \overline{\left(X - \mu_X\right)\left(Y - \mu_Y\right)^*} \\
&= R_{XY} - \mu_X \mu_Y^*
\end{aligned} \tag{2.16}$$

- Correlation coefficient

$$\rho_{XY} = \frac{\mathrm{Cov}\left(X,Y\right)}{\sigma_X \sigma_Y} \tag{2.17}$$

Definition 2.4: Uncorrelated random variables
Two random variables X and Y are uncorrelated if $\rho_{XY} = 0$.

Chapter 3

Chapter 3

3.1 Introduction to Stochastic Processes[1]

3.1.1 Definitions, distributions, and stationarity

Definition 3.1: Stochastic Process
Given a sample space, a stochastic process is an indexed collection of random variables defined for each $\omega \in \Omega$.

$$X_t(\omega) \quad , \quad t \in \mathbb{R} \tag{3.1}$$

Example 3.1
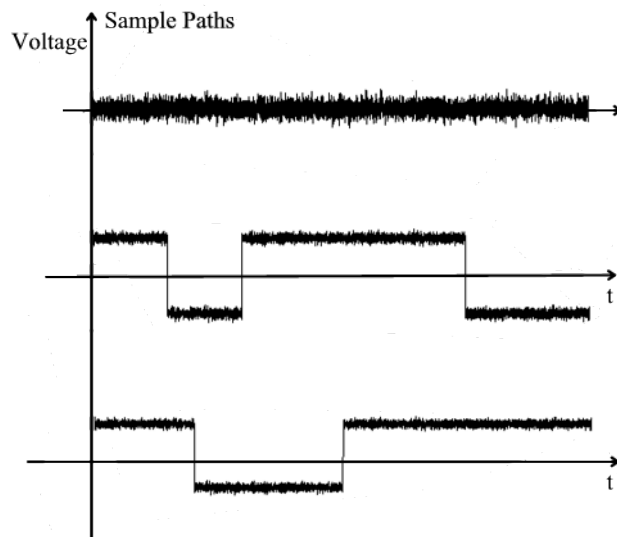Received signal at an antenna as in Figure 3.1.

Figure 3.1

For a given t, $X_t(\omega)$ is a random variable with a distribution

First-order distribution

$$
\begin{aligned}
F_{X_t}(b) &= Pr[X_t \leq b] \\
&= Pr[\{\omega \in \Omega \mid X_t(\omega) \leq b\}]
\end{aligned}
\tag{3.2}
$$

Definition 3.2: First-order stationary process
If $F_{X_t}(b)$ is not a function of time then X_t is called a first-order stationary process.

Second-order distribution

$$
F_{X_{t_1}, X_{t_2}}(b_1, b_2) = Pr[X_{t_1} \leq b_1, X_{t_2} \leq b_2]
\tag{3.3}
$$

for all $t_1 \in \mathbb{R}$, $t_2 \in \mathbb{R}$, $b_1 \in \mathbb{R}$, $b_2 \in \mathbb{R}$

Nth-order distribution

$$
F_{X_{t_1}, X_{t_2}, \ldots, X_{t_N}}(b_1, b_2, \ldots, b_N) = Pr[X_{t_1} \leq b_1, \ldots, X_{t_N} \leq b_N]
\tag{3.4}
$$

Nth-order stationary : A random process is stationary of order N if

$$
F_{X_{t_1}, X_{t_2}, \ldots, X_{t_N}}(b_1, b_2, \ldots, b_N) = F_{X_{t_1+T}, X_{t_2+T}, \ldots, X_{t_N+T}}(b_1, b_2, \ldots, b_N)
\tag{3.5}
$$

Strictly stationary : A process is strictly stationary if it is Nth order stationary for all N.

Example 3.2
$X_t = \cos(2\pi f_0 t + \Theta(\omega))$ where f_0 is the deterministic carrier frequency and $\Theta(\omega) : \Omega \rightarrow \mathbb{R}$ is a random variable defined over $[-\pi, \pi]$ and is assumed to be a uniform random variable; *i.e.,*

$$
f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & \text{if } \theta \in [-\pi, \pi] \\ 0 & \text{otherwise} \end{cases}
$$

$$
\begin{aligned}
F_{X_t}(b) &= Pr[X_t \leq b] \\
&= Pr[\cos(2\pi f_0 t + \Theta) \leq b]
\end{aligned}
\tag{3.6}
$$

$$
F_{X_t}(b) = Pr[-\pi \leq 2\pi f_0 t + \Theta \leq -\arccos(b)] + Pr[\arccos(b) \leq 2\pi f_0 t + \Theta \leq \pi]
\tag{3.7}
$$

$$
\begin{aligned}
F_{X_t}(b) &= \int_{(-\pi)-2\pi f_0 t}^{(-\arccos(b))-2\pi f_0 t} \frac{1}{2\pi} d\theta + \int_{\arccos(b)-2\pi f_0 t}^{\pi-2\pi f_0 t} \frac{1}{2\pi} d\theta \\
&= (2\pi - 2\arccos(b)) \frac{1}{2\pi}
\end{aligned}
\tag{3.8}
$$

$$
\begin{aligned}
f_{X_t}(x) &= \frac{d}{dx}\left(1 - \frac{1}{\pi}\arccos(x)\right) \\
&= \begin{cases} \frac{1}{\pi\sqrt{1-x^2}} & \text{if } |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
\tag{3.9}
$$

This process is stationary of order 1.

Plots of Cosines with different Phases and the same frequency
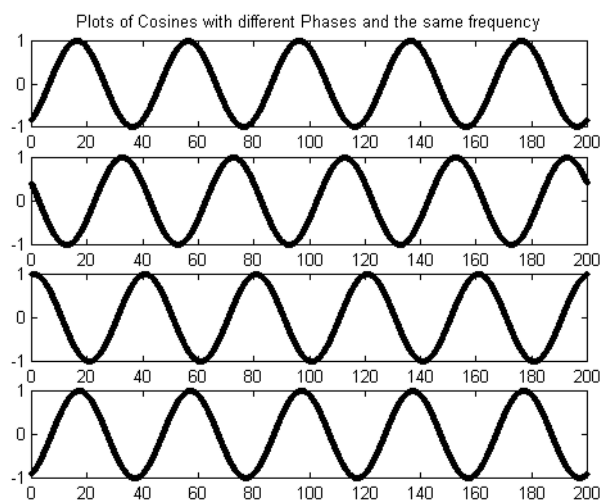
Figure 3.2

The second order stationarity can be determined by first considering conditional densities and the joint density. Recall that

$$X_t = \cos\left(2\pi f_0 t + \Theta\right) \tag{3.10}$$

Then the relevant step is to find

$$Pr\left[X_{t_2} \le b_2 \mid X_{t_1} = x_1\right] \tag{3.11}$$

Note that

$$\left(X_{t_1} = x_1 = \cos\left(2\pi f_0 t + \Theta\right)\right) \Rightarrow \left(\Theta = \arccos\left(x_1\right) - 2\pi f_0 t\right) \tag{3.12}$$

$$\begin{aligned} X_{t_2} &= \cos\left(2\pi f_0 t_2 + \arccos\left(x_1\right) - 2\pi f_0 t_1\right) \\ &= \cos\left(2\pi f_0 \left(t_2 - t_1\right) + \arccos\left(x_1\right)\right) \end{aligned} \tag{3.13}$$
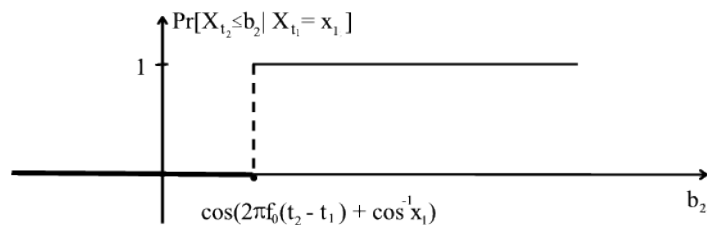
Figure 3.3

$$F_{X_{t_2},X_{t_1}}(b_2,b_1) = \int_{-\infty}^{b_1} f_{X_{t_1}}(x_1) \, Pr\left[X_{t_2} \leq b_2 \mid X_{t_1} = x_1\right] dx\, 1 \tag{3.14}$$

Note that this is only a function of $t_2 - t_1$.

Example 3.3
Every T seconds, a fair coin is tossed. If heads, then $X_t = 1$ for $nT \leq t < (n+1)T$. If tails, then $X_t = -1$ for $nT \leq t < (n+1)T$.
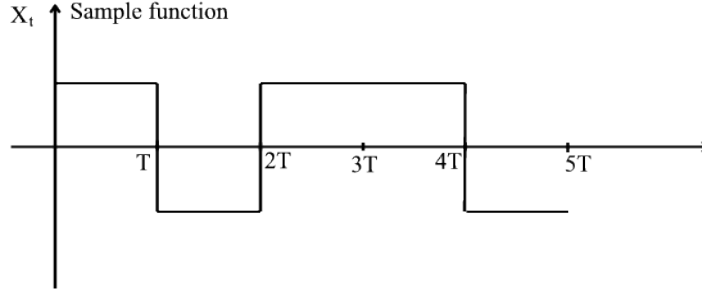
Figure 3.4

$$p_{X_t}(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = -1 \end{cases} \tag{3.15}$$

for all $t \in \mathbb{R}$. X_t is stationary of order 1.
    Second order probability mass function

$$p_{X_{t_1}X_{t_2}}(x_1,x_2) = p_{X_{t_2}|X_{t_1}}(x_2|x_1)\, p_{X_{t_1}}(x_1) \tag{3.16}$$

The conditional pmf

$$p_{X_{t_2}|X_{t_1}}(x_2|x_1) = \begin{cases} 0 & \text{if } x_2 \neq x_1 \\ 1 & \text{if } x_2 = x_1 \end{cases} \tag{3.17}$$

when $nT \leq t_1 < (n+1)T$ and $nT \leq t_2 < (n+1)T$ for some n.

$$p_{X_{t_2}|X_{t_1}}(x_2|x_1) = p_{X_{t_2}}(x_2) \tag{3.18}$$

for all x_1 and for all x_2 when $nT \leq t_1 < (n+1)T$ and $mT \leq t_2 < (m+1)T$ with $n \neq m$

$$p_{X_{t_2}X_{t_1}}(x_2,x_1) = \begin{cases} 0 & \text{if } x_2 \neq x_1 \text{for } nT \leq t_1, t_2 < (n+1)T \\ p_{X_{t_1}}(x_1) & \text{if } x_2 = x_1 \text{for } nT \leq t_1, t_2 < (n+1)T \\ p_{X_{t_1}}(x_1)\, p_{X_{t_2}}(x_2) & \text{if } n \neq m \text{for } (nT \leq t_1 < (n+1)T) \cdot (mT \leq t_2 < (m+1)T) \end{cases} \tag{3.19}$$

3.2 Second-order Description[2]

3.2.1 Second-order description

Practical and incomplete statistics

Definition 3.3: Mean
The mean function of a random process X_t is defined as the expected value of X_t for all t's.

$$
\begin{aligned}
\mu_{X_t} &= E\left[X_t\right] \\
&= \begin{cases} \int_{-\infty}^{\infty} x f_{X_t}\left(x\right) dx & \text{if continuous} \\ \sum_{k=-\infty}^{\infty} x_k p_{X_t}\left(x_k\right) & \text{if discrete} \end{cases}
\end{aligned}
\tag{3.20}
$$

Definition 3.4: Autocorrelation
The autocorrelation function of the random process X_t is defined as

$$
\begin{aligned}
R_X\left(t_2, t_1\right) &= E\left[X_{t_2} X_{t_1}{}^*\right] \\
&= \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 x_1{}^* f_{X_{t_2}, X_{t_1}}\left(x_2, x_1\right) dx\,1dx\,2 & \text{if continuous} \\ \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x_l x_k{}^* p_{X_{t_2}, X_{t_1}}\left(x_l, x_k\right) & \text{if discrete} \end{cases}
\end{aligned}
\tag{3.21}
$$

Rule 3.1:
If X_t is second-order stationary, then $R_X\left(t_2, t_1\right)$ only depends on $t_2 - t_1$.
Proof:

$$
\begin{aligned}
R_X\left(t_2, t_1\right) &= E\left[X_{t_2} X_{t_1}{}^*\right] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 x_1{}^* f_{X_{t_2}, X_{t_1}}\left(x_2, x_1\right) dx\,2dx\,1
\end{aligned}
\tag{3.22}
$$

$$
\begin{aligned}
R_X\left(t_2, t_1\right) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 x_1{}^* f_{X_{t_2-t_1}, X_0}\left(x_2, x_1\right) dx\,2dx\,1 \\
&= R_X\left(t_2 - t_1, 0\right)
\end{aligned}
\tag{3.23}
$$

If $R_X\left(t_2, t_1\right)$ depends on $t_2 - t_1$ only, then we will represent the autocorrelation with only one variable $\tau = t_2 - t_1$

$$
\begin{aligned}
R_X\left(\tau\right) &= R_X\left(t_2 - t_1\right) \\
&= R_X\left(t_2, t_1\right)
\end{aligned}
\tag{3.24}
$$

Properties

1. $R_X\left(0\right) \geq 0$
2. $R_X\left(\tau\right) = R_X\left(-\tau\right)^*$
3. $\left|R_X\left(\tau\right)\right| \leq R_X\left(0\right)$

---

[2]This content is available online at $<$http://cnx.org/content/m10236/2.13/$>$.

Example 3.4
$X_t = \cos\left(2\pi f_0 t + \Theta\left(\omega\right)\right)$ and Θ is uniformly distributed between 0 and 2π. The mean function

$$
\begin{aligned}
\mu_X\left(t\right) &= E\left[X_t\right] \\
&= E\left[\cos\left(2\pi f_0 t + \Theta\right)\right] \\
&= \int_0^{2\pi} \cos\left(2\pi f_0 t + \theta\right) \frac{1}{2\pi} d\theta \\
&= 0
\end{aligned}
\tag{3.25}
$$

The autocorrelation function

$$
\begin{aligned}
R_X\left(t+\tau,t\right) &= E\left[X_{t+\tau}X_t{}^*\right] \\
&= E\left[\cos\left(2\pi f_0\left(t+\tau\right)+\Theta\right)\cos\left(2\pi f_0 t + \Theta\right)\right] \\
&= 1/2E\left[\cos\left(2\pi f_0\tau\right)\right] + 1/2E\left[\cos\left(2\pi f_0\left(2t+\tau\right)+2\Theta\right)\right] \\
&= 1/2\cos\left(2\pi f_0\tau\right) + 1/2\int_0^{2\pi}\cos\left(2\pi f_0\left(2t+\tau\right)+2\theta\right)\frac{1}{2\pi}d\theta \\
&= 1/2\cos\left(2\pi f_0\tau\right)
\end{aligned}
\tag{3.26}
$$

Not a function of t since the second term in the right hand side of the equality in (3.26) is zero.

Example 3.5
Toss a fair coin every T seconds. Since X_t is a discrete valued random process, the statistical characteristics can be captured by the pmf and the mean function is written as

$$
\begin{aligned}
\mu_X\left(t\right) &= E\left[X_t\right] \\
&= 1/2 \times -1 + 1/2 \times 1 \\
&= 0
\end{aligned}
\tag{3.27}
$$

$$
\begin{aligned}
R_X\left(t_2,t_1\right) &= \sum_{kk}\sum_{ll} x_k x_l p_{X_{t_2},X_{t_1}}\left(x_k,x_l\right) \\
&= 1 \times 1 \times 1/2 - 1 \times -1 \times 1/2 \\
&= 1
\end{aligned}
\tag{3.28}
$$

when $nT \le t_1 < \left(n+1\right)T$ and $nT \le t_2 < \left(n+1\right)T$

$$
\begin{aligned}
R_X\left(t_2,t_1\right) &= 1 \times 1 \times 1/4 - 1 \times -1 \times 1/4 - 1 \times 1 \times 1/4 + 1 \times -1 \times 1/4 \\
&= 0
\end{aligned}
\tag{3.29}
$$

when $nT \le t_1 < \left(n+1\right)T$ and $mT \le t_2 < \left(m+1\right)T$ with $n \ne m$

$$
R_X\left(t_2,t_1\right) = \begin{cases} 1 & \text{if} \quad \left(nT \le t_1 < \left(n+1\right)T\right) \;\cdot\; \left(nT \le t_2 < \left(n+1\right)T\right) \\ 0 & \text{otherwise} \end{cases}
\tag{3.30}
$$

A function of t_1 and t_2.

Definition 3.5: Wide Sense Stationary
A process is said to be wide sense stationary if μ_X is constant and $R_X\left(t_2,t_1\right)$ is only a function of $t_2 - t_1$.

Rule 3.2:
If X_t is strictly stationary, then it is wide sense stationary. The converse is not necessarily true.

Definition 3.6: Autocovariance
Autocovariance of a random process is defined as

$$
\begin{aligned}
C_X\left(t_2, t_1\right) &= E\left[\left(X_{t_2} - \mu_X\left(t_2\right)\right)\left(X_{t_1} - \mu_X\left(t_1\right)\right)^*\right] \\
&= R_X\left(t_2, t_1\right) - \mu_X\left(t_2\right)\mu_X\left(t_1\right)^*
\end{aligned}
\tag{3.31}
$$

The variance of X_t is $\mathrm{Var}\left(X_t\right) = C_X\left(t, t\right)$
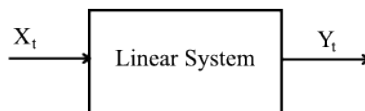Two processes defined on one experiment (Figure 3.5).

Figure 3.5

Definition 3.7: Crosscorrelation
The crosscorrelation function of a pair of random processes is defined as

$$
\begin{aligned}
R_{XY}\left(t_2, t_1\right) &= E\left[X_{t_2}Y_{t_1}{}^*\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X_{t_2}, Y_{t_1}}\left(x, y\right) dx dy
\end{aligned}
\tag{3.32}
$$

$$
C_{XY}\left(t_2, t_1\right) = R_{XY}\left(t_2, t_1\right) - \mu_X\left(t_2\right)\mu_Y\left(t_1\right)^*
\tag{3.33}
$$

Definition 3.8: Jointly Wide Sense Stationary
The random processes X_t and Y_t are said to be jointly wide sense stationary if $R_{XY}\left(t_2, t_1\right)$ is a function of $t_2 - t_1$ only and $\mu_X\left(t\right)$ and $\mu_Y\left(t\right)$ are constant.

3.3 Linear Filtering[3]

Integration

$$
Z\left(\omega\right) = \int_a^b X_t\left(\omega\right) dt
\tag{3.34}
$$

Linear Processing

$$
Y_t = \int_{-\infty}^{\infty} h\left(t, \tau\right) X_\tau d\tau
\tag{3.35}
$$

Differentiation

$$
X_t{}' = \frac{d}{dt}\left(X_t\right)
\tag{3.36}
$$

---

[3]This content is available online at $<$http://cnx.org/content/m10237/2.10/$>$.

Properties

1. $\bar{Z} = \int_a^b \bar{X_t}(\omega)\,dt = \int_a^b \mu_X(t)\,dt$

2. $\bar{Z^2} = \int_a^b X_{t_2} dt\,2 \int_a^b \bar{X_{t_1}}^* dt\,1 = \int_a^b \int_a^b R_X(t_2, t_1)\,dt\,1dt\,2$

$$X_t \longrightarrow \boxed{\begin{array}{c}\text{Linear System}\\ h(t,\tau)\end{array}} \longrightarrow Y_t$$
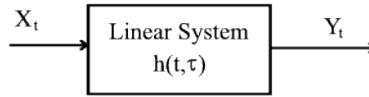
Figure 3.6

$$
\begin{aligned}
\mu_Y(t) &= \int_{-\infty}^{\infty} h\overline{(t,\tau)}X_\tau d\tau \\
&= \int_{-\infty}^{\infty} h(t,\tau)\mu_X(\tau)\,d\tau
\end{aligned}
\tag{3.37}
$$

If X_t is wide sense stationary and the linear system is time invariant

$$
\begin{aligned}
\mu_Y(t) &= \int_{-\infty}^{\infty} h(t-\tau)\mu_X d\tau \\
&= \mu_X \int_{-\infty}^{\infty} h(t')\,dt\prime \\
&= \mu_Y
\end{aligned}
\tag{3.38}
$$

$$
\begin{aligned}
R_{YX}(t_2, t_1) &= Y_{t_2}\overline{X_{t_1}}^* \\
&= \int_{-\infty}^{\infty} h(t_2 - \tau)\overline{X_\tau}d\tau X_{t_1}^* \\
&= \int_{-\infty}^{\infty} h(t_2 - \tau)R_X(\tau - t_1)\,d\tau
\end{aligned}
\tag{3.39}
$$

$$
\begin{aligned}
R_{YX}(t_2, t_1) &= \int_{-\infty}^{\infty} h(t_2 - t_1 - \tau')R_X(\tau')\,d\tau\prime \\
&= h * R_X(t_2 - t_1)
\end{aligned}
\tag{3.40}
$$

where $\tau' = \tau - t_1$.

$$
\begin{aligned}
R_Y(t_2, t_1) &= Y_{t_2}\overline{Y_{t_1}}^* \\
&= Y_{t_2} \int_{-\infty}^{\infty} h\overline{(t_1,\tau)}X_\tau^* d\tau \\
&= \int_{-\infty}^{\infty} h(t_1,\tau)R_{YX}(t_2,\tau)\,d\tau \\
&= \int_{-\infty}^{\infty} h(t_1 - \tau)R_{YX}(t_2 - \tau)\,d\tau
\end{aligned}
\tag{3.41}
$$

$$
\begin{aligned}
R_Y(t_2, t_1) &= \int_{-\infty}^{\infty} h\left(\tau' - (t_2 - t_1)\right) R_{YX}(\tau')\, d\tau\prime \\
&= R_Y(t_2 - t_1) \\
&= \widetilde{h} * R_{YX}(t_2, t_1)
\end{aligned}
\tag{3.42}
$$

where $\tau' = t_2 - \tau$ and $\widetilde{h}(\tau) = h(-\tau)$ for all $\tau \in \mathbb{R}$. Y_t is WSS if X_t is WSS and the linear system is time-invariant.
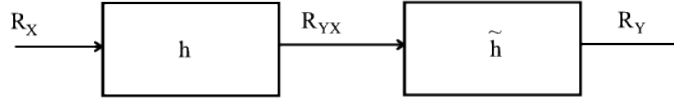

Figure 3.7

Example 3.6

X_t is a wide sense stationary process with $\mu_X = 0$, and $R_X(\tau) = \frac{N_0}{2}\delta(\tau)$. Consider the random process going through a filter with impulse response $h(t) = e^{-(at)}u(t)$. The output process is denoted by Y_t. $\mu_Y(t) = 0$ for all t.

$$
\begin{aligned}
R_Y(\tau) &= \frac{N_0}{2} \int_{-\infty}^{\infty} h(\alpha) h(\alpha - \tau)\, d\alpha \\
&= \frac{N_0}{2} \frac{e^{-(a|\tau|)}}{2a}
\end{aligned}
\tag{3.43}
$$

X_t is called a white process. Y_t is a Markov process.

Definition 3.9: Power Spectral Density

The power spectral density function of a wide sense stationary (WSS) process X_t is defined to be the Fourier transform of the autocorrelation function of X_t.

$$
S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\, e^{-(j2\pi f\tau)}\, d\tau
\tag{3.44}
$$

if X_t is WSS with autocorrelation function $R_X(\tau)$.

Properties

1. $S_X(f) = S_X(-f)$ since R_X is even and real.
2. $\mathrm{Var}(X_t) = R_X(0) = \int_{-\infty}^{\infty} S_X(f)\, df$
3. $S_X(f)$ is real and nonnegative $S_X(f) \geq 0$ for all f.

