NATIONAL OPEN UNIVERSITY OF NIGERIA

SCHOOL OF SCIENCE AND TECHNOLOGY

COURSE CODE: CIT 754

COURSE TITLE: DIGITAL COMMUNICATION

CIT 754: DIGITAL COMMUNICATIONS

COURSE GUIDE

NATIONAL OPEN UNIVERSITY OF NIGERIA

National Open University of Nigeria Headquarters, 14/16 Ahmadu Bello Way,Victoria Island Lagos.

Course writers: Greg. O. Onwodi
National Open University of Nigeria

Engr . C . Obi

Editor: Dr Adekunle Yinka (university of Ibadan)

Programme Leader: Professor Kehinde Obidairo

Course Cordinator; Greg. Onwodi

EDITED COPY

Contents

- Introduction
- What you will learn in the course
- Course Aims
- Course Objectives
- Courses Materials
- Study Units
- Assessment
- Tutor-Marked Assignment
- Course Overview
- How to get the best from the course
- Summary

Introduction

The course, digital communication is one of the core courses for students.

The overall aim of this course, is to present the basis principles that underline the analysis and design of digital communication systems. The subject of digital communication involves the transmission of information in digital form from a source that generates the information to one or more destinations. Of particular importance in the analysis and design of communication systems are the characteristics of the physical channels through which the information is transmitted. The characteristics of the channel generally affect the design of the basic building blocks of the communication system.

It a background, we presume that the reader has a thorough understanding of basic calculus and elementary linear systems theory and prior knowledge of probability and stochastic processes.

Course Aims

The overall aims and objectives of this course will help you to:

1. Design a signal for band limited channels, the optimum receiver for channels with intersymbol interference.
2. Design basic elements of a digital communication system.
3. Design basic building blocks of communication system.

4. Develop the capacity.

Course Objectives

Upon completion of the course, you should able to;

1. Describe the elements of basic digital communication.

2. Identify various linear block codes and their properties.

3. Explain types of fading channel and their statistical model.

4. Discuss multi-user communications.

Course Materials

Basically, we made use of textbooks and online materials. You are expected to search for web references and literature for further reading and understanding. Each unit has references and web references that here used to develops this material.

Online Materials

You are free to refer to the websites provided for all the online reference materials required in this course.

Study Units

Assignment

The course, digital communications, entails attending a 2-hour final examination which contributes 70% to your final grading. The final examination covers materials from all part of the course with a method identical to the tutor marked assignment (TMA).

The examination aims at testing your ability to apply the knowledge you have learned throughout the course. In preparing for the examination, it is important you receive the activities and tutor marked assignments you have completed in each unit. The other 30% will account for all the TMA's at the end of each unit.

Tutor Marked Assignment

About 20 hours of tutorial will be provided in support of this course. You will be noticed of the dates, time and locations for these tutorials, together with the name and phone number of your tutor as soon as you are allotted a tutorial group.

Your tutor will mark and comment on your assignment, keep a close watch on your progress and on any difficulties you mighty encounter and provide assistance to you during the course. You must mail your TMAs to your tutor well before the due date (at least two working days are required). They will be marked by your tutor and returned to you as soon as possible.

Do not hesitate to contact your tutor by phone, e-mail if you need help. The following mighty be circumstances in which you would find help necessary. You can also contact your tutor if:

- You do not understand any part of the study units or the assigned readings.
- You have difficulty with the TMAs.
- You have a question and problem with your tutors comments on an assignment or with the grading of an assignment or with the grading of an assignment. You should try your best to attend tutorials, since it is the only opportunity to have interaction with you. Tutor and to ask questions which are answered immediately. You can raise any problem encountered in the course of your study. To gain maximum benefit from the course tutorials, you are advised to prepare a list of questions before attending tutorial. You will learn a lot from participating in discussions actively.

Course Overview

This section proposes the number of weeks that you are expected to spend on the six modules comprising of 23 units and the assignment that follow each unit. We recommend that two units with its associated TMA be completed in one week, bringing your period to a maximum of 11 weeks.

- **How to get the most of this course**

  In order for you to learn various concepts in this course, it is essential to practice. Independent and case activities which are based on a particular scenario are presented in the units. The activities include open questions to promote discussions on the relevant topics and questions with standard answers.

  You may try to go into each unit adopting the following steps;

1. Read the study unit.
2. Read the text book, printed or online references
3. Perform the activities.
4. Participate in group discussions.
5. Complete the tutor marked assignment.
6. Participate in online discussion.

Summary

The course, digital communications is intended to present the basis principles that underline the analysis and design of digital communication systems as a text for self study and as a guide for the practicing engineer involved in the design and analysis of digital communications systems.

It is also design to serve as a text for a first-year graduate level course for students in electrical engineering. We hope that you will find the course enlightening, interesting and useful. In the longer term, we also hope you will get acquainted with the National Open University of Nigeria. We wish you success in all your future endeavours.

MODULE 1**: Digital modulation schemes**

Unit 1: Power Spectrum of Digitally Modulated Signals

1.0 Introduction

In this unit, the information about the power spectral density helps us determine the required transmission bandwidth of these modulation schemes and their bandwidth efficiency.

2.0 Objectives

At the end of this unit, you should be able to;

- Understand the power spectrum of digitally modulated signal
- Explain power efficiency
- Discuss positive spectrum with respect to pass band and low pass signals.

3.1 Power spectral Density of a digitally Modulated Signal with Memory

Power Spectrum of Digitally Modulated Signals

In this section we study the power spectral density of digitally modulated signals. The information about the power spectral density helps us determine the required transmission bandwidth of these modulation schemes and their bandwidth efficiency. We start by considering a general modulation scheme with memory in which the current transmitted signal can depend on the entire history of the information sequence and then specialize this general formulation to the cases where the modulation system has a finite memory, the case where the modulation is linear, and when the modulated signal can be determined by the state of a Markov chain. We conclude this section with the spectral characteristics of CPM and CPFSK signals.

Power Spectral Density of a Digitally Modulated Signal with Memory

Here we assume that the bandpass modulated signal is denoted by v(t) with a lowpass equivalent signal of the form

$$v_l(t) = \sum_{n=-\infty}^{\infty} s_l(t - nT; \boldsymbol{I}_n)$$

Here $s_l(t; \boldsymbol{I}_n)$ E $\{s_{1l}(t), s2a(t), \ldots, S_{M1}(t)\}$ is one of the possible M lowpass equivalent signals determined by the information sequence up to time n, denoted by $\boldsymbol{I}_n = (\ldots, \boldsymbol{I}_{n-2}, \boldsymbol{I}_{n-1}, \boldsymbol{I}_n)$. We assume that $\boldsymbol{I}_n$ is stationary process. Our goal here is to determine the power spectral density of $v_l(t)$. This is done by first deriving the power spectral density of v, (t) and using Equation 2.9-14 to obtain the power spectral density of v (t).

We first determine the autocorrelation function of $v_l(t)$

$$\begin{aligned} R_{v_l}(t+\tau, t) &= \mathrm{E}\left[v_l(t+\tau)v_l^*(t)\right] \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathrm{E}\left[s_l(t+\tau-nT; I_n)s_l^*(t-mT; I_m)\right] \end{aligned}$$

Changing t to $t + T$ does not change the mean and the autocorrelation function of $v_l(t)$, hence $v_l(t)$ is a cyclostationary process; to determine its power spectral density, we

have to average $R_{vl}(t + r, t)$ over one period T. We have (with a change of variable of k=n-m)

$$\bar{R}_{v_l}(\tau) = \frac{1}{T}\sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\int_0^T \mathrm{E}\left[s_l(t+\tau-mT-kT;\boldsymbol{I}_{m+k})s_l^*(t-mT;\boldsymbol{I}_m)\right]$$

$$\overset{(a)}{=} \frac{1}{T}\sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\int_{-mT}^{-(m-1)T} \mathrm{E}\left[s_l(u+\tau-kT;\boldsymbol{I}_k)s_l^*(u;\boldsymbol{I}_0)\right]\,du$$

$$= \frac{1}{T}\sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{E}\left[s_l(u+\tau-kT;\boldsymbol{I}_k)s_l^*(u;\boldsymbol{I}_0)\right]\,du$$

Where in (a) we have introduced a change of variable of the form u = t – *mT* and we have used the fact that the Narkov chain is in the steady state and the input process $\{\boldsymbol{I}_n\}$ is stationary. Defined

$$g_k(\tau) = \int_{-\infty}^{\infty} \mathrm{E}\left[s_l(t+\tau;\boldsymbol{I}_k)s_l^*(t;\boldsymbol{I}_0)\right]\,dt$$

we can write Equation 3.4–3 as

$$\bar{R}_{v_l}(\tau) = \frac{1}{T}\sum_{k=-\infty}^{\infty} g_k(\tau-kT)$$

The power spectral density of $v_l(t)$ which is the Fourier transform of $R_{vl}(T)$, is therefore given by

$$\mathcal{S}_{v_l}(f) = \frac{1}{T}\mathscr{F}\left[\sum_k g_k(\tau-kT)\right]$$

$$= \frac{1}{T}\sum_{k=-\infty}^{\infty} G_k(f)e^{-j2\pi kfT}$$

Where $G_k(f)$ denotes the Fourier transform of $gk(T)$.

3.2 Bandpass and low pass signals Representation

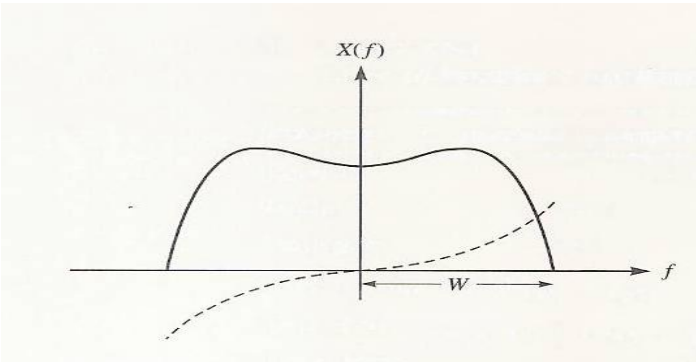
The process of communication consists of transmission of the output of an information source over a communication channel. In almost all cases, the spectral characteristics of the information sequence do not directly match the spectral characteristics of the communication channel, and hence the information signal cannot be directly transmitted over the channel. In many cases the information signal is a low frequency (baseband) signal, and the available spectrum of the

communication channel is at higher frequencies. Therefore, at the transmitter the information signal is translated to a higher frequency signal that matches the properties of the communication channel. This is the modulation process in which the baseband information signal is turned into a bandpass modulated signal. In this section we study the main properties of baseband and bandpass signals.

Bandpass and Lowpass Signals

In this section we will show that any real, narrowband, and high frequency signal - called a bandpass signal can be represented in terms of a complex low frequencysignal, called the lowpass equivalent of the original bandpass signal. This result makes it possible to work with the lowpass equivalents of bandpass signals instead of directly working with them, thus greatly simplifying the handling of bandpass signals. That is so because applying signal processing algorithms to lowpass signals is much easier due to lower required sampling rates which in turn result in lower rates of the sampled data.

The Fourier transform of a signal provides information about the frequency content, or spectrum, of the signal. The Fourier transform of a real signal x(t) has Hermitian symmetry, i.e., X (- f) = X *(f), from which we conclude that I X (- f) l = X (f) j and LX * (f) = - LX(f). In other words, for real x (t), the magnitude of X (f) is even and



its phase is odd. Because of this symmetry, all information about the signal is in the positive (or negative) frequencies, and in particular x(t) can be perfectly reconstructed by specifying X(f) for f > 0. Based on this observation, for a real signal x(t), we define the bandwidth as the smallest range of positive frequencies such that X(f) = 0 when /f/ is outside this range. It is clear that the bandwidth of a real signal is one-half of its frequency support set.

A lowpass, or baseband, signal is a signal whose spectrum is located around the zero frequency. For instance, speech, music, and video signals are all lowpass signals, although they have different spectral characteristics and bandwidths. Usually lowpass signals are low frequency signals, which means that in the time domain, they are slowly varying signals with no jumps or sudden variations. The bandwidth of a real lowpass signal is the minimum positive W such that X(f) = 0 outside [-W, +W]. For these signals the frequency support, i.e., the range of frequencies for which X (f) ≠ 0, is [-W, +W]. An example of the spectrum of a real-valued lowpass signal is shown in Fig. 2.1-1. The solid line shows the magnitude spectrum /X(f)/, and the dashed line indicates the phase spectrum ∠X(f).

We also define the positive spectrum and the negative spectrum of a signal x (t) as

$$X_+(f) = \begin{cases} X(f) & f > 0 \\ \frac{1}{2}X(0) & f = 0 \\ 0 & f < 0 \end{cases} \qquad X_-(f) = \begin{cases} X(f) & f < 0 \\ \frac{1}{2}X(0) & f = 0 \\ 0 & f > 0 \end{cases} \qquad (2.1\text{–}1)$$

3.3 A comparison of Digital Signaling Method

The digital modulation methods described in the previous sections can be compared in a number of ways. For example, one can compare them on the basis of the SNR required to achieve a specified probability of error. However, such a comparison would not be very meaningful, unless it were made on the basis of some constraint, such as a fixed data rate of transmission or, equivalently, on the basis of a fixed bandwidth.

The criterion for power efficiency of a signaling scheme is the SNR per bit that is required by that scheme to achieve a certain error probability. The error probability that is usually considered for comparison of various signaling schemes is Pe = 10-5.

The yb = No required by a signaling scheme to achieve an error probability of 10-5 is a criterion for power efficiency of that scheme. Systems requiring lower yb to achieve this error probability are more power-efficient.

To measure the bandwidth efficiency, we define a parameter r, called the spectral bit rate, or the bandwidth efficiency, as the ratio of bit rate of the signaling scheme to the bandwidth of it, i.e.,

$$r = {}^{R}/_{W} \text{ b/s/Hz}$$

A system with larger *r* is a more bandwidth-efficient system since it can transmit at a higher bit rate in each hertz of bandwidth. The parameters r and yb defined above are the two criteria we use for comparison of power and bandwidth efficiency of different modulation schemes.

4.0 Conclusion

The spectral characteristics of the communication channel, and hence the information signal cannot be directly transmitted over the channel. In many cases, the information signal is a low frequency (base band) signal, and the available spectrum of communication channel is at higher frequencies. Therefore, at the transmitter the information signal is translated to a higher frequency signal that matches the properties of the communication channel. This is the modulation process in which the base band information signal is turned into a band pass modulated signal.

5.0 Summary

In this unit, we considered the spectral characteristics of continuous-phase modulation (CPM) and continuous-phase frequency-shift keying (CPFSK). Besides, we also considered the main properties of base-band and bandpass signals.

6.0 Tutor Marked Assignment

1. Let x (t) and y (t) denote two band pass signals, and let x (t) and Yi (t) denote their low pass equivalents with respect to some frequency fo. We know that in general Xi (t) and yi (t) are complex signals.

i. Show that;

∫ x(E)y (t) dt = 1/Z Re [I xi(t) yi (t)dt]

ii. From this conclude that Y_ x = '/z Y-xc, i.e the energy in a bandpass signal is one-half the energy in its low pass equivalent

2. The information sequence {an) is a sequence taking the values -1, 2 and with probabilities $^1/_4$, $^1/4$ and $^1/z$. This information sequence is used to generate the baseband signal.

$$V(t) = Z\ an\ \text{Sinc}\left(\frac{t-nT}{T}\right)$$

i. Determine the power spectral density of v(t)

ii. Define the sequence (b_n) as b, = 9, + 9_n _1 -9,_z and generate the base-band signal.

$$u(t) = Z\ a,\ \text{Sinc}\left(\frac{t-nT}{T}\right)$$

Determine the power spectral density of u(t). What are the possible values for the b_n sequence?

6.0 References/ Further Reading

The linear representation of continuous phase modulation for binary modulation by Laurent (1986).

MODULE 2: LINEAR BLOCK CODE AND GRAPH BASED CODES

Unit 1: Linear Block Codes

Unit 2: Some Specific Linear Block Codes

Unit 3: Trellis and Graph Based Code

Unit 1: Linear Block Codes

1.0 Introduction

2.0 Objective

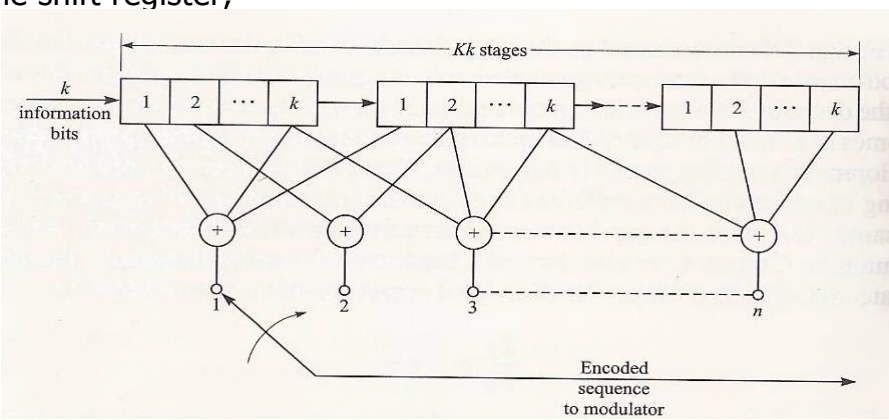
3.0 Main Contents

3.1 Basic definitions

Channel codes can be classified into two major classes, block codes and convolutional codes. In block codes one of the M = 2k messages, each representing a binary sequence of length k, called the information sequence, is mapped to a binary sequence of length n, called the codeword, where n > k. The codeword is usually transmitted over the communication channel by sending a sequence of n binary symbols, for instance, by using BPSK. QPSK and BFSK are other types of signaling schemes frequently used for transmission of a codeword. Block coding schemes are memoryless. After a codeword is encoded and transmitted, the system receives a new set of k information bits and encodes them using the mapping defined by the coding scheme. The resulting codeword depends only on the current k information bits and is independent of all the codewords transmitted before.

Convolutionalodes are described in terms of finite-state machines. In these codes, at each time instance i, k information bits enter the encoder, causing n binary symbols generated at the encoder output and changing the state of the encoder from σ_{i-1} to σ_i. The set of possible states is finite and denoted by E. The n binary symbols generated at the encoder output and the next state 6l depend on the k input bits as well as σ_{i-1}.

At each time instance, k bits enter the encoder and the contents of the shift register are shifted to the right by k memory elements. The contents of the rightmost k elements of the shift register leave the encoder. After the k bits have entered the shift register,



A convolutional encoder.

the n adders add the contents of the memory elements they are connected to (modulo-2 addition) thus generating the code sequence of length n which is sent to

the modulator. The state of this convolutional code is given by the contents of the first (K - 1)k elements of the shift register.

The code rate of a block or convolutional code is denoted by Rc and is given by

$$= Rc = {}^{k}/_{n}$$

The rate of a code represents the number of information bits sent in transmission of a binary symbol over the channel. The unit of R, is information bits per transmission. Since generally n > k, we have R, < 1.

Let us assume that a codeword of length n is transmitted using an N-dimensional constellation of size M, where M is assumed to be a power of 2 and $L = {}^{n}/_{\log 2M}$ M is assumed to be an integer representing the number of M-ary symbol transmitted per codeword. If the symbol duration is TS, then the transmission time for k bits is $T_s = L\ T_s$ and the transmission rate is given by

$$R = \frac{k}{LT_s} = \frac{k}{n} \times \frac{\log_2 M}{T_s} = R_c \frac{\log_2 M}{T_s} \text{ bits/s}$$

The dimension of the space of the encoded and modulated signals is LN, and using the dimensionality theorem as stated in Equation, we conclude that the minimum required transmission bandwidth is given by

$$W = \frac{N}{2T_s} = \frac{RN}{2R_c \log_2 M} \text{ bits/s}$$

and from Equation 7.1–3, the resulting spectral bit rate is given by

$$r = \frac{R}{W} = \frac{2\log_2 M}{N} R_c$$

These equations indicate that compared with an uncoded system that uses the same modulation scheme, the bit rate is changed by a factor of R, and the bandwidth is changed by a factor of $1\ /R_c$ i.e., there is a decrease in rate and an increase in bandwidth.

If the average energy of the constellation is denoted by E_{av}, then the energy per codeword E, is given by

$$\mathcal{E} = L\mathcal{E}_{\text{av}} = \frac{n}{\log_2 M}\mathcal{E}_{\text{av}}$$

and $\mathcal{E}_c$, energy per component of the codeword, is given by

$$\mathcal{E}_c = \frac{\mathcal{E}}{n} = \frac{\mathcal{E}_{\text{av}}}{\log_2 M}$$

The energy per transmitted bit is denoted by $\mathcal{E}_b$ and can be found from

$$\mathcal{E}_b = \frac{\mathcal{E}}{k} = \frac{\mathcal{E}_{\text{av}}}{R_c \log_2 M}$$

From Equations 7.1–6 and 7.1–7 we conclude that

$$\mathcal{E}_c = R_c \mathcal{E}_b$$

The transmitted power is given by

$$P = \frac{\mathcal{E}}{LT_s} = \frac{\mathcal{E}_{\text{av}}}{T_s} = R\frac{\mathcal{E}_{\text{av}}}{R_c \log_2 M} = R\mathcal{E}_b$$

Modulation schemes frequently used with coding are BPSK, BFSK, and QPSK. The minimum required bandwidth and the resulting spectral bit rates for these modulation scheme† are given below:

$$\text{BPSK}: \begin{cases} W = \frac{R}{R_c} \\ r = R_c \end{cases} \quad \text{BFSK}: \begin{cases} W = \frac{R}{R_c} \\ r = R_c \end{cases} \quad \text{QPSK}: \begin{cases} W = \frac{R}{2R_c} \\ r = 2R_c \end{cases} \qquad (7.1\text{–}10)$$

3.2 The Structure of Finite Fields

To further explore properties of block codes, we need to introduce the notion of a finite field and its main properties. Simply stated, a field is a collection of objects that can be added, subtracted, multiplied, and divided. To define fields, we begin by defining Abelian groups. An Abelian group is a set with a binary operation that has the basic properties of addition. A set G and a binary operation denoted by + constitute an Abelian group if the following properties hold:

1. The operation $+$ is commutative; i.e., for any $a, b \in G$, $a + b = b + a$.
2. The operation $+$ is associative; i.e., for any $a, b, c \in G$, we have $(a + b) + c = a + (b + c)$.

Addition and Multiplication Tables for GF(2)

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

3. The operation + has an identity element denoted by 0 such that for any a E G, a -1 -0=0+a=a.

4. For any a E G there exists an element -a E G such that a + (-a) = (-a) + a = 0. The element -a is called the (additive) inverse of a.

An Abelian group is usually denoted by {G, +, 01}.

A finite field or Galois field is a finite set F with two binary operations, addition and multiplication, denoted, respectively, by + and -, satisfying the following properties:

1. {F, +, 0} is an Abelian group.

2. {F – {0}, ., 1} is an Abelian group; i.e., the nonzero elements of the field constitute an Abelian group under multiplication with an identity element denoted by "1". The multiplicative inverse of a E F is denoted by a-l.

3. Multiplication is distributive with respect to addition: a - (b + c) = (b + c) - a = a-b+a-c.

A field is usually denoted by {F, +, .1. It is clear that IIB, the set of real numbers, is a field (but not a finite field) with ordinary addition and multiplication. The set F = {0, 11 with modulo-2 addition and multiplication is an example of a Galois (finite) field. This field is called the binary field and is denoted by GF(2). The addition and multiplication tables for this field are given in table above.

Characteristic of a Field and the Ground Field

A fundamental theorem of algebra states that a Galois field with q elements, denoted by GF(q), exists if and only if $q = p^m$, where p is a prime and m is a positive integer. It can also be proved that when GF(q) exists, it is unique up to isomorphism. This means that any two Galois fields of the same size can be obtained from each other after renaming the elements. For the case of q = p, the Galois field can be denoted by GF(p) = 10, 1, 2, . . ., p - 11 with modulo-p addition and multiplication. For instance GF(5) = {0, 1, 2, 3, 41 is a finite field with modulo-5 addition and multiplication. When $q = p^m$, the resulting Galois field is called an extension field of GF(p). In this case GF(p) is called the ground field of GF(pm), and p is called the characteristic of $GF(p^m)$.

Polynomials over Finite Fields

To study the structure of extension fields, we need to define polynomials over GF(p). A polynomial of degree m over GF(p) is a polynomial

$$g(X) = g_0 + g_1 X + g_2 X^2 + \cdots + g_m X^m$$

where gi , $0 \leq i \leq m$, are elements of GF(p) and $g_m \neq 0$. Addition and multiplication of polynomials follow standard addition and multiplication rules of ordinary polynomials except that addition and multiplication of the coefficients are done modulo-p. If $g_m = 1$, the polynomial is called monic. If a polynomial of degree m over GF(p) cannot be written as the product of two polynomials of lower degrees over the same Galois field, then the polynomial is called an irreducible polynomial. For instance, $X^2 + X + 1$ is an irreducible polynomial over GF(2), whereas $X^2 + 1$ is not irreducible over GF(2) because $X^2 + 1 = (X + 1)^2$. A polynomial that is both monic and irreducible is called a prime polynomial. A fundamental result of algebra states that a polynomial of degree m over GF(p) has m roots (some may be repeated), but the roots are not necessarily in GF(p). In general, the roots are in some extension field of GF(p).

The Structure of Extension Fields

From the above definitions it is clear that there exist pm polynomials of degree less than m; in particular these polynomials include two special polynomials g(X) = 0 and g(X) = 1. Now let us assume that g(X) is a prime (monic and irreducible) polynomial of degree m and consider the set of all polynomials of degree less than m over GF(p) with ordinary addition and with polynomial multiplication modulo-g(X). It can be shown that the set of these polynomials with the addition and multiplication operations defined above is a Galois field with pm elements.

EXAMPLE 7.1-1. We know that X2 + X + 1 is prime over GF(2); therefore this poly-nomial can be used to construct GF(22) = GF(4). Let us consider all polynomials of degree less than 2 over GF(2). These polynomials are 0, 1, X, and X -f- 1 with addition and multiplication tables given in table above. Note that the multiplication rule basically entails multiplying the two polynomials, dividing the product by $g(X) = X^2 + X + 1$, and finding the remainder. This is what is meant by multiplying modulo-g(X). It is

interesting to note that all nonzero elements of GF(4) can be written as powers of X; i.e,$X=X^1$, $X+1= X^2$, and $1=X^3$.

Addition and Multiplication Table for GF(4)

$+$	0	1	X	$X+1$
0	0	1	X	$X+1$
1	1	0	$X+1$	X
X	X	$X+1$	0	1
$X+1$	$X+1$	X	1	0

$\cdot$	0	1	X	$X+1$
0	0	0	0	0
1	0	1	X	$X+1$
X	0	X	$X+1$	1
$X+1$	0	$X+1$	1	X

TABLE 7.1–3
Multiplication Table for GF(8)

$\cdot$	0	1	X	$X+1$	X^2	X^2+1	X^2+X	X^2+X+1
0	0	0	0	0	0	0	0	0
1	0	1	X	$X+1$	X^2	X^2+1	X^2+X	X^2+X+1
X	0	X	X^2	X^2+X	$X+1$	1	X^2+X+1	X^2+1
$X+1$	0	$X+1$	X^2+X	X^2+1	X^2+X+1	X^2	1	X
X^2	0	X^2	$X+1$	X^2+X+1	X^2+X	X	X^2+1	1
X^2+1	0	X^2+1	1	X^2	X	$X+2+X+1$	$X+1$	X^2+X
X^2+X	0	X^2+X	X^2+X+1	1	X^2+1	$X+1$	X	X^2
X^2+X+1	0	X^2+X+1	X^2+1	X	1	X^2+X	X^2	$X+1$

To generate GF(23), we can use either of the two prime polynomials gl(X) = X3 + X + 1 or 92(X) = X3 + XZ + 1. If g(X) = X3 + X -f- l is used, the multiplication table for GF(23) is given by Table above. The addition table has a trivial structure. Here again note that X1 = X, X2 = X2, X3 = X + l, X4 = XZ + X, XS = X2 + X + 1, X6 = X2 + 1, and X7 = 1. In other words, all nonzero elements of GF(8) can be written as powers of X. The nonzero elements of the field can be expressed either as polynomials of degree less than 3 or, equivalently, as X` for $1 \leq i \leq 7$. A third method for representing the field elements is to write coefficients of the polynomial as a vector of length 3. The representation of the form X` is the appropriate representation when multiplying field elements since X - Xi = X'+j, where i + j should be reduced modulo-7 because $X^7 = 1$. The polynomial and vector representations of field elements are more appropriate when adding field elements. A table of the three representations of field elements is given in Table 7.1-4. For instance, to multiply $X^2 + X + 1$ and $X^2 + 1$, we use their power representation as XS and X6 and we have $(X^2 + X +1)(X2+1)=X11 =X^4=X^2+X$.

3.3 General properties of linear block code

A q-ary block code C consists of a set of M vectors of length n denoted by cm = (cm_1, c_{m2}, . . . , C_{mn}), $1 \leq m \leq M$, and called codewords whose components are selected from an alphabet of q symbols, or elements. When the alphabet consists of two symbols, 0 and l, the code is a binary code. It is interesting to note that when q is a power of 2, i.e., q = 2b where b is a positive integer, each q-ary symbol has an equivalent binary representatioh consisting of b bits; thus, a nonbinary code of block length N can be mapped into a binary code of block length n = bN.

There are 2n possible codewords in a binary block code of length n. From these 2' codewords, we may select M = 2k codewords (k < n) to form a code. Thus, a block of k information bits is mapped into a codeword of length n selected from the set of M = 2k codewords. We refer to the resulting block code as an (n, k) code, with rate R, = k/n. More generally, in a code having q symbols, there are qn possible codewords. A subset of M = qk codewords may be selected to transmit k-symbol blocks of information.

Besides the code rate parameter R, an important parameter of a codeword is its weight, which is simply the number of nonzero elements that it contains. In general, each codeword has its own weight. The set of all weights in a code constitutes the weight distribution of the code. When all the M codewords have equal weight, the code is called a fixed-weight code or a constant-weight code.

A subset of block codes, called linear block codes, is particularly well studied during the last few decades. The reason for the popularity of linear block codes is that linearity guarantees easier implementation and analysis of these. codes. In addition, it is remarkable that the performance of the class of linear block codes is similar to the performance of the general class of block codes. Therefore, we can limit our study to the subclass of linear block codes without sacrificing system performance.

A linear block code C is a k-dimensional subspace of an n -dimensional space which is usually called an (n, k) code. For binary codes, it follows from Problem 7.11 that a linear block code is a collection of 2k binary sequences of length n such that for any

two codewords c_l, $c_2 \in C$ we have $c_l + c_2$ E C. Obviously, 0 is a codeword of any linear block code.

Generator and Parity Check Matrices

In a linear block code, the mapping from the set of M = 2_k information sequences of length *k* to the corresponding 2_k codewords of length n can be represented by a *k x n* matrix **G** called the generator matrix as;

$$c_m = u_m G, \qquad 1 \leq m \leq 2^k$$

where $\boldsymbol{u}_m$ is a binary vector of length k denoting the information sequence and $\boldsymbol{c}_m$ is the corresponding codeword. The rows of G and denoted by gi, 1 ≤ i ≤ k, denoting the codewords corresponding to the information sequence (1, 0, …., 0), (0, 1, 0, …., 0), …, (0,….0,1).

$$G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{bmatrix}$$

and hence,

$$c_m = \sum^{k} u_{mi} g_i$$

where the summation is in GF(2), i.e., modulo-2 summation.

Two linear block codes C_l and C_2 are called equivalent if the corresponding generator matrices have the same row space, possibly after a permutation of columns.

If the generator matrix G has the following structure;

G = ($\mathbf{I}_k$ / P)

where I k is a k x k identity matrix and P is a k x (n -k) matrix, the resulting linear block code is called systematic. In systematic codes the first k components of the codeword are equal to the information sequence, and the following n - k components, called the parity check bits, provide the redundancy for protection against errors.

Since C is a k-dimensional subspace of the n-dimensional binary space, its orthogonal complement, i.e., the set of all n-dimensional binary vectors that are orthogonal to the the codewords of C, is an (n - k)-dimensional subspace of the n-dimensional space, and therefore it defines an (n, n - k) code which is denoted by C' and is called the dual code of C. The generator matrix of the dual code is an (n -

k) x n matrix whose rows are orthogonal to the rows of G, the generator matrix of C. The generator matrix of the dual code is called the parity check matrix of the original code C and is denoted by ***H.*** Since any codeword of C is orthogonal to all rows of ***H***, we concluded that for all c $\in$ C.

$$\boldsymbol{cH}^t = 0$$

Also if for some binary n-dimensional vector c we have $\boldsymbol{cH}^t = 0$, then c belongs to the orthogonal complement of H, i.e., c $\in$ C. Therefore, a necessary and sufficient condition for c $\in$ (0,1)n Since rows of G are codewords, we conclude that;

$$\boldsymbol{GH}^t = 0$$

In the special case of systematic codes, where G = [$\boldsymbol{I}_k$ / **P**], the parity check matrix is given by

$$H = [-\boldsymbol{P}^t / \boldsymbol{I}_{n-k}]$$

which obviously satisfies $\boldsymbol{GH}^t = 0$. For binary code Pt = Pt and H ($\boldsymbol{P}^t$ / $\boldsymbol{I}_{n-k}$]

Consider a (7,4) linear block code with

$$G = [I_4 \mid P] = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Obviously this is a systematic code. This parity check matric for this code is

$$H = \left[P^t \mid I_{n-k}\right] = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

If u = (u_1, u_2, u_3, u_4) is an information sequence, the corresponding codeword c = (c_1, c_2,…., c_7) is given by

$$\begin{aligned} c_1 &= u_1 \\ c_2 &= u_2 \\ c_3 &= u_3 \\ c_4 &= u_4 \\ c_5 &= u_1 + u_2 + u_3 \\ c_6 &= u_2 + u_3 + u_4 \\ c_7 &= u_1 + u_2 + u_4 \end{aligned}$$

Weight and Distance for Linear Block Codes

The weight of a codeword c E C is denoted by w(c) and is the number of nonzero components of that codeword. Since 0 is a codeword of all linear block codes, we conclude that each linear block code has one codeword of weight zero. The Hamming distance between two codewords cl, c2 E C, denoted by d(cl, c2), is the number of components at which cl and c2 differ. It is clear that the weight of a codeword is its distance from 0.

The distance between cl and c2 is the weight of cl - c2, and since in linear block codes cl - c2 is a codeword, then d(cl, c2) = w(cl - c2). We clearly see that in linear block codes there exists a one-to-one correspondence between weight and the distance between codewords. This means that the set of possible distances from any codeword c E C to all other codewords is equal to the set of weights of different codewords, and thus is independent of c. In other words, in a linear block code, looking from any codeword to all other codewords, one observes the same set of distance, regardless of the codeword one is looking from. Also note that in binary linear block codes we can substitute cl - c2 with cl + c2.

The minimum distance of a code is the minimum of all possible distances between distinct codewords of the code, i.e.,

$$d_{\min} = \min_{\substack{c_1, c_2 \in \mathcal{C} \\ c_1 \neq c_2}} d(c_1, c_2)$$

The minimum weight of a code is the minimum of the weights of all nonzero codewords which for linear block codes is equal to the minimum distance.

$$w_{\min} = \min_{\substack{c \in \mathcal{C} \\ c \neq 0}} w(c)$$

There exists a close relation between the minimum weight of a linear block code and the columns of the parity check matrix H. We have previously seen that the necessary and sufficient condition for c E {0, lln to be a codeword is that cHt = 0. If we choose c to be a codeword of minimum weight, from this relation we conclude that w„,in (or drain) columns of H are linearly dependent. On the other hand, since there exists no

codeword of weight less than d nun, no fewer than d„,ir, columns of H can be linearly dependent. Therefore, dmi„ represents the minimum number of columns of H that can be linearly dependent. In other words the column space of H has dimension d,,-,in - 1.

In certain modulation schemes there exists a close relation between Hamming distance and Euclidean distance of the codewords. In binary antipodal signaling-for instance, BPSK modulation-the 0 and 1 components of a codeword c E C are mapped to and +,I-Sc, respectively. Therefore if s is the vector corresponding to the modulat

$$s_{mj} = (2c_{mj} - 1)\sqrt{\mathcal{E}_c}, \qquad 1 \le j \le n,\ 1 \le m \le M$$

and therefore,

$$d^2_{s_m, s_{m'}} = 4\mathcal{E}_c d(\boldsymbol{c}_m, \boldsymbol{c}_{m'})$$

where d_{sm}, s_m denotes the Euclidean distance between the modulate sequences and $d(\boldsymbol{c}_m, \boldsymbol{c}_m)$ is the Hamming distance between the corresponding codewords. From the above we have

$$d^2_{\mathrm{E\,min}} = 4\mathcal{E}_c d_{\min}$$

where d_{Emin} is the minimum Euclidean distance of the BPSK modulated sequences corresponding to the codewords. Using equation 7.1-8 we conclude that

$d^2_{Emin} = 2R_c\varepsilon_b d_{min}$

For the binary orthogonal modulations, e.g binary orthogonal FSK, we similarly have

$$d^2_{Emin} = 2R_c\varepsilon_b d_{min}$$

The Distribution Polynomial

An (n, k) code has 2k codewords that can have weights between 0 and n. In any linear block code there exists one codeword of weight 0, and the weights of nonzero codewords can be between d,nin and n. The weight distribution polynomial (WEP) or weight enumeration function (WEF) of a code is a polynomial that specifies the number of codewords of different weights in a code. The weight distribution polynomial or weight enumeration function is denoted by A(Z) and is defined by

$$A(Z) = \sum_{i=0}^{n} A_i Z^i = 1 + \sum_{i=d_{\min}}^{n} A_i Z^i$$

where Ai denotes the number of codewords of weight i. The following properties of the weight enumeration function for linear block codes are straightforward.

$$A(1) = \sum_{i=0}^{n} A_i = 2^k$$

$$A(0) = 1$$

The weight enumeration function for many block codes is unknown. For low rate codes the weight enumeration function can be obtained by using a computer search. The MacWilliams identity expresses the weight enumeration function of a code in terms of the weight enumeration function of its dual code. By this identity, the weight enumeration function of a code A(Z) is related to the weight enumeration function of its dual code A_d (Z) by

$$A(Z) = 2^{-(n-k)}(1+Z)^n A_d\left(\frac{1-Z}{1+Z}\right)$$

Note that for a linear block code, the set of distances seen from any codeword to other codewords is independent of the codeword from which these distances are seen. Therefore, in linear block codes the error bound is independent of the transmitted codeword, and thus, without loss of generality, w can always assume that the all-zero codewodrd 0 is transmitted.

For orthogonal binary F

$$d_E^2(s_m) = 2\mathcal{E}_b R_c w(c_m)$$

The distance enumerator fu

$$d_E^2(s_m) = 4\mathcal{E}_b R_c w(c_m)$$

$$T(X) = \sum_{i=d_{\min}}^{n} A_i X^{4R_c\mathcal{E}_b i} = (A(Z)-1)\big|_{Z=X^{4R_c\mathcal{E}_b}}$$

and for orthogonal BFSK by

$$T(X) = \sum_{i=d_{\min}}^{n} A_i X^{2R_c\mathcal{E}_b i} = (A(Z)-1)\big|_{Z=X^{2R_c\mathcal{E}_b}}$$

Another version of the weight enumeration function provides information about the weight of the codewords as well as weight of the responding information sequences.