If $Y_t = \int_{-\infty}^{\infty} h(t - \tau) X_\tau d\tau$ then

$$
\begin{aligned}
S_Y(f) &= \mathcal{F}(R_Y(\tau)) \\
&= \mathcal{F}\left(h * \widetilde{h} * R_X(\tau)\right) \\
&= H(f)\, \widetilde{H}(f)\, S_X(f) \\
&= \left(|H(f)|\right)^2 S_X(f)
\end{aligned}
\tag{3.45}
$$

since $\widetilde{H}\left(f\right) = \int_{-\infty}^{\infty} \widetilde{h}\left(t\right) e^{-\left(j2\pi ft\right)} dt = H\left(f\right)^{*}$

Example 3.7
X_t is a white process and $h\left(t\right) = e^{-\left(at\right)} u\left(t\right)$.

$$H\left(f\right) = \frac{1}{a + j2\pi f} \tag{3.46}$$

$$S_Y\left(f\right) = \frac{\frac{N_0}{2}}{a^2 + 4\pi^2 f^2} \tag{3.47}$$

3.4 Gaussian Processes[4]

3.4.1 Gaussian Random Processes

Definition 3.10: Gaussian process
A process with mean $\mu_X\left(t\right)$ and covariance function $C_X\left(t_2, t_1\right)$ is said to be a Gaussian process if **any** $X = \left(X_{t_1}, X_{t_2}, \ldots, X_{t_N}\right)^T$ formed by **any** sampling of the process is a Gaussian random vector, that is,

$$f_X\left(x\right) = \frac{1}{\left(2\pi\right)^{\frac{N}{2}}\left(\det\Sigma_X\right)^{\frac{1}{2}}} e^{-\left(\frac{1}{2}\left(x - \mu_X\right)^T \Sigma_X^{-1}\left(x - \mu_X\right)\right)} \tag{3.48}$$

for all $x \in \mathbb{R}^n$ where

$$\mu_X = \begin{pmatrix} \mu_X\left(t_1\right) \\ \vdots \\ \mu_X\left(t_N\right) \end{pmatrix}$$

and

$$\Sigma_X = \begin{pmatrix} C_X\left(t_1, t_1\right) & \ldots & C_X\left(t_1, t_N\right) \\ \vdots & \ddots & \\ C_X\left(t_N, t_1\right) & \ldots & C_X\left(t_N, t_N\right) \end{pmatrix}$$

. The complete statistical properties of X_t can be obtained from the second-order statistics.

Properties

1. If a Gaussian process is WSS, then it is strictly stationary.
2. If two Gaussian processes are uncorrelated, then they are also statistically independent.
3. Any linear processing of a Gaussian process results in a Gaussian process.

Example 3.8
X and Y are Gaussian and zero mean and independent. $Z = X + Y$ is also Gaussian.

$$\begin{aligned} \phi_X\left(u\right) &= e^{\overline{juX}} \\ &= e^{-\left(\frac{u^2}{2}\sigma_X^2\right)} \end{aligned} \tag{3.49}$$

---

[4]This content is available online at $<$http://cnx.org/content/m10238/2.7/$>$.

for all $u \in \mathbb{R}$

$$
\begin{aligned}
\phi_Z \left(u \right) & = & e^{ju\overline{(X+Y)}} \\
& = & e^{-\left(\frac{u^2}{2} \sigma_X^2 \right)} e^{-\left(\frac{u^2}{2} \sigma_Y^2 \right)} \\
& = & e^{-\left(\frac{u^2}{2} \left(\sigma_X^2 + \sigma_Y^2 \right) \right)}
\end{aligned}
\tag{3.50}
$$

therefore Z is also Gaussian.

Chapter 4

Chapter 4

4.1 Data Transmission and Reception[1]

We will develop the idea of **data transmission** by first considering simple channels. In additional modules, we will consider more practical channels; **baseband** channels with **bandwidth** constraints and **passband** channels.
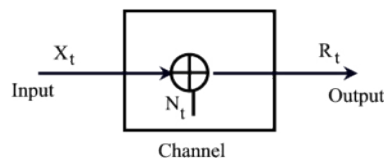

Simple additive white Gaussian channels

Figure 4.1: X_t carries data, N_t is a white Gaussian random process.

The concept of using different types of modulation for transmission of data is introduced in the module Signalling (Section 4.2). The problem of demodulation and detection of signals is discussed in Demodulation and Detection (Section 4.4).

4.2 Signalling[2]

Example 4.1
Data symbols are "1" or "0" and data rate is $\frac{1}{T}$ Hertz.

---

[1]This content is available online at <http://cnx.org/content/m10115/2.9/>.
[2]This content is available online at <http://cnx.org/content/m10116/2.11/>.
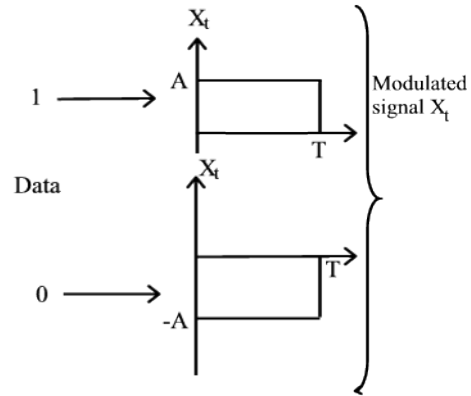
Pulse amplitude modulation (PAM)
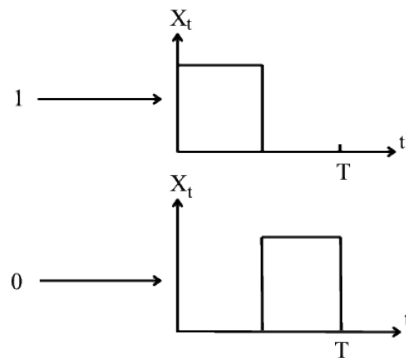

Figure 4.2

Pulse position modulation

Figure 4.3

Example 4.2: Example

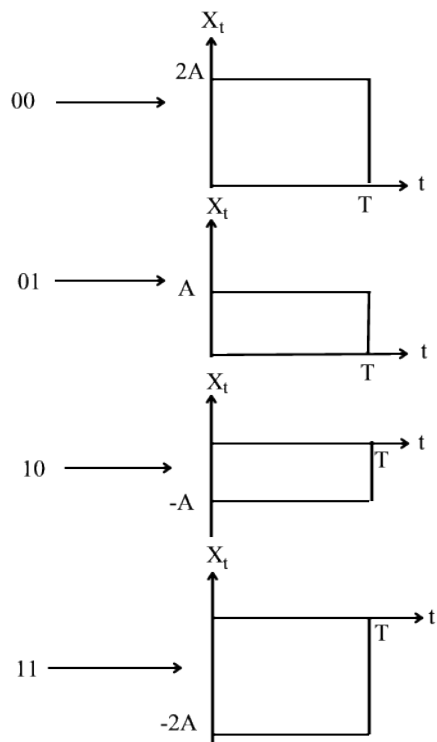Data symbols are "1" or "0" and the data rate is $\frac{2}{T}$ Hertz.

Figure 4.4

This strategy is an alternative to PAM with half the period, $\frac{T}{2}$.
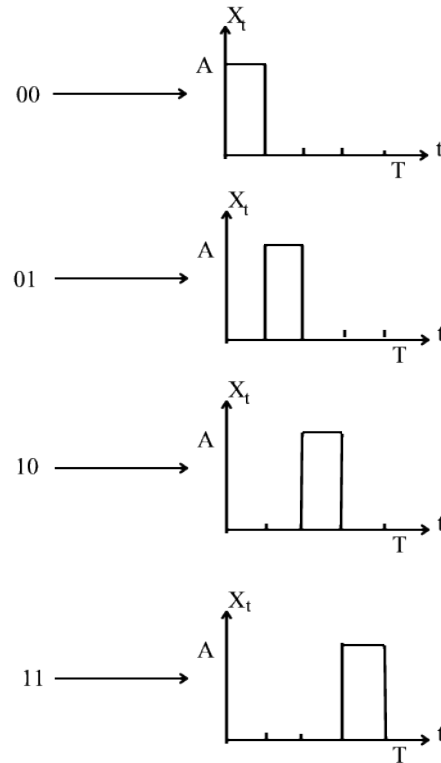
Figure 4.5

Relevant measures are energy of modulated signals

$$E_m = \int_0^T s_m{}^2(t)\, dt \quad , \quad m \in \{1, 2, \ldots, M\} \tag{4.1}$$

and how different they are in terms of inner products.

$$< s_m, s_n > = \int_0^T s_m(t)\, s_n(t)^* dt \tag{4.2}$$

for $m \in \{1, 2, \ldots, M\}$ and $n \in \{1, 2, \ldots, M\}$.

Definition 4.1: antipodal
Signals $s_1(t)$ and $s_2(t)$ are antipodal if $s_2(t) = -s_1(t) \quad , \quad t \in [0, T]$

Definition 4.2: orthogonal
Signals $s_1(t)$, $s_2(t)$,..., $s_M(t)$ are orthogonal if $< s_m, s_n > = 0$ for $m \neq n$.

Definition 4.3: biorthogonal
Signals $s_1(t)$, $s_2(t)$,..., $s_M(t)$ are biorthogonal if $s_1(t)$,..., $s_{\frac{M}{2}}(t)$ are orthogonal and $s_m(t) = -s_{\frac{M}{2}+m}(t)$ for some $m \in \left\{1, 2, \ldots, \frac{M}{2}\right\}$.

It is quite intuitive to expect that the smaller (the more negative) the inner products, $< s_m, s_n >$ for all $m \neq n$, the better the signal set.

Definition 4.4: Simplex signals

Let $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ be a set of orthogonal signals with equal energy. The signals $\tilde{s_1}(t), \ldots, \tilde{s_M}(t)$ are simplex signals if

$$\tilde{s_m}(t) = s_m(t) - \frac{1}{M}\sum_{k=1}^{M} s_k(t) \tag{4.3}$$

If the energy of orthogonal signals is denoted by

$$E_s = \int_0^T {s_m}^2(t)\, dt \quad, \quad m \in \{1, 2, ..., M\} \tag{4.4}$$

then the energy of simplex signals

$$E_{\tilde{s}} = \left(1 - \frac{1}{M}\right) E_s \tag{4.5}$$

and

$$< \tilde{s_m}, \tilde{s_n} > = \frac{-1}{M-1} E_{\tilde{s}} \quad, \quad m \neq n \tag{4.6}$$

It is conjectured that among all possible M-ary signals with equal energy, the simplex signal set results in the smallest probability of error when used to transmit information through an additive white Gaussian noise channel.

The geometric representation of signals (Section 4.3) can provide a compact description of signals and can simplify performance analysis of communication systems using the signals.

Once signals have been modulated, the receiver must detect and demodulate (Section 4.4) the signals despite interference and noise and decide which of the set of possible transmitted signals was sent.

4.3 Geometric Representation of Modulation Signals[3]

Geometric representation of signals can provide a compact characterization of signals and can simplify analysis of their performance as modulation signals.

Orthonormal bases are essential in geometry. Let $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ be a set of signals.

Define $\psi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}$ where $E_1 = \int_0^T {s_1}^2(t)\, dt$.

Define $s_{21} = < s_2, \psi_1 > = \int_0^T s_2(t)\psi_1(t)^* dt$ and $\psi_2(t) = \frac{1}{\sqrt{\hat{E_2}}}(s_2(t) - s_{21}\psi_1)$ where $\hat{E_2} = \int_0^T (s_2(t) - s_{21}\psi_1(t))^2 dt$

In general

$$\psi_k(t) = \frac{1}{\sqrt{\hat{E_k}}}\left(s_k(t) - \sum_{j=1}^{k-1} s_{kj}\psi_j(t)\right) \tag{4.7}$$

where $\hat{E_k} = \int_0^T \left(s_k(t) - \sum_{j=1}^{k-1} s_{kj}\psi_j(t)\right)^2 dt$.

The process continues until all of the M signals are exhausted. The results are N orthogonal signals with unit energy, $\{\psi_1(t), \psi_2(t), \ldots, \psi_N(t)\}$ where $N \leq M$. If the signals $\{s_1(t), \ldots, s_M(t)\}$ are linearly independent, then $N = M$.

---

[3]This content is available online at $<$http://cnx.org/content/m10035/2.13/$>$.

The M signals can be represented as

$$s_m(t) = \sum_{n=1}^{N} s_{\text{mn}} \psi_n(t) \tag{4.8}$$

with $m \in \{1, 2, \ldots, M\}$ where $s_{\text{mn}} = <s_m, \psi_n>$ and $E_m = \sum_{n=1}^{N} s_{\text{mn}}{}^2$. The signals can be represented by

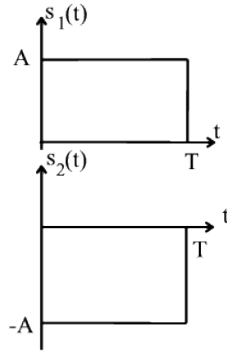$$s_m = \begin{pmatrix} s_{m1} \\ s_{m2} \\ \vdots \\ s_{\text{mN}} \end{pmatrix}$$

Example 4.3

Figure 4.6

$$\psi_1(t) = \frac{s_1(t)}{\sqrt{A^2 T}} \tag{4.9}$$

$$s_{11} = A\sqrt{T} \tag{4.10}$$

$$s_{21} = -\left(A\sqrt{T}\right) \tag{4.11}$$

$$
\begin{aligned}
\psi_2(t) &= (s_2(t) - s_{21}\psi_1(t)) \frac{1}{\sqrt{\hat{E_2}}} \\
&= \left(-A + \frac{A\sqrt{T}}{\sqrt{T}}\right) \frac{1}{\sqrt{\hat{E_2}}} \\
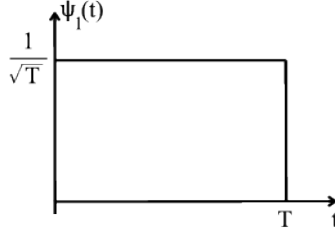&= 0
\end{aligned}
\tag{4.12}
$$

Figure 4.7

Dimension of the signal set is 1 with $E_1 = s_{11}{}^2$ and $E_2 = s_{21}{}^2$.
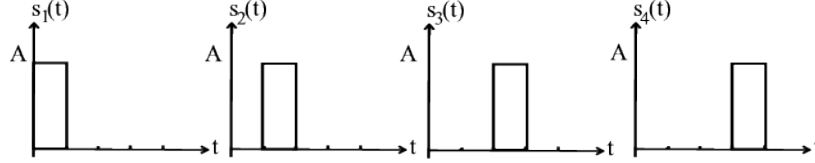
Example 4.4

Figure 4.8

$\psi_m(t) = \frac{s_m(t)}{\sqrt{E_s}}$ where $E_s = \int_0^T s_m{}^2(t)\, dt = \frac{A^2 T}{4}$

$$s_1 = \begin{pmatrix} \sqrt{E_s} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \; s_2 = \begin{pmatrix} 0 \\ \sqrt{E_s} \\ 0 \\ 0 \end{pmatrix}, \; s_3 = \begin{pmatrix} 0 \\ 0 \\ \sqrt{E_s} \\ 0 \end{pmatrix}, \text{ and } s_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sqrt{E_s} \end{pmatrix}$$

$$d_{\mathrm{mn}} = |s_m - s_n| = \sqrt{\sum_{j=1}^{N} (s_{\mathrm{mj}} - s_{\mathrm{nj}})^2} = \sqrt{2E_s} \tag{4.13}$$

is the Euclidean distance between signals.

Example 4.5
Set of 4 equal energy biorthogonal signals. $s_1(t) = s(t)$, $s_2(t) = s^\perp(t)$, $s_3(t) = -s(t)$, $s_4(t) = -s^\perp(t)$.

The orthonormal basis $\psi_1(t) = \frac{s(t)}{\sqrt{E_s}}$, $\psi_2(t) = \frac{s^\perp(t)}{\sqrt{E_s}}$ where $E_s = \int_0^T s_m{}^2(t)\, dt$

$$s_1 = \begin{pmatrix} \sqrt{E_s} \\ 0 \end{pmatrix}, \; s_2 = \begin{pmatrix} 0 \\ \sqrt{E_s} \end{pmatrix}, \; s_3 = \begin{pmatrix} -\sqrt{E_s} \\ 0 \end{pmatrix}, \; s_4 = \begin{pmatrix} 0 \\ -\sqrt{E_s} \end{pmatrix}.$$ The four signals can
be geometrically represented using the 4-vector of projection coefficients s_1, s_2, s_3, and s_4 as a set
of constellation points.
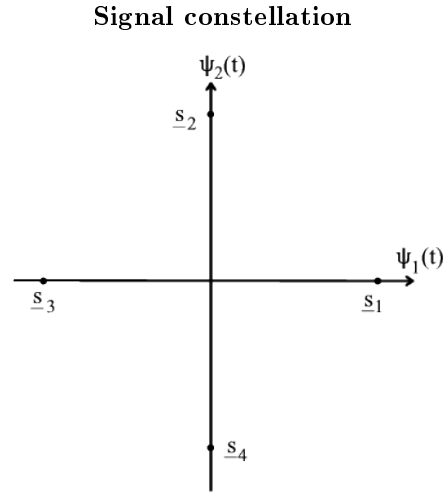
Signal constellation

Figure 4.9

$$
\begin{aligned}
d_{21} &= |s_2 - s_1| \\
&= \sqrt{2E_s}
\end{aligned}
\tag{4.14}
$$

$$
\begin{aligned}
d_{12} &= d_{23} \\
&= d_{34} \\
&= d_{14}
\end{aligned}
\tag{4.15}
$$

$$
\begin{aligned}
d_{13} &= |s_1 - s_3| \\
&= 2\sqrt{E_s}
\end{aligned}
\tag{4.16}
$$

$$
d_{13} = d_{24}
\tag{4.17}
$$

Minimum distance $d_{\min} = \sqrt{2E_s}$

4.4 Demodulation and Detection[4]

Consider the problem where signal set, $\{s_1, s_2, \ldots, s_M\}$, for $t \in [0, T]$ is used to transmit $\log_2 M$ bits. The **modulated** signal X_t could be $\{s_1, s_2, \ldots, s_M\}$ during the interval $0 \le t \le T$.

---

[4]This content is available online at $<$http://cnx.org/content/m10054/2.14/$>$.
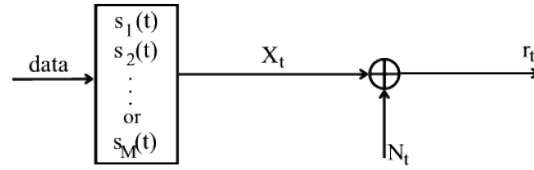
Figure 4.10: $r_t = X_t + N_t = s_m(t) + N_t$ for $0 \leq t \leq T$ for some $m \in \{1, 2, \ldots, M\}$.

Recall $s_m(t) = \sum_{n=1}^{N} s_{mn} \psi_n(t)$ for $m \in \{1, 2, \ldots, M\}$ the signals are decomposed into a set of orthonormal signals, perfectly.

Noise process can also be decomposed

$$N_t = \sum_{n=1}^{N} \eta_n \psi_n(t) + \tilde{N}_t \tag{4.18}$$

where $\eta_n = \int_0^T N_t \psi_n(t)\, dt$ is the projection onto the nth basis signal, $\tilde{N}_t$ is the left over noise.

The problem of demodulation and detection is to observe r_t for $0 \leq t \leq T$ and decide which one of the M signals were transmitted. Demodulation is covered here (Section 4.5). A discussion about detection can be found here (Section 4.6).

4.5 Demodulation[5]

4.5.1 Demodulation

Convert the continuous time received signal into a vector without loss of information (or performance).

$$r_t = s_m(t) + N_t \tag{4.19}$$

$$r_t = \sum_{n=1}^{N} s_{mn} \psi_n(t) + \sum_{n=1}^{N} \eta_n \psi_n(t) + \tilde{N}_t \tag{4.20}$$

$$r_t = \sum_{n=1}^{N} (s_{mn} + \eta_n) \psi_n(t) + \tilde{N}_t \tag{4.21}$$

$$r_t = \sum_{n=1}^{N} r_n \psi_n(t) + \tilde{N}_t \tag{4.22}$$

Proposition 4.1:
The noise projection coefficients η_n's are zero mean, Gaussian random variables and are mutually independent if N_t is a white Gaussian process.
Proof:

$$\begin{aligned} \mu_\eta(n) &= E[\eta_n] \\ &= E\left[\int_0^T N_t \psi_n(t)\, dt\right] \end{aligned} \tag{4.23}$$

$$
\begin{aligned}
\mu_\eta(n) &= \int_0^T E[N_t]\,\psi_n(t)\,dt \\
&= 0
\end{aligned}
\tag{4.24}
$$

$$
\begin{aligned}
E[\eta_k\eta_n{}^*] &= E\left[\int_0^T N_t\psi_k(t)\,dt\int_0^T N_{t'}{}^*\psi_k(t')^*\,dt\prime\right] \\
&= \int_0^T\int_0^T (N_tN_{t'})^*\psi_k(t)\,\psi_n(t')\,dtdt\prime
\end{aligned}
\tag{4.25}
$$

$$
E[\eta_k\eta_n{}^*] = \int_0^T\int_0^T R_N(t-t')\,\psi_k(t)\,\psi_n{}^*\,dtdt\prime
\tag{4.26}
$$

$$
E[\eta_k\eta_n{}^*] = \frac{N_0}{2}\int_0^T\int_0^T \delta(t-t')\,\psi_k(t)\,\psi_n(t')^*\,dtdt\prime
\tag{4.27}
$$

$$
\begin{aligned}
E[\eta_k\eta_n{}^*] &= \frac{N_0}{2}\int_0^T \psi_k(t)\,\psi_n(t)^*\,dt \\
&= \frac{N_0}{2}\delta_{kn} \\
&= \begin{cases} \frac{N_0}{2} & \text{if } k=n \\ 0 & \text{if } k\neq n \end{cases}
\end{aligned}
\tag{4.28}
$$

η_k 's are uncorrelated and since they are Gaussian they are also independent. Therefore, $\eta_k \simeq$ Gaussian $\left(0,\frac{N_0}{2}\right)$ and $R_\eta(k,n) = \frac{N_0}{2}\delta_{kn}$

Proposition 4.2:
The r_n's, the projection of the received signal r_t onto the orthonormal bases $\psi_n(t)$'s, are independent from the residual noise process $\tilde{N}_t$.
    The residual noise $\tilde{N}_t$ is irrelevant to the decision process on r_t.
Proof:
Recall $r_n = s_{mn} + \eta_n$, given $s_m(t)$ was transmitted. Therefore,

$$
\begin{aligned}
\mu_r(n) &= E[s_{mn}+\eta_n] \\
&= s_{mn}
\end{aligned}
\tag{4.29}
$$

$$
\begin{aligned}
\text{Var}(r_n) &= \text{Var}(\eta_n) \\
&= \frac{N_0}{2}
\end{aligned}
\tag{4.30}
$$

The correlation between r_n and $\tilde{N}_t$

$$
E\left[\tilde{N}_t r_n{}^*\right] = E\left[\left(N_t - \sum_{k=1}^N \eta_k\psi_k(t)\right)(s_{mn}+\eta_n)^*\right]
\tag{4.31}
$$

$$
E\left[\tilde{N}_t r_n{}^*\right] = E\left[N_t - \sum_{k=1}^N \eta_k\psi_k(t)\right]s_{mn} + E[\eta_k\eta_n{}^*] - \sum_{k=1}^N E[\eta_k\eta_n{}^*]\psi_k(t)
\tag{4.32}
$$

$$
E\left[\tilde{N}_t r_n{}^*\right] = E\left[N_t\int_0^T N_{t'}{}^*\psi_n(t')^*\,dt\prime\right] - \sum_{k=1}^N \frac{N_0}{2}\delta_{kn}\psi_k(t)
\tag{4.33}
$$

$$
E\left[\tilde{N}_t r_n{}^*\right] = \int_0^T \frac{N_0}{2}\delta(t-t')\,\psi_n(t')\,dt\prime - \frac{N_0}{2}\psi_n(t)
\tag{4.34}
$$

$$E\left[\tilde{N}_t r_n{}^*\right] = \frac{N_0}{2}\psi_n(t) - \frac{N_0}{2}\psi_n(t)$$
$$= 0 \tag{4.35}$$

Since both $\tilde{N}_t$ and r_n are Gaussian then $\tilde{N}_t$ and r_n are also independent.