The polynomial is called the input-output weight enumeration function (IOWEF), denoted by B (Y,Z) and is defined as

$$B(Y, Z) = \sum_{i=0}^{n} \sum_{j=0}^{k} B_{ij} Y^j Z^i$$

Where Bij is the number of codewords is weight i that are generated by information sequences of weight j. clearly

$$A_i = \sum_{j=0}^{k} B_{ij}$$

and for linear block codes we have $B(0, 0) = B_{00} = 1$. It is also clear that

$$A(Z) = B(Y, Z)|_{Y=1}$$

A third form of the weight enumeration function called the conditional weight enumeration function (CWEF) is defined by

$$B_j(Z) = \sum_{i=0}^{n} B_{ij} Z^i$$

and it represents the weight enumeration function of all codewords corresponding to information sequences of weight j. from equations, it is easy to see that

$$B_j(Z) = \frac{1}{j!} \frac{\partial^j}{\partial Y^j} B(Y, Z)|_{Y=0}$$

In the code discussed, there are 24 = 16 codewords with possible weights between 0 and 7. Substituting all possible information sequences of the form u = (u l, u2, u3, u4) and generating the codewords, we can verify that for this code drain = 3 and there are 7 codewords of weight 3 and 7 codewords of weight 4. There exist one codeword of weight 7 and one codeword of weight 0. Therefore,

$$A(Z) = 1 + 7Z^3 + 7Z^4 + Z^7$$

It is also easy to verify that for this code

$$B_{00} = 1 \quad B_{31} = 3 \quad B_{32} = 3 \quad B_{33} = 1$$
$$B_{41} = 1 \quad B_{42} = 3 \quad B_{43} = 3 \quad B_{74} = 1$$

Hence,

$$B(Y, Z) = 1 + 3YZ^3 + 3Y^2Z^3 + Y^3Z^3 + YZ^4 + 3Y^2Z^4 + 3Y^3Z^4 + Y^4Z^7$$

and

$$B_0(Z) = 1$$
$$B_1(Z) = 3Z^3 + Z^4$$
$$B_2(Z) = 3Z^3 + 3Z^4$$
$$B_3(Z) = Z^3 + 3Z^4$$
$$B_4(Z) = Z^7$$

Error Probability of Linear Block Codes

Two types of error probability can be studied when linear block codes are employed. The block error probability or word error probability is defined as the probability of transmitting a codeword c,n and detecting a different codeword c,,,. The second type of error probability is the bit error probability, defined as the probability of receiving a transmitted information bit in error.

Block Error Probability

Linearity of the code guarantees that the distances from c,n to all other codewords are independent of the choice of c„2. Therefore, without loss of generality we can assume that the all-zero codeword 0 is transmitted.

To determine the block (word) error probability Pe, we note that an error occurs if the receiver declares any codeword c,n :~L_ 0 as the transmitted codeword. The probability of this event is denoted by the pairwise error probability Po,m.

$$P_e \leq \sum_{\substack{c_m \in C \\ c_m \neq 0}} P_{0 \to c_m}$$

Where in general P0cm depends on the Hamming distance between **0** and $\boldsymbol{c}_m$, which is equal w($\boldsymbol{c}_m$), in a way that depends on the modulation scheme employed for transmission of the codewords. Since for codewords of equal weight we have the same $\boldsymbol{P}_0$---$\boldsymbol{c}_m$, we conclude that

$$P_e \leq \sum_{i=d_{\min}}^{n} A_i P_2(i)$$

Where $\boldsymbol{P}_2$(i) denotes the pairwise error probability (PEP) between two codewords with Hamming distance i.

From Equation, we know that;

$$P_{\mathbf{0}\to c_m} \le \prod_{i=1}^{n} \sum_{y_i \in \mathscr{Y}} \sqrt{p(y_i|0)p(y_i|c_{mi})}$$

Following Example 6.8–1 we define

$$\Delta = \sum_{y \in \mathscr{Y}} \sqrt{p(y|0)p(y|1)}$$

With this definition, Equation 7.2–35 reduces to

$$P_{\mathbf{0}\to c_m} = P_2(w(c_m)) \le \Delta^{w(c_m)}$$

Substituting this result into Equation 7.2–34 results in

$$P_e \le \sum_{i=d_{\min}}^{n} A_i \Delta^i$$

or

$$P_e \le A(\Delta) - 1$$

where $A(Z)$ is the weight enumerating function of the linear block code.

From the inequality

$$\sum_{y \in \mathscr{Y}} \left(\sqrt{p(y|0)} - \sqrt{p(y|1)}\right)^2 \ge 0$$

we easily conclude that

$$\Delta = \sum_{y \in \mathscr{Y}} \sqrt{p(y|0)p(y|1)} \le 1$$

and hence, for $i \ge d_{\min}$,

$$\Delta^i \le \Delta^{d_{\min}}$$

Using this result in equation yields the simpler, but looser, bound

$$P_e \le (2^k - 1)\Delta^{d_{\min}}$$

Bit Error Probability

In general, errors at different locations of an information sequence of length k can occur, with different probabilities. We define the average of these error probabilities as the bit error probability for a linear block code. We again assume that the all-zero sequence is transmitted; then the probability that a specific codeword of weight i will be decoded at the detector is equal to $P_2(i)$. The number of codewords of weight i that correspond to information sequences of weight j is denoted by Bid. Therefore, when 0 is transmitted, the expected number of information bits received in error is given by

$$\bar{b} \le \sum_{j=0}^{k} j \sum_{i=d_{\min}}^{n} B_{ij} P_2(i)$$

Since for $0 < i < d_{\min}$ we have $B_{ij} = 0$, we can write this as

$$\bar{b} \le \sum_{j=0}^{k} j \sum_{i=0}^{n} B_{ij} P_2(i)$$

The (average) bit error probability of the linear block code Pb is defined as the ratio of the expected number of bits received in error to the total numbe of transmited bits, i.e.

$$P_b = \frac{\bar{b}}{k}$$
$$\leq \frac{1}{k} \sum_{j=0}^{k} j \sum_{i=0}^{n} B_{ij} P_2(i)$$
$$\leq \frac{1}{k} \sum_{j=0}^{k} j \sum_{i=0}^{n} B_{ij} \Delta^i$$

from equation, we see that the last sum is simply Bj (Δ); therefore

$$P_b \leq \frac{1}{k} \sum_{j=0}^{k} jB_j(\Delta)$$

we can also express the bit error probability in terms of the IOWEF by using equation as

$$P_b \leq \frac{1}{k} \sum_{i=0}^{n} \sum_{j=0}^{k} jB_{ij} \Delta^i$$
$$= \frac{1}{k} \frac{\partial}{\partial Y} B(Y, Z)\Big|_{Y=1, Z=\Delta}$$

4.0 Conclusion

This unit is devoted to block codes whose construction is based on familiar algebraic structures such as groups, rings and fields.

5.0 Summary

We considered one of he channel codes named block code. Hard decision decoding of these codes results in a binary symmetric channel model consisting of the binary modulator, the waveform channel, and the optimum binary detector.

6.0 Tutor Marked Assignment

1. The generator matrix for a linear binary code is

$$G = \begin{Bmatrix} 0011101 \\ 0100111 \\ 1001110 \end{Bmatrix}$$

a. Express G in systematic [I/P] form.

b. Determine the party check matrix H for the code.

c. Construct the table of syndromes for the code.

d. Determine the minimum distance of the code.

e. Demonstrate that the codeword C corresponding to the information sequence 101 satisfies $cH^t = 0$.

2. Prove that;

i. Elements of the standard array of a linear block code are distinct.

ii. Two element belonging to two distinct cosset of a standard array have distinct syndromes.

7.0 References/ Further Reading

Coding techniques for noisy channels by Elias (1954,1955) and Slepian (1956).

Unit 2: Some specific linear block codes

1.0 Introduction

2.0 Objectives

3.0 Main Contents

3.1 Specific linear block codes

3.2 Optimum soft decision decoding of linear block codes

3.3 Hard decision decoding of linear block codes

3.4 Comparison of performance between hard decision and soft decision decoding

3.5 Bounds on minimum distance of linear block codes

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/Further Reading

1.0 Introduction

We are going to examine some linear block codes with their parameters. Besides, study some bounds on minimum distances of linear block codes.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain error detection and error correction capability of block codes.
- Understanding the parameters of some specific linear codes.
- Discus soft and hard decision decoding of linear block codes.

3.1 Specific linear block codes.

In this section, we briefly describe some linear block codes that are frequently encountered in practice and list their important parameters. Additional classes of linear codes are introduced in our study of cyclic codes.

Repetition Codes

A binary repetition code is an (n, 1) code with two code words of length n. One codeword is the all-zero codeword, and the other one is the all-one codeword. This code has a rate of R, = n and a minimum distance of d_{min} = n. The dual of a repetition code is an (n, n - 1) code consisting of all binary sequences of length n with even parity. The minimum distance of the dual code is clearly $d_{min} = 2$.

Hamming Codes

Hamming. codes are one of the earliest codes studied in coding theory. Hamming codes are linear block codes with parameters $n = 2^m - 1$ and $k = 2^m - m - l$, for $m > 3$. Hamming codes are best described in terms of their parity check matrix H which is an

(n - k) x n = m x (2m - 1) matrix. The 2m - 1 columns of H consist of all possible binary vectors of length m excluding the all-zero vector. The rate of a Hamming code is given by

$$\boldsymbol{R}_c = \frac{2^m - m - 1}{2^m - 1}$$

which is close to 1 for large values of m.

Since the columns of H include all nonzero sequences of length m, the sum of any two columns is another column. In other words, there always exist three columns that are linearly dependent. Therefore, for Hamming codes, independent of the value of m, $d_{min} = 3$.

The weight distribution polynomial for the class of Hamming (n, k) codes is known and is expressed as

$$A(Z) = n + 1\ [(I + Z)^n + n(1 + Z)^{(n-1)/2}\ (1 - Z)^{(n+1)/2}]$$

To generate the H matrix for a (7, 4) Hamming code (corresponding to m - 3), we have to use all nonzero sequences of length 3 as columns of H. We can arrange these columns in such a way that the resulting code is systematic as

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

This is the parity check matrix derived in Example above.

Maximum-Length Codes

Maximum-length codes are duals of Hamming codes; therefore these are a family of (2'n - 1, m) codes for m >_ 3. The generator matrix of a maximum-length code is the parity check matrix of a Hamming code, and therefore its columns are all sequences of length m with the exception of the all-zero sequence. Therefore, the weight enumeration function for these codes is given by

$$A(Z) - 1 + (2^m - 1)Z^{m-1}$$

Reed-Muller Codes

Reed-Muller codes introduced by Reed (1954) and Muller (1954) are a class of linear block codes with flexible parameters that are particularly interesting due to the existence of simple decoding algorithms for them.

A Reed-Muller code with block length n = 2m and order r < m is an (n, k) linear block code with

n - 2n2
k (7.3-5) i
d - _r min 2m

whose generator matrix is given by

$$G = \begin{bmatrix} G_0 \\ G_1 \\ G_2 \\ \vdots \\ G_r \end{bmatrix}$$

where $\boldsymbol{G}_0$ is a 1 x n matrix of all is

$\boldsymbol{G}_0$ = (1 1 1 ….. 1)

And $\boldsymbol{G}_1$ is an m x n matrix whose columns are distinct binary sequences of length m put in natural binary order.

$$G_1 = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & \cdots & 1 & 1 \\ 0 & 1 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

$\boldsymbol{G}_2$ is an $\binom{m}{2}$ x n matrix whose rows are obtained by bitwise multiplication of two row of G2 at a time. Similarly, $\boldsymbol{G}_i$ for 2 < i ≤ r is a $\binom{m}{r}$ x n matrix whose rows are obtained by bitwise multiplication got r rows of $\boldsymbol{G}_2$ at a time.

The first-order Reed-Muller code with block length 8 is an (8,4) code with generator matrix.

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

This code can be obtained from a (7,3) maximum-length code by adding one extra parity bit to make the overall weight of each codeword even. This code has a minimum distance of 4. The second-order Reed-Muller code with block length 8 has the generator matrix.

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and has a minimum distance of 2.

Hadamard Codes

A Hadamard code is obtained by selecting as codewords the rows of a Hadamard matrix. A Hadamard matrix Mn is an n x n matrix (n is an even integer) of 1 s and Os with the property that any row differs from any other row in exactly 2 positions.t One row of the matrix contains all zeros. The other rows each contain 2 zeros and 2 ones.

or n = 2, the Hadamard matrix is

$$\boldsymbol{M}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Furthermore, from Mn, we can generate the Hadamard matrix M2n according to the relation

$$\boldsymbol{M}_{2n} = \begin{matrix} \boldsymbol{M}_n & \boldsymbol{M}_n \\ \boldsymbol{M}_n & \boldsymbol{M}_n \end{matrix}$$

Where $\boldsymbol{M}_n$ denotes the complement (Os replaced by is and vice versa) of Mn. We obtain

$$M_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

The complement of M_4 is

$$\overline{M}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Now the rows of $\boldsymbol{M}_4$ and $\boldsymbol{M}_4$ form a linear binary code of block length n = 4 having 2n = 8 code words. The minimum distance of the code is $\boldsymbol{d}_{min} = {}^{n}/_{2} = 2$.

we can generate Hadamard codes with block length $n = 2^m$, $k = \log_2 2_n = \log_2 2^{m+1} = m + 1$, and $d_{min} = {}^{n}/_{2} = 2^{m-1}$, where m is a positive integer. In addition to the important special cases where $n = 2^m$, Hadamard codes of other block lengths are possible, but the resulting codes are not linear.

Golay Code

The Golay code (Golay (1949)) is a binary linear (23, 12) code with d,';" = 7. The extended Golay code is obtained by adding an overall parity bit to the (23, 12) Golay code such that each codeword has even parity. The resulting code is a binary linear (24, 12) code with dn,1n = 8. The weight distribution polynomials of Golay code and extended Golay code are known and are given by

$AG(Z) = 1 + 253Z^7 + 506Z^8 + 12882Z^{11} + 1288Z^{12} + 506Z^{15} + 253Z^{16} + Z^{23}$

$AEG(Z) = 1 + 759Z^8 + 2576Z^{12} + 759Z^{16} + Z^{24}$

We discuss the generation of the Golay code in Section 7.9-5. 7.4

3.2 OPTIMUM SOFT DECISION DECODING OF LINEAR BLOCK CODES

In this section, we derive the performance of linear binary block codes on an AWGN channel when optimum (unquantized) soft decision decoding is employed at the receiver. The bits of a codeword may be transmitted by any one of the binary signaling methods described For our purposes, we consider binary (or quaternary) coherent PSK, which is the most efficient method, and binary orthogonal FSK with either coherent detection or noncoherent detection.

we know that the optimum receiver, in the sense of minimizing the average probability of a codeword error, for the AWGN channel can be realized as a parallel bank of M = 2k filters matched to the M possible transmitted waveforms. The outputs of the M matched filters at the end of each signaling interval, which encompasses the transmission of n binary symbols in the codeword, are compared, and the codeword corresponding to the largest matched filter output is selected. Alternatively, M cross-correlators can be employed. In either case, the receiver implementation can be simplified. That is, an equivalent optimum receiver can be realized by use of a single

filter (or cross-correlator) matched to the binary PSK waveform used to transmit each bit in the codeword, followed by a decoder that forms the M decision variables corresponding to the M code words.

To be specific, let rj , j = 1, 2, . . . , n, represent the n sampled outputs of the matched filter for any particular codeword. Since the signaling is binary coherent PSK, the output rj may be expressed either as

rj = $\sqrt{\varepsilon_c}$+ nj

when the jth bit of a codeword is a 1, or as

rj = - $\sqrt{\varepsilon_c}$ + ni

when the jth bit is a 0. The variables {nj} represent additive white Gaussian noise at the sampling instants. Each nj has zero mean and variance ½ No. From knowledge of the M possible transmitted code words and upon reception of {rj}, the optimum decoder forms the M correlation metrics.

CM„, = C(r, e„,) = E(2c„,j - 1) rj, in = 1, 2, . . . , M (7.4-3) j=1

$$CM_m = C(r, c_m) = \sum_{j=1}^{n} (2c_{mj} - 1) r_j, \quad m = 1, 2, \ldots, M$$

where c_{mj} denotes the bit in the jth position of the mth codeword. Thus, if c_{mj} = 1, the weighting factor $2c^{mj}$ - 1 = 1; and if c_{mj} = 0, the weighting factor $2c_{mj}$ - 1 = -1. In this manner, the weighting $2c_{mj}$ - 1 aligns the signal components in {rj} such that the correlation metric corresponding to the actual transmitted codeword will have a mean value n $\sqrt{\varepsilon_c}$, while the other M - 1 metrics will have smaller mean values.

Although the computations involved in forming the correlation metrics for soft decision decoding according to Equation 7.4-3 are relatively simple, it may still be impractical to compute Equation 7.4-3 for all the possible codewords when the number of codewords is large, e.g., M > 2^{10}. In such a case it is still possible to implement soft decision decoding using algorithms which employ techniques for discarding improbable codewords without computing their entire correlation metriC. Several different types of soft decision decoding algorithms have been described in the technical literature. The

interested reader is referred to the papers by Forney (1966b), Weldon (1971), Chase (1972), Wainberg and Wolf (1973), Wolf (1978), and Matis and Modestino (1982).

Block and Bit Error Probability in Soft Decision Decoding

We can use the general bounds on the block error probability derived in Equations to find bounds on the block error probability for soft decision decoding. The value of A defined by Equation has to be found under the specific modulation employed to transmit codeword components. we obtain

$$P_e \leq (A(Z) - 1)\Big|_{Z=e^{-\frac{R_C \mathcal{E}_b}{N_0}}}$$

where A(Z) is the weight enumerating polynomial of the code. The simple bound of Equation under soft decision decoding reduces to

$$P_e \leq (2^k - 1)e^{-R_c d_{\min} \mathcal{E}_b / N_0}$$

It is shown that for binary orthogonal signaling, for instance, orthogonal BFSK, we have Δ = e-Ec/N_o. Using this result, we obtain the simple bound

$\boldsymbol{P}_e$ < (2k - 1)e -$R_c d_{min}$-$E_b/2N_o$

for orthogonal BFSK modulation.

Using the inequality 2k - 1 < 2k = $e^{k\ In2}$, we obtain

$$P_e \leq e^{-\gamma_b\left(R_c d_{\min} - \frac{k \ln 2}{\gamma_b}\right)} \quad \text{for BPSK}$$

and

$$P_e \leq e^{-\frac{\gamma_b}{2}\left(R_c d_{\min} - \frac{k \ln 2}{\gamma_b}\right)} \quad \text{for orthogonal BFSK}$$

Where as usual yb denotes Eb/No, the SNR per bit.

When the upper bound in Equation is compared with the performance of an uncoded binary PSK system, whish is upper-bounded as ½ exp (-yb), we find that coding yields

as gain of approximately 10log ($R_c d_{min}$ – k 1n2/yb) dB. We may call this the coding gain. We note that its values depends on the code parameters and also on the SNR per bit yb. For large values of yb, the limit of the coding gain, i.e Rcdmin, is called the asymptotic coding gain.

Similar to the block error probability, we can use equation to bound the bit error probability for BFSK and orthogonal BFSK modulation. We obtain;

$$P_b \leq \frac{1}{k}\frac{\partial}{\partial Y}B(Y,Z)\Big|_{Y=1,Z=\exp\left(-\frac{R_c\mathcal{E}_b}{N_0}\right)} \qquad \text{for BPSK}$$

$$P_b \leq \frac{1}{k}\frac{\partial}{\partial Y}B(Y,Z)\Big|_{Y=1,Z=\exp\left(-\frac{R_c\mathcal{E}_b}{2N_0}\right)} \qquad \text{for orthogonal BFSK}$$

Soft Decision Decoding with Noncoherent Detection

In noncoherent detection of binary orthogonal FSK signaling, the performance is further degraded by the noncoherent loss. Here the input variables to the decoder are;

$$\begin{cases} r_{0j} = |\sqrt{\mathcal{E}_c} + N_{0j}|^2 \\ r_{0j} = |N_{1j}|^2 \end{cases}$$

for j = 1,2,……n, where (*Noj*) and (N_{1j}) represent complex-valued mutually statistically independent Gaussian random variables with zero mean and variable 2No. The correlation metric CM_1 is given as;

$$CM_1 = \sum_{j=1}^{n} r_{0j}$$

while the correlation metric corresponding to the codeword having weight wm is statistically equivalent to the correlation metric of a codeword in which c_{mj} = 1 for 1 ≤ j ≤ wm and c_{mj} = 0 for wm + 1 ≤ j ≤ n. Hence, CMm may be expressed as

$$CM_m = \sum_{j=1}^{w_m} r_{1j} + \sum_{j=w_m+1}^{n} r_{0j}$$

The difference between CM_1 and CM_m is

$$CM_1 - CM_m = \sum_{j=1}^{w_m}(r_{0j} - r_{1j})$$

and the pairwise error probability (PEP) is simply probability that CM1 – CMm < 0. But this difference is a special case of the general quadratic form in complex-valued Gaussian random variables considered in Chapter 11 and in Appendix B. The expression for the probability of error in deciding between CM1 and CMm is

$$P_2(m) = \frac{1}{2^{2w_m-1}} \exp\left(-\frac{1}{2}\gamma_b R_c w_m\right) \sum_{i=0}^{w_m-1} K_i \left(\frac{1}{2}\gamma_b R_c w_m\right)^i$$

where, by definition,

$$K_i = \frac{1}{i!} \sum_{r=0}^{w_m-1-i} \binom{2w_m-1}{r}$$

The union bound obtained by summing $P_2(m)$ over $2 \leq m \leq M$ provides us with an upper bound on the probability of a codeword error.

As an alternative, we may use the minimum distance instead of the weight distribution to obtain the looser upper bound.

$$P_e \leq \frac{M-1}{2^{2d_{\min}-1}} \exp\left(-\frac{1}{2}\gamma_b R_c d_{\min}\right) \sum_{i=0}^{d_{\min}-1} K_i \left(\frac{1}{2}\gamma_b R_c d_{\min}\right)^i$$

As measure of the non-coherent combining loss inherent in the square-law detection and combining of the n elementary binary FSK waveforms in a codeword where dmin is used in place of L. The loss obtained is relative to the case in which the n elementary binary FSK waveforms are first detected coherently and combined, and then the sums are square-law-detected or envelope-detected to yield the M decision variables. The binary probability for the latter case is;

$$P_2(m) = \frac{1}{2} \exp\left(-\frac{1}{2}\gamma_b R_c w_m\right)$$

and hence

$$P_e \leq \sum_{m=2}^{M} P_2(m)$$

If d_{min} is used instead of the weight distribution, the union bound for the codeword error probability in the latter case is;

$$P_e \leq \frac{1}{2}(M-1) \exp\left(-\frac{1}{2}\gamma_b R_c d_{\min}\right)$$

We have previously seen in equation that the channel bandwidth required to transmit the coded waveforms, when binary PSK is used to transmit each bit, is given by;

$$W = \frac{R}{R_c}$$

From Equation, the bandwidth requirement for an uncoded BPSK scheme is R. Therefore, the bandwidth expansion factor Be for the coded waveforms is

$B_e = {}^{1}/R_c$

Comparison with Orthogonal Signaling

We are now in a position to compare the performance characteristics and bandwidth requirements of coded signaling with orthogonal signaling. Orthogonal signals are more power-efficient compared to BPSK signaling, but using them requires large bandwidth. We have also seen that using coded BPSK signals results in a moderate expansion in bandwidth and, at the same time, by providing the coding gain, improves the power efficiency of the system.

Let us consider two systems, one employing orthogonal signaling and one employing coded BPSK signals to achieve the same performance. To have equal bounds on the error probability, we must have k = 2R,dmin. Under this condition, the dimensionality of the orthogonal signals, given by N = M = 2k, is given by N = 2R,dmin. The dimensionality of the BPSK code waveform is n = k/R, = 2dmin. Since dimensionality is proportional to the bandwidth, we conclude that

$$\frac{W_{orthogonal}}{W_{coded\ BPSK}} = \frac{2^2 R_c d_{min}}{2d_{min}}$$

For example, suppose we use a (63, 30) binary code that has a minimum distance dmin = 13.

3.3 Hard Decision Decoding Of Linear Block Codes

The bounds given on the performance of coded signaling waveforms on the AWGN channel are based on the premise that the samples from the matched filter or cross-correlator are not quantized. Although this processing yields the best performance, the basic limitation is the computational burden of forming M correlation metrics and

comparing these to obtain the largest. The amount of computation becomes excessive when the number M of codewords is large.

To reduce the computational burden, the analog samples can be quantized and the decoding operations are then performed digitally. In this section, we consider the extreme situation in which each sample corresponding to a single bit of a codeword is quantized to two levels: 0 and 1. That is, a hard decision is made as to whether each transmitted bit in a codeword is a 0 or a 1. The resulting discrete-time channel (consisting of the modulator, the AWGN channel, and the modulator/demodulator) constitutes a BSC with crossover probability p. If coherent PSK is employed in transmitting and receiving the bits in each codeword, then

$$p = Q\left(\sqrt{\frac{2\mathcal{E}_c}{N_0}}\right) = Q\left(\sqrt{2\gamma_b R_c}\right)$$

On the other hand, if FSK is used to transmit the bits in each codeword, then

$$p = Q\left(\sqrt{\gamma_b R_c}\right)$$

for coherent detection and

$$p = \frac{1}{2}\exp\left(-\frac{1}{2}\gamma_b R_c\right)$$

for noncoherent detection.

Minimum-Distance (Maximum-Likelihood) Decoding

The n bits from the detector corresponding to a received codeword are passed to the decoder, which compares the received codeword with the M possible transmitted codewords and decides in favor of the codeword that is closest in Hamming distance (number of bit positions in which two codewords differ) to the received codeword. This minimum-distance decoding rule is optimum in the sense that it results in a minimum probability of a codeword error for the binary symmetric channel.

A conceptually simple, albeit computationally inefficient, method for decoding is to first add (modulo-2) the received codeword vector to all the M possible transmitted codewords c_m to obtain the error vectors e,n. Hence, e_m represents the error event that must have occurred on the channel in order to transform the codeword cm to the particular received codeword. The number of errors in transforming cm into the received codeword is just equal to the number of 1 s in e_m. Thus, if we simply compute the weight of each of the M error vectors (e_m} and decide in favor of the

codeword that results in the smallest weight error vector, we have, in effect, a realization of the minimum-distance decoding rule.

Syndrome and Standard Array

A more efficient method for hard decision decoding makes use of the parity check matrix H. To elaborate, suppose that c_m is the transmitted codeword and y is the received sequence at the output of the detector. In general, y may be expressed as

$$y = c_m + e$$

where e denotes an arbitrary error vector. The product yH^t yields.

$$s = \boldsymbol{yH^t}$$

$$= c_m\boldsymbol{H^t} + \boldsymbol{eH^t}$$

$$= \boldsymbol{eH^t}$$

where the (n - k)-dimensional vector s is called the syndrome of the error pattern. In other words, the vector s has components that are zero for all parity check equations that are satisfied and nonzero for all parity check equations that are not satisfied. Thus, s contains the pattern of failures in the parity checks.

We emphasize that the syndrome s is a characteristic of the error pattern and not of the transmitted codeword. If a syndrome is equal to zero, then the error pattern is equal to one of the codewords. In this case we have an undetected error. Therefore, an error pattern remains undetected if it is equal to one of the nonzero codewords. Hence, from the 2'2 - 1 error patterns (the all-zero sequence does not count as an error), 2k - 1 are not detectable; the remaining 2'2 - 2k nonzero error patterns can be detected, but not all can be corrected because there are only 2n-k syndromes and, consequently, different error patterns result in the same syndrome. For ML decoding we are looking for the error pattern of least weight among all possible error patterns.

Suppose we construct a decoding table in which we list all the 2k possible code-words in the first row, beginning with the all-zero codeword c 1 = 0 in the first (leftmost) column. This all-zero codeword also represents the all-zero error pattern. After com pleting the first row, we put a sequence of length n which has not been included in the first row (i.e., is not a codeword) and among all such sequences has the minimum weight in the first column of the second row, and we

call it e2. We complete the second row of the table by adding e2 to all codewords and putting the result in the column corresponding to that codeword. After the second row is complete, we look among all sequences of length n that have not been included in the first two rows and choose a sequence of minimum weight, call it e3, and put it in the first column of the third row; and complete the third row similar to the way we completed the second row. This process is continued until all sequences of length n are used in the table. We obtain an n x (n - k) table as follows:

C1 = 0	C2	C3	...	C2k		
e2			C2 + e2	C3 + e2	...	C2k + e2
e3			C2 + e3	C3 + e3	...	C2k + e3
e2n-k			C2 + e2n-k	C3 + e2n-k	...	C2k .+ e2n-k

This table is called a standard array. Each row, including the first, consists of k possible received sequences that would result from the corresponding error pattern in the first column. Each row is called a coset, and the first (leftmost) codeword (or error pattern) is called a coset leader. Therefore, a coset consists of all the possible received sequences resulting from a particular error pattern (coset leader). Also note that by construction the coset leader has the lowest weight among all coset members.

Example: Let us construct the standard array for the (5, 2) systematic code with generator matrix given by

$$G = \begin{matrix} 1\ 0\ 1\ 0\ 1 \\ 0\ 1\ 0\ 1\ 1 \end{matrix}$$

The Standard Array for Example 7.5-1

00000	01011	10101	11110
00001	01010	10100	11111
00010	01001	10111	11100
00100	01111	10001	11010
01000	00011	11101	10110
10000	11011	00101	01110
11000	10011	01101	00110
10010	11001	00111	01100

This code has a minimum distance dmin = 3. Note that in this code, the coset leaders consist of the all-zero error pattern, five error patterns of weight 1, and two error

patterns of weight 2. Although many more double error patterns exist, there is room for only two to complete the table.

Now, suppose that ej is a coset leader and that cm was the transmitted codeword. Then the error pattern ej would result in the received sequence;

$y=c_m+ei$

The syndrome is

$s = yH^t = (c_m + ei)\ H^t = c_mH^t + ei\boldsymbol{H}^t = ej\boldsymbol{H}^t$

Clearly, all received sequences in the same coset have the same syndrome, since the latter depends only on the error pattern. Furthermore, each coset has a different syndrome. This means that there exists a one-to-one correspondence between cosets (or coset leaders) and syndromes.

The process of decoding the received sequence y basically involves finding the error sequence of the lowest weight et such that s = Y H' = ejH^t. Since each syndrome s corresponds to a single coset, the error sequence ej is simply the lowest member of the coset, i.e., the coset leader. Therefore, after the syndrome is found, it is sufficient to find the coset leader corresponding to the syndrome and add the coset leader to y to obtain the most likely transmitted codeword.

The above discussion makes it clear that coset leaders are the only error patterns that are correctable. To sum up the above discussion, from all possible 2^k - 1 nonzero error patterns, 2k - 1 corresponding to nonzero codewords are not detectable, and 2' - 2k are detectable of which only 2^n-k - 1 are correctable.

Consider the (5, 2) code with the standard array.

Now suppose the actual error vector on the channel is

e=(1 0 1 0 0)

The syndrome computed for the error is s = (0 0 1). Hence, the error determined from the table is e = (0 0 0 0 1). When e is added to y, the result is a decoding

Syndromes and Coset Leaders for

Syndrome	**Error Pattern**
000	00000
001	00001
010	00010
100	00100

011	01000
101	10000
110	11000
111	10010

error. In other words, the (5, 2) code corrects all single errors and only two double errors, namely, (1 1 0 0 0) and (1 0 0 1 0).

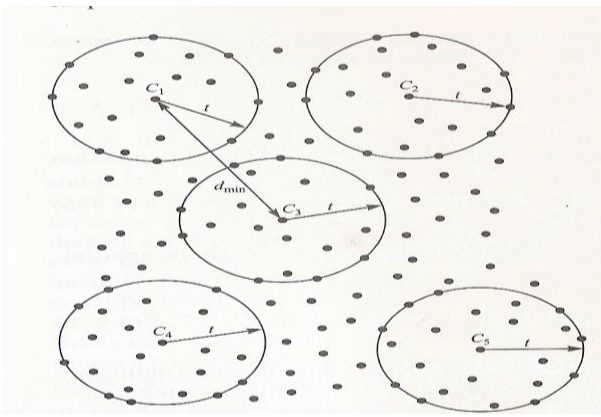
Error Detection and Error Correction Capability of Block Codes

It is clear from the discussion above that when the syndrome consists of all zeros, the received codeword is one of the 2k possible transmitted codewords. Since the minimum separation between a pair of codewords is dmin, it is possible for an error pattern of weight drain to transform one of these 2k codewords in the code to another codeword. When this happens, we have an undetected error. On the other hand, if the actual number of errors is less than d_{min}, the syndrome will have a nonzero weight. When this occurs, we have detected the presence of one or more errors on the channel. Clearly, the (n, k) block code is capable of detecting up to d_{min} - 1 errors. Error detection may be used in conjunction with an automatic repeat-request (ARQ) scheme for retransmission of the codeword.

The error correction capability of a code also depends on the minimum distance. However, the number of correctable error patterns is limited by the number of possible syndromes or coset leaders in the standard array. To determine the error correction capability of an (n, k) code, it is convenient to view the 2k codewords as points in an n-dimensional space. If each codeword is viewed as the center of a sphere of radius (Hamming distance) t, the largest value that t may have without intersection (or tangency) of any pair of the 2k spheres is t = [z (d .,i„ - 1)], where Lx] denotes the largest integer contained in x. Within each sphere lie all the possible received codewords of distance less than or equal to t from the valid codeword. Consequently, any received code vector that falls within a sphere is decoded into the valid codeword at the center of the sphere. This implies that an (n, k) code with minimum distance d_{min} is capable of correcting t = L2(d_{min} - 1)J errors.

As described above, a code may be used to detect d_{min} - 1 errors or to correct t = (½ (d_{min} - 1) errors. Clearly, to correct t errors implies that we have detected t errors.

However, it is also possible to detect more than t errors if we compromise in the error correction capability of the code. For example, a code with d_{min} = 7 can correct



up to t = 3 errors. If we wish to detect four errors, we can do so by reducing the radius of the sphere around each codeword from 3 to 2. Thus, patterns with four errors are detectable, but only patterns of two errors are correctable. In other words, when only two errors occur, these are corrected; and when three or four errors occur, the receiver may ask for a retransmission. If more than four errors occur, they will go undetected if the codeword falls within a sphere of radius 2. Similarly, for din = 7, five errors can be detected and one error corrected. In general, a code with minimum distance d_{min} can detect e_d errors and correct e, errors, where

$$e_d + e_c \leq d_{\min} - 1$$

and

$$e_c \leq e_d$$

Block and Bit Error Probability for Hard Decision Decoding

In this section we derive bounds on the probability of error for hard decision decoding of linear binary block codes based on error correction only.

From the above discussion, it is clear that the optimum decoder for a binary symmetric channel will decode correctly if (but not necessarily only if) the number of errors in a codeword is less than one-half the minimum distance din of the code. That is, any number of errors up to

$$t = \left\lfloor \frac{1}{2}(d_{\min} - 1) \right\rfloor$$

is always correctable. Since the binary symmetric channel is memoryless, the bit errors occur independently. Hence, the probability of m errors in a block of n bits is

$$P(m,n) = \binom{n}{m} p^m (1-p)^{n-m}$$

and, therefore, the probability of a codeword error is upper-bounded by the expression

$$P_e \leq \sum_{m=t+1}^{n} P(m,n)$$

For high signal-to-noise ratios, i.e., small values of p, can be approximated by its first term, and we have

$$P_e \approx \binom{n}{t+1} p^{t+1} (1-p)^{n-t-1}$$

This equation states that when 0 is transmitted, the probability of error almost entirely is equal to the probability of receiving sequences of weight *t + 1.* To derive an approximate bound on the error probability of each binary symbol in a codeword, we note that if 0 is sent and a sequence of weight *t + 1* is received, the decoder will decode the received sequence of weight *t + 1* to a codeword at a distance at most *t* from the received sequence and hence a distance of at most 2t + 1 from 0. But since the minimum weight of the code is 2t + 1, the decoded codeword has to be of weight 2t + 1. This means that for each highly probable block error we have 2t + 1 bit errors in the codeword components.

$$P_{bs} \approx \frac{2t+1}{n} \binom{n}{t+1} p^{t+1} (1-p)^{n-t-1}$$

Equality holds if the linear block code is a *perfect code. To* describe the basic characteristics of a perfect code, suppose we place a sphere of radius *t* around each of the possible transmitted codewords. Each sphere around a codeword contains the

set of all codewords of Hamming distance less than or equal to t from the codeword. Now, the number of codewords in a sphere of radius t = ½ (d_{min} - 1)] is

$$1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{t} = \sum_{i=0}^{t} \binom{n}{i}$$

Since there are M = 2^k possible transmitted codewords, there are 2^k nonoverlapping spheres, each having a radius t. The total number of codewords enclosed in the 2^k spheres cannot exceed the 2' possible received codewords. Thus, a t-error correcting code must satisfy the inequality

$$2^k \sum^{t} \binom{n}{i} \leq 2^n$$

or, equivalently,

$$2^{n-k} \geq \sum_{i=0}^{t} \binom{n}{i}$$

A perfect code has the property that all spheres of Hamming distance t = ½ (d_{min} - 1)J around the M = 2k possible transmitted codewords are disjoint and every received codeword falls in one of the spheres. Thus, every received code word is at most at a distance t from one of the possible transmitted codeword For such a code, all error patterns of weight less than or equal to t are corrected by the optimum (minimum-distance) decoder. On the other hand, any error pattern of weight t + 1 or greater cannot be corrected. The reader can easily verify that the Hamming codes, which have the parameters n = 2n-k - 1, drain = 3, and t = 1, are an example of perfect codes. The (23, 12) Golay code has parameters drain = 7 and t = 3. It can be easily verified that this code is also a perfect code. These two nontrivial codes and the trivial code consisting of two codewords of odd length n and d_{min} = n are the only perfect binary block codes.

A quasi-perfect code is characterized by the property that all spheres of Hamming radius t around the M possible transmitted codewords are disjoint and every received codeword is at most at a distance t + 1 from one of the possible transmitted codewords. For such a code, all error patterns of weight less than or equal to t and some error patterns of weight t + 1 are correctable, but any error pattern of weight t

+ 2 or greater leads to incorrect decoding of the codeword. Clearly, Equation 7.5-6 is an upper bound on the error probability, and

is a lower bound.

A more precise measure of the performance for quasi-perfect codes can be obtained by making use of the inequality in Equation . That is, the total number of codewords outside the 2k spheres of radius t is

$$N_{t+1} = 2^n - 2^k \sum_{i=0}^{t} \binom{n}{i}$$

If these codewords are equally subdivided into 2k sets and each set is associated with one of the 2k spheres, then each sphere is enlarged by the addition of

$$\beta_{t+1} = 2^{n-k} - \sum_{i=0}^{t} \binom{n}{i}$$

codewords having distance t + 1 from the transmitted codeword. Consequently, of the (t+1) error patterns of distance t + 1 from each codeword, we can correct, Bt+i error patterns. Thus, the error probability for decoding the quasi- perfect code may be expressed as

$$P_e = \sum_{m=t+2}^{n} P(m,n) + \left[\binom{n}{t+1} - \beta_{t+1} \right] p^{t+1}(1-p)^{n-t-1}$$

Another pair of upper and lower bounds is obtained by considering two codewords that differ by the minimum distance. First, we note that ***P***e cannot be less than the probability of erroneously decoding the transmitted codeword as its nearest neighbor, which is at a distance d_{min} from the transmitted codeword. That is

$$P_e \geq \sum_{m=\lfloor d_{\min}/2 \rfloor + 1}^{d_{\min}} \binom{d_{\min}}{m} p^m (1-p)^{d_{\min}-m}$$

On the other hand, Pe cannot be greater than 2k -1 times the probability of erroneously decoding the transmitted codeword as its nearest neighbor, which is at a distance d min from the transmitted codeword. That is a union bound, which is expressed as

$$P_e \leq (2^k - 1) \sum_{m=\lfloor d_{\min}/2 \rfloor + 1}^{d_{\min}} \binom{d_{\min}}{m} p^m (1-p)^{d_{\min}-m}$$

When M = 2k is large, the lower bound in Equation 7.5-15 and the upper bound in Equation 7.5-16 are very loose.

General bounds on block and bit error probabilities under hard decision decoding are obtained by using relations derived in Equations 7.2-39, 7.2-43, and 7.2-48. The value of A for hard decision decoding was found in Example 6.8-1 and is given by 0 = √4p(1-p). The results are

$$P_e \leq (A(Z) - 1)\Big|_{Z=\sqrt{4p(1-p)}}$$

$$P_e \leq (2^k - 1)\,[4p(1-p)]^{\frac{d_{\min}}{2}}$$

$$P_b \leq \frac{1}{k}\frac{\partial}{\partial Y}B(Y, Z)\Big|_{Y=1, Z=\sqrt{4p(1-p)}}$$

3.4 Comparison of Performance between Hard Decision And Soft Decision Decoding

It is both interesting and instructive to compare the bounds on the error rate performance of linear block codes for soft decision decoding and hard decision decoding on an AWGN channel. For illustrative purposes, we use the Golay (23, 12) code, which has the relatively simple weight distribution given in Equation 7.3-15. As stated previously, this code has a minimum distance d_{min} = 7.