The conjecture is to ignore $\tilde{N}_t$ and extract information from $\begin{pmatrix} r_1 \\ r_2 \\ \dots \\ r_N \end{pmatrix}$. Knowing the vector r

we can reconstruct the relevant part of random process r_t for $0 \leq t \leq T$

$$r_t = s_m(t) + N_t$$
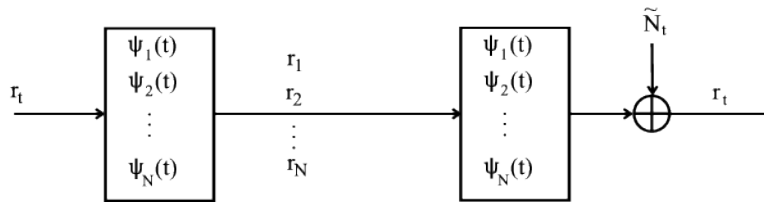$$= \sum_{n=1}^{N} r_n \psi_n(t) + \tilde{N}_t \tag{4.36}$$
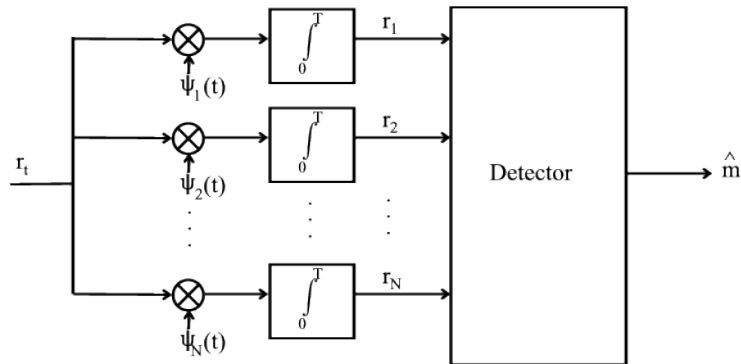

Figure 4.11

Figure 4.12

Once the received signal has been converted to a vector, the correct transmitted signal must be detected based upon observations of the input vector. Detection is covered elsewhere (Section 4.6).

4.6 Detection by Correlation[6]
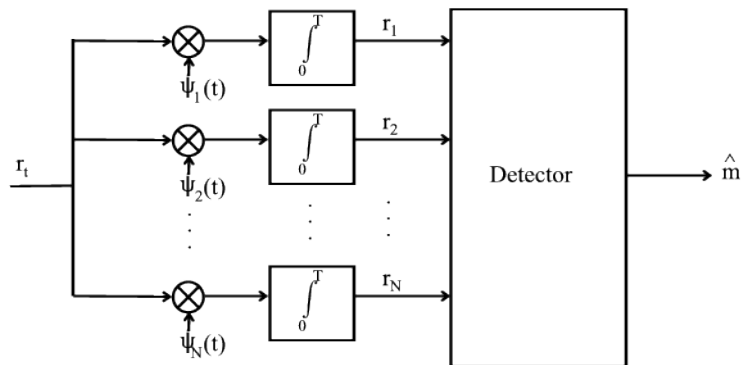
Demodulation and Detection

Figure 4.13

4.6.1 Detection

Decide which $s_m(t)$ from the set of $\{s_1(t),\ldots,s_m(t)\}$ signals was transmitted based on observing $r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{pmatrix}$, the vector composed of demodulated (Section 4.5) received signal, that is, the vector of projection of the received signal onto the N bases.

$$\hat{m} = arg \max_{1 \leq m \leq M} Pr\left[s_m(t) \text{ was transmitted } \mid r \text{ was observed}\right] \tag{4.37}$$

Note that

$$Pr\left[s_m \mid r\right] \triangleq Pr\left[\mathbf{s}_m(t) \text{ was transmitted } \mid r \text{ was observed}\right] = \frac{f_{r|\mathbf{s}_m} Pr\left[s_m\right]}{f_{\mathbf{r}}} \tag{4.38}$$

If $Pr\left[s_m \text{ was transmitted}\right] = \frac{1}{M}$, that is information symbols are equally likely to be transmitted, then

$$arg \max_{1 \leq m \leq M} Pr\left[s_m \mid r\right] = arg \max_{1 \leq m \leq M} f_{\mathbf{r}|\mathbf{s}_m} \tag{4.39}$$

Since $r(t) = s_m(t) + N_t$ for $0 \leq t \leq T$ and for some $m = \{1, 2, \ldots, M\}$ then $r = s_m + \eta$ where $\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_N \end{pmatrix}$

and η_n's are Gaussian and independent.

$$f_{\mathbf{r}|\mathbf{s}_m} = \frac{1}{\left(2\pi\frac{N_0}{2}\right)^{\frac{N}{2}}} e^{\frac{-\sum_{n=1}^{N}(r_n - s_{m,n})^2}{2\frac{N_0}{2}}} \quad , \quad r_n \in \mathbb{R} \tag{4.40}$$

$$\begin{aligned} \hat{m} &= arg \max_{1 \leq m \leq M} f_{\mathbf{r}|\mathbf{s}_m} \\ &= arg \max_{1 \leq m \leq M} \ln\left(f_{\mathbf{r}|\mathbf{s}_m}\right) \\ &= arg \max_{1 \leq m \leq M} \left(-\left(\frac{N}{2}\ln\left(\pi N_0\right)\right)\right) - \frac{1}{N_0}\sum_{n=1}^{N}(r_n - s_{m,n})^2 \\ &= arg \min_{1 \leq m \leq M} \sum_{n=1}^{N}(r_n - s_{m,n})^2 \end{aligned} \tag{4.41}$$

where $D(r, s_m)$ is the l_2 distance between vectors r and s_m defined as $D(r, s_m) \triangleq \sum_{n=1}^{N}(r_n - s_{m,n})^2$

$$\begin{aligned} \hat{m} &= arg \min_{1 \leq m \leq M} D(r, s_m) \\ &= arg \min_{1 \leq m \leq M} (\parallel r \parallel)^2 - 2 < (r, s_m) > + (\parallel s_m \parallel)^2 \end{aligned} \tag{4.42}$$

where $\parallel r \parallel$ is the l_2 norm of vector r defined as $\parallel r \parallel \triangleq \sqrt{\sum_{n=1}^{N}(r_n)^2}$

$$\hat{m} = arg \max_{1 \leq m \leq M} 2 < (r, s_m) > - (\parallel s_m \parallel)^2 \tag{4.43}$$

This type of receiver system is known as a **correlation** (or correlator-type) receiver. Examples of the use of such a system are found here (Section 4.7). Another type of receiver involves linear, time-invariant filters and is known as a matched filter (Section 4.8) receiver. An analysis of the performance of a correlator-type receiver using antipodal and orthogonal binary signals can be found in Performance Analysis (Section 4.11).

4.7 Examples of Correlation Detection[7]

The implementation and theory of correlator-type receivers can be found in Detection (Section 4.6).
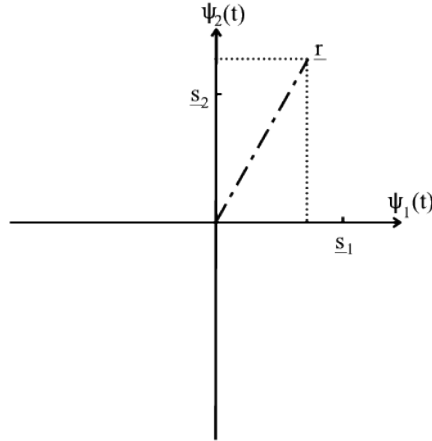
Example 4.6

Figure 4.14

$\hat{m} = 2$ since $D(r, s_1) > D(r, s_2)$ or $(\| s_1 \|)^2 = (\| s_2 \|)^2$ and $< r, s_2 >> < r, s_1 >$.
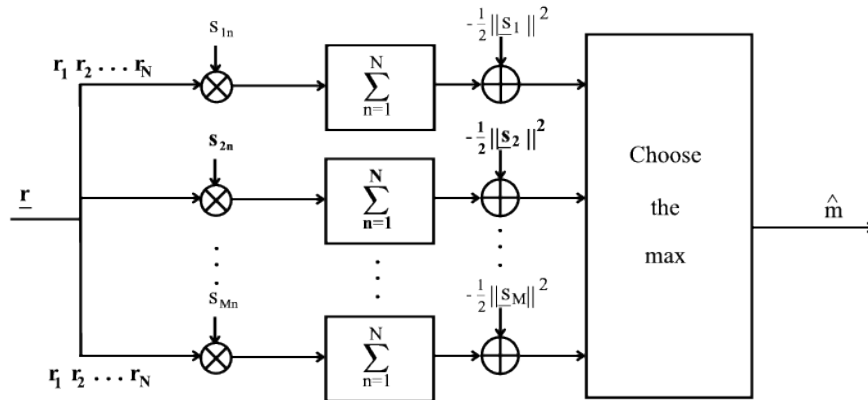

Figure 4.15

Example 4.7

Data symbols "0" or "1" with equal probability. Modulator $s_1(t) = s(t)$ for $0 \leq t \leq T$ and $s_2(t) = -s(t)$ for $0 \leq t \leq T$.

---

[7]This content is available online at <http://cnx.org/content/m10149/2.10/>.
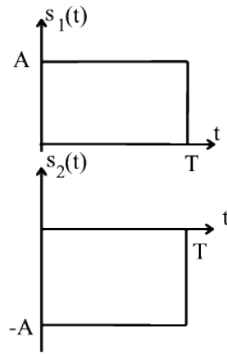
Figure 4.16

$$\psi_1\left(t\right) = \frac{s(t)}{\sqrt{A^2 T}}, \; s_{11} = A\sqrt{T}, \text{ and } s_{21} = -\left(A\sqrt{T}\right)$$

$$r_t = s_m\left(t\right) + N_t \quad , \quad m = \{1, 2\} \tag{4.44}$$
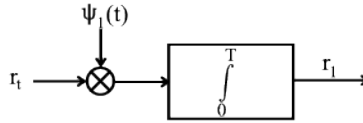
Figure 4.17

$$r_1 = A\sqrt{T} + \eta_1 \tag{4.45}$$

or

$$r_1 = -\left(A\sqrt{T}\right) + \eta_1 \tag{4.46}$$

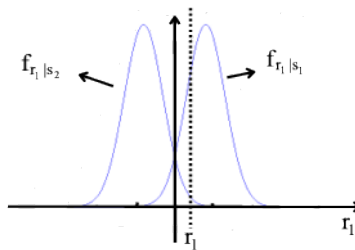η_1 is Gaussian with zero mean and variance $\frac{N_0}{2}$.

Figure 4.18

$\hat{m}= argmax \left\{ A\sqrt{T}r_1, -\left(A\sqrt{T}r_1 \right) \right\}$, since $A\sqrt{T} > 0$ and $Pr\left[s_1\right] = Pr\left[s_1\right]$ then the MAP decision rule decides.

$s_1\left(t\right)$ was transmitted if $r_1 \geq 0$
$s_2\left(t\right)$ was transmitted if $r_1 < 0$
An alternate demodulator:

$$\left(r_t = s_m\left(t\right) + N_t\right) \Rightarrow \left(r = s_m + \eta\right) \tag{4.47}$$

4.8 Matched Filters[8]

Signal to Noise Ratio (SNR) at the output of the demodulator is a measure of the quality of the demodulator.

$$\text{SNR} = \frac{\text{signal energy}}{\text{noise energy}} \tag{4.48}$$

In the correlator described earlier, $E_s = \left(|s_m|\right)^2$ and $\sigma_{\eta_n}{}^2 = \frac{N_0}{2}$. Is it possible to design a demodulator based on linear time-invariant filters with maximum signal-to-noise ratio?
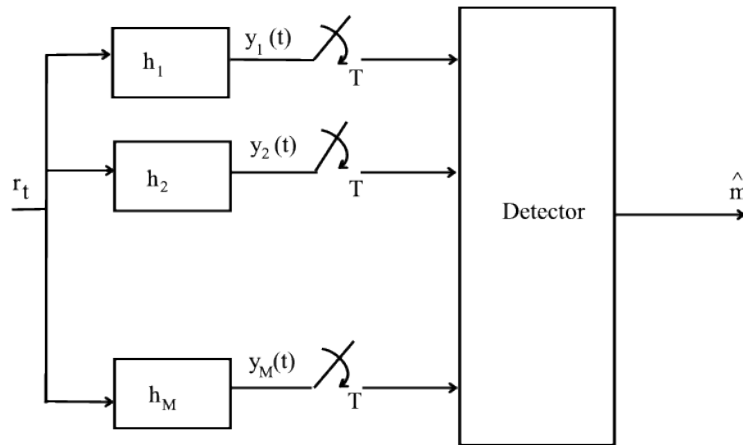
Figure 4.19

If $s_m\left(t\right)$ is the transmitted signal, then the output of the kth filter is given as

$$\begin{aligned}
y_k\left(t\right) &= \int_{-\infty}^{\infty} r_\tau h_k\left(t - \tau\right)d\tau \\
&= \int_{-\infty}^{\infty} \left(s_m\left(\tau\right) + N_\tau\right) h_k\left(t - \tau\right)d\tau \\
&= \int_{-\infty}^{\infty} s_m\left(\tau\right) h_k\left(t - \tau\right)d\tau + \int_{-\infty}^{\infty} N_\tau h_k\left(t - \tau\right)d\tau
\end{aligned} \tag{4.49}$$

[8]This content is available online at <http://cnx.org/content/m10101/2.14/>.

Sampling the output at time T yields

$$y_k\left(T\right) = \int_{-\infty}^{\infty} s_m\left(\tau\right) h_k\left(T-\tau\right) d\tau + \int_{-\infty}^{\infty} N_\tau h_k\left(T-\tau\right) d\tau \tag{4.50}$$

The noise contribution:

$$\nu_k = \int_{-\infty}^{\infty} N_\tau h_k\left(T-\tau\right) d\tau \tag{4.51}$$

The expected value of the noise component is

$$\begin{aligned} E\left[\nu_k\right] &= E\left[\int_{-\infty}^{\infty} N_\tau h_k\left(T-\tau\right) d\tau\right] \\ &= 0 \end{aligned} \tag{4.52}$$

The variance of the noise component is the second moment since the mean is zero and is given as

$$\begin{aligned} \sigma\left(\nu_k\right)^2 &= E\left[\nu_k{}^2\right] \\ &= E\left[\int_{-\infty}^{\infty} N_\tau h_k\left(T-\tau\right) d\tau \int_{-\infty}^{\infty} N_{\tau'}{}^* h_k\left(T-\tau'\right)^* d\tau'\right] \end{aligned} \tag{4.53}$$

$$\begin{aligned} E\left[\nu_k{}^2\right] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{N_0}{2} \delta\left(\tau-\tau'\right) h_k\left(T-\tau\right) h_k\left(T-\tau'\right)^* d\tau d\tau' \\ &= \frac{N_0}{2} \int_{-\infty}^{\infty} \left(|h_k\left(T-\tau\right)|\right)^2 d\tau \end{aligned} \tag{4.54}$$

Signal Energy can be written as

$$\left(\int_{-\infty}^{\infty} s_m\left(\tau\right) h_k\left(T-\tau\right) d\tau\right)^2 \tag{4.55}$$

and the signal-to-noise ratio (SNR) as

$$\text{SNR} = \frac{\left(\int_{-\infty}^{\infty} s_m\left(\tau\right) h_k\left(T-\tau\right) d\tau\right)^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} \left(|h_k\left(T-\tau\right)|\right)^2 d\tau} \tag{4.56}$$

The signal-to-noise ratio, can be maximized considering the well-known Cauchy-Schwarz Inequality

$$\left(\int_{-\infty}^{\infty} g_1\left(x\right) g_2\left(x\right)^* dx\right)^2 \leq \int_{-\infty}^{\infty} \left(|g_1\left(x\right)|\right)^2 dx \int_{-\infty}^{\infty} \left(|g_2\left(x\right)|\right)^2 dx \tag{4.57}$$

with equality when $g_1\left(x\right) = \alpha g_2\left(x\right)$. Applying the inequality directly yields an upper bound on SNR

$$\frac{\left(\int_{-\infty}^{\infty} s_m\left(\tau\right) h_k\left(T-\tau\right) d\tau\right)^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} \left(|h_k\left(T-\tau\right)|\right)^2 d\tau} \leq \frac{2}{N_0} \int_{-\infty}^{\infty} \left(|s_m\left(\tau\right)|\right)^2 d\tau \tag{4.58}$$

with equality $h_k^{\text{opt}}\left(T-\tau\right) = \alpha s_m\left(\tau\right)^*$. Therefore, the filter to examine signal m should be

Matched Filter

$$\tau h_m^{\text{opt}}\left(\tau\right) = s_m\left(T-\tau\right)^* \tag{4.59}$$

The constant factor is not relevant when one considers the signal to noise ratio. The maximum SNR is unchanged when both the numerator and denominator are scaled.

$$\frac{2}{N_0} \int_{-\infty}^{\infty} \left(|s_m\left(\tau\right)|\right)^2 d\tau = \frac{2E_s}{N_0} \tag{4.60}$$

Examples involving matched filter receivers can be found here (Section 4.9). An analysis in the frequency domain is contained in Matched Filters in the Frequency Domain (Section 4.10).

Another type of receiver system is the correlation (Section 4.6) receiver. A performance analysis of both matched filters and correlator-type receivers can be found in Performance Analysis (Section 4.11).

4.9 Examples with Matched Filters[9]

The theory and rationale behind matched filter receivers can be found in Matched Filters (Section 4.8).
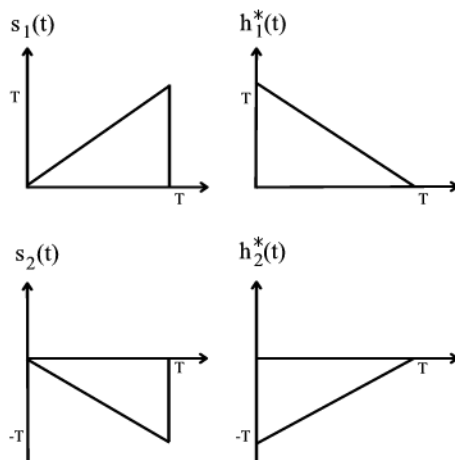
Example 4.8

Figure 4.20

$s_1(t) = t$ for $0 \leq t \leq T$
$s_2(t) = -t$ for $0 \leq t \leq T$
$h_1(t) = T - t$ for $0 \leq t \leq T$
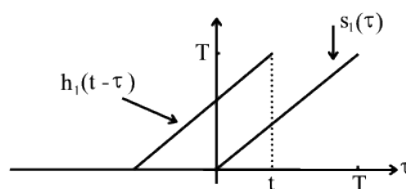$h_2(t) = -T + t$ for $0 \leq t \leq T$

Figure 4.21

$$\tilde{s_1}(t) = \int_{-\infty}^{\infty} s_1(\tau) h_1(t - \tau) d\tau \quad , \quad 0 \leq t \leq 2T \tag{4.61}$$

$$\begin{aligned} \tilde{s_1}(t) &= \int_0^t \tau (T - t + \tau) d\tau \\ &= \frac{1}{2}(T - t) \tau^2 \big|_0^t + \frac{1}{3} \tau^3 \big|_0^t \\ &= \frac{t^2}{2}\left(T - \frac{t}{3}\right) \end{aligned} \tag{4.62}$$

---

[9]This content is available online at <http://cnx.org/content/m10150/2.10/>.

$$\tilde{s}_1(T) = \frac{T^3}{3} \tag{4.63}$$

Compared to the correlator-type demodulation

$$\psi_1(t) = \frac{s_1(t)}{\sqrt{E_s}} \tag{4.64}$$

$$s_{11} = \int_0^T s_1(\tau)\,\psi_1(\tau)\,d\tau \tag{4.65}$$

$$
\begin{aligned}
\int_0^t s_1(\tau)\,\psi_1(\tau)\,d\tau &= \frac{1}{\sqrt{E_s}}\int_0^t \tau\tau\,d\tau \\
&= \frac{1}{\sqrt{E_s}}\frac{1}{3}t^3
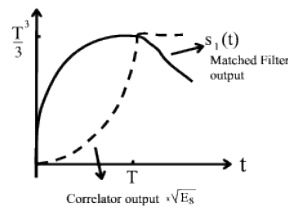\end{aligned}
\tag{4.66}
$$

Figure 4.22

Example 4.9

Assume binary data is transmitted at the rate of $\frac{1}{T}$ Hertz.

$0 \Rightarrow (b = 1) \Rightarrow (s_1(t) = s(t))$ for $0 \le t \le T$

$1 \Rightarrow (b = -1) \Rightarrow (s_2(t) = -s(t))$ for $\quad 0 \le t \le T$

$$X_t = \sum_{i=-P}^{P} b_i s(t - iT) \tag{4.67}$$

Figure 4.23

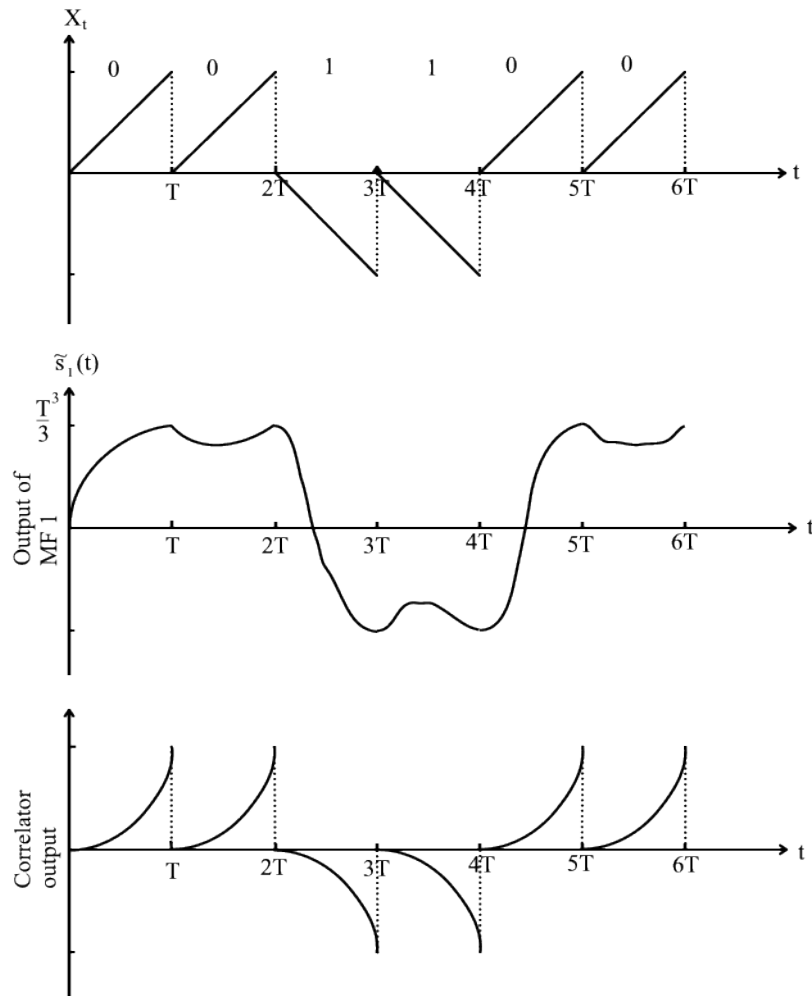## 4.10 Matched Filters in the Frequency Domain[10]

4.10.1

The time domain analysis and implementation of matched filters can be found in Matched Filters (Section 4.8).