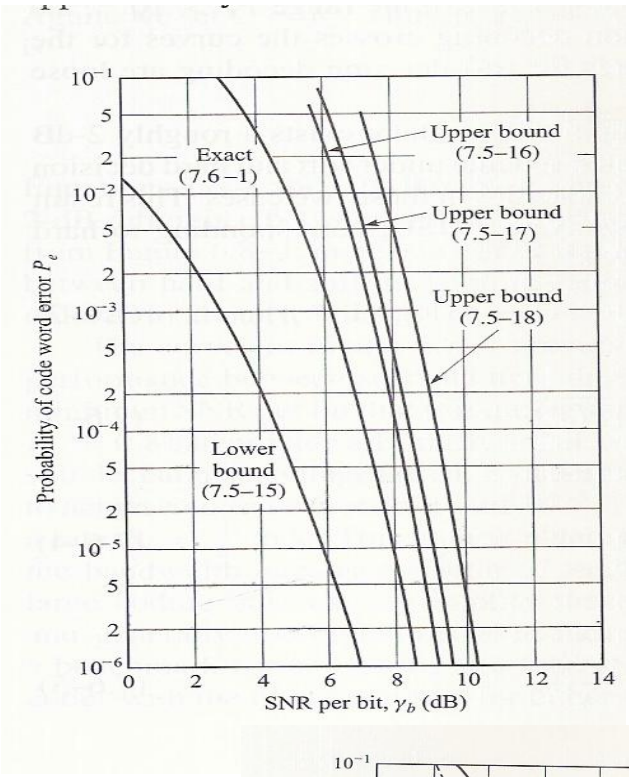
First we compute and compare the bounds on the error probability for hard decision decoding. Since the Golay (23, 12) code is a perfect code, the exact error probability for hard decision decoding is given by equation as

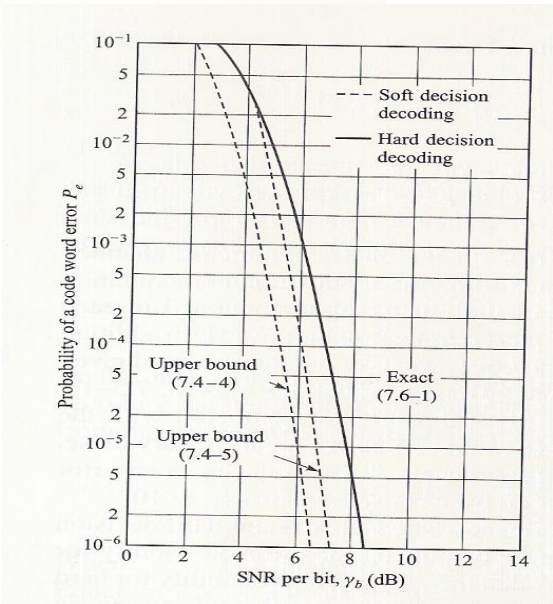
$$P_e = \sum_{m=4}^{23} \binom{23}{m} p^m (1-p)^{23-m}$$

$$= 1 - \sum_{m=0}^{3} \binom{23}{m} p^m (1-p)^{23-m}$$

where p is the probability of a binary digit error for the binary symmetric channel. Binary (or four-phase) coherent PSK is assumed to be the modulation/demodulation technique for the transmission and reception of the binary digits contained in each codeword. Thus, We observe that the lower bound is very loose. At $P_e = 10^{-5}$, the lower bound is off by approximately 2 dB from the exact error probability. All three upper bounds are very loose for error rates above P_e = 10-Z.

It is also interesting to compare the performance between soft and hard decision decoding. For this comparison, we use the upper bounds on the error probability for soft decision decoding given by Equation and the exact error probability for hard decision decoding given illustrates these performance characteristics. We observe that the two bounds for soft decision decoding differ by approximately 0.5 dB at $P_e = 10^{-6}$ and by approximately 1 dB at $P_e = 10^{-2}$. We also





observe that the difference in performance between hard and soft decision decoding is approximately 2 dB in the range 10-2 < Pe < 10-6. In the range Pe > 10-2, the curve of the error probability for hard decision_ decoding crosses the curves for the bounds. This behavior indicates that the bounds for soft decision decoding are loose when Pe > 10-2.

There exists a roughly 2-dB gap between the cutoff rates of a BPSK modulated scheme under soft and hard decision decoding. A similar gap also exits between the capacities in these two cases. This result can be shown directly by noting that the capacity of a BSC, corresponding to hard decision decoding, is given by Equation

$$C = 1 - H_2(p) = 1 + p\log_2 p + (1-p)\log_2(1-p)$$

where

$$p = Q\left(\sqrt{2\gamma_b R_c}\right)$$

For small values of R_c we can use the approximation

$$Q(\epsilon) \approx \frac{1}{2} - \frac{\epsilon}{\sqrt{2\pi}} \qquad \epsilon > 0$$

to obtain

$$p \approx \frac{1}{2} - \sqrt{\frac{\gamma_b R_c}{\pi}}$$

and using the approximation.

$$\log_2(1+x) \approx \frac{x - \frac{1}{2}x^2}{\ln 2}$$

we obtain

$$C = \frac{2}{\pi \ln 2}\gamma_b R_c$$

Now we set C = Rc. Thus, in the limit as Rc approaches zero, we obtain the result.

$$\gamma_b = \frac{1}{2}\pi \ln 2 \sim 0.37 \text{ dB}$$

The capacity of the binary-input AWGN channel wit soft decision decoding can be computed in a similar manner. The expression for the capacity in bits per code symbols, derived in equation to can be approximately for low values of Rc as

$$C \approx \frac{\gamma_b R_c}{\ln 2} \qquad (7.6\text{–}9)$$

Again, we set $C = R_c$. Thus, as $R_c \to 0$, the minimum SNR per bit to achieve capacity is

$$\gamma_b = \ln 2 \sim -1.6 \text{ dB} \qquad (7.6\text{–}10)$$

clearly shown that at low SNR values there exists roughly a 2-dB difference between the performance of hard and soft decision decoding. As seen, increasing SNR results in a decrease in the performance difference between hard and soft decision decoding. For example, at *R,* = 0.8, the difference reduces to about 1.5 dB.

The curves provide more information than just the difference in performance between soft and hard decision decoding. These curves also specify the minimum SNR per bit that is required for a given code rate. For example, a code rate of R, = 0.8 can provide arbitrarily small error probability at an SNR per bit of 2 dB, when soft decision decoding is used. By comparison, an uncoded binary PSK requires 9.6 dB to achieve an error probability of 10^{-5}. Hence, a 7.6-dB gain is possible by employing a rate R, = s code. This gain is obtained by expanding the bandwidth by 25% since the bandwidth expansion factor of such a code is *1/R,* = 1.25. To achieve such a large coding gain usually implies the use of an extremely long block length code, and generally a complex decoder. Nevertheless, the curves provide a benchmark for comparing the coding gains achieved by practically implementable codes with the ultimate limits for either soft or hard decision decoding.

3.5 Bounds on minimum distance of linear block codes

The expressions for the probability of error derived in this module for soft decision and hard decision decoding of linear binary block codes clearly indicate the importance of the minimum-distance parameter in the performance of the code. If we consider soft decision decoding, for example, the upper bound on the error probability given by Equation indicates that, for a given code rate R, = k/n, the probability of error in an AWGN channel decreases exponentially with d_{min}. When this bound is used in conjunction with the lower bound on d_{min} given below, we obtain an upper bound on ***P***e, the probability of a codeword error. Similarly, we may use the upper bound given by Equation for the probability of error for hard decision decoding in conjunction with the lower bound on d_{min} to obtain an upper bound on the error probability for linear binary block codes on the binary symmetric channel.

On the other hand, an upper bound on d_{min} can be used to determine a lower bound on the probability of error achieved by the best code. For example, suppose that hard decision decoding is employed. In this section we study some bounds on minimum distance of linear block codes.

Singleton Bound

The Singleton bound is obtained using the properties of the parity check matrix H. Recall that the minimum distance of a linear block code is equal to the minimum number of columns of H, the parity check matrix, that are linearly dependent. From this we conclude that the rank of the parity check matrix is equal to d_{min} - 1. Since the parity check matrix is an (n - k) x n matrix, its rank is at most n - k. Hence,

$d_{min} - 1 \leq n - k$
or
$d_{min} \leq n - k + 1$

The bound given in Equation is called the Singleton bound. Since dmin - 1 is approximately twice the number of errors that a code can correct, from this Equation, we conclude that the number of parity checks in a code must be at least equal to twice the number of errors a code can correct. Although the proof of the Singleton bound presented here was based on the linearity of the code, this bound applies to all block codes, linear and nonlinear, binary and nonbinary.

Codes for which the Singleton bound is satisfied with equality, i.e., codes for which dmin = n - k + 1, are called maximum-distance separable, or MDS, codes. Repetition codes, and their duals are examples of MDS codes. In fact these codes are the only binary MDS codes.

Dividing both sides of the Singleton bound by n, we have;

$$\frac{d_{\min}}{n} \leq 1 - R_c + \frac{1}{n}$$

If we define

$$\delta_n = \frac{d_{\min}}{n}$$

we have

$$\delta_n \leq 1 - R_c + \frac{1}{n}$$

Note that $d_{min}/2$ is roughly the number of errors that a code can correct. Therefore;

$$\frac{1}{2}\delta_n \approx \frac{t}{n}$$

i.e., $^{\partial n}/_{2}$ approximately represents the fraction of correctable errors in transmission of n bits.

If we define $\delta = \lim_{n\to\infty} \delta_n$, we conclude that as $n \to \infty$,

$$\delta \leq 1 - R_c$$

This is the asymptotic form of the Singleton bound.

Hamming Bound

The Hamming or sphere packing bound was previously developed in our study of the performance of hard decision decoding and is given by equation as

$$2^{n-k} \geq \sum_{i=0}^{t} \binom{n}{i}$$

Taking the logarithm and dividing by n result in

$$1 - R_c \geq \frac{1}{n} \log_2 \left[\sum_{i=0}^{t} \binom{n}{i} \right]$$

or

$$1 - R_c \geq \frac{1}{n} \log_2 \left[\sum_{i=0}^{\lfloor \frac{d_{\min}-1}{2} \rfloor} \binom{n}{i} \right]$$

This relation gives an upper bound for d_{min} in terms of n and k, known as the Hamming bound. Note that the proof of the Hamming bound is independent of the linearity of the code; therefore this bound applies to all block codes. For the q-ary block codes the Hamming bound yields.

$$1 - R_c \geq \frac{1}{n} \log_q \left[\sum_{i=0}^{t} \binom{n}{i} (q-1)^i \right]$$

It is shown that for large n the right-hand side of equation can be approximately by;

$$\sum_{i=0}^{t} \binom{n}{i} \approx 2^{nH_b(\frac{t}{n})}$$

Where Hb(.) is the binary entropy function. Using this approximately, we see that the asymptotic form of the Hamming bound for binary codes becomes.

$$H_b\left(\frac{\delta}{2}\right) \leq 1 - R_c$$

The hamming bound is tight for high-rate codes.

Plotkin Bound

The Plotkin bound due to Plotkin states that for any q-ary block code we have;

$$\frac{d_{\min}}{n} \leq \frac{q^k - q^{k-1}}{q^k - 1}$$

For binary codes this bound becomes

$$d_{\min} \leq \frac{n2^{k-1}}{2^k - 1}$$

The proof is based on noting that the minimum distance of a code cannot exceed its average codeword weight.

Another version of the Plotkin bound, given in equation forbinary codes, is tighter for higher-rate codes:

$$d_{\min} \leq \min_{1 \leq j \leq k} (n - k + j) \frac{2^{j-1}}{2^j - 1}$$

A simplified version of this bound, obtained by choosing j = 1 + $\log_2 \boldsymbol{d}_{\min}$, result in

$$2d_{\min} - 2 - \lfloor \log_2 d_{\min} \rfloor \leq n - k$$

The asymptotic form of this bound with the assumption of $\delta \leq \frac{1}{2}$ is

$$\delta \leq \frac{1}{2}(1 - R_c)$$

Elias Bound

The asymptotic form of the Elias bound states that for any binary code with δ ≤ ½ we have

$$H_b\left(\frac{1}{2}\left(1 - \sqrt{1 - 2\delta}\right)\right) \leq 1 - R_c$$

The Elias bound also applies to nonbinary codes. For nonbinary codes this bound states that for any q-ary code with $\delta \leq 1 - {}^{1}/_{q}$ we have;

$$H_q\left(\frac{q-1}{q}\left(1-\sqrt{1-\frac{q}{q-1}\delta}\right)\right) \leq 1 - R_c$$

where $H_q(\cdot)$ is defined by

$$H_q(p) = -p\log_q p - (1-p)\log_q(1-p) + p\log_q(q-1)$$

for $0 \leq p \leq 1$.

McEliece-Rodemich-Rumsey-Welch (MRRW) Bound

McEliece-Rodemich-Rumsey-Welch (MRRW) Bound derived by McEliece et al (1977) is the tightest known bound for low to moderate rates. Thos bound has two forms; the simpler form has the asymptotic form given by

for binary codes and for $\delta \leq ½$. This bound is derived based on linear programming techniques.

$$R_c \leq H_b\left(\frac{1}{2} - \sqrt{\delta(1-\delta)}\right)$$

Varshamov-Gilbert Bound

All bounds stated so far give the necessary conditions that must be stratified by the three main parameters n, k, and d for a block code. the Varshamov-Gilbert bound due to Gilbert (1952) and Varshamov (1957) gives the sufficient conditions for the existence of an (n, k) code with minimum distance d_{min}. The Varshamov-Gilbert bound in fact goes further to prove the existence of a linear block code with the given parameters.

The Varshamov-Gilbert states that if the inequality;

$$\sum_{i=0}^{d-2}\binom{n-1}{i}(q-1)^i < q^{n-k}$$

Is satisfied, the there exists a q-ary (n, k) linear block code with minimum distance $d_{min} \geq d$. For the binary case the Varshamov-Gilbert bound becomes.

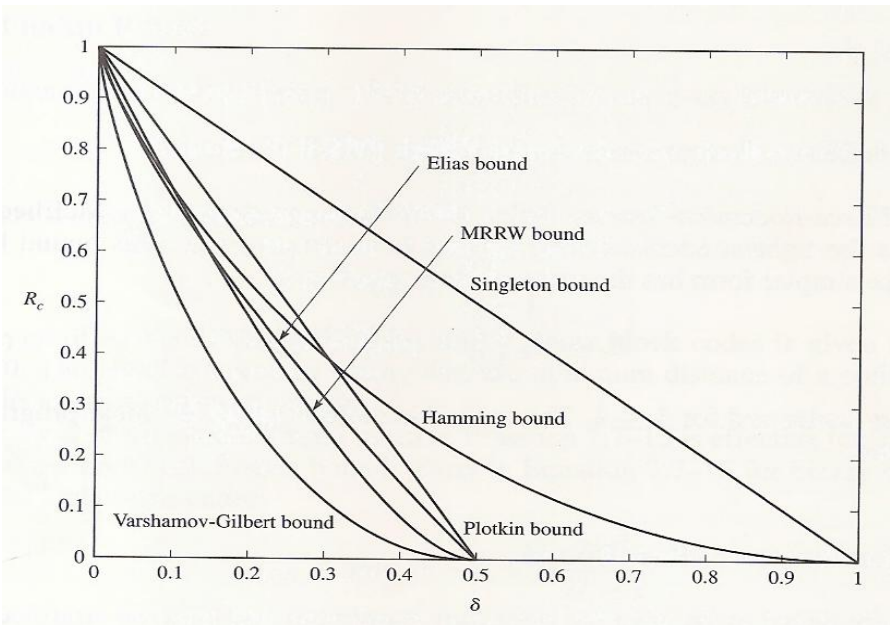
$$\sum_{i=0}^{d-2} \binom{n-1}{i} < 2^{n-k} \qquad (7.7\text{–}24)$$

The asymptotic version of the Varshamov-Gilbert bound states that if for $0 < \delta \leq 1 - \frac{1}{q}$ we have

$$H_q(\delta) < 1 - R_c \qquad (7.7\text{–}25)$$

Where Hq(.) is given , then there exists a q-ary (n, $R_c n$) linear block code with minimum distance of at lease δn.

A comparison of the asymptotic version of the bounds discussed above is shown in Figure below for the binary codes. As seen in the figure, the tightest asymptotic upper bounds are the Elias and the MRRW bounds. We add here that there exist a second



version of the MRRW bound that is better than the Elias bound at higher rates. The ordering of the bounds shown on this plot is only an indication of how these bounds compare as n --> ∞. The region between the tightest upper bound and the Varshamov-Gilbert lower bound can still be a rather wide region for certain block lengths. For instance, for a (127, 33) code the best upper bound and lower bound yield d_{min} = 48 and d_{min} = 32, respectively.

4.0 Conclusion

In this unit, we have considered linear block codes. These codes are mainly used with hard decision decoding that employs the built-in algebraic structure of the code based on the properties of finite fields. Hard decision decoding of these codes result in a binary symmetric channel model consisting of the binary modulator, the ware form channel and the optimum binary detector.

5.0 Summary

The decoder of these codes tries to find the code word at the minimum Hamming distance from the output of the BSC. The goal in designing good linear block codes is to find the code with highest minimum distance for a given n and k.

6.0 Tutor Marked Assignment

A code C consists of all binary sequences of length 6 and weight 3. 1. Is this code a linear block? Why?

2. What is the rate of this code? What is the minimum distance of this code? What is the minimum weight of this code?

3. If the code is used for error detection, how many errors can it detect?

4. If the code on a binary symmetric channel with crossover probability of P, what is the probability that an undetectable error occurs?

5. Find the smallest linear block code C, such that $C \subseteq C1$ (by the smallest code we mean the code with the fewest code words).

7.0 References and Further Reading

Key/papers in the Development of Coding Theory by Berlekamp. Coding Techniques for Noisy Channel by Elias (1954, 1955)

UNIT 3: TRELLIS AND GRAPH BASED CODES

1.0 Introduction

This is another class of codes whose structure is more conveniently described in terms of trellises or graphs. For this family of codes, soft decision decoding is possible, and in some cases performance very close to channel capacity is achievable.

2.0 Objectives

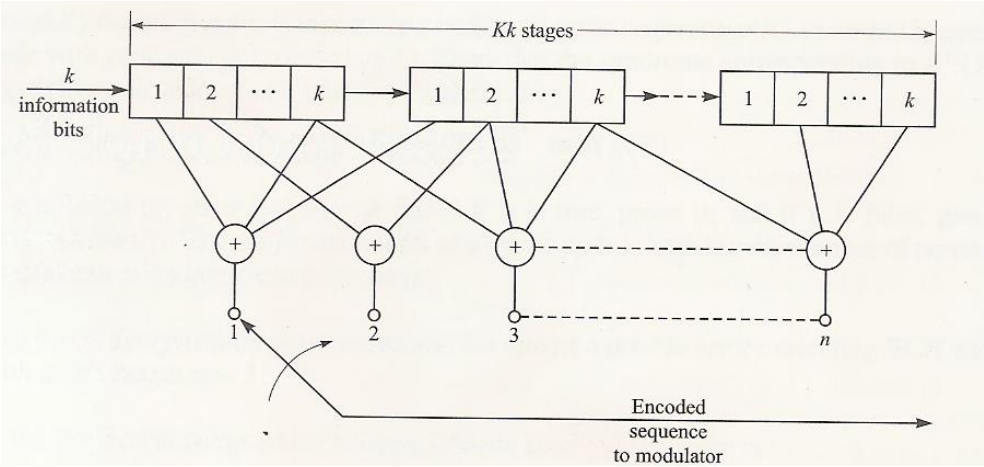
At the end of this unit, you should be able to;

- Understand the structure of convolutional codes.
- Explain decoding of convolutional codes;
- Discuss punctured convolutional codes.

3.1 The structure of convolutional codes

A convolutional code is generated bypassing the information sequence to be transmitted through a linear finite-state shift register. In general, the shift register consists of K (k-

bit) stages and n linear algebraic function generators. The input data to the encoder, which is assumed to be binary, is shifted into and along the shift register k bits at a time. The number of output bits for each k-bit input sequence is n bits. Consequently, the code rate is defined as R, = k/ n, consistent with the definition of the code rate for a block code. The parameter K is called the constraint length of the convolution code.



Convolutional encoder

One method for describing a convolutional code is to give its generator matrix, just as we did for block codes. In general, the generator matrix for a convolutional code is semi-infinite since the input sequence is semi-infinite in length. As an alternative to specifying the generator matrix, we shall use a functionally equivalent representation in which we specify a set of n vectors, one vector for each of the n modulo-2 adders. Each vector has Kk dimensions and contains the connections of the encoder to that modulo-2 adder. A 1 in the ith position of the vector indicates that the corresponding stage in the shift register is connected to the modulo-2 adder, and a 0 in a given position indicates that no connection exists between that stage and the modulo-2 adder.

To be specific, let us consider the binary convolutional encoder with constraint length K = 3, k = 1, and n = 3, which is shown in Figure 8.1-2. Initially, the shift register is assumed to be in the all-zeros state. Suppose the first input bit is a 1. Then the output sequence of 3 bits is 111. Suppose the second bit is a 0. The output sequence will then be 001. If the third bit is a l, the output will be 100, and so on. Now, suppose we number the outputs of the function generators that generate each 3-bit output sequence as 1, 2, and 3, from top to bottom, and similarly number each corresponding function

generator. Then, since only the first stage is connected to the first function generator (no modulo-2 adder is needed), the generator is

Finally,

G3 = (111)

The generators for this code are more conveniently given in octal form as (4, 5, 7). We conclude that when k = 1, we require n generators, each of dimension K to specify the encoder.

It is clear that g1, g2, and g3 are the impulse responses from the encoder input to the three outputs. Then if the input to the encoder is the information sequence u, the three outputs are given by;

$$c^{(1)} = u \star g_1$$
$$c^{(2)} = u \star g_2$$
$$c^{(3)} = u \star g_3$$

Where * denotes the convolution operation. The corresponding code sequence $\boldsymbol{c}$ is the result of interleaving $\boldsymbol{c}^{(1)}$, $\boldsymbol{c}^{(2)}$ and $\boldsymbol{c}^{(3)}$ as

$$c = \left(c_1^{(1)}, c_1^{(2)}, c_1^{(3)}, c_2^{(1)}, c_2^{(2)}, c_2^{(3)}, \ldots\right)$$

The convolutional operation is equivalent to multiplication in the transform domain. We define the D transform[+] of $\boldsymbol{u}$ as;

$$u(D) = \sum_{i=0}^{\infty} u_i D^i$$

and the transfor function for the three impulse responses g1, g2, and g3 as;

$$g_1(D) = 1$$
$$g_2(D) = 1 + D^2$$
$$g_3(D) = 1 + D + D^2$$

The output transforms are then given by

$$c^{(1)}(D) = u(D)g_1(D)$$
$$c^{(2)}(D) = u(D)g_2(D)$$
$$c^{(3)}(D) = u(D)g_3(D)$$

and the transform of the encoder output $\boldsymbol{c}$ is given by

$$c(D) = c^{(1)}(D^3) + Dc^{(2)}(D^3) + D^2c^{(3)}(D^3)$$

Let the sequence u = (100111) be the input sequence to the convolution encoder shown in figure 8.1-2. We have

$$u(D) = 1 + D^3 + D^4 + D^5$$

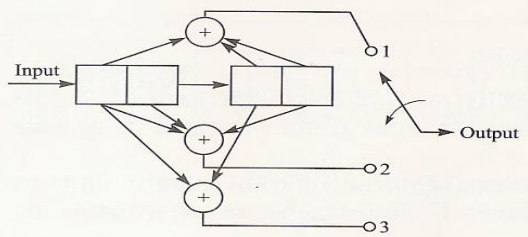


FIGURE 8.1–3
$K = 2, k = 2, n = 3$ convolutional encoder.

and

$$c^{(1)}(D) = (1 + D^3 + D^4 + D^5)(1) = 1 + D^3 + D^4 + D^5$$
$$c^{(2)}(D) = (1 + D^3 + D^4 + D^5)(1 + D^2) = 1 + D^2 + D^3 + D^4 + D^6 + D^7$$
$$c^{(3)}(D) = (1 + D^3 + D^4 + D^5)(1 + D + D^2) = 1 + D + D^2 + D^3 + D^5 + D^7$$

and

$$\begin{aligned} c(D) &= c^{(1)}(D^3) + Dc^{(2)}(D^3) + D^2c^{(3)}(D^3) \\ &= 1 + D + D^2 + D^5 + D^7 + D^8 + D^9 + D^{10} + D^{11} + D^{12} + D^{13} + D^{15} \\ &\quad + D^{17} + D^{19} + D^{22} + D^{23} \end{aligned}$$

corresponding to the code sequence

$$\boldsymbol{c} = (111001011111110101010011)$$

For a rate k/n binary convolutionary code with k > 1 and constraint length K, the n generators are ***K***k – dimensional vectors, as stated above. The following example illustrates the case in which *k* = 2 and n = 3.

The generators are;

g1 = (1011), g2 = (1101), g3 = (1010)

In octal form, these generators are (13, 15, 12).

The code shown above can be also realized by the diagram shown below. In this realization, instead a single shift register of length 4, two shift registers each of length 2 are employed. The information sequence ***u*** is split into two substream $\boldsymbol{u}^{(1)}$ and $\boldsymbol{u}^{(2)}$ using serial-to-parallel converter. Each of the two substreams.

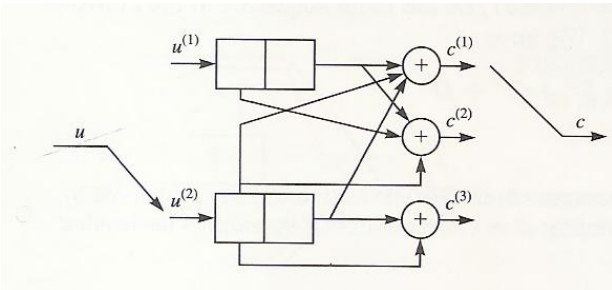


FIGURE 8.1–4
Double shift register implementation of the convolutional encoder shown in Figure 8.1–3.

is the input to one of the two shift registers. At the output, the three generated sequences, $\boldsymbol{c}^{(1)}$, $\boldsymbol{c}^{(2)}$ and $\boldsymbol{c}^{(3)}$ are interleaved to generate the code sequence c. In general, instead of one shift register with length $\boldsymbol{L} = \boldsymbol{K}$k, we can use a parallel implementation of k shift registers each of length K.

In the implementation shown in figure above, the encoder has two input sequences $\boldsymbol{u}^{(1)}$ and $\boldsymbol{u}^{(2)}$ and three output sequences $\boldsymbol{c}^{(1)}$, $\boldsymbol{c}^{(2)}$ and $\boldsymbol{c}^{(3)}$. The encoder thus can be described in terms of six impulse responses, and hence six transfer functions which are the D transforms of the impulse responses. If we denotes by $g_i^{(j)}$ the impulse response from input stream $\boldsymbol{u}^{(i)}$ to the output stream $\mathbf{c}^{(j)}$, in the encoder depicted in figure above we have;

$$g_1^{(1)} = [0\ 1] \qquad g_2^{(1)} = [1\ 1]$$
$$g_1^{(2)} = [1\ 1] \qquad g_2^{(2)} = [1\ 0]$$
$$g_1^{(3)} = [0\ 0] \qquad g_2^{(3)} = [1\ 1]$$

and the transfer functions are

$$g_1^{(1)}(D) = D \qquad g_2^{(1)}(D) = 1 + D$$
$$g_1^{(2)}(D) = 1 + D \qquad g_2^{(2)}(D) = 1$$
$$g_1^{(3)}(D) = 0 \qquad g_2^{(3)}(D) = 1 + D$$

From the transfer functions and the D transform of the input sequences we obtain the D transform of the three output sequences as;

$$c^{(1)}(D) = u^{(1)}(D)g_1^{(1)}(D) + u^{(2)}(D)g_2^{(1)}(D)$$
$$c^{(2)}(D) = u^{(1)}(D)g_1^{(2)}(D) + u^{(2)}(D)g_2^{(2)}(D)$$
$$c^{(3)}(D) = u^{(1)}(D)g_1^{(3)}(D) + u^{(2)}(D)g_2^{(3)}(D)$$

and finally

$$c(D) = c^{(1)}(D^3) + Dc^{(2)}(D^3) + D^2c^{(3)}(D^3)$$

Equation can be written in a more compact way by defining;

In general, matrix ***G***(D) is a k x n matrix whose elements are polynomials in ***D*** with degree at most K-1. This matrix is called the transform domain generator matrix of the convolutional code. For the code whose encoder is shown in figure 8.1-4.

$$c^{(1)}(D) = u^{(1)}(D)g_1^{(1)}(D) + u^{(2)}(D)g_2^{(1)}(D)$$
$$c^{(2)}(D) = u^{(1)}(D)g_1^{(2)}(D) + u^{(2)}(D)g_2^{(2)}(D)$$
$$c^{(3)}(D) = u^{(1)}(D)g_1^{(3)}(D) + u^{(2)}(D)g_2^{(3)}(D)$$

and finally

$$c(D) = c^{(1)}(D^3) + Dc^{(2)}(D^3) + D^2c^{(3)}(D^3)$$

Equation 8.1–9 can be written in a more compact way by defining

$$\boldsymbol{u}(D) = [\,u^{(1)}(D) \quad u^{(2)}(D)\,]$$

and

$$\boldsymbol{G}(D) = \begin{bmatrix} g_1^{(1)}(D) & g_1^{(2)}(D) & g_1^{(3)}(D) \\ g_2^{(1)}(D) & g_2^{(2)}(D) & g_2^{(3)}(D) \end{bmatrix}$$

By these definitions Equation 8.1–9 can be written as

$$\boldsymbol{c}(D) = \boldsymbol{u}(D)\boldsymbol{G}(D)$$

where

$$\boldsymbol{c}(D) = [\,c^{(1)}(D) \quad c^{(2)}(D) \quad c^{(3)}(D)\,]$$

we have

$$\boldsymbol{G}(D) = \begin{bmatrix} D & 1+D & 0 \\ 1+D & 1 & 1+D \end{bmatrix}$$

and for the convolutional code shown in Figure 8.1–2 we have

$$\boldsymbol{G}(D) = [\,1 \quad D^2+1 \quad D^2+D+1\,]$$

Tree, Trellis, and State Diagrams

There are three alternative methods that are often used to describe a convolutional code. These are the tree diagram, the trellis diagram, and the state diagram.

Close observation of the tree that is generated by the convolutional encoder shown in Figure below reveals that the structure repeats itself after the third stage. This behavior is consistent with the fact that the constraint length K = 3. That is, the 3-bit output sequence at each stage is determined by the input bit and the 2 previous input bits, i.e., the 2 bits contained in the first two stages of the shift register. The bit in the last stage of the shift register is shifted out at the right and does not affect the output. Thus we may say that the 3-bit output sequence for each input bit is determined by the input bit and the four possible states of the shift register, denoted as a = 00, b = 01, c = 10, d = 11.

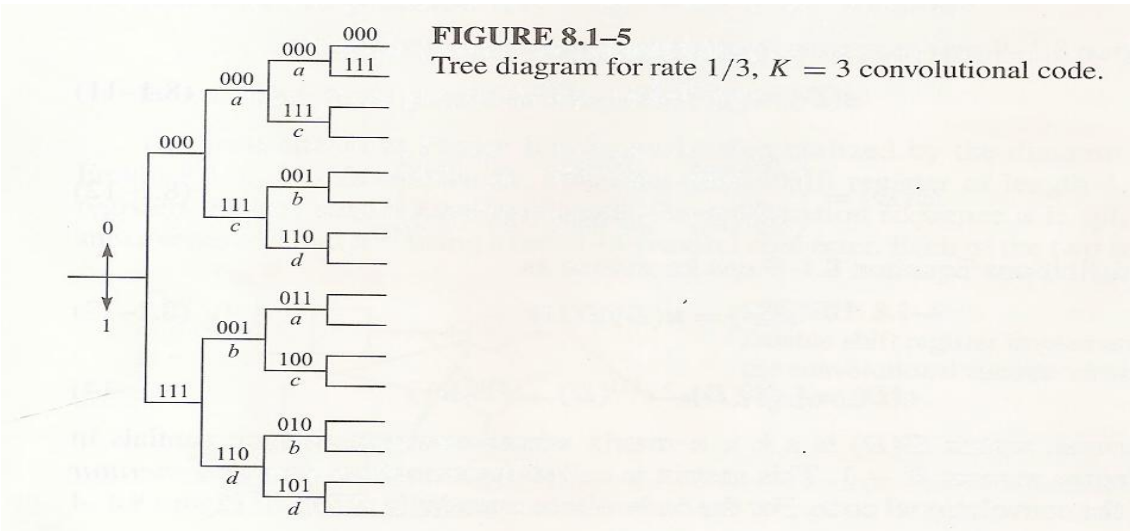
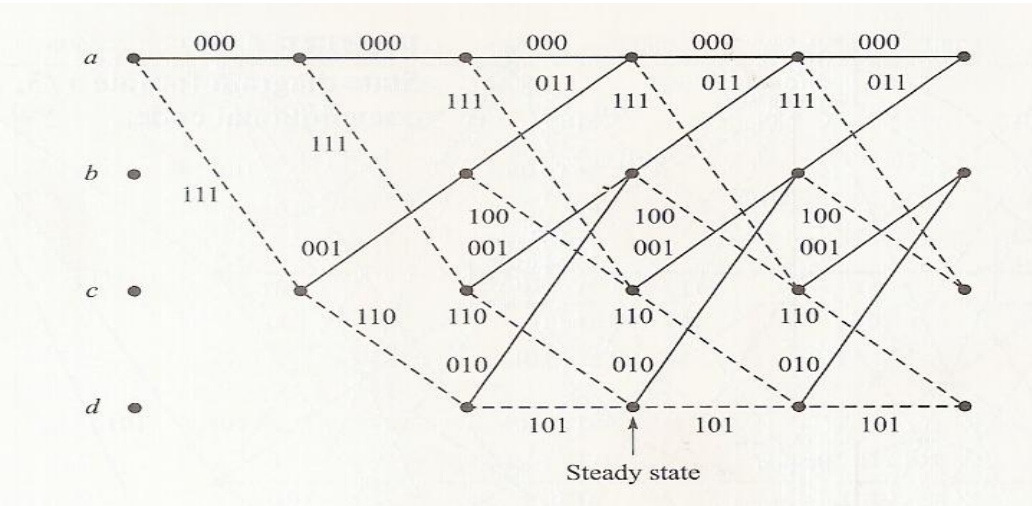


FIGURE 8.1–5
Tree diagram for rate 1/3, $K = 3$ convolutional code.



Trellis diagram for rate 113, K = 3 convolutional code.

If we label each node in the tree to correspond to the four possible states in the shift register, we find that at the third stage there are two nodes with label a, two with label b, two with label c, and two with label d. Now we observe that all branches emanating from two nodes having the same label (same state) are identical in the sense that they generate identical output sequence.

Since the output of the encoder is determined by the input and the state of the encoder, an even more compact diagram than the trellis is the state diagram. The state diagram is simply a graph of the possible states of the encoder and the possible transitions from one state to another. For example, the state diagram for the encoder

shows that the possible transitions are

$$a \xrightarrow{0} a,\ a \xrightarrow{1} c,\ b \xrightarrow{0} a,\ b \xrightarrow{1} c,\ c \xrightarrow{0} b,\ c \xrightarrow{1} d,\ d \xrightarrow{0} b,\ d \xrightarrow{1} d$$

where a $\xrightarrow{1}$ denotes the transition from state a to /3 when the input bit is a 1. The 3 bits shown next to each branch in the state diagram represent the output bits. A dotted line in the graph indicates that the input bit is a l, while the solid line indicates that the input bit is a 0.

Let us consider the k = 2, rate 2/3 convolutional code described in Example 8.1-2 and shown in Figure 8.1-3. The first two input bits may be 00, 01, 10,

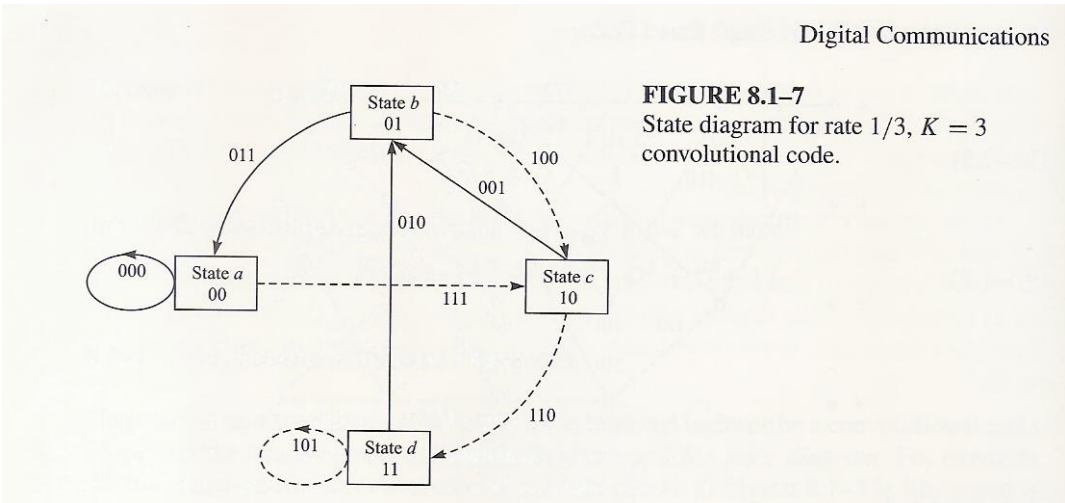
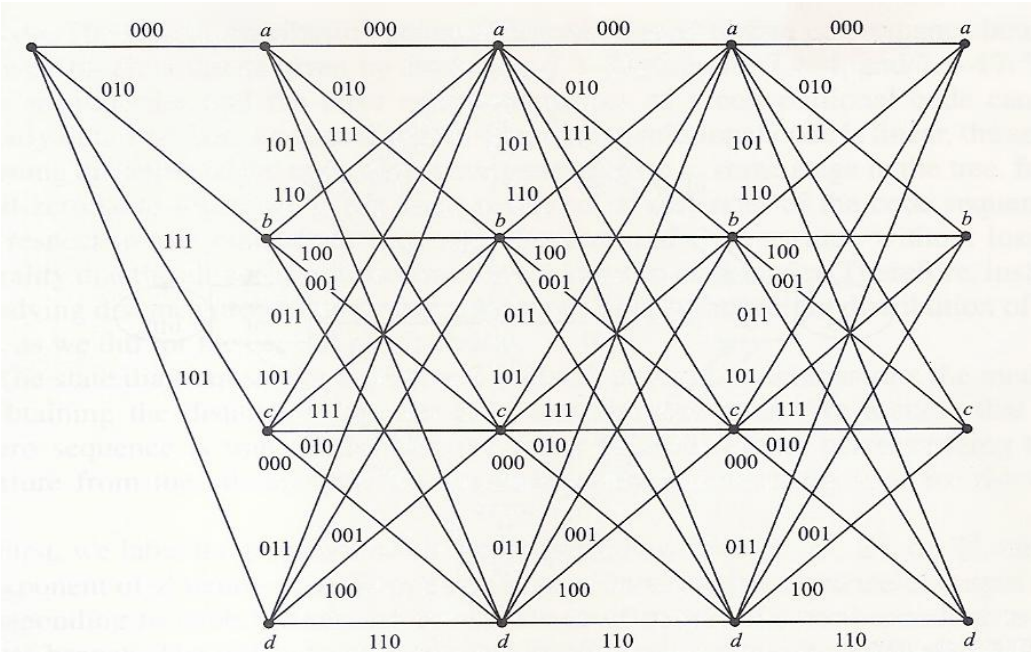


FIGURE 8.1–7
State diagram for rate 1/3, $K = 3$ convolutional code.

or 11. The corresponding output bits are 000, 010 111, 101. When the next pair of input bits enters the encoder, the first pair is shifted to the second stage.

Trellis and Graph Based Codes



Trellis diagram for K = 2, k = 2, n = 3 convolutional code.

Since the constraint length of the code is K = 2, the tree begins to repeat after the second stage.