A frequency domain interpretation of matched filters is very useful

$$\text{SNR} = \frac{\left(\int_{-\infty}^{\infty} s_m\left(\tau\right) h_m\left(T - \tau\right) d\tau\right)^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} \left(|h_m\left(T - \tau\right)|\right)^2 d\tau} \tag{4.68}$$

---

[10]This content is available online at <http://cnx.org/content/m10151/2.11/>.

For the m-th filter, h_m can be expressed as

$$
\begin{aligned}
\tilde{s_m}(T) &= \int_{-\infty}^{\infty} s_m(\tau) h_m(T - \tau) d\tau \\
&= \mathcal{F}^{-1}(H_m(f) S_m(f)) \\
&= \int_{-\infty}^{\infty} H_m(f) S_m(f) e^{j2\pi fT} df
\end{aligned}
\tag{4.69}
$$

where the second equality is because $\tilde{s}_m$ is the filter output with input S_m and filter H_m and we can now define $\hat{H}_m(f) = H_m(f)^* e^{-(j2\pi fT)}$, then

$$
\tilde{s_m}(T) = < S_m(f), \hat{H}_m(f) >
\tag{4.70}
$$

The denominator

$$
\int_{-\infty}^{\infty} (|h_m(T - \tau)|)^2 d\tau = \int_{-\infty}^{\infty} (|h_m(\tau)|)^2 d\tau
\tag{4.71}
$$

$$
\begin{aligned}
h_m * h_m(0) &= \int_{-\infty}^{\infty} (|H_m(f)|)^2 df \\
&= < H_m(f), H_m(f) >
\end{aligned}
\tag{4.72}
$$

$$
\begin{aligned}
h_m * h_m(0) &= \int_{-\infty}^{\infty} H_m(f) e^{j2\pi fT} H_m(f)^* e^{-(j2\pi fT)} df \\
&= < \hat{H}_m(f), \hat{H}_m(f) >
\end{aligned}
\tag{4.73}
$$

Therefore,

$$
\text{SNR} = \frac{\left(< S_m(f), \hat{H}_m(f) > \right)^2}{\frac{N_0}{2} < \left(\hat{H}_m(f), \hat{H}_m(f) \right) >} \leq \frac{2}{N_0} < (S_m(f), S_m(f)) >
\tag{4.74}
$$

with equality when

$$
\hat{H}_m(f) = \alpha S_m(f)
\tag{4.75}
$$

or

Matched Filter in the frequency domain

$$
H_m(f) = S_m(f)^* e^{-(j2\pi fT)}
\tag{4.76}
$$

Matched Filter
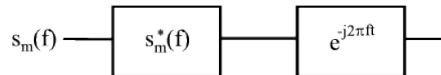
$$s_m(f) \longrightarrow \boxed{s_m^*(f)} \longrightarrow \boxed{e^{-j2\pi ft}} \longrightarrow$$

Figure 4.24

$$
\begin{aligned}
\tilde{s_m}\left(t\right) &= \quad \mathcal{F}^{-1}\left(s_m\left(f\right)s_m\left(f\right)^*\right) \\
&= \quad \int_{-\infty}^{\infty}\left(\left|s_m\left(f\right)\right|\right)^2 e^{j2\pi ft}df \\
&= \quad \int_{-\infty}^{\infty}\left(\left|s_m\left(f\right)\right|\right)^2 \cos\left(2\pi ft\right)df
\end{aligned}
\tag{4.77}
$$

where $\mathcal{F}^{-1}$ is the inverse Fourier Transform operator.

4.11 Performance Analysis[11]

In this section we will evaluate the probability of error of both correlator type receivers and matched filter receivers. We will only present the analysis for transmission of binary symbols. In the process we will demonstrate that both of these receivers have identical bit-error probabilities.

4.11.1 Antipodal Signals

$r_t = s_m\left(t\right) + N_t$ for $0 \le t \le T$ with $m = 1$ and $m = 2$ and $s_1\left(t\right) = -s_2\left(t\right)$

An analysis of the performance of correlation receivers with antipodal binary signals can be found here (Section 4.12). A similar analysis for matched filter receivers can be found here (Section 4.14).

4.11.2 Orthogonal Signals

$r_t = s_m\left(t\right) + N_t$ for $0 \le t \le T$ with $m = 1$ and $m = 2$ and $< s_1, s_2 > = 0$

An analysis of the performance of correlation receivers with orthogonal binary signals can be found here (Section 4.13). A similar analysis for matched filter receivers can be found here (Section 4.15).

It can be shown in general that correlation and matched filter receivers perform with the same symbol error probability if the detection criteria is the same for both receivers.

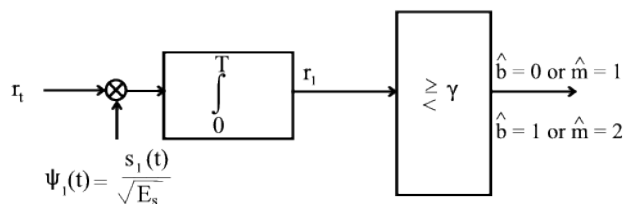# 4.12 Performance Analysis of Antipodal Binary signals with Correlation[12]

Figure 4.25

[11]This content is available online at $<$http://cnx.org/content/m10106/2.10/$>$.

[12]This content is available online at $<$http://cnx.org/content/m10152/2.11/$>$.

The bit-error probability for a correlation receiver with an antipodal signal set (Figure 4.25) can be found as follows:

$$
\begin{aligned}
P_e &= Pr\left[\hat{m} \neq m\right] \\
&= Pr\left[\hat{b} \neq b\right] \\
&= \pi_0 Pr\left[r_1 < \gamma|_{m=1}\right] + \pi_1 Pr\left[r_1 \geq \gamma|_{m=2}\right] \\
&= \pi_0 \int_{-\infty}^{\gamma} f_{r_1,s_1(t)}(r)\,dr + \pi_1 \int_{\gamma}^{\infty} f_{r_1,s_2(t)}(r)\,dr
\end{aligned}
\tag{4.78}
$$

if $\pi_0 = \pi_1 = 1/2$, then the optimum threshold is $\gamma = 0$.

$$
f_{r_1|s_1(t)}(r) = \mathcal{N}\left(\sqrt{E_s}, \frac{N_0}{2}\right)
\tag{4.79}
$$

$$
f_{r_1|s_2(t)}(r) = \mathcal{N}\left(-\sqrt{E_s}, \frac{N_0}{2}\right)
\tag{4.80}
$$

If the two symbols are equally likely to be transmitted then $\pi_0 = \pi_1 = 1/2$ and if the threshold is set to zero, then

$$
P_e = 1/2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi \frac{N_0}{2}}} e^{-\frac{(|r-\sqrt{E_s}|)^2}{N_0}}\,dr + 1/2 \int_{0}^{\infty} \frac{1}{\sqrt{2\pi \frac{N_0}{2}}} e^{-\frac{(|r+\sqrt{E_s}|)^2}{N_0}}\,dr
\tag{4.81}
$$

$$
P_e = 1/2 \int_{-\infty}^{-\sqrt{\frac{2E_s}{N_0}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(|r'|)^2}{2}}\,dr' + 1/2 \int_{\sqrt{\frac{2E_s}{N_0}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(|r''|\right)^2}{2}}\,dr''
\tag{4.82}
$$

with $r' = \frac{r-\sqrt{E_s}}{\sqrt{\frac{N_0}{2}}}$ and $r'' = \frac{r+\sqrt{E_s}}{\sqrt{\frac{N_0}{2}}}$

$$
\begin{aligned}
P_e &= \frac{1}{2}Q\left(\sqrt{\frac{2E_s}{N_0}}\right) + \frac{1}{2}Q\left(\sqrt{\frac{2E_s}{N_0}}\right) \\
&= Q\left(\sqrt{\frac{2E_s}{N_0}}\right)
\end{aligned}
\tag{4.83}
$$

where $Q(b) = \int_{b}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\,dx$.
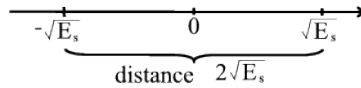
Note that

Figure 4.26

$$
P_e = Q\left(\frac{d_{12}}{\sqrt{2N_0}}\right)
\tag{4.84}
$$

where $d_{12} = 2\sqrt{E_s} = (\| s_1 - s_2 \|)^2$ is the Euclidean distance between the two constellation points (Figure 4.26).

This is exactly the same bit-error probability as for the matched filter case.

A similar bit-error analysis for matched filters can be found here (Section 4.14). For the bit-error analysis for correlation receivers with an orthogonal signal set, refer here (Section 4.13).

4.13 Performance Analysis of Binary Orthogonal Signals with Correlation[13]

Orthogonal signals with equally likely bits, $r_t = s_m(t) + N_t$ for $0 \le t \le T$, $m = 1$, $m = 2$, and $< s_1, s_2 > = 0$.

4.13.1 Correlation (correlator-type) receiver

$r_t \Rightarrow \left(r = (r_1, r_2)^T = s_m + \eta \right)$ (see Figure 4.27)
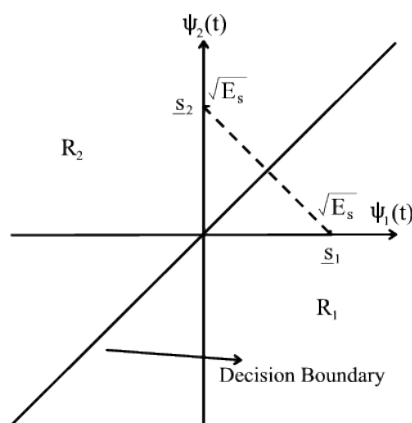

Figure 4.27

Decide $s_1(t)$ was transmitted if $r_1 \ge r_2$.

$$
\begin{aligned}
P_e &= Pr\left[\hat{m} \ne m\right] \\
    &= Pr\left[\hat{b} \ne b\right]
\end{aligned}
\tag{4.85}
$$

$P_e = 1/2 Pr\left[r \in R_2 \mid s_1(t) \text{ transmitted}\right] + 1/2 Pr\left[r \in R_1 \mid s_2(t) \text{ transmitted}\right] = \tag{4.86}$

$1/2 \int_{R_2} \int f_{r, s_1(t)}(r)\, dr\, 1 dr\, 2 \qquad + \qquad 1/2 \int_{R_1} \int f_{r, s_2(t)}(r)\, dr\, 1 dr\, 2 \qquad\qquad =$

$1/2 \int_{R_2} \int \frac{1}{\sqrt{2\pi \frac{N_0}{2}}} e^{\frac{-(|r_1 - \sqrt{E_s}|)^2}{N_0}} \frac{1}{\sqrt{\pi N_0}} e^{\frac{-(|r_2|)^2}{N_0}}\, dr\, 1 dr\, 2 + 1/2 \int_{R_1} \int \frac{1}{\sqrt{2\pi \frac{N_0}{2}}} e^{\frac{-(|r_1|)^2}{N_0}} \frac{1}{\sqrt{\pi N_0}} e^{\frac{-(|r_2 - \sqrt{E_s}|)^2}{N_0}}\, dr\, 1 dr\, 2$

Alternatively, if $s_1(t)$ is transmitted we decide on the wrong signal if $r_2 > r_1$ or $\eta_2 > \eta_1 + \sqrt{E_s}$ or when $\eta_2 - \eta_1 > \sqrt{E_s}$.

$$
\begin{aligned}
P_e &= 1/2 \int_{\sqrt{E_s}}^{\infty} \frac{1}{\sqrt{2\pi N_0}} e^{\frac{-\eta'^2}{2N_0}}\, d\eta\prime + 1/2 Pr\left[r_1 \ge r_2 \mid s_2(t) \text{ transmitted}\right] \\
    &= Q\left(\sqrt{\frac{E_s}{N_0}}\right)
\end{aligned}
\tag{4.87}
$$

---

[13] This content is available online at <http://cnx.org/content/m10154/2.11/>.

Note that the distance between s_1 and s_2 is $d_{12} = \sqrt{2E_s}$. The average bit error probability $P_e = Q\left(\frac{d_{12}}{\sqrt{2N_0}}\right)$ as we had for the antipodal case (Section 4.12). Note also that the bit-error probability is the same as for the matched filter (Section 4.15) receiver.

4.14 Performance Analysis of Binary Antipodal Signals with Matched Filters[14]

4.14.1 Matched Filter receiver

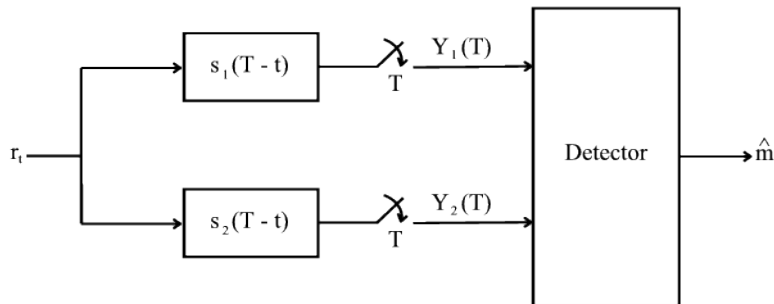Recall $r_t = s_m(t) + N_t$ where $m = 1$ or $m = 2$ and $s_1(t) = -s_2(t)$ (see Figure 4.28).

Figure 4.28

$$Y_1(T) = E_s + \nu_1 \tag{4.88}$$

$$Y_2(T) = -E_s + \nu_2 \tag{4.89}$$

since $s_1(t) = -s_2(t)$ then ν_1 is $\mathcal{N}\left(0, \frac{N_0}{2}E_s\right)$. Furthermore $\nu_2 = -\nu_1$. Given ν_1 then ν_2 is deterministic and equals $-\nu_1$. Then $Y_2(T) = -Y_1(T)$ if $s_1(t)$ is transmitted.

If $s_2(T)$ is transmitted

$$Y_1(T) = -E_s + \nu_1 \tag{4.90}$$

$$Y_2(T) = E_s + \nu_2 \tag{4.91}$$

ν_1 is $\mathcal{N}\left(0, \frac{N_0}{2}E_s\right)$ and $\nu_2 = -\nu_1$.

The receiver can be simplified to (see Figure 4.29)

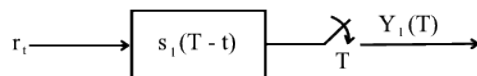[14]This content is available online at <http://cnx.org/content/m10153/2.11/>.

Figure 4.29

If $s_1(t)$ is transmitted $Y_1(T) = E_s + \nu_1$.
If $s_2(t)$ is transmitted $Y_1(T) = -E_s + \nu_1$.

$$
\begin{aligned}
P_e &= 1/2 Pr\left[Y_1(T) < 0 \mid s_1(t)\right] + 1/2 Pr\left[Y_1(T) \geq 0 \mid s_2(t)\right] \\
&= 1/2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi \frac{N_0}{2} E_s}} e^{\frac{-(|y - E_s|)^2}{N_0 E_s}} dy + 1/2 \int_{0}^{\infty} \frac{1}{\sqrt{2\pi \frac{N_0}{2} E_s}} e^{\frac{-(|y + E_s|)^2}{N_0 E_s}} dy \\
&= Q\left(\frac{E_s}{\sqrt{\frac{N_0}{2} E_s}}\right) \\
&= Q\left(\sqrt{\frac{2E_s}{N_0}}\right)
\end{aligned}
\tag{4.92}
$$

This is the exact bit-error rate of a correlation receiver (Section 4.12). For a bit-error analysis for orthogonal signals using a matched filter receiver, refer here (Section 4.15).

4.15 Performance Analysis of Orthogonal Binary Signals with Matched Filters[15]

$$
r_t \Rightarrow \left(Y = \left(\begin{array}{c} Y_1(T) \\ Y_2(T) \end{array} \right) \right)
\tag{4.93}
$$

If $s_1(t)$ is transmitted

$$
\begin{aligned}
Y_1(T) &= \int_{-\infty}^{\infty} s_1(\tau) h_1^{\mathrm{opt}}(T - \tau) d\tau + \nu_1(T) \\
&= \int_{-\infty}^{\infty} s_1(\tau) s_1^*(\tau) d\tau + \nu_1(T) \\
&= E_s + \nu_1(T)
\end{aligned}
\tag{4.94}
$$

$$
\begin{aligned}
Y_2(T) &= \int_{-\infty}^{\infty} s_1(\tau) s_2^*(\tau) d\tau + \nu_2(T) \\
&= \nu_2(T)
\end{aligned}
\tag{4.95}
$$

If $s_2(t)$ is transmitted, $Y_1(T) = \nu_1(T)$ and $Y_2(T) = E_s + \nu_2(T)$.
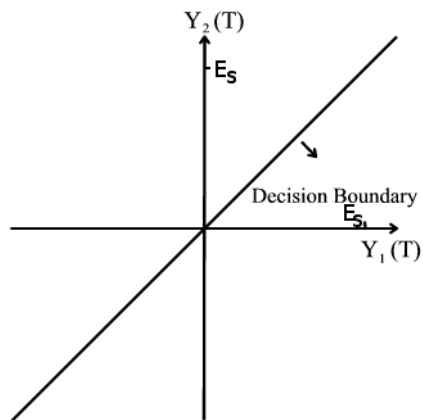
---

[15]This content is available online at $<$http://cnx.org/content/m10155/2.9/$>$.

Figure 4.30

H0

$$Y = \begin{pmatrix} E_s \\ 0 \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \tag{4.96}$$

H1

$$Y = \begin{pmatrix} 0 \\ E_s \end{pmatrix} + \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \tag{4.97}$$

where ν_1 and ν_2 are independent are Gaussian with zero mean and variance $\frac{N_0}{2} E_s$. The analysis is identical to the correlator example (Section 4.13).

$$P_e = Q\left(\sqrt{\frac{E_s}{N_0}} \right) \tag{4.98}$$

Note that the maximum likelihood detector decides based on comparing Y_1 and Y_2. If $Y_1 \geq Y_2$ then s_1 was sent; otherwise s_2 was transmitted. For a similar analysis for binary antipodal signals, refer here (Section 4.14). See Figure 4.31 or Figure 4.32.
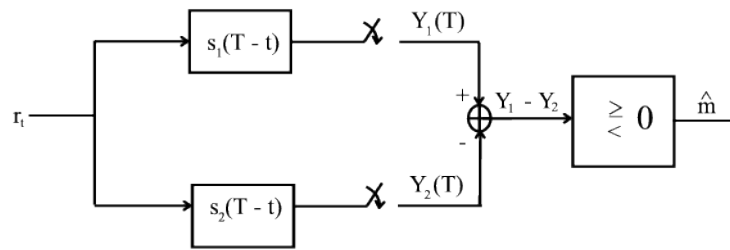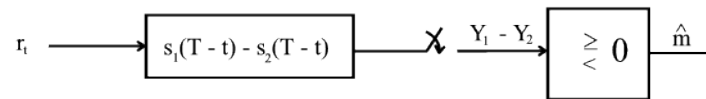
Figure 4.31

Figure 4.32

Chapter 5

Chapter 5

5.1 Digital Transmission over Baseband Channels[1]

Until this point, we have considered data transmissions over simple additive Gaussian channels that are not time or band limited. In this module we will consider channels that do have bandwidth constraints, and are limited to frequency range around zero (DC). The channel is best modified as $g(t)$ is the impulse response of the baseband channel.

Consider modulated signals $x_t = s_m(t)$ for $0 \leq t \leq T$ for some $m \in \{1, 2, \ldots, M\}$. The channel output is then

$$
\begin{aligned}
r_t &= \int_{-\infty}^{\infty} x_\tau g(t - \tau) \, d\tau + N_t \\
&= \int_{-\infty}^{\infty} S_m(\tau) g(t - \tau) \, d\tau + N_t
\end{aligned}
\tag{5.1}
$$

The signal contribution in the frequency domain is

$$
\tilde{S_m}(f) = S_m(f) G(f)
\tag{5.2}
$$

The optimum matched filter should match to the filtered signal:

$$
H_m^{\text{opt}}(f) = S_m(f)^* G(f)^* e^{(-j)2\pi ft}
\tag{5.3}
$$

This filter is indeed **optimum** (*i.e.*, it maximizes signal-to-noise ratio); however, it requires knowledge of the channel impulse response. The signal energy is changed to

$$
E_{\tilde{s}} = \int_{-\infty}^{\infty} \left(|\tilde{S_m}(f)| \right)^2 df
\tag{5.4}
$$

The band limited nature of the channel and the stream of time limited modulated signal create aliasing which is referred to as **intersymbol interference**. We will investigate ISI for a general PAM signaling.

5.2 Pulse Amplitude Modulation Through Bandlimited Channel[2]

Consider a PAM system $b_{-10}, \ldots, b_{-1}, b_0, b_1, \ldots$

This implies

$$
x_t = \sum_{n=-\infty}^{\infty} a_n s(t - nT) \quad , \quad a_n \in \{\text{M levels of amplitude}\}
\tag{5.5}
$$

---

[1]This content is available online at <http://cnx.org/content/m10056/2.12/>.
[2]This content is available online at <http://cnx.org/content/m10094/2.7/>.

The received signal is

$$
\begin{aligned}
r_t &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_n s\left(t-(\tau-nT)\right) g\left(\tau\right) d\tau + N_t \\
&= \sum_{n=-\infty}^{\infty} a_n \int_{-\infty}^{\infty} s\left(t-(\tau-nT)\right) g\left(\tau\right) d\tau + N_t \\
&= \sum_{n=-\infty}^{\infty} a_n \tilde{s}\left(t-nT\right) + N_t
\end{aligned}
\tag{5.6}
$$

Since the signals span a one-dimensional space, one filter matched to $\tilde{s}\left(t\right) = s^* g\left(t\right)$ is sufficient. The matched filter's impulse response is

$$
h^{\mathrm{opt}}\left(t\right) = s^* g\left(T-t\right)
\tag{5.7}
$$

The matched filter output is

$$
\begin{aligned}
y\left(t\right) &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_n \tilde{s}\left(t-(\tau-nT)\right) h^{\mathrm{opt}}\left(\tau\right) d\tau + \nu\left(t\right) \\
&= \sum_{n=-\infty}^{\infty} a_n \int_{-\infty}^{\infty} \tilde{s}\left(t-(\tau-nT)\right) h^{\mathrm{opt}}\left(\tau\right) d\tau + \nu\left(t\right) \\
&= \sum_{n=-\infty}^{\infty} a_n u\left(t-nT\right) + \nu\left(t\right)
\end{aligned}
\tag{5.8}
$$

The decision on the kth symbol is obtained by sampling the MF output at kT:

$$
y\left(kT\right) = \sum_{n=-\infty}^{\infty} a_n u\left(kT-nT\right) + \nu\left(kT\right)
\tag{5.9}
$$

The kth symbol is of interest:

$$
y\left(kT\right) = a_k u\left(0\right) + \sum_{n=-\infty}^{\infty} a_n u\left(kT-nT\right) + \nu\left(kT\right)
\tag{5.10}
$$

where $n \neq k$.

Since the channel is bandlimited, it provides memory for the transmission system. The effect of old symbols (possibly even future signals) lingers and affects the performance of the receiver. The effect of ISI can be eliminated or controlled by proper design of **modulation signals** or **precoding** filters at the transmitter, or by **equalizers** or **sequence detectors** at the receiver.

5.3 Precoding and Bandlimited Signals[3]

5.3.1 Precoding

The data symbols are manipulated such that

$$
y_k\left(kT\right) = a_k u\left(0\right) + \mathrm{ISI} + \nu\left(kT\right)
\tag{5.11}
$$

5.3.2 Design of Bandlimited Modulation Signals

Recall that modulation signals are

$$
X_t = \sum_{n=-\infty}^{\infty} a_n s\left(t-nT\right)
\tag{5.12}
$$

---

[3]This content is available online at $<$http://cnx.org/content/m10118/2.6/$>$.

We can design $s(t)$ such that

$$u(nT) = \begin{cases} \text{large} & \text{if } n = 0 \\ \text{zero or small} & \text{if } n \neq 0 \end{cases} \quad (5.13)$$

where $y(kT) = a_k u(0) + \sum_{n=-\infty}^{\infty} a_n u(kT - nT) + \nu(kT)$ (ISI is the sum term, and once again, $n \neq k$.) Also, $y(nT) = s^* g^* h^{\text{opt}}(nT)$ The signal $s(t)$ can be designed to have reduced ISI.

5.3.3 Design Equalizers at the Receiver

Linear equalizers or decision feedback equalizers reduce ISI in the statistic y_t

5.3.4 Maximum Likelihood Sequence Detection

$$y(kT) = \sum_{n=-\infty}^{\infty} a_n(kT - nT) + \nu(k(T)) \quad (5.14)$$

By observing $y(T), y(2T), \ldots$ the date symbols are observed frequently. Therefore, ISI can be viewed as diversity to increase performance.

Chapter 6

Chapter 6

6.1 Carrier Phase Modulation[1]

6.1.1 Phase Shift Keying (PSK)

Information is impressed on the phase of the carrier. As data changes from symbol period to symbol period, the phase shifts.

$$s_m(t) = AP_T(t)\cos\left(2\pi f_c t + \frac{2\pi(m-1)}{M}\right) \quad, \quad m \in \{1, 2, \ldots, M\} \tag{6.1}$$

Example 6.1
Binary $s_1(t)$ or $s_2(t)$

6.1.2 Representing the Signals

An orthonormal basis to represent the signals is

$$\psi_1(t) = \frac{1}{\sqrt{E_s}} AP_T(t)\cos(2\pi f_c t) \tag{6.2}$$

$$\psi_2(t) = \frac{-1}{\sqrt{E_s}} AP_T(t)\sin(2\pi f_c t) \tag{6.3}$$

The signal

$$S_m(t) = AP_T(t)\cos\left(2\pi f_c t + \frac{2\pi(m-1)}{M}\right) \tag{6.4}$$

$$S_m(t) = A\cos\left(\frac{2\pi(m-1)}{M}\right) P_T(t)\cos(2\pi f_c t) - A\sin\left(\frac{2\pi(m-1)}{M}\right) P_T(t)\sin(2\pi f_c t) \tag{6.5}$$

The signal energy

$$
\begin{aligned}
E_s &= \int_{-\infty}^{\infty} A^2 P_T{}^2(t)\cos^2\left(2\pi f_c t + \frac{2\pi(m-1)}{M}\right) dt \\
&= \int_0^T A^2\left(\frac{1}{2} + \frac{1}{2}\cos\left(4\pi f_c t + \frac{4\pi(m-1)}{M}\right)\right) dt
\end{aligned} \tag{6.6}
$$

$$E_s = \frac{A^2 T}{2} + \frac{1}{2}A^2 \int_0^T \cos\left(4\pi f_c t + \frac{4\pi\,(m-1)}{M}\right) dt \simeq \frac{A^2 T}{2} \tag{6.7}$$

(Note that in the above equation, the integral in the last step before the aproximation is very small.) Therefore,

$$\psi_1\,(t) = \sqrt{\frac{2}{T}} P_T\,(t)\cos\,(2\pi f_c t) \tag{6.8}$$

$$\psi_2\,(t) = \left(-\sqrt{\frac{2}{T}}\right) P_T\,(t)\sin\,(2\pi f_c t) \tag{6.9}$$

In general,

$$s_m\,(t) = AP_T\,(t)\cos\left(2\pi f_c t + \frac{2\pi\,(m-1)}{M}\right)\ ,\ \ m \in \{1, 2, \dots, M\} \tag{6.10}$$

and $\psi_1\,(t)$

$$\psi_1\,(t) = \sqrt{\frac{2}{T}} P_T\,(t)\cos\,(2\pi f_c t) \tag{6.11}$$

$$\psi_2\,(t) = \sqrt{\frac{2}{T}} P_T\,(t)\sin\,(2\pi f_c t) \tag{6.12}$$

$$s_m = \left(\begin{array}{c} \sqrt{E_s}\cos\left(\frac{2\pi(m-1)}{M}\right) \\ \sqrt{E_s}\sin\left(\frac{2\pi(m-1)}{M}\right) \end{array}\right) \tag{6.13}$$

6.1.3 Demodulation and Detection

$$r_t = s_m\,(t) + N_t,\ \text{for some} m \in \{1, 2, \dots, M\} \tag{6.14}$$

We must note that due to phase offset of the oscillator at the transmitter, **phase jitter** or **phase changes** occur because of propagation delay.

$$r_t = AP_T\,(t)\cos\left(2\pi f_c t + \frac{2\pi\,(m-1)}{M} + \phi\right) + N_t \tag{6.15}$$

For binary PSK, the modulation is antipodal, and the optimum receiver in AWGN has average bit-error probability

$$\begin{aligned} P_e &= Q\left(\sqrt{\frac{2(E_s)}{N_0}}\right) \\ &= Q\left(A\sqrt{\frac{T}{N_0}}\right) \end{aligned} \tag{6.16}$$

The receiver where

$$r_t = \pm\,(AP_T\,(t)\cos\,(2\pi f_c t + \phi)) + N_t \tag{6.17}$$

The statistics

$$\begin{aligned} r_1 &= \int_0^T r_t \alpha\cos\left(2\pi f_c t + \hat{\phi}\right) dt \\ &= \pm\left(\int_0^T \alpha A\cos\,(2\pi f_c t + \phi)\cos\left(2\pi f_c t + \hat{\phi}\right) dt\right) + \int_0^T \alpha\cos\left(2\pi f_c t + \hat{\phi}\right) N_t dt \end{aligned} \tag{6.18}$$

$$r_1 = \pm \left(\frac{\alpha A}{2} \int_0^T \cos\left(4\pi f_c t + \phi + \hat{\phi}\right) + \cos\left(\phi - \hat{\phi}\right) dt \right) + \eta_1 \tag{6.19}$$

$$r_1 = \pm \left(\frac{\alpha A}{2} T \cos\left(\phi - \hat{\phi}\right) \right) + \int_0^T \pm \left(\frac{\alpha A}{2} \cos\left(4\pi f_c t + \phi + \hat{\phi}\right) \right) dt + \eta_1 \pm \left(\frac{\alpha A T}{2} \cos\left(\phi - \hat{\phi}\right) \right) + \eta_1 \tag{6.20}$$

where $\eta_1 = \alpha \int_0^T N_t \cos\left(\omega_c t + \hat{\phi}\right) dt$ is zero mean Gaussian with variance $\simeq \frac{\alpha^2 N_0 T}{4}$.

Therefore,

$$
\begin{aligned}
\bar{P}_e &= Q\left(\frac{2 \frac{\alpha A T}{2} \cos\left(\phi - \hat{\phi}\right)}{2\sqrt{\frac{\alpha^2 N_0 T}{4}}} \right) \\
&= Q\left(\cos\left(\phi - \hat{\phi}\right) A \sqrt{\frac{T}{N_0}} \right)
\end{aligned}
\tag{6.21}
$$

which is not a function of α and depends strongly on phase accuracy.

$$P_e = Q\left(\cos\left(\phi - \hat{\phi}\right) \sqrt{\frac{2E_s}{N_0}} \right) \tag{6.22}$$

The above result implies that the amplitude of the local oscillator in the correlator structure does not play a role in the performance of the correlation receiver. However, the accuracy of the phase does indeed play a major role. This point can be seen in the following example:

Example 6.2

$$x_{t'} = -1^i A \cos\left(- (2\pi f_c t') + 2\pi f_c \tau\right) \tag{6.23}$$

$$x_t = -1^i A \cos\left(2\pi f_c t - (2\pi f_c \tau' - 2\pi f_c \tau + \theta')\right) \tag{6.24}$$

Local oscillator should match to phase θ.

6.2 Differential Phase Shift Keying[2]

The phase lock loop provides estimates of the phase of the incoming modulated signal. A phase ambiguity of exactly π is a common occurance in many phase lock loop (PLL) implementations.

Therefore it is possible that, $\hat{\theta} = \theta + \pi$ without the knowledge of the receiver. Even if there is no noise, if $b = 1$ then $\hat{b} = 0$ and if $b = 0$ then $\hat{b} = 1$.

In the presence of noise, an incorrect decision due to noise may results in a correct final desicion (in binary case, when there is π phase ambiguity with the probability:

$$\overline{P_e} = 1 - Q\left(\sqrt{\frac{2E_s}{N_0}} \right) \tag{6.25}$$

---

[2]This content is available online at <http://cnx.org/content/m10156/2.7/>.

Consider a stream of bits $a_n \in \{0, 1\}$ and BPSK modulated signal

$$\sum_n -1^{a_n} A P_T (t - nT) \cos (2\pi f_c t + \theta) \tag{6.26}$$

In differential PSK, the transmitted bits are first encoded $b_n = a_n \oplus b_{n-1}$ with initial symbol (e.g. b_0) chosen without loss of generality to be either 0 or 1.

Transmitted DPSK signals

$$\sum_n -1^{b_n} A P_T (t - nT) \cos (2\pi f_c t + \theta) \tag{6.27}$$

The decoder can be constructed as

$$
\begin{aligned}
b_{n-1} \oplus b_n &= b_{n-1} \oplus a_n \oplus b_{n-1} \\
&= 0 \oplus a_n \\
&= a_n
\end{aligned}
\tag{6.28}
$$

If two consecutive bits are detected correctly, if $\hat{b}_n = b_n$ and $\hat{b}_{n-1} = b_{n-1}$ then

$$
\begin{aligned}
\hat{a}_n &= \hat{b}_n \oplus \hat{b}_{n-1} \\
&= b_n \oplus b_{n-1} \\
&= a_n \oplus b_{n-1} \oplus b_{n-1} \\
&= a_n
\end{aligned}
\tag{6.29}
$$

if $\hat{b}_n = b_n \oplus 1$ and $\hat{b}_{n-1} = b_{n-1} \oplus 1$. That is, two consecutive bits are detected incorrectly. Then,

$$
\begin{aligned}
\hat{a}_n &= \hat{b}_n \oplus \hat{b}_{n-1} \\
&= b_n \oplus 1 \oplus b_{n-1} \oplus 1 \\
&= b_n \oplus b_{n-1} \oplus 1 \oplus 1 \\
&= b_n \oplus b_{n-1} \oplus 0 \\
&= b_n \oplus b_{n-1} \\
&= a_n
\end{aligned}
\tag{6.30}
$$

If $\hat{b}_n = b_n \oplus 1$ and $\hat{b}_{n-1} = b_{n-1}$, that is, one of two consecutive bits is detected in error. In this case there will be an error and the probability of that error for DPSK is

$$
\begin{aligned}
\overline{P}_e &= Pr\left[\hat{a}_n \neq a_n\right] \\
&= Pr\left[\hat{b}_n = b_n, \hat{b}_{n-1} \neq b_{n-1}\right] + Pr\left[\hat{b}_n \neq b_n, \hat{b}_{n-1} = b_{n-1}\right] \\
&= 2Q\left(\sqrt{\frac{2E_s}{N_0}}\right)\left[1 - Q\left(\sqrt{\frac{2E_s}{N_0}}\right)\right] \simeq 2Q\left(\sqrt{\frac{2E_s}{N_0}}\right)
\end{aligned}
\tag{6.31}
$$

This approximation holds if Q is small.

6.3 Carrier Frequency Modulation[3]

6.3.1 Frequency Shift Keying (FSK)

The data is impressed upon the carrier frequency. Therefore, the M different signals are

$$s_m(t) = A P_T(t) \cos(2\pi f_c t + 2\pi(m-1)\Delta(f)t + \theta_m) \tag{6.32}$$

for $m \in \{1, 2, \ldots, M\}$

The M different signals have M different carrier frequencies with possibly different phase angles since the generators of these carrier signals may be different. The carriers are

$$f_1 = f_c \tag{6.33}$$

$$f_2 = f_c + \Delta(f)$$

$$f_M = f_c - M\Delta(f)$$

Thus, the M signals may be designed to be orthogonal to each other.

$< s_m, s_n > = \int_0^T A^2 \cos(2\pi f_c t + 2\pi(m-1)\Delta(f)t + \theta_m)\cos(2\pi f_c t + 2\pi(n-1)\Delta(f)t + \theta_n)\,dt = \tag{6.34}$
$\frac{A^2}{2}\int_0^T \cos(4\pi f_c t + 2\pi(n+m-2)\Delta(f)t + \theta_m + \theta_n)\,dt \qquad\qquad +$
$\frac{A^2}{2}\int_0^T \cos(2\pi(m-n)\Delta(f)t + \theta_m - \theta_n)\,dt = \frac{A^2}{2}\frac{\sin(4\pi f_c T + 2\pi(n+m-2)\Delta(f)T + \theta_m + \theta_n) - \sin(\theta_m + \theta_n)}{4\pi f_c + 2\pi(n+m-2)\Delta(f)} +$
$\frac{A^2}{2}\left(\frac{\sin(2\pi(m-n)\Delta(f)T + \theta_m - \theta_n)}{2\pi(m-n)\Delta(f)} - \frac{\sin(\theta_m - \theta_n)}{2\pi(m-n)\Delta(f)}\right)$

If $2f_c T + (n + m - 2)\Delta(f)T$ is an integer, and if $(m - n)\Delta(f)T$ is also an integer, then $< S_m, S_n > = 0$ if $\Delta(f)T$ is an integer, then $< s_m, s_n > \simeq 0$ when f_c is much larger than $\frac{1}{T}$.

In case , $\theta_m = 0$

$$< s_m, s_n > \simeq \frac{A^2 T}{2}\text{sinc}(2(m-n)\Delta(f)T) \tag{6.35}$$

Therefore, the frequency spacing could be as small as $\Delta(f) = \frac{1}{2T}$ since $\text{sinc}(x) = 0$ if $x = \pm(1)$ or $\pm(2)$.

If the signals are designed to be orthogonal then the average probability of error for binary FSK with optimum receiver is

$$\overline{P}_e = Q\left(\sqrt{\frac{E_s}{N_0}}\right) \tag{6.36}$$

in AWGN.

Note that $\text{sinc}(x)$ takes its minimum value not at $x = \pm(1)$ but at $\pm(1.4)$ and the minimum value is -0.216. Therefore if $\Delta(f) = \frac{0.7}{T}$ then

$$\overline{P}_e = Q\left(\sqrt{\frac{1.216 E_s}{N_0}}\right) \tag{6.37}$$

which is a gain of $10 \times \log 1.216 \simeq 0.85 d\theta$ over orthogonal FSK.

---

[3]This content is available online at $<$http://cnx.org/content/m10163/2.10/$>$.

Chapter 7

Chapter 7

7.1 Information Theory and Coding[1]

In the previous chapters, we considered the problem of digital transmission over different channels. Information sources are not often digital, and in fact, many sources are analog. Although many channels are also analog, it is still more efficient to convert analog sources into digital data and transmit over analog channels using digital transmission techniques. There are two reasons why digital transmission could be more efficient and more reliable than analog transmission:

1. Analog sources could be compressed to digital form efficiently.
2. Digital data can be transmitted over noisy channels reliably.

There are several key questions that need to be addressed:

1. How can one model information?
2. How can one quantify information?
3. If information can be measured, does its information quantity relate to how much it can be compressed?
4. Is it possible to determine if a particular channel can handle transmission of a source with a particular information quantity?

Figure 7.1

Example 7.1
The information content of the following sentences: "Hello, hello, hello." and "There is an exam today." are not the same. Clearly the second one carries more information. The first one can be compressed to "Hello" without much loss of information.

In other modules, we will quantify information and find efficient representation of information (Entropy (Section 7.2)). We will also quantify how much (Section 7.5) information can be transmitted through channels, reliably. Channel coding (Section 7.9) can be used to reduce information rate and increase reliability.

---

[1]This content is available online at <http://cnx.org/content/m10162/2.10/>.

7.2 Entropy[2]

Information sources take very different forms. Since the information is not known to the destination, it is then best modeled as a random process, discrete-time or continuous time.

Here are a few examples:

- Digital data source (e.g., a text) can be modeled as a discrete-time and discrete valued random process X_1, X_2, ..., where $X_i \in \{A, B, C, D, E, \ldots\}$ with a particular $p_{X_1}(x)$, $p_{X_2}(x)$, ..., and a specific $p_{X_1 X_2}$, $p_{X_2 X_3}$, ..., and $p_{X_1 X_2 X_3}$, $p_{X_2 X_3 X_4}$, ..., *etc.*
- Video signals can be modeled as a continuous time random process. The power spectral density is bandlimited to around 5 MHz (the value depends on the standards used to raster the frames of image).
- Audio signals can be modeled as a continuous-time random process. It has been demonstrated that the power spectral density of speech signals is bandlimited between 300 Hz and 3400 Hz. For example, the speech signal can be modeled as a Gaussian process with the shown (Figure 7.2) power spectral density over a small observation period.
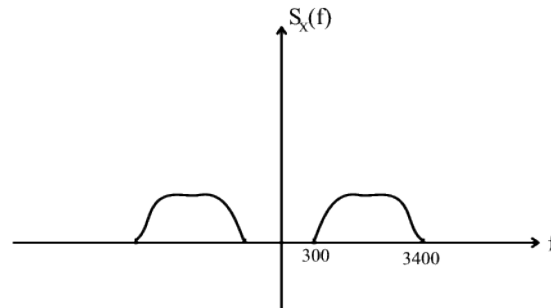
Figure 7.2

These analog information signals are bandlimited. Therefore, if sampled faster than the Nyquist rate, they can be reconstructed from their sample values.

Example 7.2
A speech signal with bandwidth of 3100 Hz can be sampled at the rate of 6.2 kHz. If the samples are quantized with a 8 level quantizer then the speech signal can be represented with a binary sequence with the rate of

$$
\begin{aligned}
6.2 \times 10^3 \log_2 8 &= 18600 \frac{\text{bits}}{\text{sample}} \frac{\text{samples}}{\text{sec}} \\
&= 18.6 \frac{\text{kbits}}{\text{sec}}
\end{aligned}
\tag{7.1}
$$

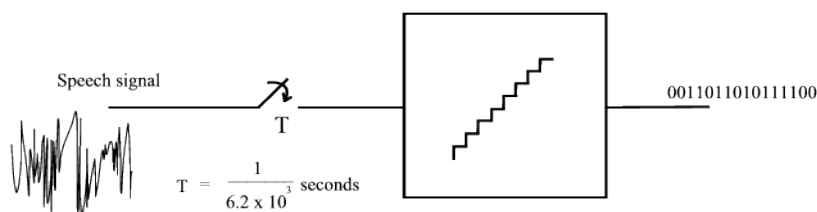[2]This content is available online at $<$http://cnx.org/content/m10164/2.16/$>$.

Figure 7.3

The sampled real values can be quantized to create a discrete-time discrete-valued random process. Since any bandlimited analog information signal can be converted to a sequence of discrete random variables, we will continue the discussion only for discrete random variables.

Example 7.3
The random variable x takes the value of 0 with probability 0.9 and the value of 1 with probability 0.1. The statement that $x = 1$ carries more information than the statement that $x = 0$. The reason is that x is expected to be 0, therefore, knowing that $x = 1$ is more surprising news!! An intuitive definition of information measure should be larger when the probability is small.

Example 7.4
The information content in the statement about the temperature and pollution level on July 15th in Chicago should be the sum of the information that July 15th in Chicago was hot and highly polluted since pollution and temperature could be independent.

$$I(\text{hot, high}) = I(\text{hot}) + I(\text{high}) \tag{7.2}$$

An intuitive and meaningful measure of information should have the following properties:

1. Self information should decrease with increasing probability.
2. Self information of two independent events should be their sum.
3. Self information should be a continuous function of the probability.

The only function satisfying the above conditions is the -log of the probability.

Definition 7.1: Entropy
1. The entropy (average self information) of a discrete random variable X is a function of its probability mass function and is defined as

$$H(X) = -\sum_{i=1}^{N} p_X(x_i) \log p_X(x_i) \tag{7.3}$$

where N is the number of possible values of X and $p_X(x_i) = Pr[X = x_i]$. If log is base 2 then the unit of entropy is bits. Entropy is a measure of uncertainty in a random variable and a measure of information it can reveal.
2. A more basic explanation of entropy is provided in another module[3].

Example 7.5
If a source produces binary information $\{0, 1\}$ with probabilities p and $1 - p$. The entropy of the source is

$$H(X) = (-(p \log_2 p)) - (1 - p) \log_2 (1 - p) \tag{7.4}$$

---

[3]"Entropy" <http://cnx.org/content/m0070/latest/>

If $p = 0$ then $H(X) = 0$, if $p = 1$ then $H(X) = 0$, if $p = 1/2$ then $H(X) = 1$ bits. The source has its largest entropy if $p = 1/2$ and the source provides no new information if $p = 0$ or $p = 1$.
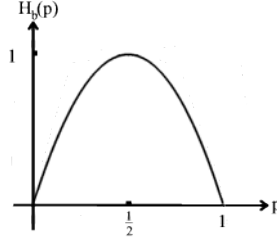

Figure 7.4

Example 7.6

An analog source is modeled as a continuous-time random process with power spectral density bandlimited to the band between 0 and 4000 Hz. The signal is sampled at the Nyquist rate. The sequence of random variables, as a result of sampling, are assumed to be independent. The samples are quantized to 5 levels $\{-2, -1, 0, 1, 2\}$. The probability of the samples taking the quantized values are $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\}$, respectively. The entropy of the random variables are

$$
\begin{aligned}
H(X) &= \left(-\left(\tfrac{1}{2}\log_2\tfrac{1}{2}\right)\right) - \tfrac{1}{4}\log_2\tfrac{1}{4} - \tfrac{1}{8}\log_2\tfrac{1}{8} - \tfrac{1}{16}\log_2\tfrac{1}{16} - \tfrac{1}{16}\log_2\tfrac{1}{16} \\
&= \tfrac{1}{2}\log_2 2 + \tfrac{1}{4}\log_2 4 + \tfrac{1}{8}\log_2 8 + \tfrac{1}{16}\log_2 16 + \tfrac{1}{16}\log_2 16 \\
&= \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{3}{8} + \tfrac{4}{8} \\
&= \tfrac{15}{8} \tfrac{\text{bits}}{\text{sample}}
\end{aligned}
\tag{7.5}
$$

There are 8000 samples per second. Therefore, the source produces $8000 \times \frac{15}{8} = 15000\frac{\text{bits}}{\text{sec}}$ of information.