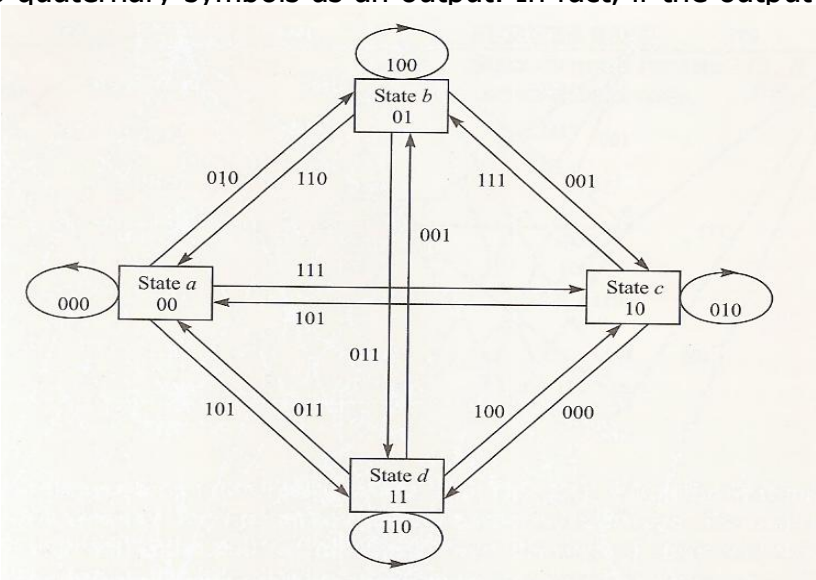
By merging the nodes having identical label the trellis, which is shown in Figure above. Finally, the state diagram for this code is shown in Figure below. To generalize, we state that a rate k/n, constraint length K, convolutional code is characterized by 2k branches emanating from each node of the tree diagram. The trellis and the state diagrams each have 2k(K-1) possible states.

Let us consider the convolutional code generated by the encoder . This code may be described as a binary convolutional code with parameters K = 2, k = 2, n = 4, R, = 1/2 and having the generators

g1= (110101) , g2= (101011), g3 = (1110), g4=(1001)

Except for the difference in rate, this code is similar in form to the rate 2/3, k = 2 convolutional code considered. Alternatively, the code generated by the encoder may

be described as a nonbinary (q = 4) code with one quaternary symbol as an input and two quaternary symbols as an output. In fact, if the output of the encoder is treated by the



State diagram for K = 2, k = 2, n = 3 convolutional code.

symbols that are transmitted over the channel by means of some M-ary (M = 4) modulation technique, the code is appropriately viewed as nonbinary. In any case, the tree, the trellis, and the state diagrams are independent of how we view the code.

We have seen that the distance properties of block codes can be expressed in terms of the weight distribution, or weight enumeration polynomial of

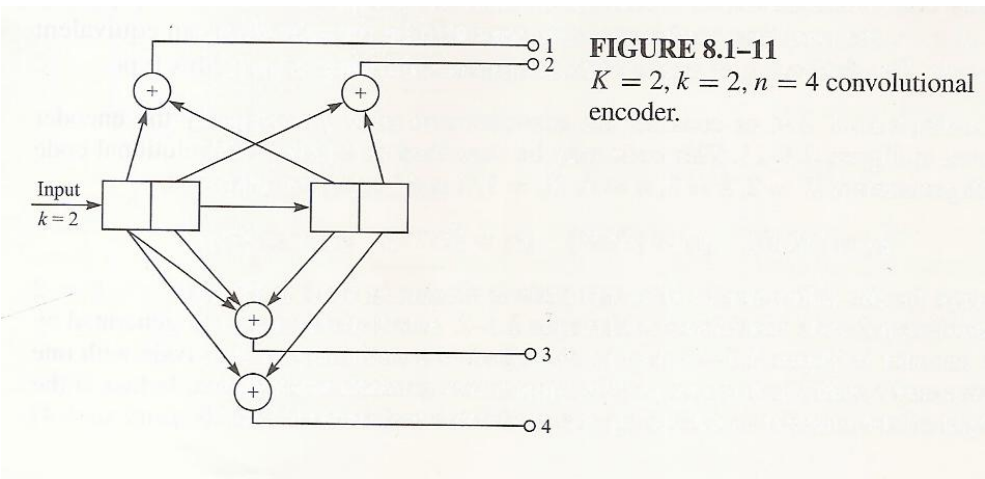


FIGURE 8.1–11
$K = 2, k = 2, n = 4$ convolutional encoder.

3.2 Decoding of convolutional codes

There exist different methods for decoding, of convolutional codes. Similar to block codes, the decoding of convolutional codes can be done either by soft decision or by hard decision decoding. In addition, the optimal decoding of convolutional codes can employ the maximum-likelihood or the maximum a posteriori principle. For convolutional codes with high constraint lengths, optimal decoding algorithms become too complex. Suboptimal decoding algorithms are usually used in such cases.

Maximum-Likelihood Decoding of Convolutional

Codes-The Viterbi Algorithm

In the decoding of a block code for a memoryless channel, we computed the distances (Hamming distance for hard-decision decoding and Euclidean distance for soft-decision decoding) between the received codeword and the 2k possible transmitted codewords. Then we selected the codeword that was closest in distance to the received codeword. This decision rule, which requires the computation of 2k metrics, is optimum in the sense that it results in a minimum probability of error for the binary symmetric channel with p < 1 and the additive white Gaussian noise channel.

Unlike a block code, which has a fixed length n, a convolutional encoder is basically a finite-state machine. Therefore, optimum decoding of a convolutional code involves a search through the trellis for the most probable sequence. Depending on whether the detector following the demodulator performs hard or soft decisions, the corresponding metric in the trellis search may be either a Hamming metric or a Euclidean metric, respectively.

A metric is defined for the jth branch of the ith path through the trellis as the logarithm of the joint probability of the sequence {rim, m = 1, 2, 31 conditioned on the transmitted sequence {$c^{(i)}_{jm}$, m= 1, 2, 3) for the *i*th path. That is;

$$\mu_j^{(i)} = \log p\left(\boldsymbol{r}_j | \boldsymbol{c}_j^{(i)}\right), \qquad j = 1, 2, 3, \ldots$$

Furthermore, a metric for the ith path consisting of B branches through the trellis is defined as

$$PM^{(i)} = \sum_{j=1}^{B} \mu_j^{(i)}$$

The criterion for deciding between two paths through the trellis is to select the one having the larger metric. This rule maximizes the probability of a correct decision, or, equivalently, it minimizes the probability of error for the sequence of information bits. The metrics for these two paths are

$$PM^{(0)} = 6\log(1-p) + 3\log p$$
$$PM^{(1)} = 4\log(1-p) + 5\log p$$

where p is the probability of a bit error. Assuming that p < z, we find that the metric PM(O) is larger than the metric PM(1). This result is consistent with the observation that the all-zero path is at Hamming distance d = 3 from the received sequence, while the i = 1 path is at Hamming distance d = 5 from the received path. Thus, the Hamming distance is an equivalent metric for hard decision decoding.

Convolutional Codes

In deriving the probability of error for convolutional codes, the linearity property for .his class of codes is employed to simplify the derivation. That is, we assume that the all-zero sequence is transmitted, and we determine the probability of error in deciding n favor of another sequence.

Since the convolutional code does not necessarily have a fixed length, we derive its performance from the probability of error for sequences that merge with the all-zero sequence for the first time at a given node in the trellis. In particular, we define the gist-event error probability as the probability that another path that merges with the all-zero path at node B has a metric that exceeds the metric of the all-zero path for .he first time.

The sequence error probability of a convolutional code is bounded by

$$P_e \leq T(Z)\Big|_{Z=\Delta}$$

where

$$\Delta = \sum_{y \in \mathscr{Y}} \sqrt{p(y|0)p(y|1)}$$

Note that unlike equation , which states is linear block codes Pe

≤ A (Δ) – 1, here we do not need to subtract 1 from T (Z) since T (Z) does not include the all-zero path. Equation can be written as

$$P_e \leq \sum_{d=d_{\text{free}}}^{\infty} a_d \Delta^d$$

The bit error probability for a convolutional code follows from equation as;

$$P_b \leq \frac{1}{k} \frac{\partial}{\partial Y} T(Y, Z)\Big|_{Y=1, Z=\Delta}$$

From example, we know that ifthe modulation is BPSK (or QPSK) and the channel is an AWGN channel with soft decision decoding, then;

$\Delta = e^{-RcYb}$

and in case of hard decision decoding, where the channel model is a binary symetric channel with crossover probability of p, we have

$$\Delta = \sqrt{4p(1-p)}$$

3.3 Distance Properties Of Binary Convolutional Codes

In this subsection, we shall tabulate the minimum free distance and the generators for several binary, short-constraint-length convolutional codes for several code rates. These binary codes are optimal in the sense that, for a given rate and a given constraint length, they have the largest possible dfree. The generators and the corresponding values of dfree tabulated below have been obtained by Odenwalder (1970), Larsen (1973), Paaske (1974), and Daut et al. (1982) using computer search methods.

Heller (1968) has derived a relatively simple upper bound on the minimum free distance of a rate l/n convolutional code. It is given by

$$d_{\text{free}} \leq \min_{l>1} \left\lfloor \frac{2^{l-1}}{2^l - 1}(K + l - 1)n \right\rfloor$$

where Lx] denotes the largest integer contained in x. For purposes of comparison, this upper bound is also given in the tables for the rate 1/n codes. For rate k/n convolutional codes, Daut et al. (1982) have given a modification to Heller's bound. The values obtained from this upper bound for k/n are also tabulated.

Punctured Convolutional Codes

In some practical applications, there is a need to employ high-rate convolutional codes, e.g., rates of (n - 1)/n. As we have observed, the trellis for such high-rate codes has 2n-1 branches that enter each state. Consequently, there are 2n-1 metric computations per state that must be performed in implementing the Viterbi algorithm and as many

Rate 1/3 Maximum Free Distance Codes

Constraint Length K	Generators in Octal			d_{free}	Upper Bound on d_{free}
3	5	7	7	8	8
4	13	15	17	10	10
5	25	33	37	12	12
6	47	53	75	13	13
7	133	145	175	15	15
8	225	331	367	16	16
9	557	663	711	18	18
10	1,117	1,365	1,633	20	20
11	2,353	2,671	3,175	22	22
12	4,767	5,723	6,265	24	24
13	10,533	10,675	17,661	24	24
14	21,645	35,661	37,133	26	26

Sources: Odenwalder (1970) and Larsen (1973).

comparisons of the updated metrics to select the best path at each state. Therefore, the implementation of the decoder of a high-rate code can be very complex.

The computational complexity inherent in the implementation of the decoder of a high-rate convolutional code can be avoided by designing the high-rate code from a lowrate code in which some of the coded bits are deleted from transmission. The deletion of selected coded bits at the output of a convolutional encoder is called puncturing.

The puncturing process may be described as periodically deleting selected bits from the output of the encoder, thus creating a periodically time-varying trellis code.

Rate 1/2 Maximum Free Distance Codes

Constraint Length K	Generators in Octal		d_{free}	Upper Bound on d_{free}
3	5	7	5	5
4	15	17	6	6
5	23	35	7	8
6	53	75	8	8
7	133	171	10	10
8	247	371	10	11
9	561	753	12	12
10	1,167	1,545	12	13
11	2,335	3,661	14	14
12	4,335	5,723	15	15
13	10,533	17,661	16	16
14	21,675	27,123	16	17

Sources: Odenwalder (1970) and Larsen (1973).

Begin with a rate 1/n parent code and define a puncturing period P, corresponding to P input information bits to the encoder. Hence, in one period, the encoder outputs nP coded bits. Associated with the n*P* encoded bits is a puncturing matrix ***P*** of the form

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1P} \\ p_{21} & p_{22} & \cdots & p_{2P} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nP} \end{bmatrix}$$

Where each column of P corresponds to the n possible output bits from the encoder for each input bit and each element of P is either 0 or 1. When ptij = 1, the corresponding output bit from the encoder is transmitted. When pig = 0, the corresponding output bit from the encoder is deleted. Thus, the code rate is determined by the period P and the number of bits deleted.

If we delete N bits out of nP, the code rate is P/(nP - N), where N may take any integer value in the range 0 to (n - 1)P - 1. Hence, the achievable code rates are

$$R_c = \frac{P}{P+M}, \qquad M = 1, 2, \ldots, (n-1)P$$

Let us construct a rate 4 code by puncturing the output of the rate 3 , K = 3 encoder shown in Figure 8.1-2. There are many choices for P and M in Equation 8.4-2 to achieve the desired rate. We may take the smallest value of P, namely, P = 3. Then out of every n P = 9 output bits, we delete N = 5 bits. Thus, we achieve a rate 4 punctured convolutional code. As the puncturing matrix, we may select P as

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Figure above, illustrates the generation of the punctured code from the rates parent code.

4.0 Conclusion

We have considered coding schemes that are best represented in terms of graphs and trellises. Besides, trellis code were described for achieving coding gains of 3-4dB.

5.0 Summary

In summary, we have examined the structure of convolutional codes and different methods for decoding of convolutional codes. Besides, convolutional codes are described in terms of finite-state machines.

6.0 Tutor Marked Assignment

A convolutional code is described by g1 = (101), g2 = (111), g3 = (111).

1. Draw the encoder corresponding to this code.
2. Draw the state-transition diagram for this code.
3. Draw the trellis diagram for this code.
4. Find the transfer function and the free distance of this code.
5. Very whether or not this code is catastrophic.

7.0 References/Further Reading

Key Papers in the Development of Coding Theory by Berlekamp (1974).

Module 3: **Spread Spectrum Signals for Digital Communications and Multi-user Communication**

Unit 1: Spread Spectrum Signals for Digital Communication.

Unit 2: Multiple Antenna Systems

Unit 3: Multi-user Communication

Unit 4: Multichannel and Multicarrier Systems.

Unit 1: Spread Spectrum Signals for Digital Communication

1.0 Introduction

Spread spectrum signal used for the transmission of digital information are distinguished by the characteristic that their bandwidth W is much greater than the information rate R in bits/s. That is, the bandwidth expansion factor Be = W/R for a spread spectrum signals is much greater than unity. The large redundancy inherent in spread spectrum signals is required to overcome the severe levels of interference that are encountered in the transmission of digital information over some radio and satellite channels.

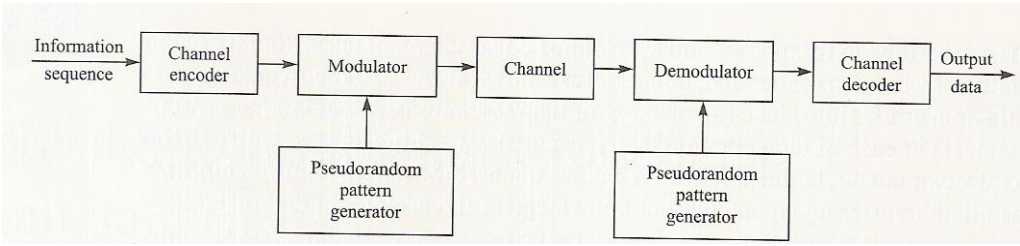
2.0 Objectives

At the end of this' unit, you should be able to;

- Understand the model of spread spectrum digital communication system.
- Describe frequency-hopped spread spectrum signals.
- Explain other types of spread spectrum signals.

3.1 Model spread spectrum digital communication system

The block diagram shown in figure below-illustrates the basic elements of a spread spectrum digital communication system with a binary information sequence at its input at the transmitting end and its output at the receiving end. The channel encoder and decoder and the modulator and demodulator are basic elements of the system.



In addition to these elements, we have two identical pseudorandom pattern generators, one that interfaces with the modulator at the transmitting end and a second that interfaces with the demodulator at the receiving end. The generators generate a pseudorandom or pseudonoise (PN) binary-valued sequence which is impressed on the transmitted signal at the modulator and removed from the received signal at the demodulator.

Synchronization of the PN sequence generated at the receiver with the PN sequence contained in the incoming received signal is required in order to demodulate the received signal. Initially, prior to the transmission of information, synchronization may be achieved by transmitting a fixed pseudorandom bit pattern that the receiver will recognize in the presence of interference with a high probability. After time synchronization of the generators is established, the transmission of information may commence.

3.2 Direct Sequence Spread Spectrum Signals

In the model shown, we assume that the information rate at the input to the encoder is R bits/s and the available channel bandwidth is W Hz. The modulation _s assumed to be binary PSK. In order to utilize the entire available channel bandwidth, _he phase of the carrier is shifted pseudorandomly according to the pattern from the PN ?enerator at a rate W times/s. The reciprocal of W, denoted by T, defines the duration of a pulse, which is called a chip; Tc is called the chip interval. The pulse is the basic element in a DS spread spectrum signal.

If we define Tb = 1/R to be the duration of a rectangular pulse corresponding to the transmission time of an information bit, the bandwidth expansion factor W/R may be expressed as

$$B_e = \frac{W}{R} = \frac{T_b}{T_c}$$

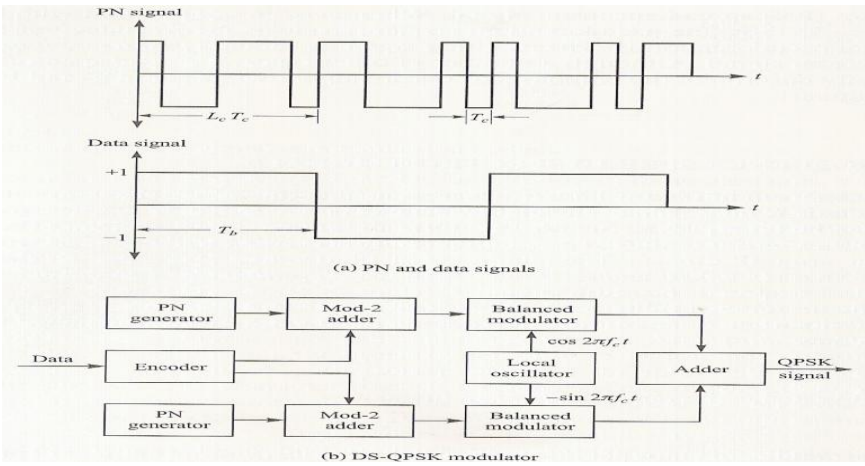
In practical systems, the ratio T_b/T_c is an integer,

$$L_c = \frac{T_b}{T_c}$$

which is the number of chips per information bit. That is, Lc is the number of phase shifts that can occur in the transmitted signal during the bit duration Tb = 1 / R.

Suppose that the encoder takes k information bits at a time and generates a binary linear (n, k) block code. The time duration available for transmitting the n code elements is kTb seconds. The number of chips that occur in this time interval is kLc. Hence, eve may select the block length of the code as n = kLc. If the encoder generates a binary convolutional code of rate k/n, the number of chips in the time interval kT_b is also n = kLc. Therefore, the following discussion applies to both block codes and onvolutional codes. We note that the code rate Rc = k/n = $1/L_c$.

One method for impressing the PN sequence on the transmitted signal is to alter directly the coded bits by modulo-2 addition with the PN sequence. Thus, each coded



The PN and data signal (a) and QPSK modulator (b) for a DS spread spectrum system.

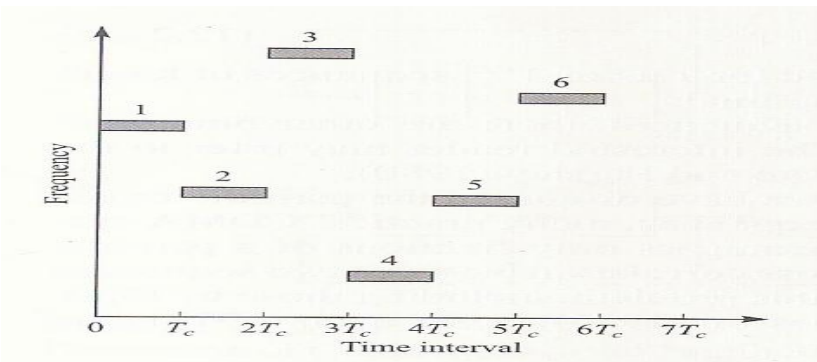
3.3 Frequency hopped spread spectrum signals

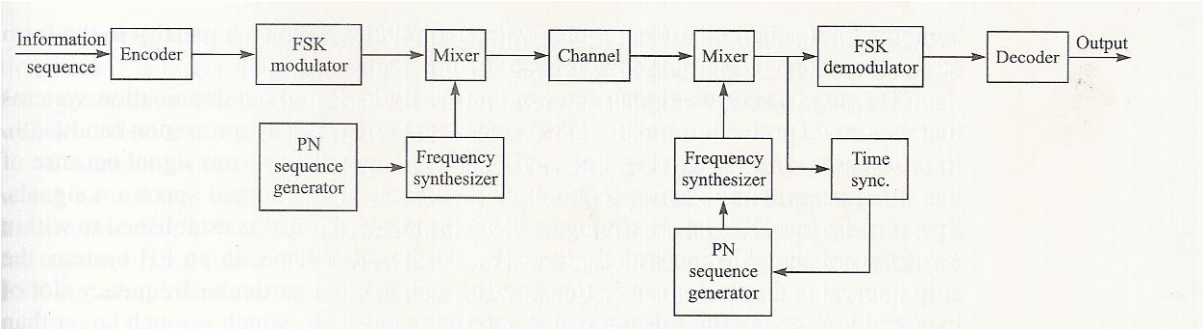
In a frequency-hopped (FH) spread spectrum communication system the available channel bandwidth is subdivided into a large number of contiguous frequency slots. In any signaling interval, the transmitted signal occupies one or more of the available frequency slots. The selection of the frequency slot(s) in each signaling interval is made pseudorandomly according to the output from a PN generator. Figure 12.3-1 illustrates a particular FH pattern in the time-frequency plane.

A block diagram of the transmitter and receiver for an FH spread spectrum system is shown in Figure 12.3-2. The modulation is usually either binary or M-ary FSK. For example, if binary FSK is employed, the modulator selects one of two frequencies corresponding to the transmission of either a 1 or a 0. The resulting FSK signal is translated in frequency by an amount that is determined by the output sequence from the PN generator, which, in turn, is used to select a frequency that is synthesized by the frequency synthesizer. This frequency is mixed with the output of the modulator and the resultant frequency-translated signal is transmitted over the channel. For example, n1 bits from the PN generator may be used to specify 2'n -1 possible frequency translations.

At the receiver, we have an identical PN generator, synchronized with the receiver signal, which is used to control the output of the frequency synthesizer. Thus, the pseudorandom frequency translation introduced at the transmitter is removed at the receiver by mixing the synthesizer output with the received signal. The resultant signal is demodulated by means of an FSK demodulator. A signal for maintaining synchronism of the PN generator with the frequency-translated received signal is usually extracted from the received signal.

Although PSK modulation gives better performance than FSK in an AWGN channel, it is sometimes difficult to maintain phase coherence in the synthesis of the frequencies used in the hopping pattern and, also, in the propagation of the signal over the channel as the signal is hopped from one frequency to another over a wide bandwidth Consequently, FSK modulation with noncoherent detection is often employed with Fl; spread spectrum signals.



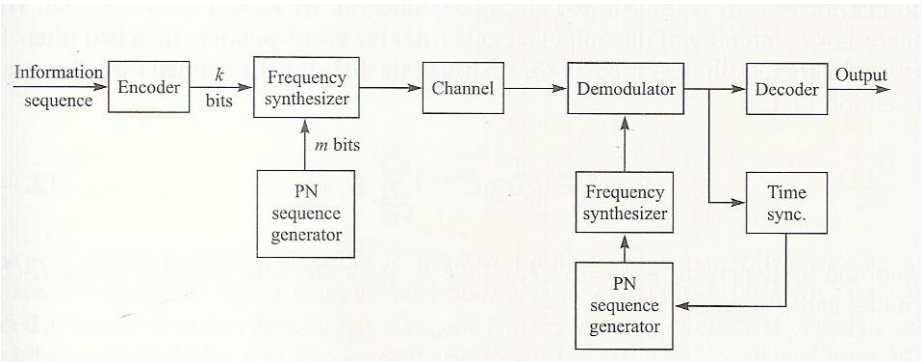


Block diagram of an FH spread spectrum system.

In the FH system depicted in Figure 12.3-2, the carrier frequency is pseudorandomly hopped in every signaling interval. The M information-bearing tones are contiguous and separated in frequency by 1/Tc, where T, is the signaling interval. This type of frequency hopping is called block hopping.

Another type of frequency hopping that is less vulnerable to some jamming strategies is independent tone hopping. In this scheme, the M possible tones from the modulator are assigned widely dispersed frequency slots. One method for accomplishing this is illustrated in Figure 12.3-3. Here, the m bits from the PN generator and the k information bits are used to specify the frequency slots for the transmitted signal. The FH rate is usually selected to be either equal to the (coded or uncoiled) symbol rate or faster than that rate. If there are multiple hops per symbol, we have a fast-hopped signal. On the other hand, if the hopping is performed at the symbol rate, we have a slow-hopped signal.

Fast frequency hopping is employed in AY applications when it is necessary to prevent a type of jammer, called a follower jammer from having sufficient time to intercept the frequency and retransmit it along with adjacent frequencies so as to create interfering signal components. However, there is a penalty incurred in subdividing a signal into several FH elements because the energy from these separate elements is



combined noncoherently. Consequently, the demodulator incurs a penalty in the form of a noncoherent combining loss as described in Section 11.1.

FH spread spectrum signals are used primarily in digital communication systems that require AJ protection and in CDMA, where many users share a common bandwidth. In most cases, an FH signal is preferred over a DS spread spectrum signal because of the stringent synchronization requirements inherent in DS spread spectrum signals. Specifically, in a DS system, timing and synchronization must be established to within a fraction of the chip interval T, zt~ 1/ W. On the other hand, in an FH system, the chip interval is the time spent in transmitting a signal in a particular frequency slot of bandwidth B << W. But this interval is approximately I/ B, which is much larger than 1/ W. Hence the timing requirements in an FH system are not as stringent as in a DS system.

In Sections 12.3-2 and 12.3-3, we shall focus on the AJ and CDMA application of FH spread spectrum signals. First, we shall determine the error rate performance m an uncoded and a coded FH signal in the presence of broadband AWGN inteference Then we shall consider a more serious type of interference that arises in AJ and CDMA applications, called partial-band interference. The benefits obtained from coding fo_ this type of interference are determined. We conclude the discussion in Section 12.3--with an example of an FH CDMA system that was designed for use by mobile user_ with a satellite serving as the channel.

3.4 Synchronization of Spread Spectrum Systems

Time synchronization of the receiver to the received spread spectrum signal may be separated into two phases. There is an initial acquisition phase and a tracking phase =tier the signal has been initially acquired.

Acquisition In a direct sequence spread spectrum system, the PN code must be time-synchronized to within a small fraction of the chip interval T, ,^; 1 / W. The problem of initial synchronization may be viewed as one in which we attempt to synchronize in time the receiver clock to the transmitter clock. Usually, extremely accurate and stable time clocks are used in spread spectrum systems. Consequently, accurate time clocks result in a reduction of the time uncertainty between the receiver and the transmitter. However, there is always an initial timing uncertainty due to range uncertainty between the

transmitter and the receiver. This is especially a problem when communication is taking place between two mobile users. In any case, the usual procedure for establishing initial synchronization is for the transmitter to send a known pseudorandom data sequence to the receiver. The receiver is continuously in a search mode looking for this sequence in order to establish initial synchronization.

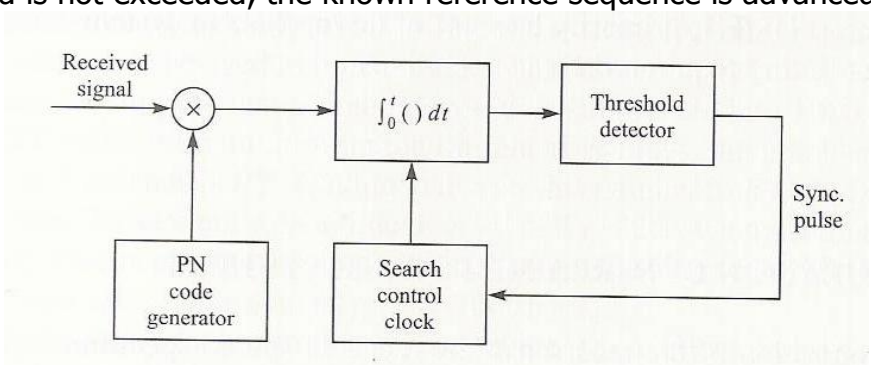
Let us suppose that the initial timing uncertainty is Tu and the chip duration is T, If initial synchronization is to take place in the presence of additive noise and other interference, it is necessary to dwell for Td = NT, in order to test synchronism at each time instant. If we search over the time uncertainty interval in (coarse) time steps of 2 T, then the time required to establish initial synchronization is

$$T_{\text{init sync}} = \frac{T_u}{\frac{1}{2}T_c} N T_c = 2NT_u$$

Clearly, the synchronization sequence transmitted to the receiver must be at least as long as 2NTu in order for the receiver to have sufficient time to perform the necessary search in a serial fashion.

In principle, matched filtering or cross correlation are optimum methods for establishing initial synchronization. A filter matched to the known data waveform generated from the known pseudorandom sequence continuously looks for exceedence of a pre determined threshold. When this occurs, initial synchronization is established and the demodulator enters the "data receive" mode.

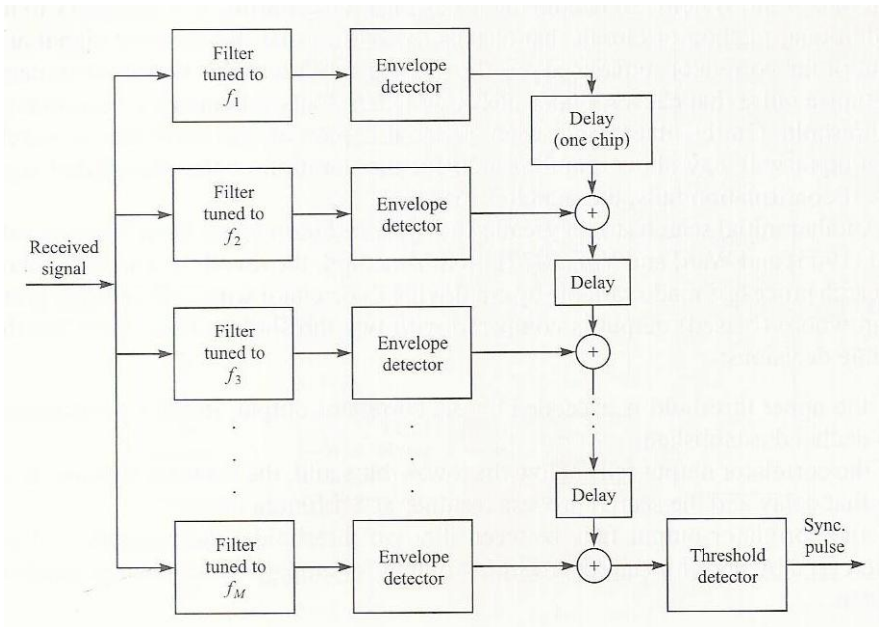
Alternatively, we may use a sliding correlator as shown in Figure 12.5-1. The correlator cycles through the time uncertainty, usually in discrete time intervals of 2 Tc, and correlates the received signal with the known synchronization sequence. The cross correlation is performed over the time interval NT, (N chips) and the correlator output is compared with a threshold to determine if the known signal sequence is present. If the threshold is not exceeded, the known reference sequence is advanced in time by



½ T_c, seconds and the correlation process is repeated. These operations are performed until a signal is detected or until the search has been performed over the time uncertainty interval T_u. In the latter case, the search process is then repeated.

A similar process may also be used for FH signals. In this case, the problem is to synchronize the Ply code that controls the hopped frequency pattern. To accomplish this initial synchronization, a known FH signal is transmitted to the receiver. The initial acquisition system at the receiver looks for this known FH signal pattern. For example, a bank of matched filters tuned to the transmitted frequencies in the known pattern may be employed. Their outputs must be properly delayed, envelope- or square-lawdetected, weighted, if necessary, and added (noncoherent integration) to produce the signal output which is compared with a threshold. A signal present is declared when the threshold is exceeded. The search process is usually performed continuously in time until a threshold is exceeded. A block diagram illustrating this signal acquisition scheme is given in Figure 12.5-2. As an alternative, a single matched-filter-envelope detector pair may be used, preceded by an FH pattern generator and followed by a postdetection integrator and a threshold detector. This configuration, shown in Figure 12.5-3, is based on a serial search and is akin to the sliding correlator for DS spread spectrum signals.

The sliding correlator for the IBS signals or its counterpart shown in Figure 12.5-3 for FH signals basically perform a serial search that is generally time-consuming. As an alternative, one may introduce some degree of parallelism by having two or more such correlators operating in parallel and searching over non-overlapping time slots. In such a case, the search time is reduced at the expense of a more complex and costly implementation.



4.0 **Conclusion**

Spread spectrum signals are used for combating or suppressing the detriment effects of interference due to jamming, interference arising from other users of the channel, and self-interference due to multipath propagation. It is also used for hiding a signal by transmitting it at low power and, this, making it difficult for an unintended listener to detect in the presence of background noise and achieving message privacy in the presence of other listeners.

5.0 Summary

Spread spectrum signals are used to obtain accurate range (time delay) and range rate (velocity) measurements in radar and navigation. Besides, the primary application of spread spectrum communications has been in the development of secure (AJ) digital communication system for military use. Expanding the bandwidth of the transmitted signal. Spatial diversity can also be achieved by using multiple antenna at the transmitter.

In this unit, we have seen that multiple antennas at the transmitter and the receiver of a wireless communication system can be used to establish multiple parallel channels for simultaneous transmission of multiple data streams in the same frequency band (spatial multiplexing) and, thus result in extremely high bandwidth efficiency.

6.0 Tutor Marked Assignment

Consider a deterministic MISCO (NT, 1) channel with AWGN and channel rector h. The received signal in any signal internal may be expressed as: y=hs+n

Where y and n are scalars

a. If the channel rector h is known as the transmitter, demonstrate that the received SNR is maximized when the information is sent in the direction of the channel rector h, s is selected as;

s=h/iihii

(The alignment of the transmit signal in the direction of the channel rector h is called transmit beam forming).

b. hat is the capacity of the MISCD channel when h is known at the transmitter?

c. Compare the capacity obtained in (b) with that of a SIMO channel, when the channel matrix h is identical for the two systems.

7.0 References/ Further Reading

Space time coding for MIMD channels by Taro kh et al (1998, 1999)

6.0 Tutor Marked Assignment

Ads binary PSK spread spectrum signals has a processing gain of 500. What is the interference margin against a continuous-tone interference if the desired error probability is 10^{-5}?

Consider the DS spread spectrum signal.

$c(t) = E\ cnp\ (t\text{-}nT_c)$

Where Cn is a periodic M sequence with a period N = 127 and p(t) is a rectangular pulse of duration Tc = lds. Determine tehpower spectral density of the signal c(t).

7.0 References/ Further Reading

Unit 2: Multiple Antenna Systems

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Channel Models for Multiple-antenna systems

3.2 Capacity of MIMO Channels

3.3 Spread spectrum signals and multimode transmission

3.4 Coding for MIMO Channels

1.0 Introduction

Multiple transmitting antennas can be used to create multiple spatial channels and this provide the capacity to increase the data rate of a wireless communication system. This method is called spatial multiplexing.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain multiple transmitting antenna systems.
- Describe the capacity of MIMO channels and its mathematical models.

3.0 main Content

3.1 Channel Models for Multiple Antenna System

A communication system employing NT transmitting antennas and NR receiving antennas is generally called a multiple-input, multiple-output (MIMO) system, and the resulting spatial channel in such a system is called a MIMO channel. The special case in which NT = NR = 1 is called a single-input, single-output (SISO) system, and the corresponding channel is called a SISO channel. A second special case is one in which NT = 1 and NR _> 2. The resulting system is called a single-input, multiple-output (SIMO) system, and the corresponding channel is called a SIMO channel. Finally, a third special case is one in which NT > 2 and NR = 1. The resulting system is called a multiple-input, single-output (MISO) system, and the corresponding channel is called a IVIISO channel.

In a MIMO system with NT transmit antennas and NR receive antennas, we denote the equivalent lowpass channel impulse response between the jth transmit antenna and the ith receive antenna as hij(τ;t), where z is the age or delay variable and t is the time variable. Thus, the randomly time-varying channel is characterized by the NR x NT

matrix $\boldsymbol{H}(\tau; t)$, defined as

$$\boldsymbol{H}(\tau; t) = \begin{bmatrix} h_{11}(\tau; t) & h_{12}(\tau; t) & \cdots & h_{1N_T}(\tau; t) \\ h_{21}(\tau; t) & h_{22}(\tau; t) & \cdots & h_{2N_T}(\tau; t) \\ \vdots & \vdots & & \vdots \\ h_{N_R1}(\tau; t) & h_{N_R2}(\tau; t) & \cdots & h_{N_RN_T}(\tau; t) \end{bmatrix}$$

Suppose that the signal transmitted from the jth transmit antenna is sj (t), j = 1, 2, …. N_T. Then the signal received at the ith antenna in the absence of noise may be expressed as;

$$\begin{aligned} r_i(t) &= \sum_{j=1}^{N_T} \int_{-\infty}^{\infty} h_{ij}(\tau; t) s_j(t-\tau)\, d\tau \\ &= \sum_{j=1}^{N_T} h_{ij}(\tau; t) * s_j(\tau), \qquad i = 1, 2, \ldots, N_R \end{aligned}$$

Where the asterisk denotes convolution. In matrix notation, Equation is expressed as

$$\boldsymbol{r}(t) = H(\tau;t) * s(\tau)$$

where s(t) is an N_T x 1 vector and r(t) is an N_R x 1 vector.

For a frequency=nonselective channel, the channel matrix H is expressed as

$$\boldsymbol{H}(t) = \begin{bmatrix} h_{11}(t) & h_{12}(t) & \cdots & h_{1N_T}(t) \\ h_{21}(t) & h_{22}(t) & \cdots & h_{2N_T}(t) \\ \vdots & \vdots & & \vdots \\ h_{N_R1}(t) & h_{N_R2}(t) & \cdots & h_{N_RN_T}(t) \end{bmatrix}$$

In this case, the signal received at the ith antenna is simply

$$r_i(t) = \sum_{j=1}^{N_T} h_{ij}(t) s_j(t), \qquad i = 1, 2, \ldots, N_R$$

and, in matrix form, the received signal vector $\boldsymbol{r}(t)$ is given as

$$\boldsymbol{r}(t) = \boldsymbol{H}(t)\boldsymbol{s}(t)$$

Furthermore, if the time variations of the channel impulse response are very slow within a time interval $0 \le t \le T$, when T may be either the symbol interval or some general time interval, equation maybe simply expressed as;

$r(t) = Hs(t)$, $0 \le t \le T$

where H is constant with the time interval $0 \le t \le T$.

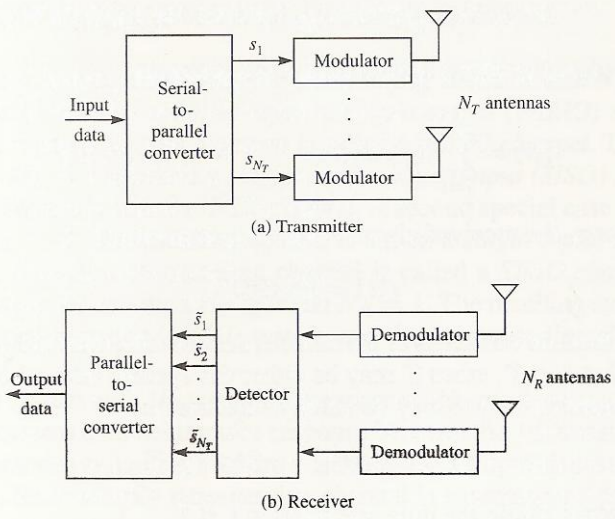
The slowly time-variant frequency-nonselective channel model embodied in equation is the simplest model for signal transmission in a MIMO channel. In the following two subsections, we employ this model to illustrate the performance characteristics of MIMO systems. At this point, we assume that the data to be transmitted are uncoiled. Coding for MIMO channels is treated in Section 15.4.

Signal Transmission through a Slow Fading Frequency-Nonselective MIMO Channel

Consider a wireless communication system that employs multiple transmitting and receiving antennas, as shown in Figure 15.1-1. We assume that there are NT transmitting antennas and NR receiving antennas. As illustrated in Figure 15.1-1, a block of NT symbols is converted from serial to parallel, and each symbol is fed to one of NT identical modulators, where each modulator is connected to a spatially separate antenna. Thus, the NT symbols are transmitted in parallel and are received on NR spatially separated receiving antennas.

In this section, we assume that each signal from a transmitting antenna to a receiving antenna undergoes frequency-nonselective Rayleigh fading. We also assume that the differences in propagation times of the signals from the NT transmitting to the NR receiving antennas are small relative to the symbol duration T, so that for all practical purposes, the signals from the NT transmitting antennas to any receiving antenna are synchronous. Hence, we can represent the equivalent lowpass received signals at the receiving antennas in a signaling interval as

$$r_m(t) = \sum_{n=1}^{N_T} s_n h_{mn} g(t) + z_m(t), \qquad 0 \le t \le T, \qquad m = 1, 2, \ldots, N_R$$



(a) Transmitter

(b) Receiver

where g(t) is the pulse shape (impulse response) of the modulation filters; hmn is the complex-valued, circular zero-mean Gaussian channel gain between the nth transmitting antenna and the mth receiving antenna; sn is the symbol transmitted on the nth antenna; and z,n (t) is a sample function of an AWGN process. The channel gains {hmn} are identically distributed and statistically independent from channel to channel. The Gaussian sample functions {z,n(t)} are identically distributed and mutually statistically independent, each having zero mean and two-sided power spectral density 2No. The information symbols {s,} are drawn from either a binary or an M-ary PSK or QAM signal constellation.