Definition 7.2: Joint Entropy

The joint entropy of two discrete random variables (X, Y) is defined by

$$
H(X, Y) = -\sum_{ii}\sum_{jj} p_{X,Y}(x_i, y_j) \log p_{X,Y}(x_i, y_j)
\tag{7.6}
$$

The joint entropy for a random vector $X = (X_1, X_2, \ldots, X_n)^T$ is defined as

$$
H(X) = -\sum_{\text{x }1x_1}\sum_{\text{x }2x_2} \cdots \sum_{\text{x }nx_n} p_X(x_1, x_2, \ldots, x_n) \log p_X(x_1, x_2, \ldots, x_n)
\tag{7.7}
$$

Definition 7.3: Conditional Entropy

The conditional entropy of the random variable X given the random variable Y is defined by

$$
H(X|Y) = -\sum_{ii}\sum_{jj} p_{X,Y}(x_i, y_j) \log p_{X|Y}(x_i|y_j)
\tag{7.8}
$$

It is easy to show that

$$H\left(X\right) = H\left(X_1\right) + H\left(X_2|X_1\right) + \cdots + H\left(X_n|X_1 X_2 \ldots X_{n-1}\right) \tag{7.9}$$

and

$$\begin{aligned} H\left(X,Y\right) &= H\left(Y\right) + H\left(X|Y\right) \\ &= H\left(X\right) + H\left(Y|X\right) \end{aligned} \tag{7.10}$$

If X_1, X_2, ..., X_n are mutually independent it is easy to show that

$$H\left(X\right) = \sum_{i=1}^{n} H\left(X_i\right) \tag{7.11}$$

Definition 7.4: Entropy Rate
The entropy rate of a stationary discrete-time random process is defined by

$$H = \lim_{n \to \infty} H\left(X_n|X_1 X_2 \ldots X_n\right) \tag{7.12}$$

The limit exists and is equal to

$$H = \lim_{n \to \infty} \frac{1}{n} H\left(X_1, X_2, \ldots, X_n\right) \tag{7.13}$$

The entropy rate is a measure of the uncertainty of information content per output symbol of the source.

Entropy is closely tied to source coding (Section 7.3). The extent to which a source can be compressed is related to its entropy. In 1948, Claude E. Shannon introduced a theorem which related the entropy to the number of bits per second required to represent a source without much loss.

7.3 Source Coding[4]

As mentioned earlier, how much a source can be compressed should be related to its entropy (Section 7.2). In 1948, Claude E. Shannon introduced three theorems and developed very rigorous mathematics for digital communications. In one of the three theorems, Shannon relates entropy to the minimum number of bits per second required to represent a source without much loss (or distortion).

Consider a source that is modeled by a discrete-time and discrete-valued random process X_1, X_2, ..., X_n, ... where $x_i \in \{a_1, a_2, \ldots, a_N\}$ and define $p_{X_i}\left(x_i = a_j\right) = p_j$ for $j = 1, 2, \ldots, N$, where it is assumed that X_1, X_2,... X_n are mutually independent and identically distributed.

Consider a sequence of length n

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \tag{7.14}$$

The symbol a_1 can occur with probability p_1. Therefore, in a sequence of length n, on the average, a_1 will appear np_1 times with high probabilities if n is very large.

Therefore,

$$P\left(X = x\right) = p_{X_1}\left(x_1\right) p_{X_2}\left(x_2\right) \ldots p_{X_n}\left(x_n\right) \tag{7.15}$$

---

[4]This content is available online at <http://cnx.org/content/m10175/2.10/>.

$$P\left(X=x\right) \simeq p_1{}^{np_1} p_2{}^{np_2} \ldots p_N{}^{np_N} = \prod_{i=1}^{N} p_i{}^{np_i} \tag{7.16}$$

where $p_i = P\left(X_j = a_i\right)$ for all j and for all i.

A typical sequence X may look like

$$X = \begin{pmatrix} a_2 \\ \vdots \\ a_1 \\ a_N \\ a_2 \\ a_5 \\ \vdots \\ a_1 \\ \vdots \\ a_N \\ a_6 \end{pmatrix} \tag{7.17}$$

where a_i appears np_i times with large probability. This is referred to as a **typical sequence**. The probability of X being a typical sequence is

$$
\begin{aligned}
P\left(X=x\right) \simeq \prod_{i=1}^{N} p_i{}^{np_i} &= \prod_{i=1}^{N} \left(2^{\log_2 p_i}\right)^{np_i} \\
&= \prod_{i=1}^{N} 2^{np_i \log_2 p_i} \\
&= 2^{n \sum_{i=1}^{N} p_i \log_2 p_i} \\
&= 2^{-(nH(X))}
\end{aligned} \tag{7.18}
$$

where $H\left(X\right)$ is the entropy of the random variables X_1, X_2,..., X_n.

For large n, almost all the output sequences of length n of the source are equally probably with probability $\simeq 2^{-(nH(X))}$. These are typical sequences. The probability of nontypical sequences are negligible. There are N^n different sequences of length n with alphabet of size N. The probability of typical sequences is almost 1.

$$\sum_{k=1}^{\text{\# of typical seq.}} 2^{-(nH(X))} = 1 \tag{7.19}$$
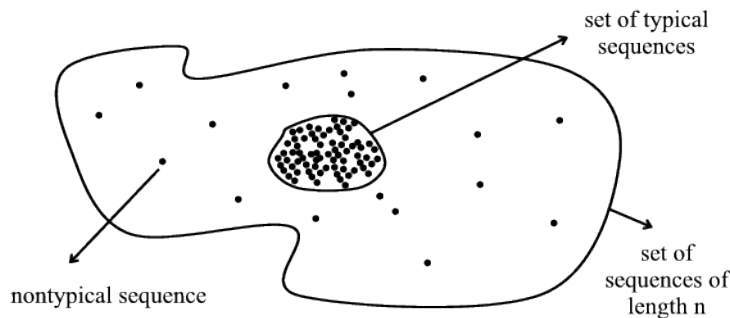
Figure 7.5

Example 7.7

Consider a source with alphabet {A,B,C,D} with probabilities $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$. Assume X_1, X_2,..., X_8 is an independent and identically distributed sequence with $X_i \in \{A, B, C, D\}$ with the above probabilities.

$$
\begin{aligned}
H(X) &= \left(-\left(\tfrac{1}{2}\log_2\tfrac{1}{2}\right)\right) - \tfrac{1}{4}\log_2\tfrac{1}{4} - \tfrac{1}{8}\log_2\tfrac{1}{8} - \tfrac{1}{8}\log_2\tfrac{1}{8} \\
&= \tfrac{1}{2} + \tfrac{2}{4} + \tfrac{3}{8} + \tfrac{3}{8} \\
&= \tfrac{4+4+6}{8} \\
&= \tfrac{14}{8}
\end{aligned}
\tag{7.20}
$$

The number of typical sequences of length 8

$$
2^{8 \times \frac{14}{8}} = 2^{14}
\tag{7.21}
$$

The number of nontypical sequences $4^8 - 2^{14} = 2^{16} - 2^{14} = 2^{14}(4-1) = 3 \times 2^{14}$

Examples of typical sequences include those with A appearing $8 \times \frac{1}{2} = 4$ times, B appearing $8 \times \frac{1}{4} = 2$ times, *etc.* {A,D,B,B,A,A,C,A}, {A,A,A,A,C,D,B,B} and much more.

Examples of nontypical sequences of length 8: {D,D,B,C,C,A,B,D}, {C,C,C,C,C,B,C,C} and much more. Indeed, these definitions and arguments are valid when n is very large. The probability of a source output to be in the set of typical sequences is 1 when $n \to \infty$. The probability of a source output to be in the set of nontypical sequences approaches 0 as $n \to \infty$.

The essence of source coding or data compression is that as $n \to \infty$, nontypical sequences never appear as the output of the source. Therefore, one only needs to be able to represent typical sequences as binary codes and ignore nontypical sequences. Since there are only $2^{nH(X)}$ typical sequences of length n, it takes $nH(X)$ bits to represent them on the average. On the average it takes $H(X)$ bits per source output to represent a simple source that produces independent and identically distributed outputs.

Theorem 7.1: Shannon's Source-Coding

A source that produced independent and identically distributed random variables with entropy H can be encoded with arbitrarily small error probability at any rate R in bits per source output if $R \geq H$. Conversely, if $R < H$, the error probability will be bounded away from zero, independent of the complexity of coder and decoder.

The source coding theorem proves existence of source coding techniques that achieve rates close to the entropy but does not provide any algorithms or ways to construct such codes.

If the source is not i.i.d. (independent and identically distributed), but it is stationary with memory, then a similar theorem applies with the entropy $H(X)$ replaced with the entropy rate $H = \lim_{n \to \infty} H(X_n | X_1 X_2 \ldots X_{n-1})$

In the case of a source with memory, the more the source produces outputs the more one knows about the source and the more one can compress.

Example 7.8
The English language has 26 letters, with space it becomes an alphabet of size 27. If modeled as a memoryless source (no dependency between letters in a word) then the entropy is $H(X) = 4.03$ bits/letter.

If the dependency between letters in a text is captured in a model the entropy rate can be derived to be $H = 1.3$ bits/letter. Note that a non-information theoretic representation of a text may require 5 bits/letter since 2^5 is the closest power of 2 to 27. Shannon's results indicate that there may be a compression algorithm with the rate of 1.3 bits/letter.

Although Shannon's results are not constructive, there are a number of source coding algorithms for discrete time discrete valued sources that come close to Shannon's bound. One such algorithm is the Huffman source coding algorithm (Section 7.4). Another is the Lempel and Ziv algorithm.

Huffman codes and Lempel and Ziv apply to compression problems where the source produces discrete time and discrete valued outputs. For cases where the source is analog there are powerful compression algorithms that specify all the steps from sampling, quantizations, and binary representation. These are referred to as waveform coders. JPEG, MPEG, vocoders are a few examples for image, video, and voice, respectively.

7.4 Huffman Coding[5]

One particular source coding (Section 7.3) algorithm is the Huffman encoding algorithm. It is a source coding algorithm which approaches, and sometimes achieves, Shannon's bound for source compression. A brief discussion of the algorithm is also given in another module[6].

7.4.1 Huffman encoding algorithm

1. Sort source outputs in decreasing order of their probabilities
2. Merge the two least-probable outputs into a single output whose probability is the sum of the corresponding probabilities.
3. If the number of remaining outputs is more than 2, then go to step 1.
4. Arbitrarily assign 0 and 1 as codewords for the two remaining outputs.
5. If an output is the result of the merger of two outputs in a preceding step, append the current codeword with a 0 and a 1 to obtain the codeword the the preceding outputs and repeat step 5. If no output is preceded by another output in a preceding step, then stop.

Example 7.9
$X \in \{A, B, C, D\}$ with probabilities $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$

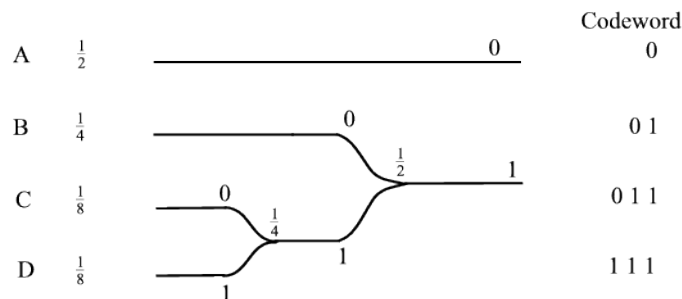|			Codeword
A	$\frac{1}{2}$	0	0
B	$\frac{1}{4}$	0	0 1
C	$\frac{1}{8}$	0	0 1 1
D	$\frac{1}{8}$	1	1 1 1

Figure 7.6

Average length $= \frac{1}{2}1 + \frac{1}{4}2 + \frac{1}{8}3 + \frac{1}{8}3 = \frac{14}{8}$. As you may recall, the entropy of the source was also $H(X) = \frac{14}{8}$. In this case, the Huffman code achieves the lower bound of $\frac{14}{8}\frac{\text{bits}}{\text{output}}$.

In general, we can define average code length as

$$\overline{\ell} = \sum_{x \in \overline{X}} p_X(x)\ell(x) \tag{7.22}$$

where $\overline{X}$ is the set of possible values of x.

It is not very hard to show that

$$H(X) \geq \overline{\ell} > H(X) + 1 \tag{7.23}$$

For compressing single source output at a time, Huffman codes provide nearly optimum code lengths.

The drawbacks of Huffman coding

1. Codes are variable length.
2. The algorithm requires the knowledge of the probabilities, $p_X(x)$ for all $x \in \overline{X}$.

Another powerful source coder that does not have the above shortcomings is Lempel and Ziv.

7.5 Channel Capacity[7]

In the previous section, we discussed information sources and quantified information. We also discussed how to represent (and compress) information sources in binary symbols in an efficient manner. In this section, we consider channels and will find out how much information can be sent through the channel reliably.

We will first consider simple channels where the input is a discrete random variable and the output is also a discrete random variable. These discrete channels could represent analog channels with modulation and demodulation and detection.

---

[7]This content is available online at <http://cnx.org/content/m10173/2.8/>.
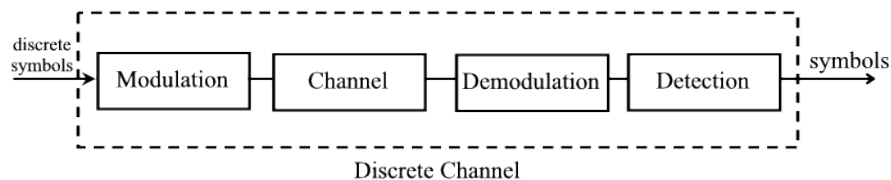
Figure 7.7

Let us denote the input sequence to the channel as

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \tag{7.24}$$

where $X_i \in \overline{X}$ a discrete symbol set or input alphabet.

The channel output

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{pmatrix} \tag{7.25}$$

where $Y_i \in \overline{Y}$ a discrete symbol set or output alphabet.

The statistical properties of a channel are determined if one finds $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ for all $y \in \overline{Y}^n$ and for all $x \in \overline{X}^n$. A discrete channel is called a **discrete memoryless channel** if

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p_{Y_i|X_i}(y_i|x_i) \tag{7.26}$$

for all $y \in \overline{Y}^n$ and for all $x \in \overline{X}^n$.

Example 7.10

A binary symmetric channel (BSC) is a discrete memoryless channel with binary input and binary output and $p_{Y|X}(y=0|x=1) = p_{Y|X}(y=1|x=0)$. As an example, a white Gaussian channel with antipodal signaling and matched filter receiver has probability of error of $Q\left(\sqrt{\frac{2E_s}{N_0}}\right)$. Since the error is symmetric with respect to the transmitted bit, then

$$\begin{aligned} p_{Y|X}(0|1) &= p_{Y|X}(1|0) \\ &= Q\left(\sqrt{\frac{2E_s}{N_0}}\right) \\ &= \epsilon \end{aligned} \tag{7.27}$$
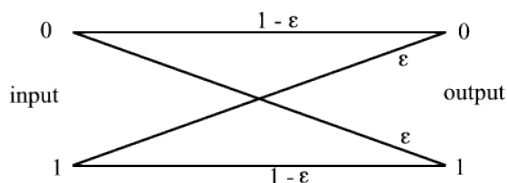
Figure 7.8

It is interesting to note that every time a BSC is used one bit is sent across the channel with probability of error of ϵ. The question is how much information or how many bits can be sent per channel use, reliably. Before we consider the above question a few definitions are essential. These are discussed in mutual information (Section 7.6).

7.6 Mutual Information[8]

Recall that

$$H(X,Y) = -\sum_{xx}\sum_{yy} p_{X,Y}(x,y)\log p_{X,Y}(x,y) \tag{7.28}$$

$$H(Y) + H(X|Y) = H(X) + H(Y|X) \tag{7.29}$$

Definition 7.5: Mutual Information
The mutual information between two discrete random variables is denoted by $\mathcal{I}(X;Y)$ and defined as

$$\mathcal{I}(X;Y) = H(X) - H(X|Y) \tag{7.30}$$

Mutual information is a useful concept to measure the amount of information shared between input and output of noisy channels.

In our previous discussions it became clear that when the channel is noisy there may not be reliable communications. Therefore, the limiting factor could very well be reliability when one considers noisy channels. Claude E. Shannon in 1948 changed this paradigm and stated a theorem that presents the rate (speed of communication) as the limiting factor as opposed to reliability.

Example 7.11
Consider a discrete memoryless channel with four possible inputs and outputs.

---

[8]This content is available online at <http://cnx.org/content/m10178/2.9/>.
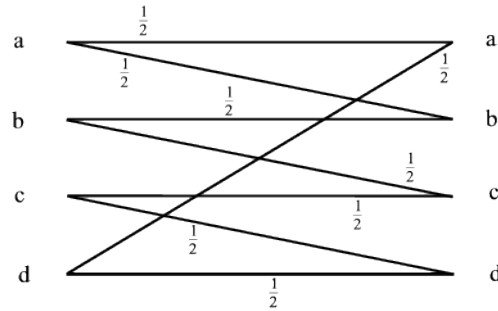
Figure 7.9

Every time the channel is used, one of the four symbols will be transmitted. Therefore, 2 bits are sent per channel use. The system, however, is very unreliable. For example, if "a" is received, the receiver can not determine, reliably, if "a" was transmitted or "d". However, if the transmitter and receiver agree to only use symbols "a" and "c" and never use "b" and "d", then the transmission will always be reliable, but 1 bit is sent per channel use. Therefore, the rate of transmission was the limiting factor and not reliability.

This is the essence of Shannon's noisy channel coding theorem, *i.e.*, using only those inputs whose corresponding outputs are disjoint (*e.g.*, far apart). The concept is appealing, but does not seem possible with binary channels since the input is either zero or one. It may work if one considers a vector of binary inputs referred to as the extension channel.

$$\text{X input vector} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \in \overline{X}^n = \{0,1\}^n$$

$$\text{Y output vector} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \in \overline{Y}^n = \{0,1\}^n$$
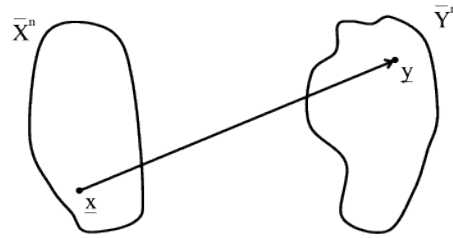
Figure 7.10

This module provides a description of the basic information necessary to understand Shannon's Noisy Channel Coding Theorem (Section 7.8). However, for additional information on typical sequences, please refer to Typical Sequences (Section 7.7).

7.7 Typical Sequences[9]

If the binary symmetric channel has crossover probability ϵ then if x is transmitted then by the Law of Large Numbers the output y is different from x in $n\epsilon$ places if n is very large.

$$d_H(x,y) \simeq n\epsilon \tag{7.31}$$

The number of sequences of length n that are different from x of length n at $n\epsilon$ is

$$\left(\begin{array}{c} n \\ n\epsilon \end{array} \right) = \frac{n!}{(n\epsilon)!\,(n - n\epsilon)!} \tag{7.32}$$

> **Example 7.12**
> $x = (0,0,0)^T$ and $\epsilon = \frac{1}{3}$ and $n\epsilon = 3 \times \frac{1}{3}$. The number of output sequences different from x by one element: $\frac{3!}{1!2!} = \frac{3 \times 2 \times 1}{1 \times 2} = 3$ given by $(1,0,1)^T$, $(0,1,1)^T$, and $(0,0,0)^T$.

Using Stirling's approximation

$$n! \simeq n^n e^{-n} \sqrt{2\pi n} \tag{7.33}$$

we can approximate

$$\left(\begin{array}{c} n \\ n\epsilon \end{array} \right) \simeq 2^{n((-(\epsilon \log_2 \epsilon)) - (1-\epsilon)\log_2(1-\epsilon))} = 2^{nH_b(\epsilon)} \tag{7.34}$$

where $H_b(\epsilon) \equiv (-(\epsilon \log_2 \epsilon)) - (1-\epsilon)\log_2(1-\epsilon)$ is the entropy of a binary memoryless source. For any x there are $2^{nH_b(\epsilon)}$ highly probable outputs that correspond to this input.

Consider the output vector Y as a very long random vector with entropy $nH(Y)$. As discussed earlier (Example 7.1), the number of typical sequences (or highly probably) is roughly $2^{nH(Y)}$. Therefore, 2^n is the total number of binary sequences, $2^{nH(Y)}$ is the number of typical sequences, and $2^{nH_b(\epsilon)}$ is the number of elements in a group of possible outputs for one input vector. The maximum number of input sequences that produce nonoverlapping output sequences

$$\begin{aligned} M &= \frac{2^{nH(Y)}}{2^{nH_b(\epsilon)}} \\ &= 2^{n(H(Y) - H_b(\epsilon))} \end{aligned} \tag{7.35}$$
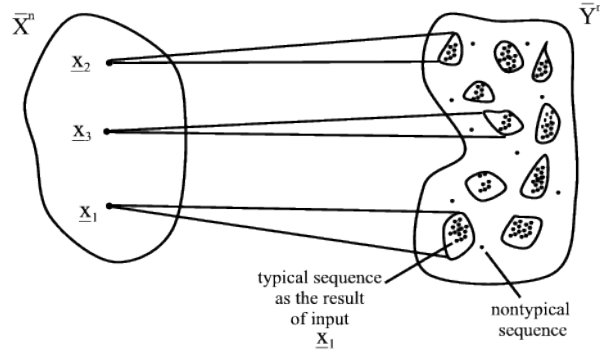

Figure 7.11

The number of distinguishable input sequences of length n is

$$2^{n(H(Y)-H_b(\epsilon))} \tag{7.36}$$

The number of information bits that can be sent across the channel reliably per n channel uses $n(H(Y) - H_b(\epsilon))$ The maximum reliable transmission rate per channel use

$$
\begin{aligned}
R &= \frac{\log_2 M}{n} \\
&= \frac{n(H(Y)-H_b(\epsilon))}{n} \\
&= H(Y) - H_b(\epsilon)
\end{aligned} \tag{7.37}
$$

The maximum rate can be increased by increasing $H(Y)$. Note that $H_b(\epsilon)$ is only a function of the crossover probability and can not be minimized any further.

The entropy of the channel output is the entropy of a binary random variable. If the input is chosen to be uniformly distributed with $p_X(0) = p_X(1) = \frac{1}{2}$.

Then

$$
\begin{aligned}
p_Y(0) &= 1 p_X(0) + \epsilon p_X(1) \\
&= \tfrac{1}{2}
\end{aligned} \tag{7.38}
$$

and

$$
\begin{aligned}
p_Y(1) &= 1 p_X(1) + \epsilon p_X(0) \\
&= \tfrac{1}{2}
\end{aligned} \tag{7.39}
$$

Then, $H(Y)$ takes its maximum value of 1. Resulting in a maximum rate $R = 1 - H_b(\epsilon)$ when $p_X(0) = p_X(1) = \frac{1}{2}$. This result says that ordinarily one bit is transmitted across a BSC with reliability $1 - \epsilon$. If one needs to have probability of error to reach zero then one should reduce transmission of information to $1 - H_b(\epsilon)$ and add redundancy.

Recall that for Binary Symmetric Channels (BSC)

$$
\begin{aligned}
H\left(Y|X\right) &= p_x\left(0\right)H\left(Y|X=0\right) + p_x\left(1\right)H\left(Y|X=1\right) \\
&= p_x\left(0\right)\left(-\left(\left(1-\epsilon\right)\log_2\left(1-\epsilon\right) - \epsilon\log_2\epsilon\right)\right) + p_x\left(1\right)\left(-\left(\left(1-\epsilon\right)\log_2\left(1-\epsilon\right) - \epsilon\log_2\epsilon\right)\right) \\
&= \left(-\left(\left(1-\epsilon\right)\log_2\left(1-\epsilon\right)\right)\right) - \epsilon\log_2\epsilon \\
&= H_b\left(\epsilon\right)
\end{aligned}
\tag{7.40}
$$

Therefore, the maximum rate indeed was

$$
\begin{aligned}
R &= H\left(Y\right) - H\left(Y|X\right) \\
&= \mathcal{I}\left(X;Y\right)
\end{aligned}
\tag{7.41}
$$

Example 7.13

The maximum reliable rate for a BSC is $1 - H_b\left(\epsilon\right)$. The rate is 1 when $\epsilon = 0$ or $\epsilon = 1$. The rate is 0 when $\epsilon = \frac{1}{2}$
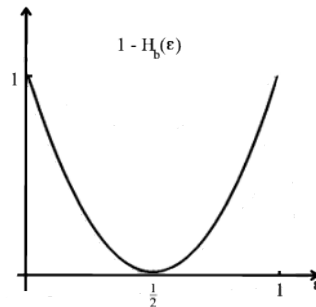
Figure 7.12

This module provides background information necessary for an understanding of Shannon's Noisy Channel Coding Theorem (Section 7.8). It is also closely related to material presented in Mutual Information (Section 7.6).