The demodulator for the signal at each of the NR receiving antennas consists of a matched filter to the pulse g(t), whose output is sampled at the end of each symbol interval. The output of the demodulator corresponding to the mth receiving antenna can be represented as

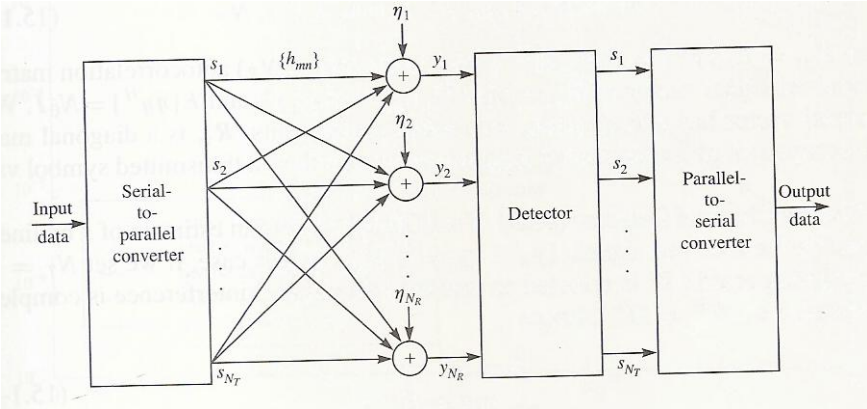
$$y_m = \sum_{n=1}^{N_T} s_n h_{mn} + \eta_m, \qquad m = 1, 2, \ldots, N_R$$

where the energy of the signal pulse g(t) is normalized to unity and r]m is the additive Gaussian noise component. The NR soft outputs from the demodulators are passed to the signal detector. For mathematical convenience, Equation 15.1-9 may be expressed in matrix form as

y = Hs + n

where y = (YI Y2 ... YNR]`, s = [s1 s2 ... SNT]r, n = [n1 n2 ... nN_R]t, and H is the NR x NT matrix of channel gains. Figure 15.1-2 illustrates the discrete-time model for the multiple transmitter and receiver signals in each signaling interval.

In the formulation of a MIMO system as described above, we observe that the transmitted symbols on the NT transmitting antennas overlap totally in both time and frequency. As a consequence, there is interchannel interference in the signals (y_m, $1 \leq m \leq N_R$} received from the spatial channel. In the following subsection, we consider three different detectors for recovering the transmitted data symbols in a MIMO system.



3.2 Capacity of MIMO Channels

In this section, we evaluate the capacity of MIM® channel models. For mathematical convenience, we limit our treatment to frequency-nonselective channels which are assumed to be known to the receiver. Thus, the channel is characterized by an NR x NT channel matrix H with elements {hid }. In any signal interval, the elements {hi.1 } are complex-valued random variables. In the special case of a Rayleigh fading channel, the (hij) are zero-mean complex-valued Gaussian random variables with uncorrelated real and imaginary components (circularly symmetric). When the {hij} are statistically -.dependent and identically distributed complex-valued Gaussian random variables, MIMQ channel is spatially white.

Mathematical Preliminaries

By using a singular value decomposition (SVD), the channel matrix H with rank r may -= expressed as

$H = U\Sigma V^H$

where U is an NR x r matrix, V is an NT x r matrix, and E' is an r x r diagonal matrix with diagonal elements the singular values σ1, σ2 …… σr, of the channel. The singular values {σi} are strictly positive and are ordered in decreasing order, i.e., of >_ of e column vectors of U and V are orthonormal. Hence UHU = I and VHV = IY, - here I,. is an r x r identity matrix. Therefore, the SVD of the channel matrix H may - expressed as

$$H = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H$$

where {ui } are the column vectors of U, which are called the left singular vectors of H, and (vi} are the column vectors of V, which are called the right singular vectors of H. We also consider the decomposition of the NR x NR square matrix HHH. This matrix may be decomposed as

$HH^H = Q\lambda QH$

Where Q is the N_R X N_R modal matrix with orthonormal column vectors (eigenvectors), i.e., $Q^HQ = \boldsymbol{I}_{NR}$, and $\wedge$ is an N_R X N_R diagonal elements (λi, i = 1, 2, ….. N_R), which are the eigenvalues of HH^H. With the eigenvalues numbers in decreasing order (λi > λi + 1), it can be easily demonstrated that the eigenvalues of HH^H are related to the singular values in the SVD of H as follows:

$$\lambda_i = \begin{cases} \sigma_i^2 & i = 1, 2, \ldots, r \\ 0 & i = r+1, \ldots, N_R \end{cases}$$

A useful metric is the Frobenius norm of $\boldsymbol{H}$, which is defined as

$$\| \boldsymbol{H} \|_F = \sqrt{\sum_{i=1}^{N_R} \sum_{j=1}^{N_T} |h_{ij}|^2}$$

$$= \sqrt{\text{trace}\,(\boldsymbol{H}\boldsymbol{H}^H)}$$

$$= \sqrt{\sum_{i=1}^{N_R} \lambda_i}$$

We shall observed below that the squared Frobenius norm $//H//^2_F$ is a parameter that determines the performance of MIMO communication systems. The statistical properties of $//H//^2_F$ can be determined for various fading channel conditions. For example, in the case of Rayleigh fading, $/hij/^2$ is a chi-squared random variable with two degrees of freedom. When the {hij) are iid (spatially white MIMO channel) with unit variable, the probability density function of $//H//^2_F$ is chi-squared with $2N_RN_T$ degrees of freedom; i.e If X = $//H//^2_F$,

$$p(x) = \frac{x^{n-1}}{(n-1)!} e^{-x}, \qquad x \geq 0$$

where $n = N_R N_T$.

Capacity of a Frequency-Non-selective Deterministic MIMO Channel

Let us consider a frequency-nonselective AWGN MIMO channel characterized by the matrix H. Let ***s*** denote the N_T X 1 transmitted signal vector, which is statistically stationary and has zero mean and autocovariance matrix $\mathbf{R}_{ss}$. In the presence of AWGN, the N_R x 1 receive signal vector y may be expressed as;

$$y = Hs + n$$

Where n is the N_R x 1 zero-mean Gaussian noise vector covariance matrix Rnn = $N_0 I_{NR}$. Although H is a realization of a random matrix, in this section we treat H as deterministic and known to the receiver.

To determine the capacity of the MIMO channel, we first compute the mutual information between the transmitted signal vector **s** and the received vector **y**, denoted as I (**s**;**y**) and then determine the probability distribution of the signal vector **s** that maximizes I **(s;y)**. thus;

$$\boldsymbol{C} = \max_{p(s)} I\,(s;y)$$

Where C is the channel capacity in bits per second per hertz (bps/Hz). It can be shown (see Telatar (1999) and Neeser and Massey (1993) that I (s;y) is maximized when **s** is a zero-mean, circularly symmetric, complex Guassian vector; hence, C is only dependent on the covariance of the signal vector. the resulting capacity of the MIMO channel is

$$C = \max_{\mathrm{tr}(\boldsymbol{R}_{ss})=\mathcal{E}_s} \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{1}{N_0}\boldsymbol{H}\boldsymbol{R}_{ss}\boldsymbol{H}^H\right) \qquad \text{bps/Hz}$$

Where tr(**R**ss) denotes the trace of the signal covariance Rss. This is the maximum rate per hertz that can be transmitted reliably (without errors) over the MIMO channel for any given realization of the channel matrix H.

In the important practical case where the signals among the NT transmitters are statistically independent symbols with energy per symbols equal to Es/N_T, the signal covariance matrix is diagonal, i.e

$$\boldsymbol{R}_{ss} = \frac{\mathcal{E}_s}{N_T}\boldsymbol{I}_{N_T}$$

and trace ($\boldsymbol{R}_{ss}$) = Es. In this case, the expression for the capacity of the MIMO channel simplifies to

$$C = \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\mathcal{E}_s}{N_T N_0}\boldsymbol{H}\boldsymbol{H}^H\right) \qquad \text{bps/Hz}$$

The capacity formula in Equation can also be expressed in terms of the eigenvalues of HHH by using the decomposition $\boldsymbol{HH^H}$ = Q∧QH. Thus;

$$\begin{aligned} C &= \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\mathcal{E}_s}{N_T N_0}\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H\right) \\ &= \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\mathcal{E}_s}{N_T N_0}\boldsymbol{Q}^H\boldsymbol{Q}\boldsymbol{\Lambda}\right) \\ &= \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\mathcal{E}_s}{N_T N_0}\boldsymbol{\Lambda}\right) \\ &= \sum_{i=1}^{r} \log_2\left(1 + \frac{\mathcal{E}_s}{N_T N_0}\lambda_i\right) \end{aligned}$$

Where $\boldsymbol{r}$ is the rank of the channel matrix H.

3.3 Spread Spectrum Signals and Multicode Transmission

In Section 15.1 we demonstrated that a MIMO system transmitting in a frequency nonselective fading channel can employ identical narrowband signals for data transmission. The signals from the NT transmit antennas were assumed to arrive at the NR receive antennas via NTNR independently fading propagation paths. By knowing the channel matrix H, the receiver is able to separate and detect the NT transmitted symbols in each signaling interval. Thus, the use of narrowband signals provided a data rate increase (spatial multiplexing gain) of NT relative to a single-antenna system and, simultaneously, a signal diversity of order NR, where NR > NT, when the maximum-likelihood detector is employed.

In this section we consider a similar MIMO system with the exception that the transmitted signals on the NT transmit antennas will be wideband, i.e., spread spectrum signals.

Orthogonal Spreading Sequences

The MIMO system under consideration is illustrated in Figure 15.3-1(a). The data symbols {sj, 1 ≤ j ≤ NT} are each multiplied (spread) by a binary sequence f cjk, I ≤ k ≤ Lc, 1 ≤ j ≤ NT} consisting of Lc bits, where each bit takes a value of either -f-I or -1. These binary sequences are assumed to be orthogonal, i.e.,

$$\sum_{k=1}^{L_c} c_{jk} c_{ik} = 0, \qquad j \neq i$$

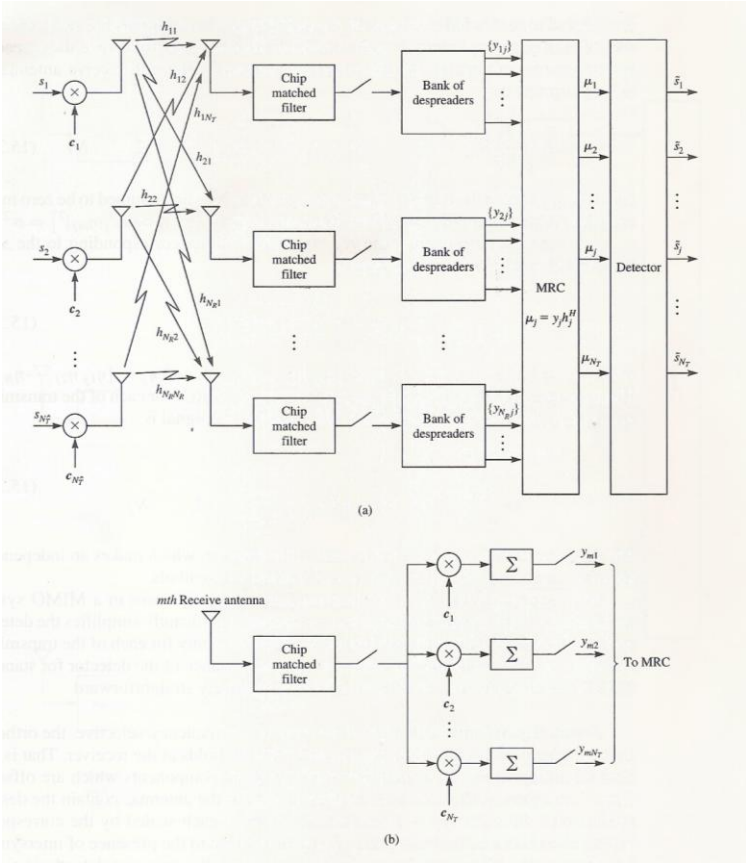
For example, the orthogonal sequences may be generated from NT Hadamard code-words of block length L, where a 0 in the Hadamard codeword is mapped into a -1 and a 1 is mapped into a +1. The resulting orthogonal sequences are usually called Walsh-Hadamard sequences.

The transmitted signal on the jth transmit antenna may be expressed as;

$$s_j(t) = s_j \sqrt{\frac{\mathcal{E}_s}{N_T}} \sum_{k=1}^{L_c} c_{jk} g(t - kT_c), \qquad 0 \leq t \leq T; \quad j = 1, 2, \ldots, N_T$$

where Ss/NT is the energy per transmitted symbol, T is the symbol duration, Tc = T/L, and g(t) is a signal pulse of duration T, and energy 1/L, The pulse g(t) is usually called a chip, and Lc is the number of chips per information symbol. Thus, the bandwidth of the information symbols, which is approximately 1 / T, is expanded by the factor L, so that the transmitted signal on each antenna occupies a bandwidth of approximately 1/Tc.

The MIMO channel is assumed to be frequency-nonselective and characterized by the matrix H, which is known to the receiver. At each receiving terminal, the received signal is passed through a chip matched filter and matched to the chip pulse g(t), and



its sampled output is fed to a bank of NT correlators whose outputs are sampled at the end of each signaling interval, as illustrated in Figure 15.3-1(b). Since the spreading sequences are orthogonal, the NT correlator outputs at the mth receive antenna are simply expressed as

$$y_{mj} = s_j \sqrt{\frac{\mathcal{E}_s}{N_T}} h_{mj} + \eta_{mj}, \qquad m = 1, 2, \ldots, N_R; \quad j = 1, 2, \ldots, N_T$$

where $\{n_{mj}\}$ denote the additive noise components, which are assumed to be zero mean, complex-valued circularly symmetric Gaussian iid with variance E $[n_{mj}\ I^2] = \sigma^2$.

It is convenient to express the NR correlator outputs corresponding to the same transmitted symbol sj in vector form as

$$\boldsymbol{y}_j = \sqrt{\frac{\mathcal{E}_s}{N_T}} s_j \boldsymbol{h}_j + \boldsymbol{\eta}_j$$

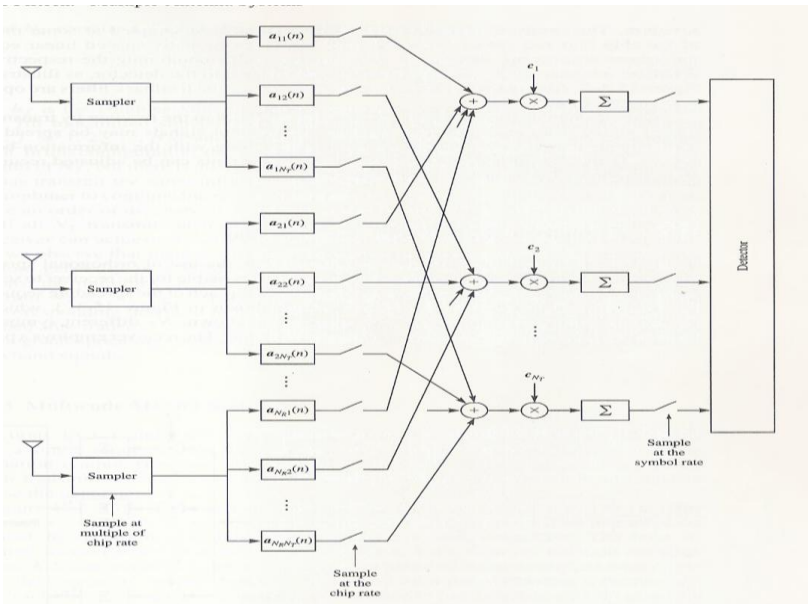
where yi = [y1j y2j … $yN_{Rj}]^t$, $\boldsymbol{h}$j = $[h_{1j}\ h_{2j}$ …. $hN_{Rj}]t$, and = $(n_{1j}\ n_{2j}$ … $n_{NRj})^t$. The optimum combiner is a maximal ratio combiner (MRC) for each of the transmitted symbols (sj). Thus, the output of the MRC for the jth signal is

$$\begin{aligned} \mu_j &= \boldsymbol{h}_j^H \boldsymbol{y} j \\ &= \sqrt{\frac{\mathcal{E}_s}{N_T}} s_j \, \| \boldsymbol{h}_j \|_F^2 + \boldsymbol{h}_j^H \boldsymbol{\eta}_j, \qquad j = 1, 2, \ldots, N_T \end{aligned}$$

The decision metrics {uj} are the inputs to the detector, which makes an independent decision on each symbol in the set {sj} of transmitted symbols.

We observe that the use of orthogonal spreading sequences in a MIMO system transmitting over a frequency-nonselective channel significantly simplifies the detector and, for a spatially white channel, yields NR-order diversity for each of the transmitted symbols {sj}. The evaluation of the error rate performance of the detector for standard signal constellations such as PSK and QAM is relatively straightforward. Frequency-Selective Channel If the channel is frequency-selective, the orthogonality property of the spreading sequences no longer holds at the receiver. That is, the channel multipath results in multiple received signal components which are offset in time. Consequently, the correlator outputs at each of the antennas contain the desired symbol plus the other NT - 1 transmitted symbols, each scaled by the corresponding cross-correlations between pairs of sequences. Due to the presence of intersymbol interference, the MRC is no longer optimum. Instead, the optimum detector is a joint maximum-likelihood detector for the NT transmitted symbols received at the NR receive antennas.

In general, the implementation complexity of the optimum detector in a frequency-selective channel is extremely high. In such channels, a suboptimum receiver may be employed. A receiver structure that is readily implemented in a MINIO frequency selective channel employs adaptive equalizers at each of the NR receivers prior to despreading the spread spectrum signals. Figure 15.3-2 illustrates the basic receiver



structure. The received signal at each receive antenna is sampled at some multiple of the chip rate and fed to a parallel bank of NT fractionally spaced linear equalizers, whose outputs are sampled at the chip rate. After combining the respective NR equalizer outputs, the NT signals are despread and fed to the detector, as illustrated in Figure 15.3-2. Alternatively DFEs may be used, where the feedback filters are operated at the symbol rate.

Training signals for the equalizers may be provided to the receiver by transmitting a pilot signal from each transmit antenna. These pilot signals may be spread spectrum signals that are simultaneously transmitted along with the information-bearing signals. Using the pilot signals, the equalizer coefficients can be adjusted recursively by employing a LMS- or RLS-type algorithm.

3.4 Coding for MIMO Channel

In this section we describe two different approaches to code design for MIMO channels and evaluate their performance for frequency-nonselective Rayleigh fading channels. The first approach is based on using conventional block or convolutional codes with interleaving to achieve signal diversity. The second approach is based on code design that is tailored for multiple-antenna systems. The resulting codes are called space-time codes. We begin by recapping the error rate performance of coded SISO systems in Rayleigh fading channels.

Performance of Temporally Coded SISO Systems in Rayleigh Fading Channels

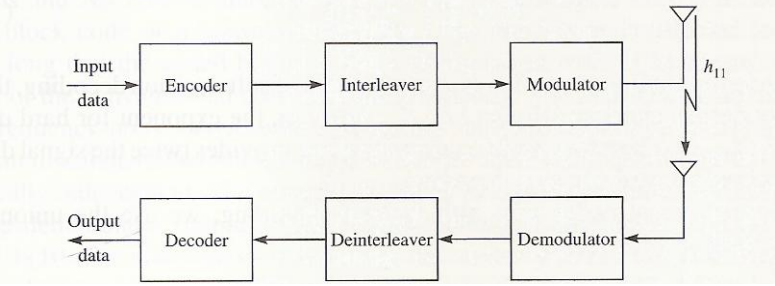
Let us consider a SISO system, as shown in Figure 15.4-1, where the fading channel is frequency-nonselective and the fading process is Rayleigh-distributed. The encoder generates either an (n, k) linear binary block code or an (n, k) binary convolutional code. The interleaver is assumed to be sufficiently long that the transmitted signals conveying the coded bits fade independently. The modulation is binary PSK, DPSK, or FSK.

The error probabilities for the coded SISO channel with Rayleigh fading are given in Sections 14.4 and 14.7. Let us consider linear block codes first. From Section 7.2-4, the union bound on the codeword error probability for soft decision decoding is

$$P_e < \sum_{m=2}^{M} P_2(w_m) < (M-1)P_2(d_{\min}) < 2^k P_2(d_{\min})$$

where P2(wm) is the pairwise error probability -given by the expression (see Section 14.7-1)

$$P_2(w_m) = \left(\frac{1-\psi}{2}\right)^{w_m} \sum_{k=0}^{w_{m-1}} \binom{w_m - 1 + k}{k} \left(\frac{1+\psi}{2}\right)^k$$



and

$$\psi = \begin{cases} \sqrt{\dfrac{\bar{\gamma}_b R_c}{1+\bar{\gamma}_b R_c}} & \text{BPSK} \\ \bar{\gamma}_b R_c/(1+\bar{\gamma}_b R_c) & \text{DPSK} \\ \bar{\gamma}_b R_c/(2+\bar{\gamma}_b R_c) & \text{FSK (noncoherent detection)} \end{cases}$$

For simplicity, we will use the simpler (looser) upper bound obtained by assuming tha

$$P_e < 2^k P_2(d_{\min})$$
$$< 2^k \binom{2d_{\min} - 1}{d_{\min}} \left(\frac{1}{q R_c \bar{\gamma}_b}\right)^{d_{\min}}$$

where

$$q = \begin{cases} 4 & \text{BPSK} \\ 2 & \text{DPSK} \\ 1 & \text{FSK (noncoherent detection)} \end{cases}$$

We observe that for soft decision decoding, the error probability decays exponentially as $1/y_b R_c$, where the exponent is equal to dmin, the minimum Hamming distance of the block codes.

For hard decision decoding, we employ the Chernov bound given in section 14.4, which may be expressed as;

$$P_e < 2^k [4p(1-p)]^{d_{\min}/2}$$

where the error probability per coded bit is given as

$$p = \frac{1-\psi}{2}$$

And Ψ is defined in equation 15.4-3. For yb >> 1, the Chernov bound simplifies to

$$P_e < 2^k \left(\frac{4}{q R_c \bar{\gamma}_b} \right)^{d_{\min}/2}$$

Where q is defined in equation 15.4-5. As in the case of soft decision decoding, the error probability decays exponentially as $1/y_b \boldsymbol{R}_c$; however, the exponent for hard decision decoding is dmin/2. Therefore, soft decision decoding provides twice the signal diversity that is obtained by hard decision decoding.

For convolutional codes with soft decision decoding, we use the union bound derived in section 14.3, namely

$$P_b < \sum_{d=d_{\text{free}}}^{\infty} \beta_d P_2(d)$$

where P2(d) is given by Equation 15.4-2 and Ψ is defined by Equation 15.4-3. If pb >> 1, we obtain the simpler form for the pairwise error probability, i.e.,

$$P_2(d) \approx \binom{2d-1}{d} \left(\frac{1}{q R_c \bar{\gamma}_b} \right)^d$$

where q is defined by Equation 15.4-5. We observe that the leading term in Equation 15.4-9 has an exponent of d = dfree. Hence, for soft decision decoding, the leading term in the error probability decays exponentially as 1/pbR,, where the exponent is dfree, the free distance of the convolutional code.

For hard decision decoding, we again use the Chemov bound for the pairwise error probability

$$P_2(d) < [4p(1-p)]^{d/2}$$

where p is defined by Equation 15.4-7 and * is defined by Equation 15.4-3. Hence, with Pb >>, P2(d) simplifies to

$$P_2(d) < \left(\frac{4}{q R_c \bar{\gamma}_b} \right)^{d/2}$$

and the bit error probability is upper-bounded as

$$P_b < \sum_{d=d_{\text{free}}}^{\infty} \beta_d \left(\frac{4}{q R_c \bar{\gamma}_b} \right)^{d/2}$$

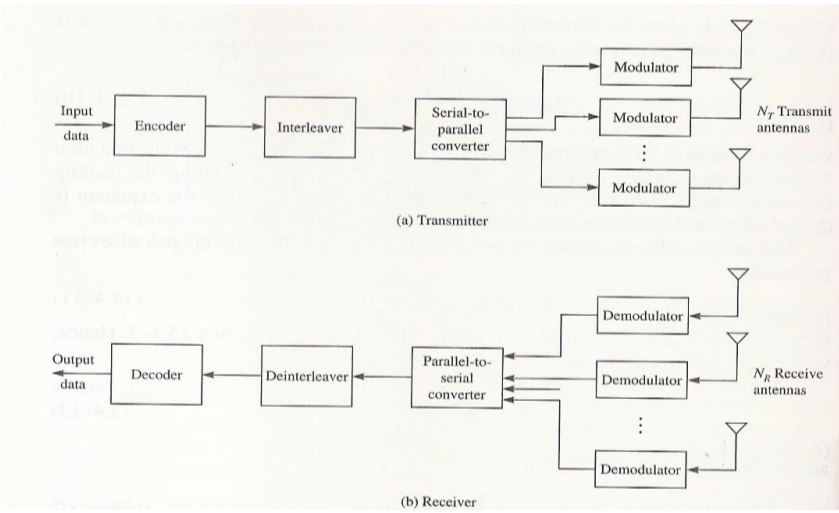
As in the case of block codes, we observe that with hard decision decoding, the signal diversity achieved by the code is reduced by a factor of 2 compared with soft decision decoding.

With this background on the performance of coded SISO systems, we now consider the performance of coded MIMO systems.

Bit-Interleaved Temporal Coding for MIMO Channels

We consider the MIMO system as shown in Figure 15.4-2, which has NT transmit antennas and NR receive antennas (NR >_ NT). The encoder may generate either a binary block code or a convolutional code. The interleaver is selected to be sufficiently long that the coded bits in a block of the block code or in several constraint lengths of the convolutional code fade independently. The MIMO channel is assumed to be frequency-nonselective with zero-mean, complex-valued, circularly symmetric Gaussian distributed coefficients (hlj), which are identically distributed and mutually statistically independent. The channel metrix H is assumed to have full rank.

The demodulator output in each signal interval is the vector y given by Equation 15.1-10. For hard decision decoding, the vector y is fed to the detector, which may employ any of the three detection algorithms (MLD, MMSE, ICD) described in Section 15.1-2 to make the hard decisions on the transmitted bits. For soft decision decoding, the vector y, after deinterleaving, is fed to the decoder. Similarly, for hard



decision decoding, the bits from the detector output are deinterleaved and fed to the decoder.

Let us consider the amount of signal diversity that is achieved in the MIMO system that employs spatial multiplexing of NT. Recall from Section 15.1-2 that with hard decision detection in an uncoded system, we achieved (NR - NT + 1)-order signal diversity with linear detection and NR-order signal diversity with the optimum maximum-likelihood detector (MLD). From our discussion in Section 15.4-1, we observed that the code provides a diversity of order dmi„/2 or dfree/2. Therefore, in a coded MIMO system, the total signal diversity achieved with a linear detector and a hard decision decoder is (NR - NT + 1)dmin/2 or (NR - NT + I)dfree/2. On the other hand, if soft decision decoding is employed, the total diversity order is NRdmin or NRdfree.

We demonstrate the additional diversity achieved with coding and bit-interleaving by computer simulation of the MIMO system shown in Figure 15.4-2 for a rate R, _ 1/2 convolutional code with dfree = 5 and BPSK modulation. Figures 15.4-3 and 15.4-4 illustrate the performance of the MIMO system for binary PSK with hard decision decoding and soft decision decoding, for (NT, NR) = (2, 2) and (NT, NR) = (2, 3). We observe that coding with interleaving improves the performance of the MIMO system relative to the performance of the uncoded system at the cost of a reduction in the data throughput rate by the reciprocal of the code rate. For (NT, NR) = (2, 3) and hard decision decoding, the MMSE detector with coding performs almost as well as the MLD

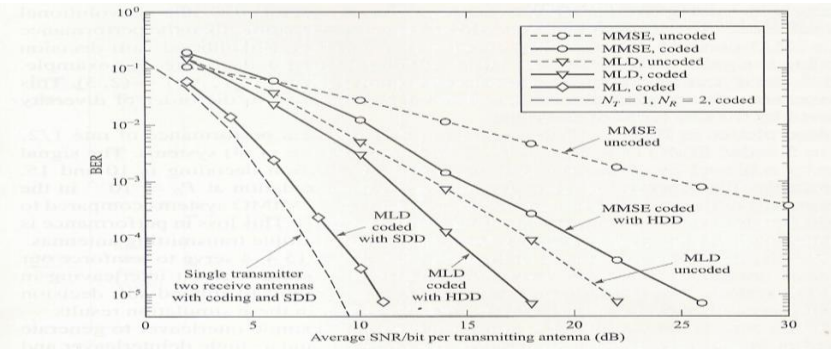


FIGURE 15.4–3
Performance of coded ($R_c = 1/2$, $d_{\text{free}} = 5$) systems with $N_T = N_R = 2$.

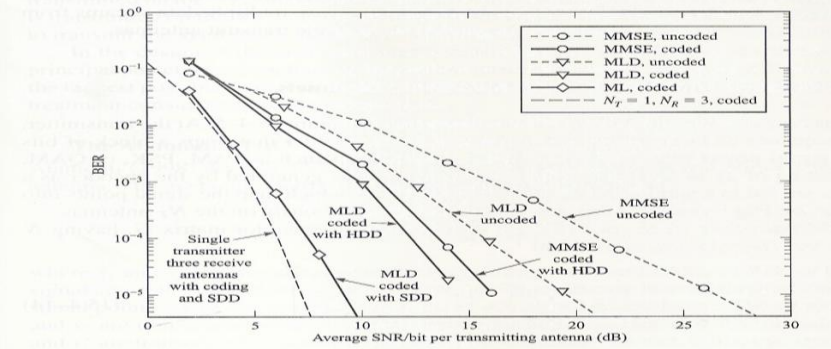


FIGURE 15.4–4
Performance of coded ($R_c = 1/2$, $d_{\text{free}} = 5$) systems with $N_T = 2$, $N_R = 3$.

detector with coding. In this case, the signal diversity provided by the convolutional code enhances the performance of the MMSE detected data more than the performance of the MLD detected data. We also observe that maximum-likelihood, soft decision decoding is significantly better than MLD with hard decision decoding. For example, at 10-5, the difference in performance is more than 5 dB for (NT, NR) = (2, 3). This performance advantage is due to the factor of 2 difference in the order of diversity achieved by the two types of decoders.

Also plotted in Figures 15.4-3 and 15.4-4 is the ideal performance of rate 1/2, dfree = 5 coded SIMO (NT, NR) = (l, 2) and (NT, NR) = (1, 3) systems. The signal diversity achieved by these two systems with soft decision decoding is 10 and 15, respectively. We observe that there is about a 2-dB degradation at Pb = 10-5 in the performance of the soft decision decoded (2, 2) and (2, 3) MIMO systems compared to the ideal performance of the corresponding SIMO systems. This loss in performance is attributed to the interference resulting from the use of multiple transmitting antennas.

The simulation results shown in Figures 15.4-3 and 15.4-4 serve to reinforce our analytical results on the signal diversity provided by coding with bit interleaving in a MIMO system. The performance superiority of maximum-likelihood soft decision decoding over hard decision decoding is clearly evident in these simulation results.

In this section we employed a single encoder and a single interleaver to generate the coded symbols for transmission on the NT antennas and a single deinterleaver and decoder at the receiver. An alternative approach that has been considered in the litera ture is to employ separate but identical encoding and interleaving on the dimultiplexed streams fed to each of the transmit antennas. This approach requires NT parallel encoders and interleavers at the transmitter and NT parallel decoders and deinterleavers at the receiver. It is especially suitable for situations where multiple data streams from different users are to be transmitted in parallel on multiple transmit antennas.

Space-Time Block Codes for MIMO Channels

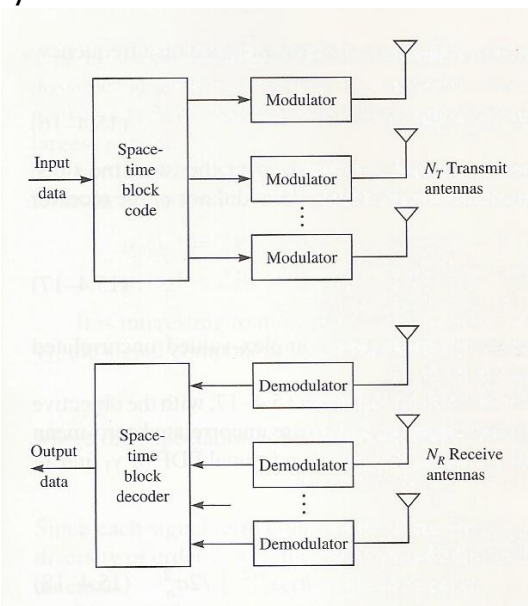
Let us now consider the MIMO system illustrated in Figure 15.4-5. At the transmitter, the sequence of information bits is fed to a block encoder that maps a block of bits

into signal points selected from a signal constellation such as PAM, PSK, or QAM, consisting of M = 2b signal points. The signal points generated by the encoder as a block are fed to a parallel set of identical modulators which map the signal points into corresponding waveforms that are transmitted simultaneously on the NT antennas.

A space-time block code (STBC) is defined by a generator matrix G, having N rows and NT columns, of the form

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1N_T} \\ g_{21} & g_{22} & \cdots & g_{2N_T} \\ \vdots & \vdots & & \vdots \\ g_{N1} & g_{N2} & \cdots & g_{NN_T} \end{bmatrix}$$

In which the elements f gig } are signal points resulting from a mapping of information bits to corresponding signal points from a binary or M-ary signal constellation. By



employing NT transmit antennas, each row of G consisting of NT signal points (symbols) is transmitted on the NT antennas in a time slot. Thus, the first row of NT symbols is transmitted on the NT antennas in the first time slot, the second row of NT symbols is transmitted on the NT antennas in the second time slot, and the Nth row of NT symbols is transmitted on the NT antennas in the Nth time slot. Therefore, N time slots are used to transmit the symbols in the N rows of the generator matrix G.

In the design of the generator matrix of a STBC, it is desirable to focus on three principal objectives: (1) achieving the highest possible diversity of NT NR, (2) achieving the highest possible spatial rate, and (3) minimizing the complexity of the decoder. Our treatment considers these three objectives.

The Alamouti STBC

Alamouti (1998) devised a STBC for NT = 2 transmit antennas and NR = 1 receive antenna. The generator matrix for the Alamouti code is given as

$$G = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$$

where sl and s2 are two signal points selected from an M-ary PAM, or PSK or QAM signal constellation with M = 2b signal points. Thus, 2b data bits are mapped into two signal points (symbols) sl and s2 from the M-ary signal constellation. The symbols sl and s2 are transmitted on the two antennas in the first time slot, and the symbols -s2 and si are transmitted on the two antennas in the second time slot. Thus, two symbols, sl and S2, are transmitted in two time slots. Consequently, the spatial code rate R,, = 1 for the Alamouti code. This is the highest possible rate for a (orthogonal) STBC.

4.0 Conclusion

The use of multiple antennas at the receiver of a communication system is a standard method for achieving spatial diversity to combat fading without.

UNIT 3: MULTIUSER COMMUNICATION

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Introduction to multiple access techniques

3.2 Capacity of multiple access methods

3.3 Multiuser detection in CDMA systems

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/Further Reading

1.0 Introduction

This unit focuses on multiple users and multiple communication links. We explore the various ways in which multiple users access a common channel to transmit information. The multiple access methods that are described in this unit form the basis for current and future wireless and wireless communication networks, such as satellite networks, cellular and mobile communication networks and underwater acoustic networks.

2.0 Objectives

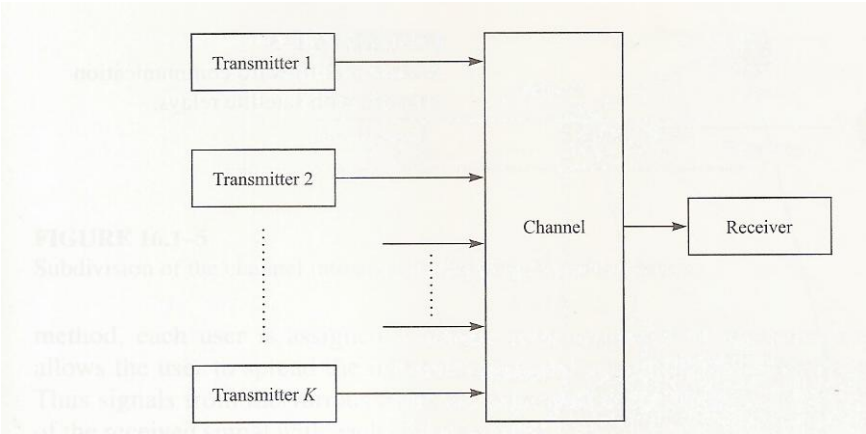
At the end of this unit, you should be able to;

- Explain the multiple access techniques.
- Discuses multiuser detection in CDMA systems.
- Understand the capacity of multiple access methods.

3.1 Introduction to multiple access techniques

It is instructive to distinguish among several types of multiuser communication systems. One type is a multiple access system in which a large number of users share a common communication channel to transmit information to a receiver. A model of such a system is depicted in Figure 16.1-1. The common channel may represent the uplink in either a cellular or a satellite communication system, or a cable to which are connected a number of terminals that access a central computer. For example, in a mobile cellular communication system, the users are the mobile terminals in any particular cell of the system, and the receiver resides in the base station of the particular cell.

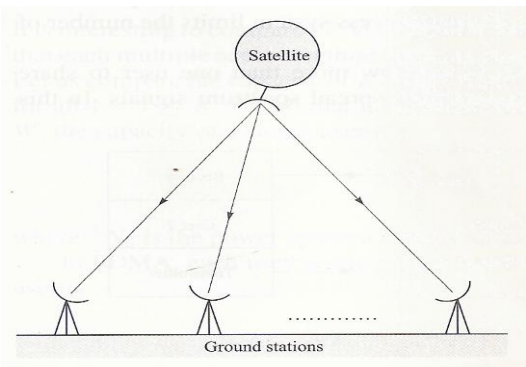
A second type of multiuser communication system is a broadcast network in which a single transmitter sends information to multiple receivers, as depicted in Figure 16.1-2. Examples of broadcast systems include the common radio and TV broadcast systems as well as the downlinks in cellular and satellite communication systems.



The multiple access and broadcast systems are the most common multiuser communication systems. A third type of multiuser system is a store-and-forward network, as depicted in Figure 16.1-3. Yet a fourth type is the two-way communication system shown in Figure 16.1-4.

In this chapter, we focus on multiple access and broadcast methods for multiuser communications. In a multiple access system, there are several different ways in which multiple users can send information through the communication channel to the receiver. One simple method is to subdivide the available channel bandwidth into a number, say K, of frequency non-overlapping subchannels, as shown in Figure 16.1-5, and to assign a subchannel to each user upon request by the users. This method is generally called frequency-division multiple access (FNMA) and is commonly used in wireline channels to accommodate multiple users for voice and data transmission.

Another method for creating multiple subchannels for multiple access is to subdivide the duration T f, called the frame duration, into, say, K non-overlapping subintervals, each of duration Tf/K. Then each user who wishes to transmit information



3.2 Capacity of multiple access method

It is interesting to compare FDMA, TDMA and CDMA in terms of the information rate that each multiple access method achieves in an ideal AWGN channel of bandwidth W. Let us compare the capacity of Y users, where each user has an average power Pi = P, for all 1 ≤ I ≤ K. Recall that in an ideal band-limited AWGN channel of bandwidth w, the capacity of a single user is

$$C = W \log_2 \left(1 + \frac{P}{W N_0}\right)$$

Where ½ No is the power spectral density of he additive noise.

In FDMA, each user is allocated a bandwidth W/K. Hence, the capacity of each user is

$$C_K = \frac{W}{K} \log_2 \left[1 + \frac{P}{(W/K) N_0}\right]$$

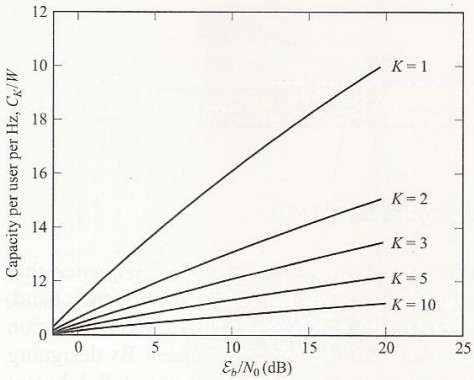


FIGURE
Normaliz
$\mathcal{E}_b/N_0$ for

and the total capacity for the K users is

$$K C_K = W \log_2 \left(1 + \frac{K P}{W N_0}\right)$$

Therefore, the total capacity is equivalent to that of a single user with average power P_{av} = K P.

It is interesting to note that for a fixed bandwidth W, the total capacity goes to infinity as the number of users increases linearly with K. One the other hand, as K increases, each user is allocated a smaller bandwidth (W/K) and, consequently, the capacity per user decreases. Figure 16.2-1 illustrates the capacity Ck per user normalized by the channel bandwidth W, as a function of Eb/No, with K as a parameter. This expression is given as

$$\frac{C_K}{W} = \frac{1}{K} \log_2 \left[1 + K \frac{C_K}{W} \left(\frac{\mathcal{E}_b}{N_0}\right)\right]$$

A more compact form of Equation 16.2-4 is obtained by defining the normalized total capacity Cn = KCk/W, which is the total bit rate for all K users per unit of bandwidth. Thus, equation 16-2-4 may be expressed as;

$$C_n = \log_2\left(1 + C_n \frac{\mathcal{E}_b}{N_0}\right)$$

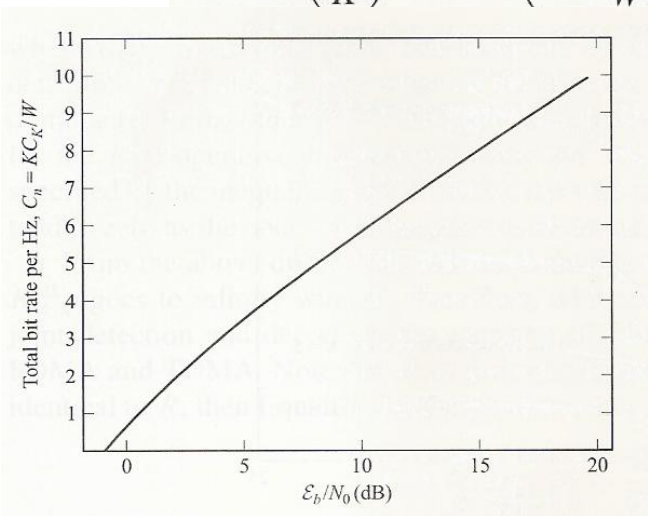
or, equivalently,

$$\frac{\mathcal{E}_b}{N_0} = \frac{2^{C_n} - 1}{C_n}$$

The graph of Cn versus Eb/No is shown in figure 16-2-2. We observe that Cn increase as Eb/No increases above the minimum value of 1n2.

In a TDMA system, each user transmits for 1/k of the time through the channel of bandwidth W, with average power KP. Therefore, the capacity per user is

$$C_K = \left(\frac{1}{K}\right) W \log_2\left(1 + \frac{KP}{WN_0}\right)$$



which is identical to the capacity of an FDMA system. However, from a practical stand-point, we should emphasize that, in TDMA, it may not be possible for the transmitters to sustain a transmitter power of K P when K is very large. Hence, there is a practical limit beyond which the transmitter power cannot be increased as K is increased.

In a CDMA system, each user transmits a pseudorandom signal of a bandwidth W and average power P. The capacity of the system depends on the level of cooperation among the K users. At one extreme is noncooperative CDMA, in which the receiver for each user signal does not know the codes and spreading waveforms of the other users, or chooses to ignore them in the demodulation process. Hence, the other users'

signals appear as interference at the receiver of each user. In this case, the multiuser receiver consists of a bank of K single-user matched filters. This is called single-user detection. If we assume that each user's pseudorandom signal waveform is Gaussian, then each user signal is corrupted by Gaussian interference of power (K - 1) P and additive Gaussian noise of power W NO. Therefore, the capacity per user for single-user detection is

$$C_K = W \log_2 \left[1 + \frac{P}{WN_0 + (K-1)P} \right]$$

or, equivalently,

$$\frac{C_K}{W} = \log_2 \left[1 + \frac{C_K}{W} \frac{\mathcal{E}_b/N_0}{1 + (K-1)(C_K/W)(\mathcal{E}_b/N_0)} \right]$$

Figure 16.2-3 illustrates the graph of Ck/W versus Eb/No, with K as a parameter.

For a large number of users, we may use the approximation 1n (1 + x) ≤ x. Hence,

$$\frac{C_K}{W} \le \frac{C_K}{W} \frac{\mathcal{E}_b/N_0}{1 + K(C_K/W)(\mathcal{E}_b/N_0)} \log_2 e$$

or, equivalently, the normalized total capacity $C_n = KC_K/W$ is

$$C_n \le \log_2 e - \frac{1}{\mathcal{E}_b/N_0}$$

$$\le \frac{1}{\ln 2} - \frac{1}{\mathcal{E}_b/N_0} < \frac{1}{\ln 2}$$

3.3 Multiuser detection in CDMA systems

As we have observed, TDMA and FDMA are multiple access methods in which the channel is partitioned into independent, single-user subchannels, i.e., non-overlapping time slots or frequency bands, respectively. In CDMA, each user is assigned a distinct signature sequence (or waveform), which the user employs to modulate and spread the information-bearing signal. The signature sequences also allow the receiver to demodulate the message transmitted by multiple users of the channel, who transmit simultaneously and, generally, asynchronously.

In this section, we treat the demodulation and detection of multiuser uncoded CDMA signals. We shall see that the optimum maximum-likelihood detector has a computational complexity that grows exponentially with the number of users. Such a

high complexity serves as a motivation to devise suboptimum detectors having lower computational complexities. Finally, we consider the performance characteristics of the various detectors.

CDMA Signal and Channel Models

Let us consider a CDMA channel that is shared by K simultaneous users. Each user is assigned a signature waveform gk(t) of duration T, where T is the symbol interval. A signature waveform may be expressed as

$$g_k(t) = \sum_{n=0}^{L-1} a_k(n) p(t - nT_c), \qquad 0 \le t \le T$$

where {ak(n), 0 < n < L - 1} is a pseudonoise (PN) code sequence consisting of L chips that take values {fl}, p (t) is a pulse of duration T, and T, is the chip interval. Thus, we have L chips per symbol and T = LT, Without loss of generality, we assume that all K signature waveforms have unit energy, i.e.,

$$\int_0^T g_k^2(t)\, dt = 1$$

The cross correlations between pairs of signature waveforms play an important role in the metrics for the signal detector and on its performance. We define the following cross correlations, where $0 \le z \le \tau$ and $i \le j$,

$$\rho_{ij}(\tau) = \int_\tau^T g_i(t) g_j(t - \tau)\, dt$$

$$\rho_{ji}(\tau) = \int_0^\tau g_i(t) g_j(t + T + \tau)\, dt$$

The cross correlations in Equations 16.3-3 and 16.3-4 apply to asynchronous transmissions among the K users. For synchronous transmission, we need only pi j (0).

For simplicity, we assume that binary antipodal signals are used to transmit the information from each user. Hence, let the information sequence of the kth user be denoted by {bk(In)}, where the value of each information bit may be ±1. It is convenient to consider the transmission of a block of bits of some arbitrary length, say N. Then, the data block from the kth user is

$\boldsymbol{b}$k = [bk(1) ... bk(N)^t

and the corresponding equivalent lowpass, transmitted waveform may be expressed as

$$s_k(t) = \sqrt{\mathcal{E}_k} \sum_{i=1}^{N} b_k(i) g_k(t - iT)$$

where £k is the signal energy per bit. The composite transmitted signal for the K users may be expressed as

$$s(t) = \sum_{k=1}^{K} s_k(t - \tau_k)$$
$$= \sum_{k=1}^{K} \sqrt{\mathcal{E}_k} \sum_{i=1}^{N} b_k(i) g_k(t - iT - \tau_k)$$

4.0 Conclusion

Frequency-division multiple access (FDMA) was the dominant multiple access scheme that has been used for decades in telephone communication systems for analog voice transmission. With the advent of digital speech transmission using PCM, DPCM and other speech coding methods, TDMA has replaced FDMA as the dominant multiple access scheme in telecommunications. CDMA and random access methods, in general, have been developed over the past three decades, primarily for use in wireless signal transmission and in local area wire line networks.

5.0 Summary

Multiuser information theory deals with basic information, theoretical limits in source coding for multiple sources and, channel coding and modulation for multiple access channels.

6.0 Tutor Marked Assignment

Consider a two-user, synchronous CDMA transmission system, where the received signal is;

r(t) = ε1big1(t) + ε2b2g2(t) + n (t), o ≤ t ≤ T

and (bl, b2) = (±1, ±1). The noise process n (t) is zero-mean

Gaussian and white, with spectral density No/2. the demodulator for r (t) is shown in figure 6.1 below.

a. Show that the correlator outputs rl and r2 at t = T may be expressed as; rl = c1bl + c2pb2 + n1

r2=F1b1p+c2b2+n2

b. Determine the variance of nl and n2 and the covariance of nl and n2.

c. Determine the joint PDF p(rl, r2/bl, b2).

2. Consider the two-user, synchronous CDMA transmission system described in problem 6.1 P(bl =1) = P(b2 =1) = '/2 and P(bl, b2) = P(bl) P (b2). The jointly optimum detector makes decisions based on the maximum a posteriori probability (MAP) criterion. That is, the detector computes.

Max P (bl, b2/r(t), o < t <_ T)

a. For the equally likely information b its (bl, b2) show that the MAP criterion is equivalent to the maximum - likelihood (ML) criterion max P[r(t), o _< t <_/bl, b2]

b. Show that the ML criterion in (a) leads to the jointly optimum detector that makes decisions on bl and b2 according to the following rule;

Max (clblrl + c2b2r2 - clc2pblb2

3. Consider the two0user, synchronous CDMA transmission system described in problem 6.1. the conventional single-user detector for the information bits bl and b2 gives the outputs.

bl = sgn (rl) b2 = sgn (r2) Assuming that p (bl =1) = ^{1}/2, and bl and b2 are statistically independent, determine the probability of error for this detector.

7.0 References/ Further Reading

Precoding and signal shaping for multichannel digital transmission by Fischer (2002).

UNIT 4: MULTICHANNEL AND CARRIER SYSTEMS

1.0 Introduction

This mode of transmission is used primarily in situations where there is a high probability that one or more of the channels will be unreliable from time to time. Multichannel signaling is sometimes employed in wireless communication systems as a means of overcoming the effects of interference of the transmitted signal.

2.0 Objectives

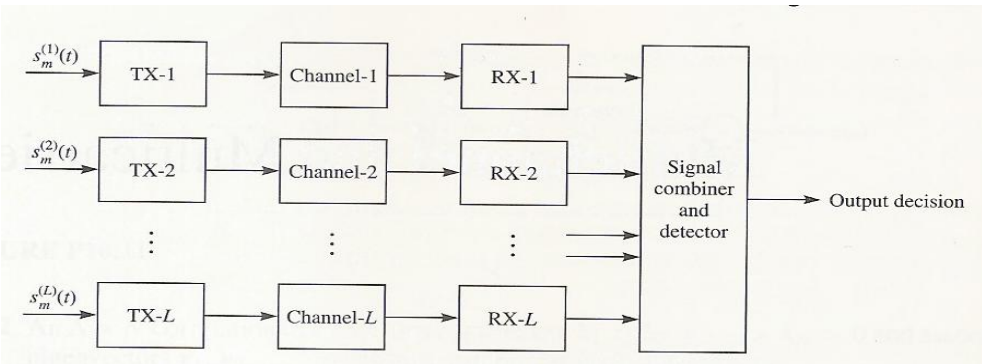
At the end of this unit, you should be able to;

- Understand multichannel digital communications in AWGN channels
- Explain multicarrier communications

3.1 Multichannel digital communications in AWGN channels

In this section, we confine our attention to multichannel signaling over fixed channels that differ only in attenuation and phase shift. The specific model for the multichannel digital signaling system is illustrated in Figure 11.1-1 and may be described as follows. The signal waveforms, in general, are expressed as

$$s_m^{(n)}(t) = \mathrm{Re}\left[s_{lm}^{(n)}(t)e^{j2\pi f_c t}\right], \qquad 0 \le t \le T$$
$$n = 1, 2, \ldots, L, \qquad m = 1, 2, \ldots, M \quad (11.1\text{–}1)$$



where L is the number of channels and M is the number of waveforms. The waveforms $\{s^{(n)}{}_m)(t)\}$ transmitted over the L channels are scaled by the attenuation factors {an}. phase-shifted by $\{\phi_n\}$, and corrupted by additive noise. The equivalent low-pass signals received from the L channels may be expressed as

$$r_l^{(n)}(t) = \alpha_n e^{j\phi_n} s_{lm}^{(n)}(t) + z_n(t), \quad 0 \le t \le T$$
$$n = 1, 2, \ldots, L, \quad m = 1, 2, \ldots, M \quad (11.1\text{–}2)$$

where {sl)(t)} are the equivalent lowpass transmitted waveforms and {zn(t)) represent the additive noise processes on the L channels. We assume that {zn(t)} are mutually statistically independent and identically distributed Gaussian noise random processes. We consider two types of processing at the receiver, namely, coherent detection and noncoherent detection. The receiver for coherent detection estimates the channel parameters fan} and f wn} and uses the estimates in computing the decision variables. Suppose we define gn = anej$^{\phi n}$ and let gn be the estimate of gn. The multichannel receiver correlates each of the L received signals with a replica of the corresponding transmitted signals, multiplies each of the correlator outputs by the corresponding estimates $\{g^*{}_n\}$, and sums the resulting signals. Thus, the decision variables for coherent detection are the correlation metrics

$$CM_m = \sum_{n=1}^{L} \mathrm{Re}\left[\hat{g}_n^* \int_0^T r_l^{(n)}(t) s_{lm}^{(n)*}(t)\, dt\right], \qquad m = 1, 2, \ldots, M$$

In noncoherent detection, no attempt is made to estimate the channel parameters. The demodulator may base its decision either on the sum of the envelopes (envelope detection) or the sum of the squared envelopes (square-law detection) of the matched filter outputs. In general, the performance obtained with envelope detection

differs little from the performance obtained with square-law detection in AWGN. However, squarelaw detection of multichannel signaling in AWGN channels is considerably easier to analyze than envelope detection. Therefore, we confine our attention to squarelaw detection of the received signals of the L channels, which produces the decision variables

$$CM_m = \sum_{n=1}^{L} \left| \int_0^T r_l^{(n)}(t) s_{lm}^{(n)*}(t)\, dt \right|^2, \qquad m = 1, 2, \ldots, M$$

Let us consider binary signaling first, and assume that *sli)(t), n = 1, 2, . . . , L,* are the L transmitted waveforms. Then an error is committed if C112 > CMI, or, equivalently, if the difference *D = CM, - CMZ* < 0. For noncoherent detection, this difference may be expressed as

$$D = \tfrac{1}{2} \sum_{n=1}^{L} (X_n Y_n^* + X_n^* Y_n)$$

where, by definition,

$$Y_n = \hat{g}_n, \qquad n = 1, 2, \ldots, L$$

$$X_n = \int_0^T r_l^{(n)}(t) \left[s_{l1}^{(n)*}(t) - s_{l2}^{(n)*}(t) \right] dt$$

The $\{X_n\}$ are mutually independent and identically distributed complex Gaussian random variables. The same statement applies to the variables *(Yn1.* However, for any *n,* X_n and Y_n may be correlated. For coherent detection, the difference *D = CM, - CMZ* may be expressed as

$$D = \sum_{n=1}^{L} (|X_n|^2 - |Y_n|^2)$$

where the variables $\{X_n\}$ and $\{Y_n\}$ are defined as

$$X_n = \int_0^T r_l^{(n)}(t) s_{l1}^{(n)*}(t)\, dt, \qquad n = 1, 2, \ldots, L$$

$$Y_n = \int_0^T r_l^{(n)}(t) s_{l2}^{(n)*}(t)\, dt, \qquad n = 1, 2, \ldots, L$$

If the estimates f *gn1* are obtained from observation of the received signal over one or more signaling intervals, as described in Appendix C, their statistical characteristics are described by the Gaussian distribution. Then the f Y_n I are characterized as mutually independent and identically distributed Gaussian random variables. The same

statement applies to the variables {X_n1. As in noncoherent detection, we allow for correlation between X_n and Y_n, but not between $X_{„z}$ and Y_n for $m \neq n$.

3.2 Multicarrier Communications

From our treatment of nonideal linear filter channels in Chapters 9 and 10, we have observed that such channels introduce ISI, which degrades performance compared with the ideal channel. The degree of performance degradation depends on the frequency response characteristics. Furthermore, the complexity of the receiver increases as the span of the ISI increases.

In this section, we consider the transmission of information on multiple carriers contained within the allocated channel bandwidth. The primary motivation for transmitting the data on multiple carriers is to reduce ISI and, thus, eliminate the performance degradation that is incurred in single carrier modulation.

Single-Carrier Versus Multicarrier Modulation

Given a particular channel characteristic, the communication system designer must decide how to efficiently utilize the available channel bandwidth in order to transmit the information reliably within the transmitter power constraint and receiver complexity constraints. For a nonideal linear filter channel, one option is to employ a single-carrier system in which the information sequence is transmitted serially at some specified rate R symbols/s. In such a channel, the time dispersion is generally much greater than the reciprocal of the symbol rate, and, hence, ISI results from the nonideal frequencyresponse characteristics of the channel. As we have observed, an equalizer is necessary to compensate for the channel distortion.

As an example of such an approach, we cite the modems designed to transmit data through voice-band channels in the switched telephone network, which are based on the International Telecommunications Union (ITU) standard V34. Such modems employ QAM impressed on a single carrier that is selected along with the symbol rate from a small set of specified values to obtain the maximum throughout at the desired level of performance (error rate). The channel frequency-response characteristics are measured upon initial setup of the telephone circuit, and the symbol rate and carrier frequency are selected based on this measurement.

An alternative approach to the design of a bandwidth-efficient communication system in the presence of channel distortion is to subdivide the available channel bandwidth into a number of subchannels, such that each subchannel is nearly ideal. To elaborate, suppose that C(f) is the frequency response of a nonideal, band-limited channel with a bandwidth W, and that the power spectral density of the additive Gaussian noise is Snn(f). Then we divide the bandwidth W into N = W/Af subbands of width Af, where A f is chosen sufficiently small that p f)1 Z/S"n(f) is approximately a constant within each subband. Furthermore, we select the transmitted signal power to be distributed in frequency as P (f), subject to the constraint that

$\int_W^P(f)\ df \leq P_{av}$

J W

where PQU is the available average power of the transmitter. Then we transmit the data on these N subchannels. Before proceeding further with this approach, we evaluate the capacity of the nonideal additive Gaussian noise channel.

4.0 Conclusion

Apart from multichannel communication, we considered multiple carrier transmission, where the frequency band of the channel is sub-divided into a number of sub-channels and information is transmitted on each of the channels.

5.0 Summary

In this unit, we considered both multichannel signal transmission and multicarrier transmission. The focus is on the performance of such systems in AWGN channels. Besides, multichannel signal transmission is commonly used on time-varying channels to overcome the effects of signal fading.

6.0 Tutor Marked Assignment

A binary communication system transmits the same information on two diversity channels. The two received signals are;

rl = ± Fb + n1

r2= F-b+n2

Where E (nl) = E (n2) = 0, E (n 21) = $r^2$1 and E (n^2_2) = r^2_2, and nl and n_z are uncorrelated Gaussian variables. The detector bases its decision on the linear combination of r_1 and r_z, i.e r = rl + kr2

a. Determine the values of k that minimizes the probability of error.

b. Plot the probability of error for Y'21 = 1, $y^2$2, = 3 and either K = 1 or K is the optimum value found in (a). Compare the results.

7.0 References/ Further Reading

Application of Multicarrier Modulation for Digital Transmission on Digital Subscriber lines by Starr et al (1999) and Bingham (2000).

Module 4: Digital Communication through band-limited channels and adaptive equalization

Unit 1: Adaptive Equalization

Unit 2: Digital Communication through Band-limited Channels

Unit 3: Carrier and Symbol Synchronization

Unit 4: An Introduction to Information Theory

Unit 1: Adaptive Equalization

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Adaptive Linear Equalizer

3.2 Adaptive Decision Feedback Equalizer

3.3 Adaptive Equalization of Trellis-Coded Signals

3.4 Self-recovering (blind) Equalization

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/Further Reading

1.0 Introduction

In this unit, we present algorithms for automatically adjusting the equalizer co-efficient to optimize a specified performance index and to adaptively compensate for time variations in the channel characteristics.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain adaptive linear equalizer
- Discuss adaptive decision-feedback equalizer
- Understand adaptive equalization of Trellis-Coded Signals

3.0 Adaptive Linear equalizer

In the case of the linear equalizer, recall that we considered two different criteria for determining the values of the equalizer coefficients {ck}. One criterion was based on the minimization of the peak distortion at the output of the equalizer, which is defined by Equation 9.4-22. The other criterion was based on the minimization of the mean square error at the output of the equalizer, which is defined by Equation 9.4-42. Below, we describe two algorithms for performing the optimization automatically and adaptively.

The Zero-Forcing Algorithm

In the peak-distortion criterion, the peak distortion D(c), given by Equation 9.4-22, is minimized by selecting the equalizer coefficients {ck}. In general, there is no simple computational algorithm for performing this optimization, except in the special case where the peak distortion at the input to the equalizer, defined as Do

in Equation 9.4-23, is less than unity. When Do < 1, the distortion D(c) at the output of the equalizer is minimized by forcing the equalizer response qn = 0, for 1 <_ In I <_ K, and qo = 1. In this case, there is a simple computational algorithm, called the zero-forcing algorithm, that achieves these conditions.

The zero-forcing solution is achieved by forcing the cross correlation between the error sequence Ek = Ik - Ik and the desired information sequence {Ik} to be zero for shifts in the range 0 <_ In I <_ K. The demonstration that this leads to the desired solution is quite simple. We have

$$\begin{aligned} E(\varepsilon_k I_{k-j}^*) &= E[(I_k - \hat{I}_k)I_{k-j}^*] \\ &= E(I_k I_{k-j}^*) - E(\hat{I}_k I_{k-j}^*), \qquad j = -K, \ldots, K \end{aligned}$$

We assume that the information symbols are uncorrelated, i.e., E (IkIj*) = Skj, and that the information sequence {Ik} is uncorrelated with the additive noise sequence NJFor 1k, we use the expression given in Equation 9.4-41. Then, after taking the expected values in Equation 10.1-1, we obtain

$$E(\varepsilon_k I_{k-j}^*) = \delta_{j0} - q_j, \qquad j = -K, \ldots, K$$

Therefore, the conditions

$$E(\varepsilon_k I_{k-j}^*) = 0, \qquad j = -K, \ldots, K$$

are fulfilled when $q_0 = 1$ and $q_n = 0$, $1 < |n| < K$.

training sequence {Ik) to the receiver, which can be used to estimate the cross correlation by substituting time averages for the ensemble averages given in Equation 10.1-1. After the initial training, which will require the transmission of a training sequence of some predetermined length that equals or exceeds the equalizer length, the equalizer coefficients that satisfy Equation 10.1-3 can be determined.

A simple recursive algorithm for adjusting the equalizer coefficients is

$$c_j^{(k+1)} = c_j^{(k)} + \Delta\varepsilon_k I_{k-j}^*, \qquad j = -K, \ldots, -1, 0, 1, \ldots, K$$

where $c^{(k)}_j$ is the value of the j th coefficient at time t = kT , Ek = Ik - I k is the error signal at time t = kT, and 0 is a scale factor that controls the rate of adjustment, as will be explained later in this section. This is the zero forcing algorithm. The term $\varepsilon k I^*_k$ -j

is an estimate of the cross correlation (ensemble average) E(Ek1k j). The averaging operation of the cross correlation is accomplished by means of the recursive first-order difference equation algorithm in Equation 10.1-4, which represents a simple discretetime integrator.

3.1 Adaptive Decision-Feedback Equalizer

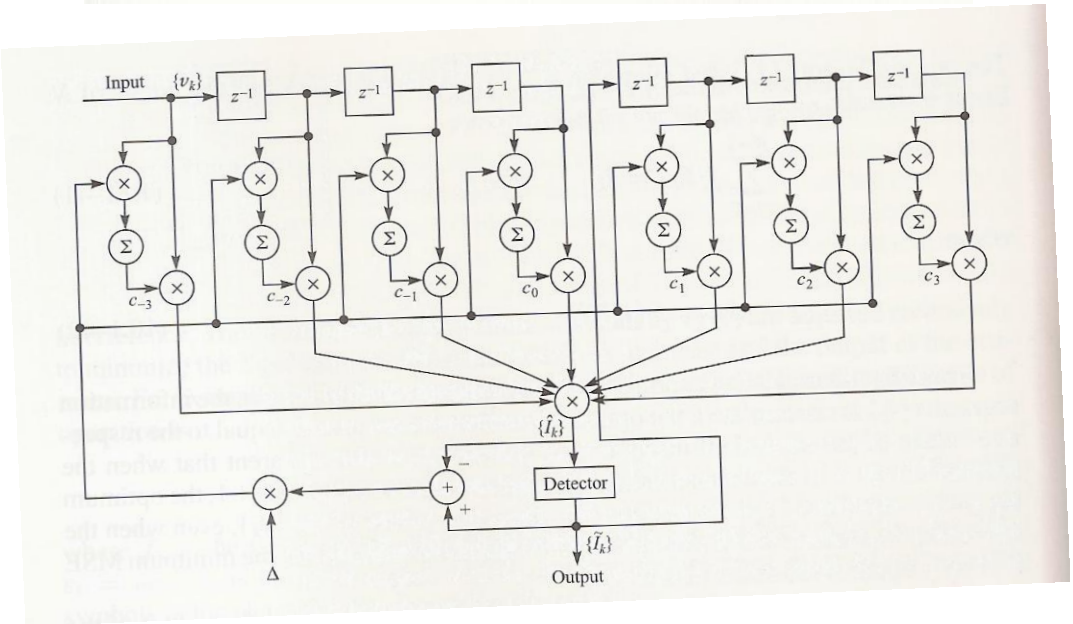
As in the case of the linear adaptive equalizer, the coefficients of the feedforward filter and the feedback filter in a decision-feedback equalizer (DFE) may be adjusted recursively, instead of inverting a matrix as implied by Equation 9.5-3. Based on the minimization of the MSE at the output of the DFE, the steepest-descent algorithm takes the form

$$C_{k+1} = C_k + \Delta E\left(\varepsilon_k V_k^*\right)$$

where Ck is the vector of equalizer coefficients in the kth signal interval, E (EkV k) is the cross correlation of the error signal Ek = Ik -I k with V k = [vk+x, ... vk Ik-1 ... Ik-x2l `, representing the signal values in the feedforward and feedback filters at time t = kT. The MSE is minimized when the cross-correlation vector E (£kV k) = 0 as k ---> oc. Since the exact cross-correlation vector is unknown at any time instant, we use as an estimate the vector Sk Vk and average out the noise in the estimate through the recursive equation.

$$\hat{C}_{k+1} = \hat{C}_k + \Delta\varepsilon_k V_k^*$$

This is the LMS algorithm for the DFE.



As in the case of a linear equalizer, we may use a training sequence to adjust the coefficients of the DFE initially. Upon convergence to the (near-) optimum coefficients decisions at the output of the detector are used in forming the error signal '-k and fed to the feedback filter. This is the adaptive mode of the DFE, which is illustrated in Figure 10.2-1. In this case, the recursive equation for adjusting the equalizer coefficient is

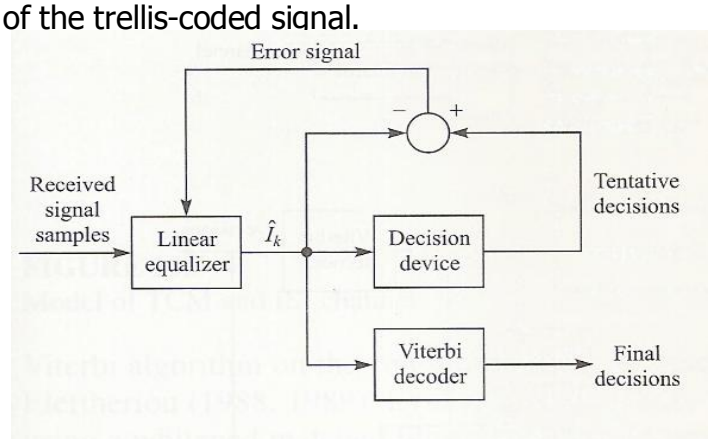
$$\tilde{C}_{k+1} = \tilde{C}_k + \Delta\tilde{\varepsilon}_k V_k^*$$

where $\tilde{\varepsilon}_k = \tilde{I}_k - \hat{I}_k$ and $V_k = [v_{k+K_1} \cdots v_k \tilde{I}_{k-1} \cdots \tilde{I}_{k-K_2}]^t$.

The performance characteristics of the LMS algorithm for the DFE are basically the same as the development given in Sections 10. 1-3 and 10. 1-4 for the linear adaptive

3.2 Adaptive Equalization of Trellis-Coded Signals

Bandwidth efficient trellis-coded modulation that was described in Section 8.12 is freSNR per bit for achieving a specified error rate. Channel distortion of the trellis-coded signal forces us to use adaptive equalization in order to reduce the intersymbol interference. The output of the equalizer is then fed to the Viterbi decoder, which performs a soft-decision decoding of the trellis-coded signal.



The question that arises regarding such a receiver is, how do we adapt the equalizer in a data transmission mode? One possibility is to have the equalizer make its own decisions at its output solely for the purpose of generating an error signal for adjusting its tap coefficients, as shown in the block diagram in Figure 10.3-1. The problem with this approach is that such decisions are generally unreliable, since the pre-decoding coded symbol SNR is relatively low. A high error rate would cause a significant degradation in the operation of the equalizer, which would ultimately affect the reliability of the decisions at the output of the decoder. The more desirable alternative is to use the post-decoding decisions from the Viterbi decoder, which are much more

reliable, to continuously adapt the equalizer. This approach is certainly preferable and viable when a linear equalizer is used prior to the Viterbi decoder. The decoding delay inherent in the Viterbi decoder can be overcome by introducing an identical delay in the tap weight adjustment of the equalizer coefficients as shown in Figure 10.3-2. The major price that must be paid for the added delay is that the step-size parameter in the LMS algorithm must be reduced, as described by Long et al. (1987, 1989), in order to achieve stability in the algorithm.

In channels with severe ISI, the linear equalizer is no longer adequate for compensating the channel intersymbol interference. Instead, we would like to use a DFE. But the DFE requires reliable decisions in its feedback filter in order to cancel out the intersymbol interference from previously detected symbols. Tentative decisions prior to decoding would be highly unreliable and, hence, inappropriate.

4.0 Conclusion

In conclusion, we have provided an overviewed of three classes of blind equalization algorithms that find applications in digital communications of the there families of algorithms described, those based on the maximumlikelihood criterion for jointly estimating the channel impulse response and the data sequence are optional and require relatively few received signal samples for performing channel estimation.

5.0 Summary

Adaptive equalization for digital communications was developed by Lucky between 1965-1966. His algorithms was based on the peak distortion criterion and led to the zero-forcing algorithm.

6.0 Tutor Marked Assignment

1. Show that the gradient vector in the minimization of the MSE may be expressed as:

Gk = - E (EK Vk)

Where the error £K = Ik - I_R, and the estimate of G_k, i.e, Gk = - £K VR

Satieties the condition that E (Gk) = Gk

7.0 References/ Further Reading

PLS Lattice Algorithms for general signal estimation application by Morf (1977), Morfand Lee (1978)

Unit 2: Digital Communication through band-limited Channel

1.0 Introduction

In this unit, we consider the problem of signal design when the channel is band limited to some specified bandwidth of W Hz. Under this condition, the channel may be modeled as a linear filter having an equivalent low pass t frequency response (ff) that is zero for /f/ > w.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain the characteristics of band limited channel.
- Understand signal design for band limited channels

- Explain optimum receiver for channels with ISI and AWGN

3.1 Characteristics of Band-Limited channel

Of the various channels available for digital communications, telephone channels are by far the most widely used. Such channels are characterized as band-limited linear filters. This is certainly the proper characterization when frequency-division multiplexing (FDM) is used as a means for establishing channels in the telephone network. Modern telephone networks employ pulse-code modulation (PCM) for digitizing and encoding the analog signal and time-division multiplexing (TDM) for establishing multiple channels. Nevertheless, filtering is still used on the analog signal prior to sampling and encoding. Consequently, even though the present telephone network employs a mixture of FDM and TDM for transmission, the linear filter model for telephone channels is still appropriate.

For our purposes, a bandlimited channel such as a telephone channel will be characterized as a linear filter having an equivalent lowpass frequency-response characteristic C(f). Its equivalent lowpass impulse response is denoted by c(t). Then, if a signal of the form

$$s(t) = \mathrm{Re}\,[v(t)e^{j2\pi f_c t}]$$

is transmitted over a bandpass telephone channel, the equivalent low-pass received signal is

$$r(t) = \int_{-\infty}^{\infty} v(\tau)c(t-\tau)\,d\tau + z(t)$$

where the integral represents the convolution of c(t) with v(t), and z(t) denotes the additive noise. Alternatively, the signal term can be represented in the frequency domain as V (f)C(f), where V (f) is the Fourier transform of v(t).

If the channel is band-limited to W Hz, then C(f) = 0 for I f I > W. As a consequence, any frequency components in V (f) above I f I = W will not be passed by the channel. For this reason, we limit the bandwidth of the transmitted signal to W Hz also. Within the bandwidth of the channel, we may express the frequency response C(f) as

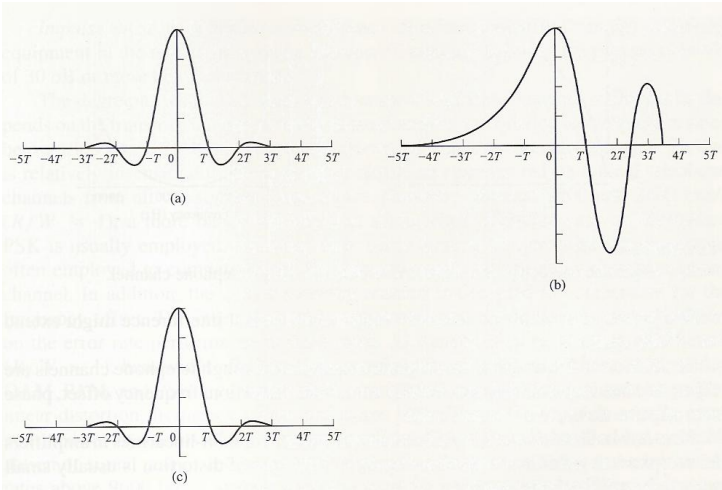
$$C(f) = |C(f)|e^{j\theta(f)}$$

where /C(f)/ I is the amplitude-response characteristic and θ(f) is the phase-response characteristic. Furthermore, the envelope delay characteristic is defined as

$$\tau(f) = -\frac{1}{2\pi}\frac{d\theta(f)}{df}$$

A channel is said to be nondistorting or ideal if the amplitude response /C(f) I is constant for all /f/ ≤ W and 0(f) is a linear function of frequency, i.e., r (f) is a constant for all /f/ ≤ W. On the other hand, if I C (f) I is not constant for all I f I < W, we say that the channel distorts the transmitted signal V (f) in amplitude, and, if r (f) is not constant for all /f/ ≤ W, we say that the channel distorts the signal V(f) in delay.

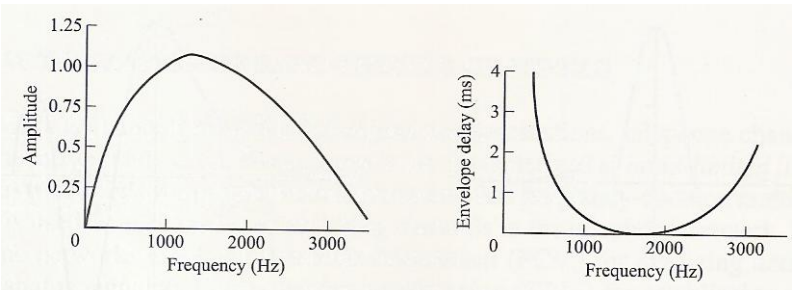
As a result of the amplitude and delay distortion caused by the nonideal channel frequency-response characteristic C(f), a succession of pulses transmitted through the channel at rates comparable to the bandwidth W are smeared to the point that they are



no longer distinguishable as well-defined pulses at the receiving terminal. Instead, they overlap, and, thus, we have intersymbol interference. As an example of the effect of delay distortion on a transmitted pulse, Figure 9.1-1a illustrates a band-limited pulse having zeros periodically spaced in time at points labeled ::LT, ::L2T, etc. If information is conveyed by the pulse amplitude, as in PAM, for example, then one can transmit a sequence of pulses, each of which has a peak at the periodic zeros of the other pulses. However, transmission of the pulse through a channel modeled as having a linear envelope delay characteristic -c (f) (quadratic phase B(f)) results in the received pulse shown in Figure 9.1-lb having zero-crossings that are no longer periodically spaced. Consequently, a sequence of successive pulses would be smeared into one another and the peaks of the pulses would no longer be

distinguishable. Thus, the channel delay distortion results in intesymbol interference. As will be discussed in this chapter, it is possible to compensate for the nonideal frequency-response characteristic of the channel by use of a filter or equalizer at the demodulator. Figure 9.1-1c illustrates the output of a linear equalizer that compensates for the linear distortion in the channel.

The extent of the intersymbol interference on a telephone channel can be appreciated by observing a frequency-response characteristic of the channel. Figure 9.1-2 illustrates the measured average amplitude and delay as functions of frequency for a medium-range (180-725 mi) telephone channel of the switched telecommunications network as given by Duffy and Tratcher (1971). We observe that the usable band of the channel extends from about 300 Hz to about 3000 Hz. The corresponding impulse response of this average channel is shown in Figure 9.1-3. Its duration is about 10 ms. In comparison, the transmitted symbol rates on such a channel may be of the order



$$r(t) = \sum_n I_n h(t - nT) + z(t)$$

Average amplitude and delay characteristics of medium-range telephone channel.

of 2500 pulses or symbols per second. Hence, intersymbol interference might extend over 20-30 symbols.

In addition to linear distortion, signals transmitted through telephone channels are subject to other impairments, specifically non-linear distortion, frequency offset, phase jitter, impulse noise, and thermal noise.

Non-linear distortion in telephone channels arises from non-linearities in amplifiers and compandors used in the telephone system. This type of distortion is usually small and it is very difficult to correct.

A small frequency offset, usually less than 5 Hz, results from the use of carrier equipment in the telephone channel. Such an offset cannot be tolerated in high-speed digital transmission systems that use synchronous phase-coherent demodulation. The offset is usually compensated for by the carrier recovery loop in the demodulator.

Phase fitter is basically a low-index frequency modulation of the transmitted signal with the low-frequency harmonics of the power line frequency (50-60 Hz). Phase jitter poses a serious problem in digital transmission at high rates. However, it can be tracked and compensated for, to some extent, at the demodulator.

3.2 Optimum receiver for channels with ISI and AWGN

In this section, we derive the structure of the optimum demodulator and detector for digital transmission through a nonideal band-limited channel with additive Gaussian noise. We begin with the transmitted (equivalent lowpass) signal given by Equation 9.2-1. The received (equivalent lowpass) signal is expressed as

$$r(t) = \sum_n I_n h(t - nT) + z(t)$$

where h(t) represents the response of the channel to the input signal pulse g(t) and z(t) represents the additive white Gaussian noise.

First we demonstrate that the optimum demodulator can be realized as a filter matched to h (t), followed by a sampler operating at the symbol rate 1/T and a subsequent processing algorithm for estimating the information sequence {In} from the sample values. Consequently, the samples at the output of the matched filter are sufficient for the estimation of the sequence $\{\mathbf{I}_n\}$.