7.8 Shannon's Noisy Channel Coding Theorem[10]

It is highly recommended that the information presented in Mutual Information (Section 7.6) and in Typical Sequences (Section 7.7) be reviewed before proceeding with this document. An introductory module on the theorem is available at Noisy Channel Theorems [11].

Theorem 7.2: Shannon's Noisy Channel Coding
The capacity of a discrete-memoryless channel is given by

$$
C = max_{p_{X}x}\left\{\mathcal{I}\left(X;Y\right) \mid p_X\left(x\right)\right\}
\tag{7.42}
$$

where $\mathcal{I}\left(X;Y\right)$ is the mutual information between the channel input X and the output Y. If the transmission rate R is less than C, then for any $\epsilon > 0$ there exists a code with block length n large

---

[10]This content is available online at <http://cnx.org/content/m10180/2.10/>.
[11]"Noisy Channel Coding Theorem" <http://cnx.org/content/m0073/latest/>

enough whose error probability is less than ϵ. If $R > C$, the error probability of any code with any block length is bounded away from zero.

Example 7.14

If we have a binary symmetric channel with cross over probability 0.1, then the capacity $C \simeq 0.5$ bits per transmission. Therefore, it is possible to send 0.4 bits per channel through the channel reliably. This means that we can take 400 information bits and map them into a code of length 1000 bits. Then the whole code can be transmitted over the channels. One hundred of those bits may be detected incorrectly but the 400 information bits may be decoded correctly.

Before we consider continuous-time additive white Gaussian channels, let's concentrate on discrete-time Gaussian channels

$$Y_i = X_i + \eta_i \tag{7.43}$$

where the X_i's are information bearing random variables and η_i is a Gaussian random variable with variance σ_η^2. The input X_i's are constrained to have power less than P

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \le P \tag{7.44}$$

Consider an output block of size n

$$Y = X + \eta \tag{7.45}$$

For large n, by the Law of Large Numbers,

$$\frac{1}{n}\sum_{i=1}^{n} \eta_i^2 = \frac{1}{n}\sum_{i=1}^{n} \left(|y_i - x_i|\right)^2 \le \sigma_\eta^2 \tag{7.46}$$

This indicates that with large probability as n approaches infinity, Y will be located in an n-dimensional sphere of radius $\sqrt{n\sigma_\eta^2}$ centered about X since $\left(|y - x|\right)^2 \le n\sigma_\eta^2$

On the other hand since X_i's are power constrained and η_i and X_i's are independent

$$\frac{1}{n}\sum_{i=1}^{n} y_i^2 \le P + \sigma_\eta^2 \tag{7.47}$$

$$|Y| \le n\left(P + \sigma_\eta^2\right) \tag{7.48}$$

This mean Y is in a sphere of radius $\sqrt{n\left(P + \sigma_\eta^2\right)}$ centered around the origin.

How many X's can we transmit to have nonoverlapping Y spheres in the output domain? The question is how many spheres of radius $\sqrt{n\sigma_\eta^2}$ fit in a sphere of radius $\sqrt{n\left(P + \sigma_\eta^2\right)}$.

$$\begin{aligned} M &= \frac{\left(\sqrt{n(\sigma_\eta^2 + P)}\right)^n}{\left(\sqrt{n\sigma_\eta^2}\right)^n} \\ &= \left(1 + \frac{P}{\sigma_\eta^2}\right)^{\frac{n}{2}} \end{aligned} \tag{7.49}$$
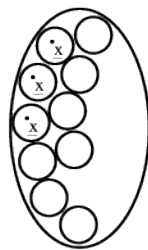
Figure 7.13

How many bits of information can one send in n uses of the channel?

The capacity of a discrete-time Gaussian channel $C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_\eta^2}\right)$ bits per channel use.

When the channel is a continuous-time, bandlimited, additive white Gaussian with noise power spectral density $\frac{N_0}{2}$ and input power constraint P and bandwidth W. The system can be sampled at the Nyquist rate to provide power per sample P and noise power

$$
\begin{aligned}
\sigma_\eta^2 &= \int_{-W}^{W} \frac{N_0}{2} df \\
&= W N_0
\end{aligned}
\tag{7.50}
$$

The channel capacity $\frac{1}{2}\log_2\left(1 + \frac{P}{N_0 W}\right)$ bits per transmission. Since the sampling rate is $2W$, then

$$
C = \frac{2W}{2}\log_2\left(1 + \frac{P}{N_0 W}\right) \text{ bits/trans. x trans./sec}
\tag{7.51}
$$

$$
C = W\log_2\left(1 + \frac{P}{N_0 W}\right)\frac{\text{bits}}{\text{sec}}
\tag{7.52}
$$

Example 7.15

The capacity of the voice band of a telephone channel can be determined using the Gaussian model. The bandwidth is 3000 Hz and the signal to noise ratio is often 30 dB. Therefore,

$$
C = 3000\log_2\left(1 + 1000\right) \simeq 30000\frac{\text{bits}}{\text{sec}}
\tag{7.53}
$$

One should not expect to design modems faster than 30 Kbs using this model of telephone channels. It is also interesting to note that since the signal to noise ratio is large, we are expecting to transmit 10 bits/second/Hertz across telephone channels.

7.9 Channel Coding[12]

Channel coding is a viable method to reduce information rate through the channel and increase reliability. This goal is achieved by adding redundancy to the information symbol vector resulting in a longer coded

---
[12]This content is available online at <http://cnx.org/content/m10174/2.11/>.

vector of symbols that are distinguishable at the output of the channel. Another brief explanation of channel coding is offered in Channel Coding and the Repetition Code[13]. We consider only two classes of codes, block codes (Section 7.9.1: Block codes) and convolutional codes (Section 7.10).

7.9.1 Block codes

The information sequence is divided into blocks of length k. Each block is mapped into channel inputs of length n. The mapping is independent from previous blocks, that is, there is no memory from one block to another.

Example 7.16
$k = 2$ and $n = 5$

$$00 \rightarrow 00000 \tag{7.54}$$

$$01 \rightarrow 10100 \tag{7.55}$$

$$10 \rightarrow 01111 \tag{7.56}$$

$$11 \rightarrow 11011 \tag{7.57}$$

information sequence $\Rightarrow$ codeword (channel input)

A binary block code is completely defined by 2^k binary sequences of length n called codewords.

$$C = \{c_1, c_2, \ldots, c_{2^k}\} \tag{7.58}$$

$$c_i \in \{0, 1\}^n \tag{7.59}$$

There are three key questions,

1. How can one find "good" codewords?
2. How can one systematically map information sequences into codewords?
3. How can one systematically find the corresponding information sequences from a codeword, *i.e.*, how can we decode?

These can be done if we concentrate on linear codes and utilize finite field algebra.

A block code is linear if $c_i \in C$ and $c_j \in C$ implies $c_i \oplus c_j \in C$ where $\oplus$ is an elementwise modulo 2 addition.

Hamming distance is a useful measure of codeword properties

$$d_H(c_i, c_j) = \# \text{ of places that they are different} \tag{7.60}$$

Denote the codeword for information sequence $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ by g_1 and $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$ by $g_2, \ldots$, and $e_k =$

[13]"Repetition Codes" <http://cnx.org/content/m0071/latest/>

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$ by g_k. Then any information sequence can be expressed as

$$
\begin{aligned}
u &= \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix} \\
&= \sum_{i=1}^{k} u_i e_i
\end{aligned}
\tag{7.61}
$$

and the corresponding codeword could be

$$c = \sum_{i=1}^{k} u_i g_i \tag{7.62}$$

Therefore

$$c = uG \tag{7.63}$$

with $c = \{0,1\}^n$ and $u \in \{0,1\}^k$ where $G = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{pmatrix}$, a kxn matrix and all operations are modulo 2.

Example 7.17
In Example 7.16 with

$$00 \rightarrow 00000 \tag{7.64}$$

$$01 \rightarrow 10100 \tag{7.65}$$

$$10 \rightarrow 01111 \tag{7.66}$$

$$11 \rightarrow 11011 \tag{7.67}$$

$g_1 = (0,1,1,1,1)^T$ and $g_2 = (1,0,1,0,0)^T$ and $G = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}$

Additional information about coding efficiency and error are provided in Block Channel Coding[14].

Examples of good linear codes include Hamming codes, BCH codes, Reed-Solomon codes, and many more. The rate of these codes is defined as $\frac{k}{n}$ and these codes have different error correction and error detection properties.

---

[14]"Block Channel Coding" $<$http://cnx.org/content/m0094/latest/$>$

7.10 Convolutional Codes[15]

Convolutional codes are one type of code used for channel coding (Section 7.9). Another type of code used is block coding (Section 7.9.1: Block codes).

7.10.1 Convolutional codes

In convolutional codes, each block of k bits is mapped into a block of n bits but these n bits are not only determined by the present k information bits but also by the previous information bits. This dependence can be captured by a finite state machine.

Example 7.18
A rate $\frac{1}{2}$ convolutional coder $k = 1$, $n = 2$ with memory length 2 and constraint length 3.
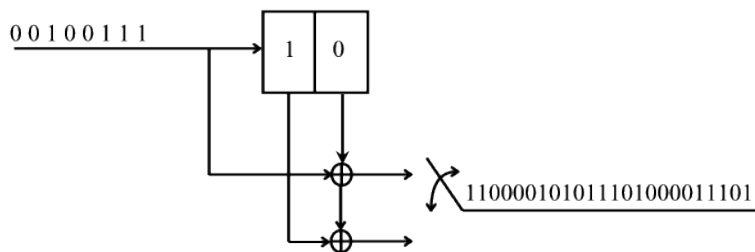
Figure 7.14

Since the length of the shift register is 2, there are 4 different rates. The behavior of the convolutional coder can be captured by a 4 state machine. States: 00, 01, 10, 11,

For example, arrival of information bit 0 transitions from state 10 to state 01.

The encoding and the decoding process can be realized in trellis structure.

---

[15]This content is available online at <http://cnx.org/content/m10181/2.7/>.
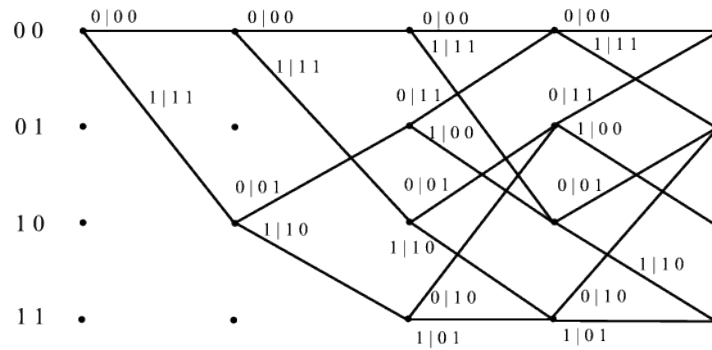
Figure 7.15

If the input sequence is

    1 1 0 0

the output sequence would be

    11 10 10 11

The transmitted codeword is then 11 10 10 11. If there is one error on the channel 11 00 10 11
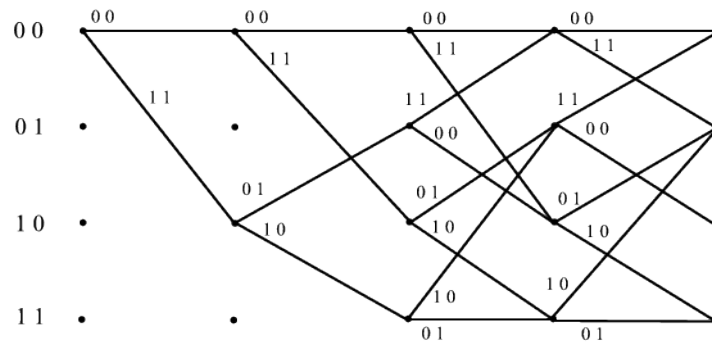
Figure 7.16

Starting from state 00 the Hamming distance between the possible paths and the received sequence is measured. At the end, the path with minimum distance to the received sequence is chosen as the correct trellis path. The information sequence will then be determined.

Convolutional coding lends itself to very efficient trellis based encoding and decoding. They are very practical and powerful codes.

Solutions to Exercises in Chapter 7

Solution to Exercise 7.1 (p. 73)

$$\log_2\left(1+\frac{P}{\sigma_\eta{}^2}\right)^{\frac{n}{2}} \tag{7.68}$$

Chapter 8

Homework

Chapter 9

Homework 1 of Elec 430[1]

Elec 430 homework set 1. Rice University Department of Electrical and Computer Engineering.

Exercise 9.1

The current I in a semiconductor diode is related to the voltage V by the relation $I = e^V - 1$. If V is a random variable with density function $f_V(x) = \frac{1}{2}e^{-|x|}$ for $-\infty < x < \infty$, find $f_I(y)$; the density function of I.

Exercise 9.2

9.1

Show that if $AB = \{\}$ then $Pr[A] \leq Pr[B^c]$

9.2

Show that for any A, B, C we have $Pr[A \cup B \cup C] = Pr[A] + Pr[B] + Pr[C] - Pr[A \cap B] - Pr[A \cap C] - Pr[B \cap C] + Pr[A \cap B \cap C]$

9.3

Show that if A and B are independent the $Pr[A \cap B^c] = Pr[A] Pr[B^c]$ which means A and B^c are also independent.

Exercise 9.3

Suppose X is a discrete random variable taking values $\{0, 1, 2, \ldots, n\}$ with the following probability mass function $p_X(k) = \begin{cases} \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k} & \text{if } k = \{0, 1, 2, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$ with parameter $\theta \in [0, 1]$

9.1

Find the characteristic function of X.

9.2

Find $\bar{X}$ and σ_X^2

NOTE: See problems 3.14 and 3.15 in Proakis and Salehi

Exercise 9.4

Consider outcomes of a fair dice $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Define events $A = \{\omega, \omega \mid$ an even number appears$\}$ and $B = \{\omega, \omega \mid$ a number less than 5 appears$\}$. Are these events disjoint? Are they independent? (Show your work!)

Exercise 9.5

This is problem 3.5 in Proakis and Salehi.

An information source produces 0 and 1 with probabilities 0.3 and 0.7, respectively. The output of the source is transmitted via a channel that has a probability of error (turning a 1 into a 0 or a 0 into a 1) equal to 0.2.

9.1

What is the probability that at the output a 1 is observed?

9.2

What is the probability that a 1 was the output of the source if at the output of the channel a 1 is observed?

Exercise 9.6

Suppose X and Y are each Gaussian random variables with means μ_X and μ_Y and variances σ_X^2 and σ_Y^2. Assume that they are also independent. Show that $Z = X + Y$ is also Gaussian. Find the mean and variance of Z.

Chapter 10

Homework 2 of Elec 430[1]

Elec 430 homework set 2. Rice University Department of Electrical and Computer Engineering.

10.1 Problem 1

Suppose A and B are two Gaussian random variables each zero mean with $\overline{A^2} < \infty$ and $\overline{B^2} < \infty$. The correlation between them is denoted by $\overline{AB}$. Define the random process $X_t = A + Bt$ and $Y_t = B + At$.

- a) Find the mean, autocorrelation, and crosscorrelation functions of X_t and Y_t.
- b) Find the 1st order density of X_t, $f_{X_t}(x)$
- c) Find the conditional density of X_{t_2} given X_{t_1}, $f_{X_{t_2}|X_{t_1}}(x_2|x_1)$. Assume $t_2 > t_1$

    NOTE: see Proakis and Salehi problem 3.28

- d) Is X_t wide sense stationary?

10.2 Problem 2

Show that if X_t is second-order stationary, then it is also first-order stationary.

10.3 Problem 3

Let a stochastic process X_t be defined by $X_t = \cos(\Omega t + \Theta)$ where Ω and Θ are statistically independent random variables. Θ is uniformaly distributed over $[-\pi, \pi]$ and Ω has an unknown density $f_\Omega(\omega)$.

- a) Compute the expected value of X_t.
- b) Find an expression for the correlation function of X_t.
- c) Is X_t wide sense stationary? Show your reasoning.
- d) Find the first-order density function $f_{X_t}(x)$.

---

[1]This content is available online at <http://cnx.org/content/m10374/2.4/>.

Chapter 11

Homework 5 of Elec 430[1]

11.1 Problem 1

Consider a ternary communication system where the source produces three possible symbols: 0, 1, 2.

a) Assign three modulation signals $s_1(t)$, $s_2(t)$, and $s_3(t)$ defined on $t \in [0, T]$ to these symbols, 0, 1, and 2, respectively. Make sure that these signals are not orthogonal and assume that the symbols have an equal probability of being generated.

b) Consider an orthonormal basis $\psi_1(t)$, $\psi_2(t)$, ..., $\psi_N(t)$ to represent these three signals. Obviously N could be either 1, 2, or 3.
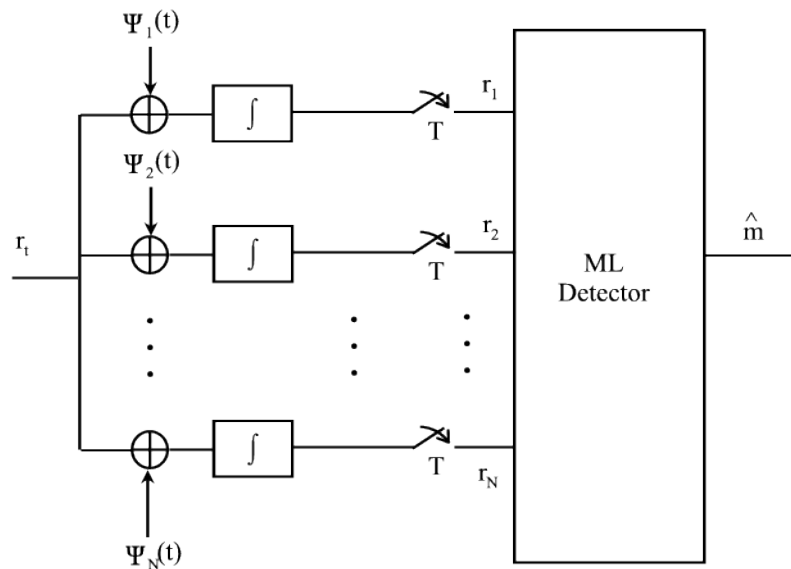
Figure 11.1

[1]This content is available online at <http://cnx.org/content/m10450/2.5/>.

Now consider two different receivers to decide which one of the symbols were transmitted when $r_t = s_m(t) + N_t$ is received where $m = \{1, 2, 3\}$ and N_t is a zero mean white Gaussian process with $S_N(f) = \frac{N_0}{2}$ for all f. What is $f_{\mathbf{r}|s_m(t)}$ and what is $f_{\mathbf{Y}|s_m(t)}$?
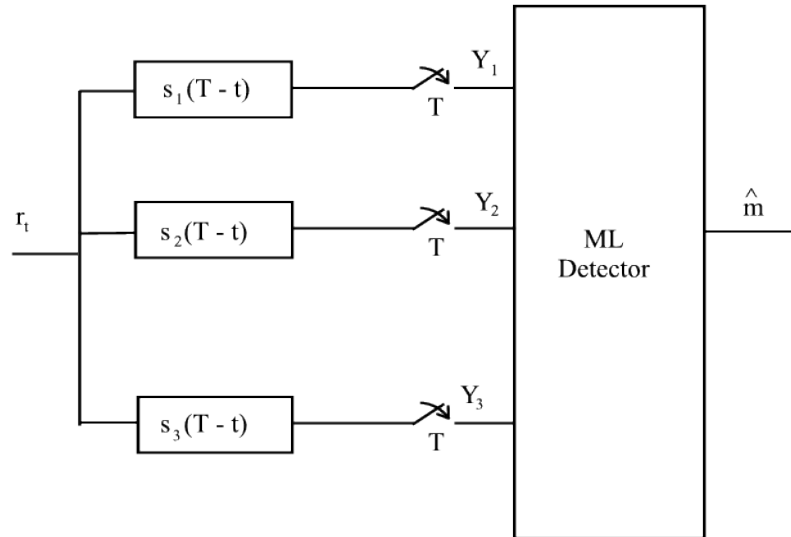
Figure 11.2

Find the probability that $\hat{m} \neq m$ for both receivers. $P_e = Pr\left[\hat{m} \neq m\right]$.

11.2 Problem 2

Proakis and Salehi problems 7.18, 7.26, and 7.32

11.3 Problem 3

Suppose our modulation signals are $s_1(t)$ and $s_2(t)$ where $s_1(t) = e^{-t^2}$ for all t and $s_2(t) = -s_1(t)$. The channel noise is AWGN with zero mean and spectral height $\frac{N_0}{2}$. The signals are transmitted equally likely.
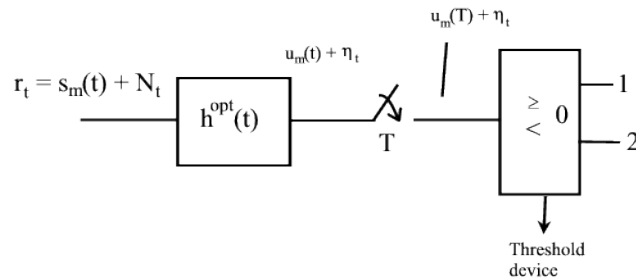
Figure 11.3

Find the impulse response of the optimum filter. Find the signal component of the output of the matched filter at $t = T$ where $s_1(t)$ is transmitted; *i.e.*, $u_1(t)$. Find the probability of error $Pr\left[\hat{m} \neq m\right]$.

In this part, assume that the power spectral density of the noise is not flat and in fact is

$$S_N(f) = \frac{1}{(2\pi f)^2 + \alpha^2} \tag{11.1}$$

for all f, where α is real and positive. Can you show that the optimum filter in this case is a cascade of two filters, one to whiten the noise and one to match to the signal at the output of the whitening filter?
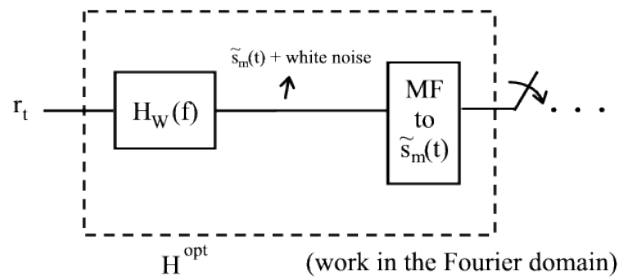
Figure 11.4

c) Find an expression for the probability of error.

Chapter 12

Homework 3 of Elec 430[1]

Exercise 12.1

Suppose that a white Gaussian noise X_t is input to a linear system with transfer function given by

$$H(f) = \begin{cases} 1 & \text{if } |f| \le 2 \\ 0 & \text{if } |f| > 2 \end{cases} \quad (12.1)$$

Suppose further that the input process is zero mean and has spectral height $\frac{N_0}{2} = 5$. Let Y_t denote the resulting output process.

1. Find the power spectral density of Y_t. Find the autocorrelation of Y_t (i.e., $R_Y(\tau)$).
2. Form a discrete-time process (that is a sequence of random variables) by sampling Y_t at time instants T seconds apart. Find a value for T such that these samples are uncorrelated. Are these samples also independent?
3. What is the variance of each sample of the output process?
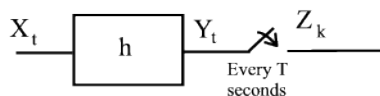
$$Z_k = Y_{kT} \text{ for } k = \dots -1, 0, 1, 2, \dots$$

Figure 12.1

Exercise 12.2

Suppose that X_t is a zero mean white Gaussian process with spectral height $\frac{N_0}{2} = 5$. Denote Y_t as the output of an integrator when the input is Y_t.