Optimum Maximum-Likelihood Receiver

Using the Karhunen-Loeve expansion, we expand the received signal r1(t) in the series

$$r_l(t) = \lim_{N \to \infty} \sum_{k=1}^{N} r_k \phi_k(t)$$

where {5k(t)1 is a complete set of orthonormal functions and Irk} are the observable random variables obtained by projecting r1(t) onto the set {ϕk(t)}. It is easily shown that

$$r_k = \sum I_n h_{kn} + z_k, \qquad k = 1, 2, \ldots$$

where h_{kn} is the value obtained from projecting h(t - nT) onto 6k(t), and zk is the value obtained from projecting z(t) onto Ok(t). The sequence tzkI is Gaussian with zero-mean and covariance

$$E\left(z_k^* z_m\right) = 2N_0 \delta_{km}$$

The joint probability density function of the random variables rN - [r1 r2 ... rN] conditioned on the transmitted sequence Ip - [*I*1 *I*2 ... *I*p], where p ≤ N, is

$$p(\boldsymbol{r}_N | \boldsymbol{I}_p) = \left(\frac{1}{2\pi N_0}\right)^N \exp\left(-\frac{1}{2N_0}\sum_{k=1}^{N}\left|r_k - \sum_n I_n h_{kn}\right|^2\right)$$

4.0 Conclusion

We treated the design of the signal pulse g(t) in a linearly modulated signal, represented as;

V(t) = ∑Ing (t-nT)

that efficiently utilizes the total available channel bandwidth w. lastly, we considered the of the receiver in the presence of inter-symbol interference and AWGN.

5.0 Summary

In this unit, we saw that when the channel is ideal for /f/ s w, a signal pulse could be designed that allow us to transmit at symbol rates comparable to or exceeding the channel bandwidth W. On the other hand, when the channel is not ideal, signal transmission at a symbol rate equal to or exceeding W results in inter-symbol interference (ISI) among a number of adjacent symbols.

6.0 Tutor Marked Assignment

1. What is inter-symbol interference?

2. A channel is said to be distortionless if the response y(t) to an input x (t) is Kx (t-to), where K and to are constants. Show that if the frequency response of the channel is A(f) ejo(f), where A (f) and t(f) are real, the necessary and sufficient conditions for distortionless transmission are A(f) = K and t (f) = 2nfto ± nTc, n = 0, 1, 2.

7.0 References/ Further Reading

Turbo Equalization by Raphaeli and Zarai (1998) and Douillard et al (1995).

Unit 3: Carrier and Symbol Synchronization

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Signal Parameter Estimation

3.2 Carrier Phase Estimation

3.3 Symbol Timing Estimation

3.4 Joint Estimation of Carrier Phase and Symbol timing

3.5 Performance characteristics of ML Estimator

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/ Further Reading

1.0 Introduction

We have observed that in a digital communication system, the output of the demodulator must be sampled periodically, once para-symbol internal, in order to recover the transmitted information. Since the propagation delay from the transmitter to the receiver is generally unknown at the receiver, symbol timing must be derived from the received signal in order to synchronously sample the output of the demodulator.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain signal parameter estimation. - Discuss carrier phase and symbol timing estimation.

- Understanding the performance characteristics of ML estimators

3.1 Signal parameter estimation

We have observed that in a digital communication system, the output of the demodulator must be sampled periodically, once per symbol interval, in order to recover the transmitted information. Since the propagation delay from the transmitter to the receiver is generally unknown at the receiver. symbol timing must be derived from the received signal in order to synchronously sample the output of the demodulator.

The propagation delay in the transmitted signal also results in a carrier offset, which must be estimated at the receiver if the detector is phase-coherent. In this chapter, we consider methods for deriving carrier and symbol synchronization at the receiver.

Signal Parameter Estimation

Let us begin by developing a mathematical model for the signal at the input to the receiver. We assume that the channel delays the signals transmitted through it and corrupts them by the addition of Gaussian noise. Hence, the received signal may be expressed as

$$r(t) = s(t - \tau) + n(t)$$

where

$$s(t) = \mathrm{Re}\left[s_l(t)e^{j2\pi f_c t}\right]$$

and where r is the propagation delay and s, (t) is the equivalent low-pass signal. The received signal may be expressed as

$$r(t) = \mathrm{Re}\{[s_l(t-\tau)e^{j\phi} + z(t)]e^{j2\pi f_c t}\}$$

where the carrier phase 0, due to the propagation delay t, is 0 = -2n f,z. Now, from this formulation, it may appear that there is only one signal parameter to be estimated, namely, the propagation delay, since one can determine ϕ from knowledge of f, and i. However, this is not the case. First of all, the oscillator that generates the carrier signal for demodulation at the receiver is generally not synchronous in phase with that at the transmitter. Furthermore, the two oscillators may be drifting slowly with time, perhaps in different directions. Consequently, the received carrier phase is not only dependent on the time delay T. Furthermore, the precision to which one must synchronize in time for the purpose of demodulating the received signal depends on the symbol interval T. Usually, the estimation error in estimating T must be a relatively small fraction of T. For example, ± 1 percent of T is adequate for practical applications. However, this level of precision is generally inadequate for estimating the carrier phase, even if 0 depends only on T. This is due to the fact that f, is generally large, and, hence, a small estimation error in T causes a large phase error.

In effect, we must estimate both parameters T and 0 in order to demodulate and coherently detect the received signal. Hence, we may express the received signal as

$$r(t) = s(t; \phi, \tau) + n(t)$$

where 0 and T represent the signal parameters to be estimated. To simplify the notation, we let 8 denote the parameter vector {o, T}, so that s(t; 4, T) is simply denoted by s(t; B).

There are basically two criteria that are widely applied to signal parameter estimation: the maximum-likelihood (ML) criterion and the maximum a posteriori probability (MAP) criterion. In the MAP criterion, the signal parameter vector 9 is modeled as random and characterized by an a priori probability density function p(e). In the maximum-likelihood criterion, the signal parameter vector 8 is treated as deterministic but unknown.

By performing an orthonormal expansion of r(t) using N orthonormal functions {0,z(t)}, we may represent r (t) by the vector of coefficients (rl r2 ... rN) =__ r. The joint PDF of the random variables (rl r2 ... rN) in the expansion can be expressed as p (r 10). Then, the ML estimate of ® is the value that maximizes p(r I B). On the other hand, the MAP estimate is the value of 9 that maximizes the a posteriori probability density function

$$p(\boldsymbol{\theta}|\boldsymbol{r}) = \frac{p(\boldsymbol{r}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{r})}$$

We note that if there is no prior knowledge of the parameter vector θ, we may assume that p(θ) is uniform (constant) over the range of values of the parameters. In such a case, the value of B that maximizes p (r/θ) also maximizes p(r/θ). Therefore, the MAP and ML estimates are identical.

In our treatment of parameter estimation given below, we view the parameters 4 and T as unknown, but deterministic. Hence, we adopt the ML criterion for estimating them. In the ML estimation of signal parameters, we require that the receiver extract the estimate by observing the received signal over a time interval $T_o \geq T$, which is called the observation interval. Estimates obtained from a single observation interval are sometimes called one-shot estimates. In practice, however, the estimation is performed on a continuous basis by using tracking loops (either analog or digital) that continuously update the estimates. Nevertheless, one-shot estimates yield insight for tracking loop implementation. In addition, they prove useful in the analysis of the performance of ML estimation, and their performance can be related to that obtained with a tracking loop.

3.2 Symbol timing estimation

In a digital communication system, the output of the demodulator must be sampled periodically at the symbol rate, at the precise sampling time instants tm = mT +r, where T is the symbol interval and r is a nominal time delay that accounts for the propagation time of the signal from the transmitter to the receiver. To perform this periodic sampling, we require a clock signal at the receiver. The process of extracting

such a clock signal at the receiver is usually called symbol synchronization or timing recovery.

Timing recovery is one of the most critical functions that is performed at the receiver of a synchronous digital communication system. We should note that the receiver must know not only the frequency (1/T) at which the outputs of the matched filters or correlators are sampled, but also where to take the samples within each symbol interval. The choice of sampling instant within the symbol interval of duration T is called the timing phase.

Symbol synchronization can be accomplished in one of several ways. In some communication systems, the transmitter and receiver clocks are synchronized to a master clock, which provides a very precise timing signal. In this case, the receiver must estimate and compensate for the relative time delay between the transmitted and received signals. Such may be the case for radio communication systems that operate in the very low frequency (VLF) band (below 30 kHz), where precise clock signals are transmitted from a master radio station.

Another method for achieving symbol synchronization is for the transmitter to simultaneously transmit the clock frequency 1/T or a multiple of 1/T along with the information signal. The receiver may simply employ a narrowband filter tuned to the transmitted clock frequency and, thus, extract the clock signal for sampling. This approach has the advantage of being simple to implement. There are several disadvantages, however. One is that the transmitter must allocate some of its available power to the transmission of the clock signal. Another is that some small fraction of the available channel bandwidth must be allocated for the transmission of the clock signal. In spite of these disadvantages, this method is frequently used in telephone transmission systems that employ large bandwidths to transmit the signals of many users. In such a case, the transmission of a clock signal is shared in the demodulation of the signals among the many users. Through this shared use of the clock signal, the penalty in the transmitter power and in bandwidth allocation is reduced proportionally by the number of users. A clock signal can also be extracted from the received data signal. There are a number of different methods that can be used at the receiver to achieve self-synchronization. In this section, we treat both decision-directed and non-decision-directed methods.

Maximum-Likelihood Timing Estimation

Let us begin by obtaining the ML estimate of the time delay r. If the signal is a base band PAM waveform, it is represented as

$$r(t) = s(t;\tau) + n(t)$$

where

$$s(t;\tau) = \sum_n I_n g(t - nT - \tau)$$

As in the case of ML phase estimation, we distinguish between two types of timing estimators, decision-directed timing estimators and non-decision-directed estimators. In the former, the information symbols from the output of the demodulator are treated as the known transmitted sequence. In this case, the log-likelihood function has the form

$$\Lambda_L(\tau) = C_L \int_{T_0} r(t)s(t;\tau)\,dt$$

If we substitute Equation 5.3–2 into Equation 5.3–3, we obtain

$$\Lambda_L(\tau) = C_L \sum_n I_n \int_{T_0} r(t)g(t - nT - \tau)\,dt$$

$$= C_L \sum_n I_n y_n(\tau)$$

where $y_n(t)$ is defined as

$$y_n(\tau) = \int_{T_0} r(t)g(t - nT - \tau)\,dt$$

A necessary condition for $\hat{\tau}$ to be the ML estimate of τ is that

$$\frac{d\Lambda_L(\tau)}{d\tau} = \sum_n I_n \frac{d}{d\tau} \int_{T_0} r(t)g(t - nT - \tau)\,dt$$

$$= \sum_n I_n \frac{d}{d\tau}[y_n(\tau)] = 0$$

3.3 Conclusion

This unit considered the methods of deriving carrier and symbol synchronization at the receiver. Besides, the propagation delay in the transmitted signal result in carrier offset, which be estimated at the receiver if the detector is phase-coherent.

5.0 Summary

In this unit, we considered the mathematics model for the signal at the input to the receiver. Besides, the ML method for signal parameter estimation was presented and

applied to the estimator of the carrier phase and symbol timing. Furthermore, described their performance characteristics.

6.0 Tutor Marked Assignment

1. Sketch the equivalent realization of the binary PSK receiver in unit 1 that employs a matched filter instead of a correlator.

2. Determine the joint ML estimate of r and 0 for PAM signal.

7.0 References/ Further Reading

Carrier Phase Recovery and Time Synchronization Techniques by Stiffler (1971), Meyret et al (1998).

Unit 4: An Introduction to Information Theory

1.0 Introduction

2.0 Objectives

3.0 Main Contents

3.1 Mathematical Models of Information Sources

3.2 A Logarithmic Measure of Information

3.3 Lossless Coding of Information Sources

3.4 Channel Models and Channel Capacity

3.5 Channel Reliability fun function

3.6 The Channel Cutoff Rate

4.0 Conclusion

5.0 Summary

6.0 Tutor marked Assignment

7.0 References/Further Reading

1.0 Introduction

This unit deals with fundamental limits on communication by fundamental limits we means the study of conditions under which the two fundamental tasks in communications - compression and transmission are possible for some important source and channel models, we can precisely state the limits for compression and transmission of information.

2.0 Objectives

At the end of this unit, you should be able to:

- Understand the mathematics models for information sources
- Explain logarithmic measure of information.
- Explain lossless coding of information sources

3.1 Mathematics models for information sources

Any information source produces an output that is random; i.e., the source output is characterized in statistical terms. Otherwise, if the source output were known exactly, there would be no need to transmit it. In this section, we consider both discrete and analog information sources, and we postulate mathematical models for each type of source.

The simplest type of a discrete source is one that emits a sequence of letters selected from a finite alphabet. For example, a binary source emits a binary sequence of the form 100101110 - - - , where the alphabet consists of the two letters {0, 1}. More generally, a discrete information source with an alphabet of L possible letters, say {x1, x2, . . . , xL}, emits a sequence of letters selected from the alphabet.

To construct a mathematical model for a discrete source, we assume that each letter in the alphabet {x1, x2, . . . , XLI has a given probability Pk of occurrence. That is,

$$p_k = \mathrm{P}[X = x_k], \qquad 1 \le k \le L$$

where

$$\sum_{k=1}^{L} p_k = 1$$

We consider two mathematical models of discrete sources. In the first, we assume that the output sequence from the source is statistically independent. That is, the current

output letter is statistically independent of all past and future outputs. A source whose output satisfies the condition of statistical independence among output letters is said to be memoryless. If the source is discrete, it is called a discrete memoryless source (DMS). The mathematical model for a DMS is a sequence of iid random variables {Xi}.

If the output of the discrete source is statistically dependent, such as English text, we may construct a mathematical model based on statistical stationarity. By definition, a discrete source is said to be stationary if the joint probabilities of two sequences of length n, say, a1, a2, . . . , an and a1+m, a2+m, . . . , an+m, are identical for all n > 1 and for all shifts m. In other words, the joint probabilities for any arbitrary length sequence of source outputs are invariant under a shift in the time origin.

An analog source has an output waveform x(t) that is a sample function of a stochastic process X(t). We assume that X(t) is a stationary stochastic process with autocorrelation function Rx(r) and power spectral density Sx(f). When X(t) is a band-limited stochastic process, i.e., Sx(f) = 0 for I f I > W, the sampling theorem may be used to represent X(t) as

$$X(t) = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2W}\right) \operatorname{sinc}\left[2W\left(t - \frac{n}{2W}\right)\right]$$

where {X(n/2W)} denote the samples of the process X(t) taken at the sampling (Nyquist) rate of f = 2W samples/s. Thus, by applying the sampling theorem, we may convert the output of an analog source to an equivalent discrete-time source. Then the source output is characterized statistically by the joint PBF P(xi, x2, . . . , x") for all m > 1, where Xn = X(n/2W), 1 < n _< m, are the random variables corresponding to the samples of X(t).

We note that the output samples {X(n/2W)} from the stationary sources are generally continuous, and hence they cannot be represented in digital form without some loss in precision. For example, we may quantize each sample to a set of discrete values, but the quantization process results in loss of precision, and consequently the original signal cannot be reconstructed exactly from the quantized sample values. Later in this chapter, we shall consider_ the distortion resulting from quantization of the samples from an analog source.

3.2 Lossless coding of information sources

The goal of data compression is to represent a source with the fewest bits such that best recovery of the source from the compressed data is possible. Data compression can be broadly classified into lossless and lossy compression. In lossless compression the goal is to minimize the number of bits in such a way that perfect (lossless) reconstruction of the source from compressed data is possible. In lossy data compression the data are compressed subject to a maximum tolerable distortion. In this section we study the fundamental bounds for lossless compression as well as some common lossless compression algorithms.

Let us assume that a DMS is represented by independent replicas of random variable X taking values in the set d' = f al , a2, . . . , aN] with corresponding probabilities P1, p2, . . . , PN. Let x denote an output sequence of length n for this source, where n is assumed to be large. We call this sequence a typical sequence if the number of occurrences of each ai in x is roughly npi for 1 <_ i <_ N. The set of typical sequences is denoted by .A.

The law of large numbers, reviewed in Section 2.5, states that with high probability approaching 1 as n -+ oc, outputs of any DMS will be typical. Since the number of occurrences of ai in x is roughly npi and the source is memoryless, we have

$$\begin{aligned}\log \mathrm{P}[X = x] &\approx \log \prod_{i=1}^{N} (p_i)^{np_i} \\ &= \sum_{i=1}^{N} np_i \log p_i \\ &= -nH(X)\end{aligned}$$

Hence,

$$\mathrm{P}[X = x] \approx 2^{-nH(X)}$$

This states that all typical sequences have roughly the same probability, and this common probability is $2^{-nH(X)}$.

Since the probability of the typical sequences, for large n, is very close to 1, we conclude that the number of typical sequences, i.e., the cardinality of .A, is roughly

$$|\mathcal{A}| \approx 2^{nH(X)}$$

This discussion shows that for large n, a subset of all possible sequences, called the typical sequences, is almost certain to occur. Therefore, for transmission of source outputs it is sufficient to consider only this subset. Since the number of typical sequences is 2nH(X), for their transmission nH(X) bits are sufficient, and therefore the number of required bits per source output, i.e., the transmission rate, is given by

$$R \approx \frac{nH(X)}{n} = H(X) \qquad \text{bits per transmission}$$

The informal argument given above can be made rigorous (see the books by Cover and Thomas (2006) and Callager (1968)) in the following theorem first stated by Shannon (1948).

SHANNON'S FIRST THEOREM (LOSSLESS SOURCE CODING THEOREM) Let X denote a DMS with entropy X. There exists a lossless source code for this source at any rate R if R > H(X). There exists no lossless code for this source at rates less than H(X). This theorem sets a fundamental limit on lossless-source coding and shows that the entropy of a DMS, which was defined previously based on intuitive reasoning, plays a fundamental role in lossless compression of information sources.

3.3 Channel models and channel capacity

In the model of a digital communication system described in Chapter l, we recall that the transmitter building blocks consist of the discrete-input, discrete-output channel encoder followed by the modulator. The function of the discrete channel encoder is to introduce, in a controlled manner, some redundancy in the binary information sequence, which can be used at the receiver to overcome the effects of noise and interference encountered in the transmission of the signal through the channel. The encoding process generally involves taking k information bits at a time and mapping each k-bit sequence into a unique n-bit sequence, called a codeword. The amount of redundancy introduced by the encoding of the data in this manner is measured by the ratio n/ k. The reciprocal of the ratio, namely k/n, is called the code rate and denoted by Rc.

The binary sequence at the output of the channel encoder is fed to the modulator, which serves as the interface to the communication channel. As we have discussed, the

modulator may simply map each binary digit into one of two possible waveforms; i.e., a 0 is mapped into sl (t) and a 1 is mapped into SAO. Alternatively, the modulator may transmit q -bit blocks at a time by using M = 2q possible waveforms.

At the receiving end of the digital communication system, the demodulator processes the channel-corrupted waveform and reduces each waveform to a scalar or a vector that represents an estimate of the transmitted data symbol (binary or M-ary). The detector, which follows the demodulator, may decide whether the transmitted bit is a 0 or a 1. In such a case, the detector has made a hard decision. If we view the decision process at the detector as a form of quantization, we observe that a hard decision corresponds to binary quantization of the demodulator output. More generally, we may consider a detector that quantizes to Q > 2 levels, i.e., a Q-ary detector. If M-ary signals are used, then Q > M. In the extreme case when no quantization is performed, Q = cc. In the case where Q > M, we say that the detector has made a soft decision.

The quantized output from the detector is then fed to the channel decoder, which exploits the available redundancy to correct for channel disturbances.

In the following sections, we describe three channel models that will be used to establish the maximum achievable bit rate for the channel.

Channel Models

In this section we describe channel models that will be useful in the design of codes. A general communication channel is described in terms of its set of possible inputs, denoted by 3' and called the input alphabet; the set of possible channel outputs, denoted by and called the output alphabet; and the conditional probability that relates the input and output sequences of any length n, which is denoted by P [yi, y2, . . . , yn Ixl, x2, . . . , xn 1, where x = (XI, x2, . . . , xn) and y = (yl, y2, . . . , yn) represent input and output sequences of length n, respectively. A channel is called memoryless if we have

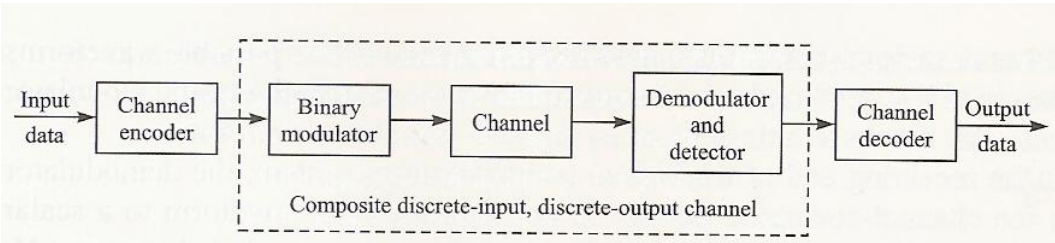
$$P[y|x] = \prod_{i=1}^{n} P[y_i|x_i] \qquad \text{for all } n$$

In other words, a channel is memoryless if the output at time i depends only on the input at time i.

The simplest channel model is the binary symmetric channel, which corresponds to the case with R= ~J = {0, 11. This is an appropriate channel model for binary modulation and hard decisions at the detector.

The Binary Symmetric Channel (BSC) Model

Let us consider an additive noise channel and let the modulator and the demodulator/detector be included as parts of the channel. If the modulator employs binary waveforms and the detector makes hard decisions, then the composite channel, shown in Figure 6.5-1, has a discrete-time binary input sequence and a discrete-time binary output sequence. Such a composite channel is characterized by the set X = {0, 1} of



A composite discrete input, discrete output channel formed by including the modulator and the demodulator as part of the channel.

possible inputs, the set of Gf = {0, 1] of possible outputs, and a set of conditional probabilities that relate the possible outputs to the possible inputs. If the channel noise and other disturbances cause statistically independent errors in the transmitted binary sequence with average probability p, then

$$P[Y=0|X=1]=P[Y=1|X=0]=p$$
$$P[Y=1|X=1]=P[Y=0|X=0]=1-p$$

Thus, we have reduced the cascade of the binary modulator, the waveform channel, and the binary demodulator and detector to an equivalent discrete-time channel which is represented by the diagram shown in Figure 6.5-2. This binary input, binary output, symmetric channel is simply called a binary symmetric channel (BSC). Since each output bit from the channel depends only on the corresponding input bit, we say that the channel is memoryless.

The Discrete memoryless Channel (DMC)

The BSC is a special case of a more general discrete input, discrete output channel. The discrete memoryless channel is a channel model in which the input and output alphabets X and Y are discrete sets and the channel is memoryless. For instance, this is the case when the channel uses an M-ary memoryless modulation scheme and the output of the detector consists of Q-ary symbols. The composite channel consists of modulatorchannel-detector as shown in Figure 6.5-1, and its input-output characteristics are described by a set of MQ conditional probabilities

$$P[y|x] \qquad \text{for } x \in \mathscr{X},\ y \in \mathscr{Y}$$

The graphical representation of a DMC is shown in Figure 6.5–3.

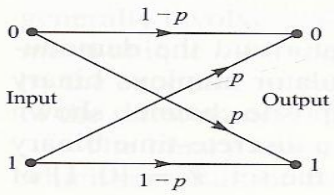
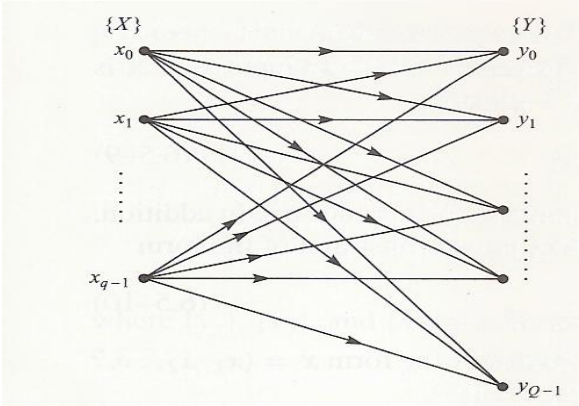


FIGURE 6.5–2
Binary symmetric channel.



In general, the conditional probabilities {P [y Ix]} that characterize a DMC can be arranged in an 19' I x 14' I matrix of the form P = [pij], 1 < i < I X I,1 < j < 19 I, P is called the probability transition matrix for the channel.

The Discrete-Input, Continuous-Output Channel

Now, suppose that the input to the modulator comprises symbols selected from a finite and discrete input alphabet X, with /x/ = M, and the output of the detector is unquantized, i.e., y = R. This leads us to define a composite discrete-time memoryless

channel that is characterized by the discrete input X, the continuous output Y, and the set of conditional probability density functions

The most important channel of this type is the additive white Gaussian noise (AWGN) channel, for which

Y=X+N

where N is a zero-mean Gaussian random variable with variance 62. For a given X = x, it follows that Y is Gaussian with mean x and variance 62. That is,

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}}$$

For any given input sequence Xi, i = l, 2, . . . , n, there is a corresponding output sequence

$$Y_i = X_i + N_i, \qquad i = 1, 2, \ldots, n$$

The condition that the channel is memoryless may be expressed as

$$p(y_1, y_2, \ldots, y_n | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(y_i | x_i)$$

4.0 Conclusion

Conclusively, this unit begins with a study of information sources and source coding communication systems are designed to transmit the information generated by a source to some destination. Information sources may take a variety of different forms. For example, in radio broadcasting, the source is generally an audio voice (voice or music). In TV broadcasting, the information source is a video source whose output is a moving image. The output of these sources are analog signals and, hence, the sources are called analog sources. In contrast, computers and storage device, such as magnetic or optional disks, produce discrete outputs (usually binary or ASCII characters) and hence, are called discrete sources.

5.0 Summary

Whether a source is analog or discrete, a digital communication system is designed to transmit information in digital form. Consequently, the outputs of the source must be converted to a format that can be transmitted digitally. Besides, we focus on communication channels and transmission of information, develop mathematical models for important channels and introduce two important parameters for communication channels-channel capacity and channel cutoff rate and elaborate on their meaning and significance.

6.0 Tutor Marked Assignment

1. X and Y are two discrete random variables with probabilities p(X= x, Y = y) = P(x,y)

Show that I (X,Y) 0, with equality if and only if X and Y are statistically independent. Hint: use the inequality In U <_ U-1, for 0 < U < 1, to show that - I (X,Y) <_ 0

2. Prove that In U <_ U-1 and also demonstrate the validity of this inequality by plotting in U and u-1 on the same graph.

7.0 References/ Further Reading

Treatment of Rate Distortion Theory by Blahut (1987) and Gray (1990)

MODULE 5: FADING CHANNEL I: CHARACTERIZATION AND SIGNALING

Unit 1: Characterization of Fading Multipath Channel

Unit 2: The Effect of Signal Characteristics on the Choice of a Channel Model

Unit 3: Diversity Technique for Fading Multipath Channels

Unit 4: Signaling Over a Frequency-Selective, Slowly Fading Channel: The rake Demodulator

Unit 5: MultiCarrier Modulation (OFDM)

UNIT 1: CHARACTERIZATION OF FADING MULTIPATH CHANNELS

1.0 Introduction

In this unit, we shall consider the signal design, receiver structure and receiver performance for more complex channels, namely, channels, having randomly time variant impulse responses. This characterization serves as a model for signal transmission over many radio channels; such as short ware ionospheric radio communication in the 3-30 MHz frequency band (Hf), tropospheric scatter (beyond - the horizon) radio communications in the 3003000 MHz frequency band (UHf) and 3000-30,000) MHz frequency bend (SHf) and ionosphere forward scatter in the 30-300 MHz frequency and (VHf).

2.0 Objectives

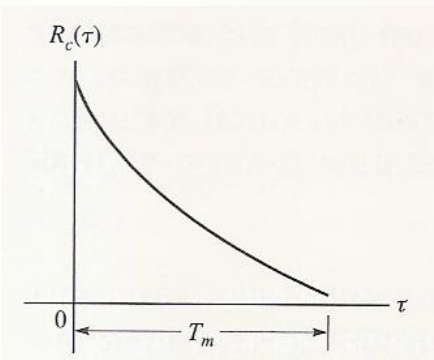
At the end of this unit, you should be able to;

- Explain the characterization of fading multipath channels;
- Explain the effect signal characteristics on the choice of channel model.
- Understanding the diversity techniques for fading multpath channels.

3.1 Channel correlation functions and power spactra.

We shall now develop a number of useful correlation functions and power spectral density functions that define the characteristics of a fading multipath channel. Our starting point is the equivalent lowpass impulse response c(-c; t), which is characterized as a complex-valued random process in the t variable. We assume that c(T; t) is widesense-stationary. Then we define the autocorrelation function of c(T; t) as

$$R_c(\tau_2, \tau_1; \Delta t) = E\left[c^*(\tau_1; t)c(\tau_2; t + \Delta t)\right]$$



In most radio transmission media, the attentuation and phase shift of the channel associated with path delay Tl is uncorrelated with the attenuation and phase shift associated with path delay T2. This is usually called uncorrelated scattering. We make the assumption that the scattering at two different delays is uncorrelated and incorporate it into Equation 13.1-10 to obtain

$$E\left[c^*(\tau_1; t)c(\tau_2; t + \Delta t)\right] = R_c(\tau_1; \Delta t)\delta(\tau_2 - \tau_1)$$

If we let Δt = 0, the resulting autocorrelation function RJT; 0) - Rjr) is simply the average power output of the channel as a function of the time delay t. For this reason, RjT) is called the multipath intensity profile or the delay power spectrum of the channel. In general, R,(T; At) gives the average power output as a function of the time delay T and the difference At in observation time.

In practice, the function $R_c(\tau; \Delta t)$ is measured by transmitting very narrow pulses or, equivalently, a wideband signal and cross-correlating the received signal with a delayed version of itself. Typically, the measured function R,(t) may appear as shown in Figure 13.1-2. The range of values of T over which Rc(τ;t) is essentially nonzero is called the multipath spread of the channel and is denoted by T_m.

A completely analogous characterization of the time-variant multipath channel begins in the frequency domain. By taking the Fourier transform of c(T; t), we obtain the time-variant transfer function C(f;t), where f is the frequency variable. Thus,

$$C(f; t) = \int_{-\infty}^{\infty} c(\tau; t)e^{-j2\pi f\tau}\, d\tau$$

If c(T; t) is modeled as a complex-valued zero-mean Gaussian random process in the variable, it follows that C (f ; t) also has the same statistics. Under the assumption that the channel is wide-sense-stationary, we define the autocorrelation function

$$R_C(f_2, f_1; \Delta t) = E\left[C^*(f_1; t)C(f_2; t + \Delta t)\right]$$

3.2 Statistical models for fading channels.

There are several probability distributions that can be considered in attempting to model the statistical characteristics of the fading channel. When there are a large number of scatters in the channel that contribute to the signal at the receiver, as is the case in

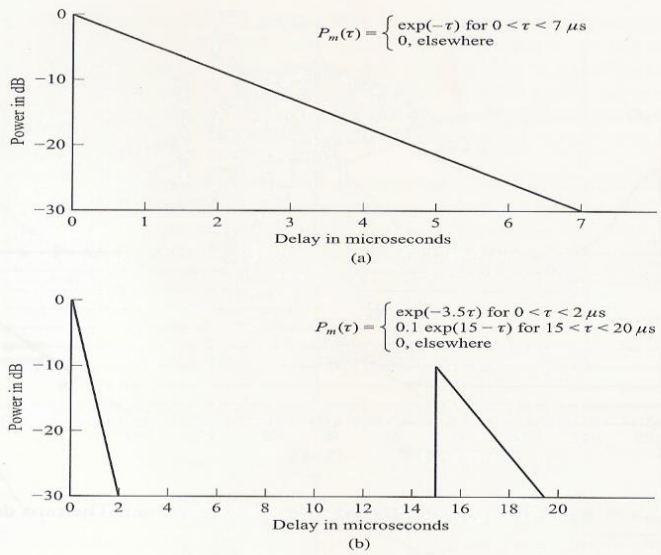


FIGURE 13.1–7
Cost 207 average power delay profiles: (a) typical delay profile for suburban and urban areas; (b) typical "bad"-case delay profile for hilly terrain. [*From Cost 207 Document 207 TD (86)51 rev 3.*]

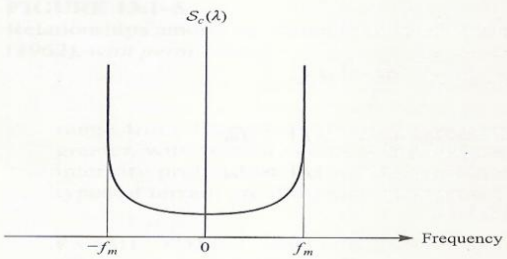


FIGURE 13.1–8
Model of Doppler spectrum for a mobile radio channel.

Ionospheric or tropospheric signal propagation, application of the central limit theorem leads to a Guassian process model for the channel impulse response. If the process is zero-mean, then the envelop of the channel response at any time instance has a Rayleigh probability distribution and the phase is uniformly distributed in the internal (0.2π).

That is

$$p_R(r) = \frac{2r}{\Omega} e^{-r^2/\Omega}, \qquad r \geq 0$$

where

$$\Omega = E(R^2)$$

We observe that the Rayleigh distribution is characterized by the single parameter E(R2).

An alternative statistical model for the envelope of the channel response is the Nakagami-m distribution given by the PDF in Equation 2.3-67. In contrast to the Rayleigh distribution, which has a single parameter that can be used to match the fad ing channel statistics, the Nakagami-m is a two-parameter distribution, involving the parameter m and the second moment SZ = E(R2). As a consequence, this distribution provides more flexibility and accuracy in matching the observed signal statistics. The Nakagami-m distribution can be used to model fading channel conditions that are either more or less severe than the Rayleigh distribution, and it includes the Rayleigh distribution as a special case (m = 1). For example, Turin et al. (1972) and Suzuki (1977) have shown that the Nakagami-m distribution provides the best fit for data signals received in urban radio multipath channels.

The Rice distribution is also a two-parameter distribution. It may be expressed by the PDF given in Equation 2.3-56, where the parameters are s and a2, where s2 is called the noncentrality parameter in the equivalent chi-square distribution. It represents the power in the nonfading signal components, sometimes called specular components, of the received signal.

There are many radio channels in which fading is encountered that are basically lineof-sight (LOS) communication links with multipath components arising from secondary reflections, or signal paths, from surrounding terrain. In such channels, the number of multipath components is small, and, hence, the channel may be modeled in a somewhat simpler form. We cite two channel models as examples.

As the first example, let us consider an airplane to ground communication link in which there is the direct path and a single multipath component at a delay to relative to the direct path. The impulse response of such a channel may be modeled as

$$c(\tau; t) = \alpha\delta(\tau) + \beta(t)\delta[\tau - \tau_0(t)]$$

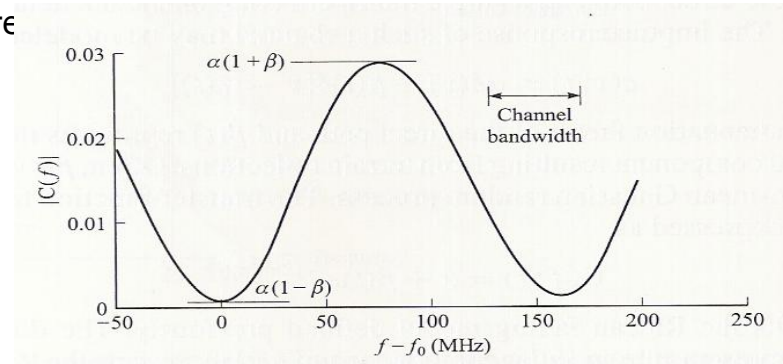
where a is the attenuation factor of the direct path and ,B(t) represents the time-variant multipath signal component resulting from terrain reflections. Often, ,B (t) can

be characterized as a zero-mean Gaussian random process. The transfer function for this channel model may be expressed as

This channel fits the Ricean fading model defined previously. The direct path with attenuation a represents the specular component and P (t) represents the Rayleigh fading component.

3.3 Propagation models for mobile radio channels.

In the link budget calculations that were described in section 4.10-2, we had characterized the path loss of radio waves propagating through free space as being inversely proportional to d^2, where d is the distance between the transmitter and the re



channel, propagation is generally neither free space nor line of sight. The mean path loss encountered in mobile radio channels may be characterized as being inversely proportional to dp, where 2 _< p < 4, with d4 being a worst-case model. Consequently, the path loss is usually much more severe compared to that of free space.

There are a number of factors affecting the path loss in mobile radio communications. Among these factors are base station antenna height, mobile antenna height, operating frequency, atmospheric conditions, and presence or absence of buildings and trees. Various mean path loss models have been developed that incorporate such factors. For example, a model for a large city in an urban area is the Hata model, in which the mean path loss is expressed as

$$\text{Loss in dB} = 69.55 + 26.16 \log_{10} f - 13.82 \log_{10} h_t - a(h_r) + (44.9 - 6.55 \log_{10} h_t) \log_{10} d$$

where f is the operating frequency in MHz (150 < f < 1500), ht is the transmitter antenna height in meters (30 < ht < 200), h, is the receiver antenna height in meters (1 < h, < 10), d is the distance between transmitter and receiver in km (1 < d < 20), and

$$a(h_r) = 3.2(\log_{10} 11.75h_r)^2 - 4.97, \qquad f \geq 400\,\text{MHz}$$

Another problem with mobile radio propagation is the effect of shadowing of the signal due to large obstructions, such as large buildings, trees, and hilly terrain between the transmitter and the receiver. Shadowing is usually modeled as a multiplicative and, generally, slowly time varying random process. That is, the received signal may be characterized mathematically as

$$r(t) = A_0 g(t) s(t)$$

where AO represents the mean path loss, s(t) is the transmitted signal, and g(t) is a random process that represents the shadowing effect. At any time instant, the shadowing process is modeled statistically as lognormally distributed. The probability density function for the lognormal distribution is

$$p(g) = \begin{cases} \dfrac{1}{\sqrt{2\pi\sigma^2}\, g} e^{-(\ln g - \mu)^2/2\sigma^2} & (g \geq 0) \\ 0 & (g < 0) \end{cases}$$

If we define a new random variable X as $X = \ln g$, then

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty$$

The random variable X represents the path loss measured in dB, /t is the mean path loss in dB, and o is the standard deviation of the path loss in dB. For typical cellular and microcellular environments, 6 is in the range of 5-12 dB.

4.0 Conclusion

We began the treatment of digital significantly over fading multipath channel by first developing a statistical characterization of the channel. Besides, evaluated the performance of several basic digital signaling techniques for communication over such channels.

5.0 Summary

One characteristics of a multipath medium is the time spread introduced in the signal that is transmitted through the channel. Besides, we examined the effects of the channel and transmitted signal.

6.0 Tutor Marked Assignment

Explain the meaning of the following;

i. The channel is frequency-non selective

ii. The channel is slowly fading.

iii. The channel is frequency selective

7.0 References/ Further Reading

A General Treatment of Wireless Communication by Rapport (1996) and Stuber (2000).

Unit 2: The Effect of Signal Characteristics on the Choice of a Channel Model

1.0 Introduction

2.0 Objectives

3.0 Main Contents

3.1 Effect of Signal Characteristics on the Selection of a Channel Model

3.2 Frequency Non-Selective, Slowly Fading Channel

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/Further Reading

1.0 Introduction

In this unit, we shall consider the effect of signal characteristics on the selection of a channel model that is appropriate for the specified signal. The effect of the channel on the transmitted signal is a function of our choice of signal bandwidth and signal duration.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain the effect of signal characteristic on the selection of a channel model.
- Explain frequency non-selective, slowly fading channel.

3.1 Effects of signal characteristics on the selection of a channel model.

Having discussed the statistical characterization of time-variant multipath channels generally in terms of the correlation functions describe in Section 13.1, we now consider the effect of signal characteristics on the selection of a channel model that is appropriate for the specified signal. Thus, let s, (t) be the equivalent lowpass signal transmitted over the channel and let Sl(f) denote its frequency content. Then the equivalent lowpass received signal, exclusive of additive noise, may be expressed either in terms of the time-domain variables c(τ;t) and sl(t) as

$$r_l(t) = \int_{-\infty}^{\infty} c(\tau; t) s_l(t-\tau)\, d\tau$$

or in terms of the frequency functions $C(f; t)$ and $S_l(f)$ as

$$r_l(t) = \int_{-\infty}^{\infty} C(f; t) S_l(f) e^{j2\pi f t}\, df$$

Suppose we are transmitting digital information over the channel by modulating (either in amplitude, or in phase, or both) the basic pulse sl (t) at a rate 1/T, where T is the signaling interval. It is apparent from Equation 13.2-2 that the time-variant channel characterized by the transfer function C(τ;t) distorts the signal SI(f). If SI (f) has a bandwidth W greater than the coherence bandwidth $(\Delta f)_c$ of the channel, SI(f) is subjected to different gains and phase shifts across the band. In such a case, the channel is said to be frequency-selective. Additional distortion is caused by the time

variations in C(f ; t). This type of distortion is evidenced as a variation in the received signal strength, and has been termed fading. It should be emphasized that the frequency selectivity and fading are viewed as two different types of distortion. The former depends on the multipath spread or, equivalently, on the coherence bandwidth of the channel relative to the transmitted signal bandwidth W. The latter depends on the time variations of the channel, which are grossly characterized by the coherence time (At), or, equivalently, by the Doppler spread Bd.