---

[1] This content is available online at <http://cnx.org/content/m10375/2.8/>.
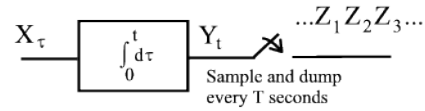
Figure 12.2

1. Find the mean function of Y_t. Find the autocorrelation function of Y_t, $R_Y(t + \tau, t)$
2. Let Z_k be a sequence of random variables that have been obtained by sampling Y_t at every T seconds and dumping the samples, that is

$$Z_k = \int_{(k-1)T}^{kT} X_\tau d\tau \tag{12.2}$$

   Find the autocorrelation of the discrete-time processes Z_k's, that is, $R_Z(k + m, k) = E(Z_{k+m}Z_k)$
3. Is Z_k a wide sense stationary process?

Exercise 12.3
Proakis and Salehi, problem 3.63, parts 1, 3, and 4

Exercise 12.4
Proakis and Salehi, problem 3.54

Exercise 12.5
Proakis and Salehi, problem 3.62

Chapter 13

Exercises on Systems and Density[1]

Exercise 13.1
Consider the following system
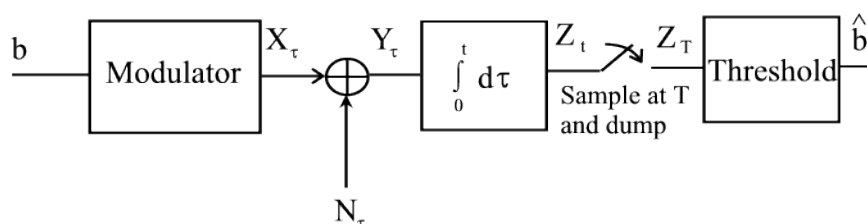
Figure 13.1

Assume that N_τ is a white Gaussian process with zero mean and spectral height $\frac{N_0}{2}$.

If b is "0" then $X_\tau = Ap_T(\tau)$ and if b is "1" then $X_\tau = (-A)p_T(\tau)$ where $p_T(\tau) = \begin{cases} 1 & \text{if } 0 \le \tau \le T \\ 0 & \text{otherwise} \end{cases}$. Suppose $Pr[b=1] = Pr[b=0] = 1/2$.

1. Find the probability density function Z_T when bit "0" is transmitted and also when bit "1" is transmitted. Refer to these two densities as $f_{Z_T, H_0}(z)$ and $f_{Z_T, H_1}(z)$, where H_0 denotes the hypothesis that bit "0" is transmitted and H_1 denotes the hypothesis that bit "1" is transmitted.

2. Consider the ratio of the above two densities; *i.e.*,

$$\Lambda(z) = \frac{f_{Z_T, H_0}(z)}{f_{Z_T, H_1}(z)} \tag{13.1}$$

and its natural log $\ln(\Lambda(z))$. A reasonable scheme to decide which bit was actually transmitted is to compare $\ln(\Lambda(z))$ to a fixed threshold γ. ($\Lambda(z)$ is often referred to as the likelihood

function and $\ln\left(\Lambda\left(z\right)\right)$ as the log likelihood function). Given threshold γ is used to decide $\hat{b}=0$ when $\ln\left(\Lambda\left(z\right)\right) \geq \gamma$ then find $Pr\left[\hat{b}\neq b\right]$ (note that we will say $\hat{b}=1$ when $\ln\left(\Lambda\left(z\right)\right) < \gamma$).

3. Find a γ that minimizes $Pr\left[\hat{b}\neq b\right]$.

Exercise 13.2
Proakis and Salehi, problems 7.7, 7.17, and 7.19

Exercise 13.3
Proakis and Salehi, problem 7.20, 7.28, and 7.23

Chapter 14

Homework 6[1]

Homework set 6 of ELEC 430, Rice University, Department of Electrical and Computer Engineering

14.1 Problem 1

Consider the following modulation system

$$s_0(t) = AP_T(t) - 1 \tag{14.1}$$

and

$$s_1(t) = (-(AP_T(t))) - 1 \tag{14.2}$$

for $0 \leq t \leq T$ where $P_T(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq T \\ 0 & \text{otherwise} \end{cases}$
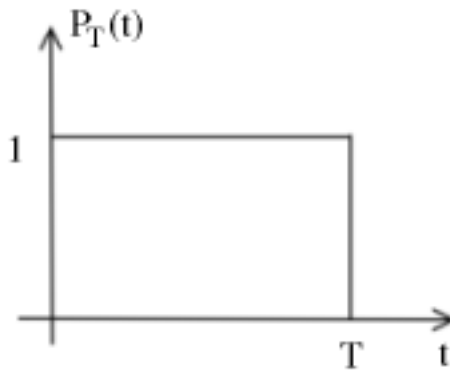
Figure 14.1

The channel is ideal with Gaussian noise which is $\mu_N(t) = 1$ for all t, wide sense stationary with $R_N(\tau) = b^2 e^{-|\tau|}$ for all $\tau \in \mathbb{R}$. Consider the following receiver structure

$$\boxed{r_t = s_m(t) + N_t}$$
(a)

(b)
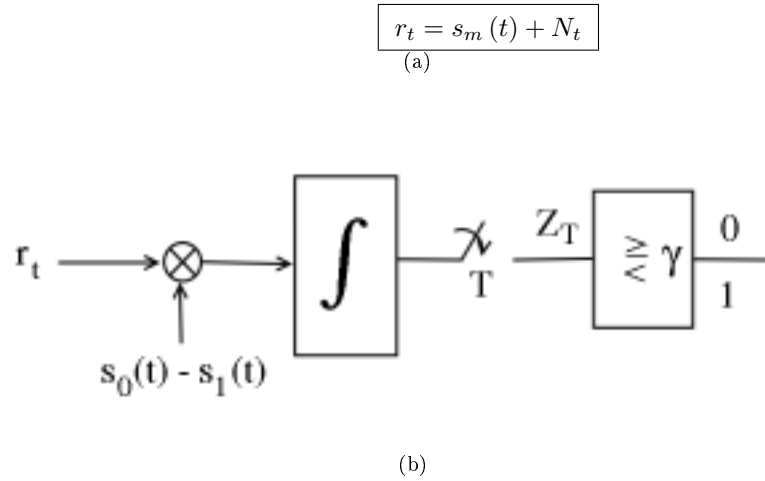
Figure 14.2

- a) Find the optimum value of the threshold for the system (e.g., γ that minimizes the P_e). Assume that $\pi_0 = \pi_1$
- b) Find the error probability when this threshold is used.

14.2 Problem 2

Consider a PAM system where symbols a_1, a_2, a_3, a_4 are transmitted where $a_n \in \{2A, A, -A, -(2A)\}$. The transmitted signal is

$$X_t = \sum_{n=1}^{4} a_n s(t - nT) \tag{14.3}$$

where $s(t)$ is a rectangular pulse of duration T and height of 1. Assume that we have a channel with impulse response $g(t)$ which is a rectangular pulse of duration T and height 1, with white Gaussian noise with $S_N(f) = \frac{N_0}{2}$ for all f.

- a) Draw a typical sample path (realization) of X_t and of the received signal r_t (do not forget to add a bit of noise!)
- b) Assume that the receiver knows $g(t)$. Design a matched filter for this transmission system.
- c) Draw a typical sample path of Y_t, the output of the matched filter (do not forget to add a bit of noise!)
- d) Find an expression (or draw) $u(nT)$ where $u(t) = s * g * h^{\mathrm{opt}}(t)$.

14.3 Problem 3

Proakis and Salehi, problem 7.35

14.4 Problem 4

Proakis and Salehi, problem 7.39

Chapter 15

Homework 7[1]

Exercise 15.1
Consider an On-Off Keying system where $s_1(t) = A\cos(2\pi f_c t + \theta)$ for $0 \le t \le T$ and $s_2(t) = 0$ for $0 \le t \le T$. The channel is ideal AWGN with zero mean and spectral height $\frac{N_0}{2}$.

1. Assume θ is known at the receiver. What is the average probability of bit-error using an optimum receiver?

2. Assume that we estimate the receiver phase to be $\hat{\theta}$ and that $\hat{\theta} \ne \theta$. Analyze the performance of the matched filter with the wrong phase, that is, examine $\overline{P_e}$ as a function of the phase error.

3. When does noncoherent become preferable? (You can find an expression for the $\overline{P_e}$ of non-coherent receivers for OOK in your textbook.) That is, how big should the phase error be before you would switch to noncoherent?

Exercise 15.2
Proakis and Salehi, Problems 9.4 and 9.14

Exercise 15.3
A **coherent** phase-shift keyed system operating over an AWGN channel with two sided power spectral density $\frac{N_0}{2}$ uses $s_0(t) = A p_T(t) \cos(\omega_c t + \theta_0)$ and $s_1(t) = A p_T(t) \cos(\omega_c t + \theta_1)$ where $|\theta_i| \le \frac{\pi}{3}$, $i \in \{0, 1\}$, are constants and that $f_c T = $ integer with $\omega_c = 2\pi f_c$.

1. Suppose θ_0 and θ_1 are **known** constants and that the optimum receiver uses filters matched to $s_0(t)$ and $s_1(t)$. What are the values of P_{e0} and P_{e1}?

2. Suppose θ_0 and θ_1 are **unknown** constants and that the receiver filters are matched to
$\hat{s}_0(t) = A p_T(t) \cos(\omega_c t)$ and $\hat{s}_1(t) = A p_T(t) \cos(\omega_c t + \pi)$ and the threshold is zero.

NOTE: Use a correlation receiver structure.

What are P_{e0} and P_{e1} now? What are the minimum values of P_{e0} and P_{e1} (as a function of θ_0 and θ_1)?

Chapter 16

Homework 8[1]

Exercise 16.1
 Proakis and Salehi, Problems 9.15 and 9.16

Exercise 16.2
 Proakis and Salehi, Problem 9.21

Exercise 16.3
 Proakis and Salehi, Problems 4.1, 4.2, and 4.3

Exercise 16.4
 Proakis and Salehi, Problems 4.5 and 4.6

---

[1]This content is available online at <http://cnx.org/content/m10453/2.3/>.

Chapter 17

Homework 9[1]

Exercise 17.1
Proakis and Salehi, Problems 4.22 and 4.28

Exercise 17.2
Proakis and Salehi, Problems 4.21 and 4.25

Exercise 17.3
Proakis and Salehi, Problems 10.1 and 10.6

Exercise 17.4
Proakis and Salehi, Problems 10.8 and 10.9

Exercise 17.5
For this problem of the homework, please either make up a problem relevant to chapters 6 or 7 of the notes or find one from your text book or other books on Digital Communication, state the problem clearly and carefully and then solve.

NOTE: If you would like to choose one from your textbook, please reserve your problem on the white board in my office. (You may not pick a problem that has already been reserved.)

Please write the problem and its solution on separate pieces of paper so that I can easily reproduce and distribute them to others in the class.

---

[1]This content is available online at <http://cnx.org/content/m10455/2.3/>.

Glossary

A antipodal

Signals $s_1(t)$ and $s_2(t)$ are antipodal if $s_2(t) = -s_1(t)$, $t \in [0, T]$

Autocorrelation

The autocorrelation function of the random process X_t is defined as

$$
\begin{aligned}
R_X(t_2, t_1) &= E\left[X_{t_2} X_{t_1}{}^*\right] \\
&= \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 x_1{}^* f_{X_{t_2}, X_{t_1}}(x_2, x_1) \, dx_1 dx_2 & \text{if continuous} \\ \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x_l x_k{}^* p_{X_{t_2}, X_{t_1}}(x_l, x_k) & \text{if discrete} \end{cases}
\end{aligned} \tag{3.21}
$$

Autocovariance

Autocovariance of a random process is defined as

$$
\begin{aligned}
C_X(t_2, t_1) &= E\left[\left(X_{t_2} - \mu_X(t_2)\right)\left(X_{t_1} - \mu_X(t_1)\right)^*\right] \\
&= R_X(t_2, t_1) - \mu_X(t_2) \mu_X(t_1)^*
\end{aligned} \tag{3.31}
$$

B biorthogonal

Signals $s_1(t)$, $s_2(t)$,..., $s_M(t)$ are biorthogonal if $s_1(t)$,..., $s_{\frac{M}{2}}(t)$ are orthogonal and $s_m(t) = -s_{\frac{M}{2}+m}(t)$ for some $m \in \left\{1, 2, \ldots, \frac{M}{2}\right\}$.

C Conditional Entropy

The conditional entropy of the random variable X given the random variable Y is defined by

$$
H(X|Y) = -\sum_{ii} \sum_{jj} p_{X,Y}(x_i, y_j) \log p_{X|Y}(x_i|y_j) \tag{7.8}
$$

Continuous Random Variable

A random variable X is continuous if the cumulative distribution function can be written in an integral form, or

$$
F_X(b) = \int_{-\infty}^{b} f_X(x) \, dx \tag{2.2}
$$

and $f_X(x)$ is the probability density function (pdf) (e.g., $F_X(x)$ is differentiable and $f_X(x) = \frac{d}{dx}(F_X(x))$)

Crosscorrelation

The crosscorrelation function of a pair of random processes is defined as

$$
\begin{aligned}
R_{XY}(t_2, t_1) &= E\left[X_{t_2} Y_{t_1}{}^*\right] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X_{t_2}, Y_{t_1}}(x, y) \, dxdy
\end{aligned} \tag{3.32}
$$

$$
C_{XY}(t_2, t_1) = R_{XY}(t_2, t_1) - \mu_X(t_2) \mu_Y(t_1)^* \tag{3.33}
$$

Cumulative distribution

The cumulative distribution function of a random variable X is a function F_X $(\mathbb{R} \mapsto \mathbb{R})$ such that

$$
\begin{aligned}
F_X \ (b) \ &= \ Pr \left[X \le b \right] \\
&= \ Pr \left[\{ \omega \in \Omega \mid X \left(\omega \right) \le b \} \right]
\end{aligned}
\tag{2.1}
$$

D Discrete Random Variable

A random variable X is discrete if it only takes at most countably many points (*i.e.,* F_X $(\cdot)$ is piecewise constant). The probability mass function (pmf) is defined as

$$
\begin{aligned}
p_X \ (x_k) \ &= \ Pr \left[X = x_k \right] \\
&= \ F_X \ (x_k) - \lim_{x(x \to x_k) \ \cdot \ (x < x_k)} F_X \ (x)
\end{aligned}
\tag{2.3}
$$

E Entropy Rate

The entropy rate of a stationary discrete-time random process is defined by

$$
H = \lim_{n \to \infty} H \left(X_n | X_1 X_2 \dots X_n \right)
\tag{7.12}
$$

The limit exists and is equal to

$$
H = \lim_{n \to \infty} \frac{1}{n} H \left(X_1, X_2, \dots, X_n \right)
\tag{7.13}
$$

The entropy rate is a measure of the uncertainty of information content per output symbol of the source.

Entropy

1. The entropy (average self information) of a discrete random variable X is a function of its probability mass function and is defined as

$$
H \left(X \right) = - \sum_{i=1}^{N} p_X \ (x_i) \log p_X \ (x_i)
\tag{7.3}
$$

   where N is the number of possible values of X and $p_X \ (x_i) = Pr \left[X = x_i \right]$. If log is base 2 then the unit of entropy is bits. Entropy is a measure of uncertainty in a random variable and a measure of information it can reveal.

2. A more basic explanation of entropy is provided in another module[2].

F First-order stationary process

If $F_{X_t} (b)$ is not a function of time then X_t is called a first-order stationary process.

G Gaussian process

A process with mean $\mu_X (t)$ and covariance function $C_X (t_2, t_1)$ is said to be a Gaussian process if **any** $X = (X_{t_1}, X_{t_2}, \dots, X_{t_N})^T$ formed by **any** sampling of the process is a Gaussian random vector, that is,

$$
f_X \left(x \right) = \frac{1}{(2\pi)^{\frac{N}{2}} \left(\det \Sigma_X \right)^{\frac{1}{2}}} e^{-\left(\frac{1}{2} (x - \mu_X)^T \Sigma_X^{-1} (x - \mu_X) \right)}
\tag{3.48}
$$

---

[2]"Entropy" $<$http://cnx.org/content/m0070/latest/$>$

for all $x \in \mathbb{R}^n$ where

$$\mu_X = \begin{pmatrix} \mu_X(t_1) \\ \vdots \\ \mu_X(t_N) \end{pmatrix}$$

and

$$\Sigma_X = \begin{pmatrix} C_X(t_1, t_1) & \ldots & C_X(t_1, t_N) \\ \vdots & \ddots & \\ C_X(t_N, t_1) & \ldots & C_X(t_N, t_N) \end{pmatrix}$$

. The complete statistical properties of X_t can be obtained from the second-order statistics.

J Joint Entropy

The joint entropy of two discrete random variables (X, Y) is defined by

$$H(X, Y) = -\sum_{ii} \sum_{jj} p_{X,Y}(x_i, y_j) \log p_{X,Y}(x_i, y_j) \tag{7.6}$$

Jointly Wide Sense Stationary

The random processes X_t and Y_t are said to be jointly wide sense stationary if $R_{XY}(t_2, t_1)$ is a function of $t_2 - t_1$ only and $\mu_X(t)$ and $\mu_Y(t)$ are constant.

M Mean

The mean function of a random process X_t is defined as the expected value of X_t for all t's.

$$\begin{aligned} \mu_{X_t} &= E[X_t] \\ &= \begin{cases} \int_{-\infty}^{\infty} x f_{X_t}(x)\, dx & \text{if continuous} \\ \sum_{k=-\infty}^{\infty} x_k p_{X_t}(x_k) & \text{if discrete} \end{cases} \end{aligned} \tag{3.20}$$

Mutual Information

The mutual information between two discrete random variables is denoted by $\mathcal{I}(X; Y)$ and defined as

$$\mathcal{I}(X; Y) = H(X) - H(X|Y) \tag{7.30}$$

Mutual information is a useful concept to measure the amount of information shared between input and output of noisy channels.

O orthogonal

Signals $s_1(t)$, $s_2(t)$,..., $s_M(t)$ are orthogonal if $< s_m, s_n >= 0$ for $m \neq n$.

P Power Spectral Density

The power spectral density function of a wide sense stationary (WSS) process X_t is defined to be the Fourier transform of the autocorrelation function of X_t.

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-(j2\pi f\tau)} d\tau \tag{3.44}$$

if X_t is WSS with autocorrelation function $R_X(\tau)$.

S Simplex signals

Let $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ be a set of orthogonal signals with equal energy. The signals $\tilde{s_1}(t), \ldots, \tilde{s_M}(t)$ are simplex signals if

$$\tilde{s_m}(t) = s_m(t) - \frac{1}{M}\sum_{k=1}^{M} s_k(t) \tag{4.3}$$

Stochastic Process

Given a sample space, a stochastic process is an indexed collection of random variables defined for each $\omega \in \Omega$.

$$X_t(\omega) \quad , \quad t \in \mathbb{R} \tag{3.1}$$

U Uncorrelated random variables

Two random variables X and Y are uncorrelated if $\rho_{XY} = 0$.

W Wide Sense Stationary

A process is said to be wide sense stationary if μ_X is constant and $R_X(t_2, t_1)$ is only a function of $t_2 - t_1$.

Index of Keywords and Terms

Keywords are listed by the section with that keyword (page numbers are in parentheses). Keywords do not necessarily appear in the text of the page. They are merely associated with that section. *Ex.* apples, § 1.1 (1) **Terms** are referenced by the page they appear on. *Ex.* apples, 1

Attributions

Collection: *Digital Communication Systems*
Edited by: Behnaam Aazhang
URL: http://cnx.org/content/col10134/1.3/
License: http://creativecommons.org/licenses/by/1.0

Module: "Review of Probability Theory"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10224/2.16/
Pages: 3-6
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Introduction to Stochastic Processes"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10235/2.15/
Pages: 7-10
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Second-order Description"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10236/2.13/
Pages: 11-13
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Linear Filtering"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10237/2.10/
Pages: 13-16
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Gaussian Processes"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10238/2.7/
Pages: 16-17
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Data Transmission and Reception"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10115/2.9/
Page: 19
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Signalling"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10116/2.11/
Pages: 19-23
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Geometric Representation of Modulation Signals"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10035/2.13/
Pages: 23-26
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Demodulation and Detection"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10054/2.14/
Pages: 26-27
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Demodulation"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10141/2.13/
Pages: 27-30
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Detection by Correlation"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10091/2.15/
Pages: 30-31
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Examples of Correlation Detection"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10149/2.10/
Pages: 32-34
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Matched Filters"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10101/2.14/
Pages: 34-35
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Examples with Matched Filters"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10150/2.10/
Pages: 36-38
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Matched Filters in the Frequency Domain"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10151/2.11/
Pages: 38-40
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Performance Analysis"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10106/2.10/
Page: 40
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Performance Analysis of Antipodal Binary signals with Correlation"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10152/2.11/
Pages: 40-41
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Performance Analysis of Binary Orthogonal Signals with Correlation"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10154/2.11/
Pages: 42-43
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Performance Analysis of Binary Antipodal Signals with Matched Filters"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10153/2.11/
Pages: 43-44
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Performance Analysis of Orthogonal Binary Signals with Matched Filters"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10155/2.9/
Pages: 44-46
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Digital Transmission over Baseband Channels"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10056/2.12/
Page: 47
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Pulse Amplitude Modulation Through Bandlimited Channel"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10094/2.7/
Pages: 47-48
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Precoding and Bandlimited Signals"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10118/2.6/
Pages: 48-49
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Carrier Phase Modulation"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10128/2.10/
Pages: 51-53
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Differential Phase Shift Keying"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10156/2.7/
Pages: 53-54
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Carrier Frequency Modulation"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10163/2.10/
Page: 55
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Information Theory and Coding"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10162/2.10/
Page: 57
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Entropy"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10164/2.16/
Pages: 58-61
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Source Coding"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10175/2.10/
Pages: 61-64
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Huffman Coding"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10176/2.10/
Pages: 64-65
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Channel Capacity"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10173/2.8/
Pages: 65-67
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Mutual Information"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10178/2.9/
Pages: 67-69
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Typical Sequences"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10179/2.10/
Pages: 69-71
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Shannon's Noisy Channel Coding Theorem"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10180/2.10/
Pages: 71-73
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Channel Coding"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10174/2.11/
Pages: 73-75
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Convolutional Codes"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10181/2.7/
Pages: 76-77
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 1 of Elec 430"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10370/2.6/
Pages: 81-82
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 2 of Elec 430"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10374/2.4/
Page: 83
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 5 of Elec 430"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10450/2.5/
Pages: 85-87
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 3 of Elec 430"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10375/2.8/
Pages: 89-90
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Exercises on Systems and Density"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10449/2.6/
Pages: 91-92
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 6 of Elec 430"
Used here as: "Homework 6"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10451/2.3/
Pages: 93-95
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 7 of Elec 430"
Used here as: "Homework 7"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10452/2.3/
Page: 97
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 8 of Elec 430"
Used here as: "Homework 8"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10453/2.3/
Page: 99
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

Module: "Homework 9 of Elec 430"
Used here as: "Homework 9"
By: Behnaam Aazhang
URL: http://cnx.org/content/m10455/2.3/
Page: 101
Copyright: Behnaam Aazhang
License: http://creativecommons.org/licenses/by/1.0

About Connexions

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities.

Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai. Connexions is part of an exciting new information distribution system that allows for **Print on Demand Books**. Connexions has partnered with innovative on-demand publisher QOOP to accelerate the delivery of printed course materials and textbooks into classrooms worldwide at lower prices than traditional academic publishers.