The effect of the channel on the transmitted signal s, (t) is a function of our choice of signal bandwidth and signal duration. For example, if we select the signaling interval T to satisfy the condition T >> Tm, the channel introduces a negligible amount of intersymbol interference. If the bandwidth of the signal pulse s, (t) is W ~ 1/T, the condition T >> T_m implies that

$$W \ll \frac{1}{T_m} \approx (\Delta f)_c$$

That is, the signal bandwidth W is much smaller than the coherence bandwidth of the channel. Hence, the channel is frequency-nonselective. In other words, all the frequency components in SI(f) undergo the same attenuation and phase shift in transmission through the channel. But this implies that, within the bandwidth occupied by $S_l(f)$, the time-variant transfer function C(f ; t) of the channel is a complex-valued constant in the frequency variable. Since Sl (f) has its frequency content concentrated in the vicinity of f = 0, C(f ; t) = C(0; t). Consequently, Equation 13.2-2 reduces to

$$\begin{aligned} r_l(t) &= C(0;t)\int_{-\infty}^{\infty} S_l(f)e^{j2\pi f t}\,df \\ &= C(0;t)s_l(t) \end{aligned}$$

Thus, when the signal bandwidth W is much smaller than the coherence bandwidth (Af)c of the channel, the received signal is simply the transmitted signal multiplied by a complex-valued random process C(0; t), which represents the time-variant characteristics of the channel. In this case, we say that the multipath components in the received are not resolvable because W << $(Af)_c$.

$$C(0;t) = \alpha(t)e^{j\phi(t)}$$

The transfer function C(0; t) for a frequency-nonselective channel may be expressed in the form

where a(t) represents the envelope and q5(t) represents the phase of the equivalent lowpass channel. When C(0; t) is modeled as a zero-mean complex-valued Gaussian random process, the envelope a (t) is Rayleigh-distributed for any fixed value of t and 0(t) is uniformly distributed over the interval (-.7,,7). The rapidity of the fading on the frequency-nonselective channel is determined either from the correlation function RC (At) or from the Doppler power spectrum Sc(;,). Alternatively, either of the channel parameters (At), or Bd can be used to characterize the rapidity of the fading.

For example, suppose it is possible to select the signal bandwidth W to satisfy the condition W << (Af), and the signaling interval T to satisfy the condition T << (At)c. Since T is smaller than the coherence time of the channel, the channel attenuation and phase shift are essentially fixed for the duration of at least one signaling interval. When this condition holds, we call the channel a slowly fading channel. Furthermore, when W ≈ 1 / T, the conditions that the channel be frequency-nonselective and slowly fading imply that the product of T_m and B_d must satisfy the condition T_mB_d < l.

The product T_mB_d is called the spread factor of the channel. If T_mB_d < 1, the channel is said to be underspread; otherwise, it is overspread. The multipath spread, the Doppler spread, and the spread factor are listed in Table 13.2-1 for several channels.

3.2 Frequency nonselective, slowly fading channel

In this section, we derive the error rate performance of binary PSK and binary FSK when these signals are transmitted over a frequency-nonselective, slowly fading channel. As described in Section 13.2, the frequency-nonselective channel results in multiplicative distortion of the transmitted signal sl(t). Furthermore, the condition that the channel fades slowly implies that the multiplicative process may be regarded as a constant during at least one signaling interval. Consequently, if the transmitted signal is sl(t), the received equivalent lowpass signal in one signaling interval is

$$r_l(t) = \alpha e^{j\phi} s_l(t) + z(t), \qquad 0 \le t \le T$$

where z(t) represents the complex-valued white Gaussian noise process corrupting the signal.

Let us assume that the channel fading is sufficiently slow that the phase shift q5 can be estimated from the received signal without error. In that case, we can achieve ideal coherent detection of the received signal. Thus, the received signal can be processed by passing it through a matched filter in the case of binary PSK or through a pair of matched filters in the case of binary FSK. One method that we can use to determine the performance of the binary communication systems is to evaluate the decision variables and from these determine the probability of error. However, we have already done this for a fixed (time-invariant) channel. That is, for a fixed attenuation a, we know the probability of error for binary PSK and binary FSK. From Equation 4.3-13, the

4.0 Conclusion

We have considered a number of number subtopic concerned with digital communications over a fading multipath channel. We began with a statistical characterization of the channel and then described the ramifications of the channel characteristics on the design of digital signals and on their performance.

5.0 Summary

The treatment of digital communication over fading channels focused primarily on the Rayleigh fading channel model. Although, other statistical models, such as the Ricean fading model of the Nakagami fading models may be more appropriate for characterizing fading on some real channels.

6.0 Tutor Marked Assignment

1. Consider a binary communication system for transmitting a binary sequence over a fading channel. The modulation is orthogonal FSK with third-order frequency diversity (L=3). The demodulator consists of matched fitters followed by square-law detectors. Assume that the FSK carrierss fade independently and identically according to a Rayleigh envelop distribution. The additive noises on the diversity signals are

zero-mean Gaussian with autocorrelation functions E [zk (t)zk (t + [ ) = 2NS(T). the noise processes are mutually statistically independent.

a. Evaluate Pbh for = 100 and 1000.

b. Evaluate the error rate lbs 'for Yc = 100 and 1000 if the decoder employs soft decision coding.

7.0 References/ Further Reading

A General treatment of Winder Communications by Rappaport (1996) and Stuber (2000).

UNIT 3: IVERSITY TECHNIQUES FOR FADING MULTIPATH CHANNELS

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Diversity Techniques

3.2 Binary Signals

3.3 Multiphase Signals

3.4 M-ary Orthogonal Signals

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/ Further Reading

1.0 Introduction

In this unit, we will consider errors that occur in reception when the channel attenuation is large. Besides, determine the error rate performance for a binary digital communication system with diversity.

2.0 Objectives

At the end of this unit, you should be able to;

- Discuss diversity techniques.
- Explain binary signals and multiphase signals

3.1 Diversity Techniques for Fading Multipath Channels

Diversity techniques are based on the notion that errors occur in reception when the channel attenuation is large, i.e., when the channel is in a deep fade. If we can supply to the receiver several replicas of the same information signal transmitted over independently fading channels, the probability that all the signal components will fade simultaneously is reduced considerably. That is, if p is the probability that any one signal will fade below some critical value, then p' is the probability that all L independently fading replicas of the same signal,will fade below the critical value. There are several ways in which we can provide the receiver with L independently fading replicas of the same information-bearing signal.

One method is to employ frequency diversity. That is, the same information-bearing signal is transmitted on L carriers, where the separation between successive carriers equals or exceeds the coherence bandwidth (Af), of the channel.

A second method for achieving L independently fading versions of the same information-bearing signal is to transmit the signal in L different time slots, where the separation between successive time slots equals or exceeds the coherence time (At), of the channel. This method is called time diversity.

Note that the fading channel fits the model of a bursty error channel. Furthermore, we may view the transmission of the same information either at different frequencies

or in difference time slots (or both) as a simple form of repetition coding. The separation of the diversity transmissions in time by (At), or in frequency by (Af), is basically a form of block-interleaving the bits in the repetition code in an attempt to break up the error bursts and, thus, to obtain independent errors. Later in the chapter, we shall demonstrate that, in general, repetition coding is wasteful of bandwidth when compared with nontrivial coding.

Another commonly used method for achieving diversity employs multiple antennas. For example, we may employ a single transmitting antenna and multiple receiving antennas. The latter must be spaced sufficiently far apart that the multipath components in the signal have significantly different propagation delays at the antennas. Usually a separation of a few wavelengths is required between two antennas in order to obtain signals that fade independently.

A more sophisticated method for obtaining diversity is based on the use of a signal having a bandwidth much greater than the coherence bandwidth (Af), of the channel. Such a signal with bandwidth W will resolve the multipath components and, thus, provide the receiver with several independently fading signal paths. The time resolution is 11W. Consequently, with a multipath spread of T,, seconds, there are TmW resolvable signal components. Since T„, = 1/(Af),, the number of resolvable signal components may also be expressed as W/(Af),. Thus, the use of a wideband signal may be viewed as just another method for obtaining frequency diversity of order L ;:~, W/(Af),. The optimum demodulator for processing the wideband signal will be derived in Section 13.5. It is called a RAKE correlator or a RAKE matched filter and was invented by Price and Green (1958).

There are other diversity techniques that have received some consideration in practice, such as angle-of-arrival diversity and polarization diversity. However, these have not been as widely used as those described above.

3.2 Binary Signals

We shall now determine the error rate performance for a binary digital communication system with diversity. We begin by describing the mathematical model for the communication system with diversity. First of all, we assume that there are L diversity channels, carrying the same information-bearing signal. Each

channel is assumed to be frequency-nonselective and slowly fading with Rayleigh-distributed envelope statistics. The fading processes among the L diversity channels are assumed to be mutually statistically independent. The signal in each channel is corrupted by an additive zero-mean white Gaussian noise process. The noise processes in the L channels are assumed to be mutually statistically independent, with identical autocorrelation functions. Thus, the equivalent low-pass received signals for the L channels can be expressed in the form

$r_{/k}(t) = \alpha k e^{j\phi k} s_{km}(t) + zk(t)$, k = 1, 2, …… L, m = 1, 2

where {akei0k} represent the attenuation factors and phase shifts for the L channels, skm(t) denotes the mth signal transmitted on the kth channel, and zk(t) denotes the additive white Gaussian noise on the kth channel. All signals in the set {skm(t)} have the same energy.

The optimum demodulator for the signal received from the kth channel consists of two matched filters, one having the impulse response

$$b_{k1}(t) = s_{k1}^*(T - t)$$

and the other having the impulse response

$$b_{k2}(t) = s_{k2}^*(T - t)$$

Of course, if binary PSK is the modulation method used to transmit the information, then sk1(t) = -sk2(t). Consequently, only a single matched filter is required for binary PSK. Following the matched filters is a combiner that forms the two decision variables. The combiner that achieves the best performance is one in which each matched filter output is multiplied by the corresponding complex-valued (conjugate) channel gain ake-Iok. The effect of this multiplication is to compensate for the phase shift in the channel and to weight the signal by a factor that is proportional to the signal strength. Thus, a strong signal carries a larger weight than a weak signal. After the complex-valued weighting operation is performed, two sums are formed. One consists of the real parts of the weighted outputs from the matched filters corresponding to a transmitted 0. The second consists of the real part of the outputs from the matched filters corresponding to a transmitted 1. This optimum combiner is called a maximal ratio combiner by Brennan (1959). Of course, the realization of this optimum combiner is based on the assumption that the channel attenuations {ak}

and the phase shifts {q5kl are known perfectly. That is, the estimates of the parameters {ak} and l0kl contain no noise. (The effect of noisy estimates on the error rate performance of multiphase PSK is considered in Appendix C.)

A block diagram illustrating the model for the binary digital communication system described above is shown in Figure 13.4-1.

Let us first consider the performance of binary PSK with Lth-order diversity. The output of the maximal ratio combiner can be expressed as a single decision variable in the form

$$U = \mathrm{Re}\left(2\mathcal{E}\sum_{k=1}^{L}\alpha_k^2 + \sum_{k=1}^{L}\alpha_k N_k\right)$$
$$= 2\mathcal{E}\sum_{k=1}^{L}\alpha_k^2 + \sum_{k=1}^{L}\alpha_k N_{kr} \qquad (1$$

where N_{kr} denotes the real part of the complex-valued Gaussian noise variable

$$N_k = e^{-j\phi_k}\int_0^T z_k(t)s_k^*(t)\,dt \qquad (1$$

4.0 Conclusion

In this unit, we have considered diversity techniques as a troll that enhances the reliability of communication system. Besides, we examine the error rate performance for a binary digital communication that the performance of M-ary Orthogonal signals transmitted over a Rayleigh fading channel.

5.0 Summary

We have examine or seen the general approach in two design of reliable communication presented in this unit. Furthermore, have presented a unified approach to evaluating the error rate performance of digital communication systems for various fading channel models.

6.0 Tutor Marked Assignment

1. Suppose that we have a frequency allocation (bandwidth) of 10 KHz and we wish to transmit at a rate of 100 bits over this channel design a binary communication system with frequency diversity. In particular, specify;

i. The type of modulation

ii. The number of sub-channels.

iii. The frequency separation between adjacent carriers.

iv. The signaling interval used in your design. Justify your choice of parameters.

7.0 References/ Further Reading

Diversity transmission and diversity combing techniques of different channel conditions by Lindsey (1964) and Pierie and Stein (1960).

Unit 4: Signaling over a Frequency Selective, Slowly Fading Channel: The Rake Demodulator

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 A tapped-Delay-Line Channel Model

3.2 The Rake Demodulator

3.3 Performance of Rake Demodulator

3.4 Generalized Rake Demodulator

3.5 Receiver Structures for Channels with Inter-symbol Interference

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 Reference/ Further Reading

1.0 Introduction

When the spread factor of the channel satisfies the condition TmBd << 1, it is possible to select signals having a bandwidth W <<(Δf)c and a signal duration T<<(At)c. Thus, the channel is frequency nonselective and slowly fading. In such a channel, diversity techniques can be employed to overcome the severe consequence of fading.

2.0 At the end of this unit, you should be able to;

- Illustrate a tapped-delay-line channel model.
- Explain the Rake Demodulator
- Discuss the performances of Rake Demodulator

3.1 A Tapped-Delay-Line Channel Model

As we shall now demonstrate, a more direct method for achieving basically the same results is to employ a wideband signal covering the bandwidth W. The channel is still assumed to be slowly fading by virtue of the assumption that T << $(\Delta t)_c$. Now suppose that W is the bandwidth occupied by the band-pass signal. Then the band occupancy of the equivalent low-pass signal sl(t) is /f/ ≤ ½W. Since sl(t) is band-limited to /f/≤ ½W, application of the sampling theorem results in the ignal representation

$$s_l(t) = \sum_{n=-\infty}^{\infty} s_l\left(\frac{n}{W}\right) \frac{\sin[\pi W(t - n/W)]}{\pi W(t - n/W)} \qquad (13.5\text{–}1)$$

The Fourier transform of $s_l(t)$ is

$$S_l(f) = \begin{cases} \dfrac{1}{W} \displaystyle\sum_{n=-\infty}^{\infty} s_l(n/W)e^{-j2\pi fn/W} & |f| \le \tfrac{1}{2}W \\ 0 & |f| > \tfrac{1}{2}W \end{cases} \qquad (13.5\text{–}2)$$

The noiseless received signal from a frequency-selective channel was previously expressed in the form

$$r_l(t) = \int_{-\infty}^{\infty} C(f;t)S_l(f)e^{j2\pi ft}\,df \qquad (13.5\text{–}3)$$

Where C (f;t) is the time-variant transfer function. Substitution for Sl(f) from equation 13.5-2 into 13.5-3 yields.

$$r_l(t) = \frac{1}{W}\sum_{n=-\infty}^{\infty} s_l(n/W)\int_{-\infty}^{\infty} C(f;t)e^{j2\pi f(t-n/W)}\,df$$
$$= \frac{1}{W}\sum_{n=-\infty}^{\infty} s_l(n/W)c(t-n/W;t)$$

where c(τ;t) is the time-variant impulse response. We observe that equation 13.5-4 has the form of a convolution sum. Hence, it can also be expressed in the alternative form

$$r_l(t) = \frac{1}{W}\sum_{n=-\infty}^{\infty} s_l(t-n/W)c(n/W;t)$$

It is convenient to define a set of time-variable channel coefficients as

$$c_n(t) = \frac{1}{W}c\left(\frac{n}{W};t\right)$$

then equation 13.5-5 expressed in terms of these channel coefficients becomes

$$r_l(t) = \sum_{n=-\infty}^{\infty} c_n(t)s_l(t-n/W)$$

the form for the received signal in equation 13.5-7 implies that the time-variant frequency-selective channel can be modeled or represented as a tapped delay line with tap spacing 1/w and tap weight coefficients {cn(t)}. In fact, we deduce from equation 13.5-7 that the low-pass impulse response for the channel is

$$c(\tau;t) = \sum_{n=-\infty}^{\infty} c_n(t)\delta(\tau-n/W)$$

and the corresponding time-variant transfer function is

$$C(f;t) = \sum_{n=-\infty}^{\infty} c_n(t)e^{-j2\pi fn/W}$$

thus, with an equivalent low-pas-signal having a bandwidth ½ W, where W >> $(\Delta f)_c$, we achieve a resolution of 1/W in the multipath delay profile. Since the total multipath spread is Tm, for all practical purposes the tapped delay line model for the channel can be truncated at L= [T_mW] + taps. Then, the noiseless received signal can be expressed in the form

3.2 The Rake Demodulator

We now consider the problem of digital signaling over a frequency-selective channel that is modeled by a tapped delay line with statistically independent time-variant tap weights {c,(t)}. It is apparent at the outset, however, that the tapped delay line model with statistically independent tap weights provides us with L replicas of the same transmitted signal at the receiver. Hence, a receiver that processes the received signal in an optimum manner will achieve the performance of an equivalent Lth-order diversity communication system.

Let us consider binary signaling over the channel. We have two equal-energy signals sll (t) and sl2(t), which are either antipodal or orthogonal. Their time duration T is selected to satisfy the condition T >> T, Thus, we may neglect any intersymbol interference due to multipath. Since the bandwidth of the signal exceeds the coherent bandwidth of the channel, the received signal is expressed as

$$\begin{aligned} r_l(t) &= \sum_{k=1}^{L} c_k(t) s_{li}(t - k/W) + z(t) \\ &= v_i(t) + z(t), \qquad 0 \le t \le T, \qquad i = 1, 2 \end{aligned}$$

where z(t) is a complex-valued zero-mean white Gaussian noise process. Assume for the moment that the channel tap weights are known. Then the optimum demodulator consists of two filters matched to vl (t) and V2 (t). The demodulator output is sampled at the symbol rate and the samples are passed to a decision circuit that selects the signal corresponding to the largest output. An equivalent optimum demodulator employs cross correlation instead of matched filtering. In either case, the decision variables for coherent detection of the binary signals can be expressed as

$$\begin{aligned} U_m &= \mathrm{Re}\left[\int_0^T r_l(t) v_m^*(t)\, dt\right] \\ &= \mathrm{Re}\left[\sum_{k=1}^{L} \int_0^T r_l(t) c_k^*(t) s_m^*(t - k/W)\, dt\right], \qquad m = 1, 2 \end{aligned}$$

Figure 13.5-2 illustrates the operations involved in the computation of the decision variables. In this realization of the optimum receiver, the two reference signals are delayed and correlated with the received signal rl(t).

An alternative realization of the optimum demodulator employs a single delay line through which is passed the received signal rl(t). The signal at each tap is correlated with ck(t)sim(t), where k = l, 2, . . ., L and m = 1, 2. This receiver structure is shown in Figure 13.5-3. In effect, the tapped delay line demodulator attempts to collect the signal energy from all the received signal paths that fall within the span of the delay line and carry the same information. Its action is somewhat analogous to an ordinary garden rake and, consequently, the name "RAKE demodulator" has been coined for this demodulator structure by Price and Green (1958). The taps on the RAKE demodulator are often called "RAKE fingers."

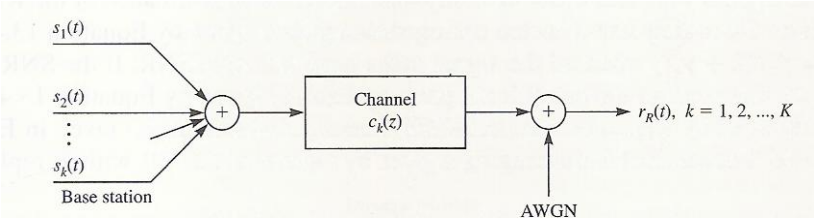
3.3 Generalized Rake Demodulator

The RAKE demodulator described above is the optimum demodulator when the additive noise is white and Gaussian. However, there are communication scenarios in which additive interference from other users of the channel results in colored additive noise. This is the case, for example, in the downlink of a cellular communication system employing CDMA as a multiple access method. In this case, the spread spectrum signals transmitted from a base station to the mobile receivers carry information on synchronously transmitted orthogonal spreading codes. However, in transmission over a frequency-selective channel, the orthogonality of the code sequences is destroyed by the channel time dispersion due to multipath. As a consequence, the RAKE demodulator for any given mobile receiver must demodulate its desired signal in the presence of additional additive interference resulting from the cross-correlations of its desired spreading code sequence with the multipath corrupted code sequences that are assigned to the other mobile users. This additional interference is generally characterized as colored Gaussian noise, as shown by Bottomley,(1993) and Klein (1997).

A model for the downlink transmission in a CDMA cellular communication system is illustrated in Figure 13.5-8. The base station transmits the combined signal.

$$s(t) = \sum_{k=1}^{K} s_k(t)$$

to the K mobile terminals, where each sk(t) is a spread spectrum signal intended for the kth user and the corresponding spreading code for the kth user is orthogonal with each of the spreading codes of the other K - 1 users. We assume that the signals propagate through a channel characterized by the baseband equivalent lowpass, time-invariant

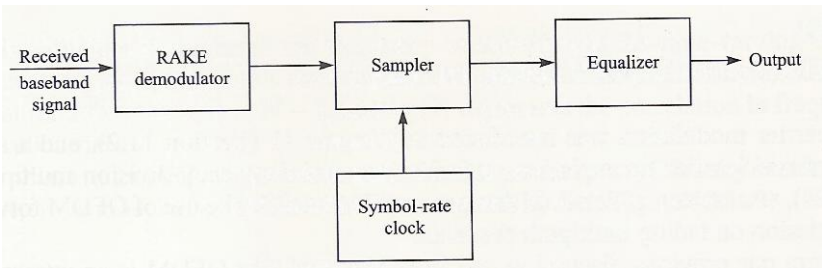


3.4 Receiver Structure for Channels with Intersymbol Interference

As described above, the wideband signal waveforms that are transmitted through the multipath channels resolve the multipath components with a time resolution of 1/ W, where W is the signal bandwidth. Usually, such wideband signals are generated as direct sequence spread spectrum signals, in which the PN spreading sequences are the outputs of linear feedback shift registers, e.g., maximum-length linear feedback shift registers. The modulation impressed on the sequences may be binary PSK, QPSK, DPSK, or binary orthogonal. The desired bit rate determines the bit interval or symbol interval.

The RAKE demodulator that we described above is the optimum demodulator based on the condition that the bit interval Tb >> T„ Z, i.e., there is negligible ISI. When this condition is not satisfied, the RAKE demodulator output is corrupted by ISI. In such a case, an equalizer is required to suppress the ISI.

To be specific, we assume that binary PSK modulation is used and spread by a PN sequence. The bandwidth of the transmitted signal is sufficiently broad to resolve two or more multipath components. At the receiver, after the signal is demodulated to baseband, it may be processed by the RAKE, which is the matched filter to the channel response, followed by an equalizer to suppress the ISI. The RAKE output is sampled at the bit rate, and these samples are passed to the equalizer. An appropriate equalizer, in this case, would be a maximum-likelihood sequence estimator implemented by use

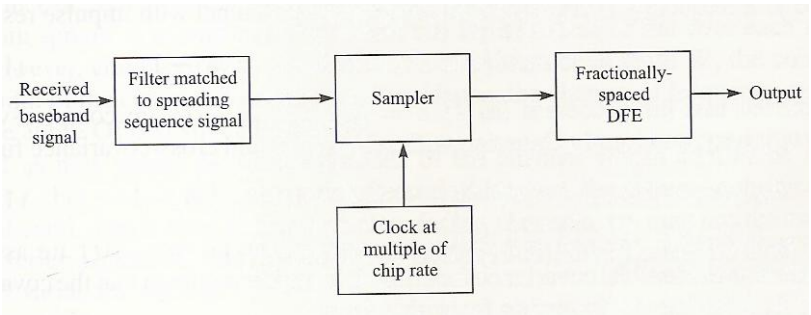


of the Viterbi algorithm or a decision feedback equalizer (DFE). This demodulator structure is shown in Figure 13.5-10.

Other receiver structures are also possible. If the period of the PN sequence is equal to the bit interval, i.e., LT, = Tb, where T, is the chip interval and L is the number of chips per bit, a fixed filter matched to the spreading sequence may be used to process the received signal and followed by an adaptive equalizer, such as a fractionally spaced DFE, as shown in Figure 13.5-11. In this case, the matched filter output is sampled at some multiple of the chip rate, e.g., twice the chip rate, and fed to the fractionally spaced DFE. The feedback filter in the DFE would have taps spaced at the bit interval. The adaptive DFE would require a training sequence for adjustment of its coefficients to the channel multipath structure.

An even simpler receiver structure is one in which the spread spectrum matched filter is replaced by a low-pass filter whose bandwidth is matched to the transmitted signal bandwidth. The output of such a filter may be sampled at an integer multiple of the chip rate and the samples are passed to an adaptive fractionally spaced DFE. In this case, the coefficients of the feedback filter in the DFE, with the aid of a training sequence, will adapt to the combination of the spreading sequence and the channel multipath. Abdulrahman et al. (1994) consider the use of a DFE to suppress ISI in a CDMA system in which each user employs a wideband direct sequence spread spectrum signal.

The paper by Taylor et al. (1998) provides a broad survey of equalization techniques and their performance for wireless channels.



4.0 Conclusion

We considered the signal representation of a Tapped-Delay-Line Channel Model. Besides, have examined the Rake Demodulator and its performance under the conditions that the fading is sufficiency slow to allow us to estimate perfectly (without noise).

5.0 Summary

We have considered the transmission of digital information through timedispersive channels and described the Rake demodulator, which is the matched fitter for the channel.

6.0 Tutor Marked Assignment

A multipath fading channel has a multipath spread of Tm = 1s and a Dappler spread Bd = 0.01 Hz. The total channel bandwidth at bandpass available for signal transmission is W = 5 Hz. To reduce the effects of intersymbol interference, the signal designer selects a pulse duration T = 10s.

1. Determine the coherence bandwidth and the coherence time.

2. Is the channel frequency selective? Explain

3. Is the channel fading slowly or rapidly? Explain

7.0 References/ Further Reading

The Effect of ICT in Orthogonal-Division Multiplexing (OFDM) for Mobile Communication by Robertson and Kaiser (1999) and Wang et al (2006).

UNIT 5: MULTICARRIER MODULATION (OFDM)

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Performance Degradation of an OFDM System Due to Perplex Spreading

1.0 Introduction

In this unit, we consider the use of OFDM for digital transmission on fading multipath channels. OFDM is an attraction alternative to single-carrier modulation for use in time-dispersioe channels. we shall also consider the use of OFDM for mobile communications and on the performance of an OFDM system.

2.0 Objectives

At the end of this unit, you should be able to;

- Discuss the performance degradation of an OFDM system due to Doppler spreading.

- Explain suppression of ICI in OFDM systems

3.1 Performance Degradation of an OFDM System due to Doppler Spreading

Let us consider an ®FDM system with N subcarriers {ei2nfkt}, where each subcarrier employs either M-ary QAM or PSK modulation. The subcarriers are orthogonal over the symbol duration T, i.e., fk = k/T, k = 1, 2, . . ., N, so that

$$\frac{1}{T}\int_0^T e^{j2\pi f_i t}\, e^{-j2\pi f_k t}\, dt = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}$$

The channel is modeled as a frequency-selective randomly varying channel with impulse response c(i; t). Within the frequency band of each subcarrier, the channel is modeled as a frequency-nonselective Rayleigh fading channel with impulse response.

$$c_k(\tau; t) = \alpha_k(t)\delta(t), \qquad k = 0, 1, \ldots, N-1$$

It is assumed that the processes {ak(t), k = 0, 1, . . . , N - 1} are complex-valued, jointly stationary, and jointly Gaussian with zero means and cross-covariance function

$$R_{\alpha_k\alpha_i}(\tau) = E[\alpha_k(t+\tau)\alpha_t^*(t)], \qquad k, i = 0, 1, \ldots, N-1$$

For each fixed k, the real and imaginary parts of the process ak(t) are assumed independent with identical covariance function. It is further assumed that the covariance function $R_{\alpha k \alpha i}$ (-c) has the following factorable form

$$R_{\alpha_k\alpha_i}(\tau) = R_1(\tau)R_2(k-i)$$

which is sufficient to represent the frequency selectivity and the time-varying effects of the channel. RI(r) represents the temporal correlation of the process αk(t), which is identical for all k = 0, 1, . . . , N - 1, and $\boldsymbol{R}_2$(k) represents the correlation in frequency across subcarriers.

To obtain numerical results, we assume that the power spectral density corresponding to R1 (r) is modeled as in lakes (1974) and given by (see Figure 13.1-8)

$$\mathcal{S}(f) = \begin{cases} \dfrac{1}{\pi f_m \sqrt{1-(f/f_m)^2}} & |f| \le f_m \\ 0 & \text{otherwise} \end{cases}$$

where F_d is the maximum Doppler frequency. We note that

$$R_1(\tau) = J_0(2\pi f_m \tau)$$

where $\boldsymbol{J}_0(\tau)$ is the zero-order Bessel function of the first kind. To specify the correlation in frequency across the subcarriers, we model the multipath power intensity profile as an exponential of the form

$$R_c(\tau) = \beta e^{-\beta\tau}, \qquad \tau > 0, \quad \beta > 0$$

where β is a parameter that controls the coherence bandwidth of the channel. The Fourier transform of RJr) yields

$$R_C(f) = \frac{\beta}{\beta + j2\pi f}$$

which provides a measure of the correlation of the fading across the subcarriers, as shown in Figure 13.6-1. Hence, R2(k) = Rc (k/ T) is the frequency separation between two adjacent subcarriers. The 3-dB bandwidth of RC (f) maybe defined as the coherence bandwidth of the channel and is easily shown to be $\sqrt{3}\beta/2\pi$.

The channel model described above is suitable for modeling OFDM signal transmission in mobile radio systems, such as cellular systems and radio broadcasting systems. Since the symbol duration T is usually selected to be much larger than the channel multipath spread, it is reasonable to model the signal fading as flat over each subcarrier. However, compared with the entire OFDM system bandwidth W, the coherence bandwidth of the channel is usually smaller. Hence, the channel is frequency-selective over the entire OFDM signal bandwidth.

Let us now model the time variations of the channel within an OFDM symbol interval T. For mobile radio channels of practical interest, the channel coherence time is significantly larger than T. For such slow fading channels, we may use the two-term Taylor series expansion, first introduced by Bello (1963), to represent the time-varying channel variations ak(t) as

$$\alpha_k(t) = \alpha_k(t_0) + \alpha_k'(t_0)(t - t_0), \qquad t_0 = \frac{T}{2}, \quad 0 \le t \le T$$

3.2 Suppression of ICI in OFDM Systems

The distortion caused by ICI in an OFDM system is akin to the distortion caused by ISI in a single-carrier system. Recall that a linear time-domain equalizer based on the minimum mean-square-error (MMSE) criterion is an effective method for suppressing ISI. In a similar manner, we may apply the MMSE criterion to suppress the ICI in the frequency domain. Thus, we begin with the N frequency samples at the output of the discrete Fourier transform (DFT) processor, which we denote by the vector R(m) for the mth frame. Then we form the estimate of the symbol sk(m) as

where bk(m) is the coefficient vector of size N x 1. This vector is selected to minimize the MSE

$$E\left[|s_k(m) - \hat{s}_k(m)|^2\right] = E\left[|s_k(m) - \boldsymbol{b}_k^H(m)\boldsymbol{R}(m)|^2\right]$$

$$\hat{s}_k(m) = \boldsymbol{b}_k^H(m)\boldsymbol{R}(m), \qquad k = 0, 1, \ldots, N-1$$

where the expectation is taken with respect to the signal and noise statistics. By applying the orthogonality principle, the optimum coefficient vector is obtained as

$$\boldsymbol{b}_k(m) = \left[\boldsymbol{G}(m)\boldsymbol{G}^H(m) + \sigma^2 \boldsymbol{I}_N\right]^{-1} \boldsymbol{g}_k(m), \qquad k = 0, 1, \ldots, N-1 \qquad (13.6\text{–}24)$$

where

$$E\left[\boldsymbol{R}(m)\boldsymbol{R}^H(m)\right] = \boldsymbol{G}(m)\boldsymbol{G}^H(m) + \sigma^2 \boldsymbol{I}_N$$

$$E\left[\boldsymbol{R}(m)s_k^H(m)\right] = \boldsymbol{g}_k(m)$$

and G(m) is related to the channel impulse response matrix H(m) through the DFT relation (see Problem 13.16)

$$\boldsymbol{G}(m) = \boldsymbol{W}^H \boldsymbol{H}(m)\boldsymbol{W}$$

where W is the orthonormal (IDFT) transformation matrix. The vector 9k(M) is the kth column of the matrix G(m), and Q.2 is the variance of the additive noise component. It is easily shown that the minimum MSE for the signal on the kth subcarrier may be expressed as

$$E\left[|s_k(m) - \hat{s}_k(m)|^2\right] = 1 - \boldsymbol{g}_k^H(m)(\boldsymbol{G}(m)\boldsymbol{G}^H(m) + \sigma^2 \boldsymbol{I}_N)^{-1}\boldsymbol{g}_k(m)$$

We observe that the optimum weight vectors {bk(m)) require knowledge of the channel impulse response. In practice, the channel response may be estimated by periodically transmitting pilot signals on each of the subcarriers and by employing a decision-directed method when data are transmitted on the N subcarriers. In a slowly fading channel, the coefficient vectors {bk(m)} may also be adjusted recursively by employing either an LMS- or an RLS-type algorithm, as previously described in the context of equalization for suppression of ISI.

4.0 Conclusion

By selecting the symbol duration in an OFDM system to be significantly larger than the channel dispersion, intersymbol interference (ISI) can be rendered negligible and completely eliminated by use of a time guard band, or equivalently, by the use of a cyclic prefix embedded in the OFDM signal. The elimination of ISI due to multipath dispersion without the use of complex equalizers, is a basic motivation for use of OFDM for digital communication in fading multipath channels.

5.0 Summary

We have considered the transmission of digital information through time-dispersive channels and described the Rake demodulator, which is the matched fitter for the channel.

6.0 Tutor Marked Assignment

A multipath fading channel has a multipath spread of Tm = 1s and a Dappler spread Bd = 0.01 Hz. The total channel bandwith at bandpass available for signal transmission is W = 5 Hz. To reduce the effects of intersymbol interference, the signal designer selects a pulse duration T = 10s.

1. Determine the coherence bandwidth and the coherence time.

2. Is the channel frequency selective? Explain

3. Is the channel fading slowly or rapidly? Explain

7.0 References/ Further Reading

The effect of ICT in Orthogonal-Division Multiplexing (OFDM) for Mobile Communication by Robertson and Kaiser (1999) and Wang et al (2006).

Unit5:Multicarrier Modulation (OFDM)

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Performance Degradation of an OFDM System Due to Perplex Spreading

3.2 Suppression of ICI in OFDM System

4.0 Conclusion

5.0 Summary

We considered that the OFDM is especially vulnerable to Doppler spread resulting from time variations in the channel impulse response, as in the case in mobile communication systems. The Doppler spreading destroys the orthogonality of the OFDM subcarriers and results in intercarrier interference (ICI) which can severely degrade the performance of OFDM system. Lastly, we evaluate the effect of a Doppler spread on the performance of OFDM.

6.0 Tutor Marked Assignment

1. The scattering function S(T;0) for a fading multipath channel is nonzero for the range of values $O \leq T \leq 1$ ms and $-0.1\text{Hz} \leq \lambda \leq 0.1\text{Hz}$. Assume that the scattering function is appropriately uniform in the two variables;

   i. The multipath spread of the channel.

   ii. The Doppler spread of the channel.

iv. The coherence time of the channel.

v. The spread factor of the channel.

7.0 References/Further Reading

Diversity Transmission Diversity Community Techniques Under a variety of Channel Conditions by Lindsey (1964).

MODULE 6: FADING CHANNELS CAPACITY AND CODING

Unit 1: Capacity of Fading Channels

Unit 2: Ergodic and Outage Capacity

Unit 3: Coding for and Performance of Code Systems in fading Channel

Unit 4: Trellis-Coded Modulation for Fading Channels

Unit 5: Bit-Interleaved Coded Modulation

Unit 1: Capacity of Fading Channels

1.0 Introduction

This unit focuses on capacity and coding aspects of fading channels and opportunities that are different from the standard additive white Gaussian noise channels. besides, the metrics that determine the performance of coding schemes over fading channels are different from the standard metrics used to compare the performance of difference coding schemes over additive white Gaussian noise channels.

2.0 Objectives

At the end of this unit, you should be able to;

- Explain capacity of fading channels.
- Explain the parameters that affect the capacity of fading channels.

3.1 The capacity of a channel

The capacity of a channel is defined as the supremum of the rates at which reliable communication over the channel is possible. Reliable communication at rate R is possible if there exists a sequence of codes with rate R for which the average error

probability tends to zero as the block length of the code increases. In other words, at any rate less than capacity we can find a code whose error probability is less than any specified,5 > 0. In Chapter 6 we gave a general expression for the capacity of a discrete memoryless channel in the form

$$C = \max_{p(x)} I(X; Y)$$

where the maximum is taken over all channel input probability density functions. For a power-constrained discrete-time AWGN channel, the capacity can be expressed as

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N}\right)$$

where P is the signal power, N is the noise power, and C is the capacity in bits per transmission, or bits per (real) dimension. For a complex-input complex-output channel with circular complex Gaussian noiset with noise variance No, or No/2 per real and imaginary components, the capacity is given by

bits per complex dimen

$$C = W \log \left(1 + \frac{P}{N_0 W}\right)$$

The capacity of an ideal band-limited, power-limited additive white Gaussian waveform channel is given by

$$C = W \log \left(1 + \frac{P}{N_0 W}\right)$$

where W denotes the bandwidth, P denotes the signal power, and No/2 is the noise power spectral density. The capacity C in this case is given in bits per second. For an infinite-bandwidth channel in which the signal-to-noise ratio P/(NOW) tends to zero, the capacity is given in Equation 6.5-44 as

$$C = \frac{1}{\ln 2} \frac{P}{N_0} \approx 1.44 \frac{P}{N_0}$$

The capacity in bits/sec/Hz (or bits per complex dimension) which determines the highest achievable spectral bit rate is given by

$$C = \log (1 + \mathrm{SNR})$$

where SNR denotes the signal-to-noise ratio defined as

$$\mathrm{SNR} = \frac{P}{N_0 W}$$

Note that since W ^Ts , where T,r is the symbol duration, the above expression for SNR can be written as SNR = PT_s/N_0 = E_s/N_0 = where E_s indicates energy per symbol. In an AWGN channel the capacity is achieved by using a Gaussian input probability density function. At low values of SNR we have

$$y(t) = x(t) \star c(t) + n(t)$$

The notion of capacity for a band-limited additive white Gaussian noise channel can be extended to a nonideal channel in which the channel frequency response is denoted by C(f). In this case the channel is described by the input-output relation of the form

$$C \approx \frac{1}{\ln 2} \text{ SNR} \approx 1.44 \text{ SNR}$$

where c(t) denotes the channel impulse response and C(f) = C-,~T[c(t)] is the channel frequency response. The noise is Gaussian with a power spectral density of S,(f). It was shown in Chapter 11 that the capacity of this channel is given by

$$C = \frac{1}{2}\int_{-\infty}^{\infty} \log\left(1 + \frac{P(f)|C(f)|^2}{\mathcal{S}_n(f)}\right) df$$

where $P(f)$, the the input power spectral density, is selected such that

$$P(f) = \left(K - \frac{\mathcal{S}_n(f)}{|C(f)|^2}\right)^+$$

where x^+ is defined by

$$x^+ = \max\{0, x\}$$

and K is selected such that

$$\int_{-\infty}^{\infty} P(f)\, df = P$$

The water-filling interpretation of this result states that the input power should be allocated to different frequencies in such a way that more power is transmitted at those frequencies of which the channel exhibits a higher signal-to-noise ratio and less power is sent at the frequencies with poor signal-to-noise ratio. A graphical interpretation of the water-filling process is shown in Figure 14.1-1.

The water-filling argument can be also applied to communication over parallel channels. If N parallel discrete-time AWGN channels have noise powers Ni, 1 <_ i <_ N, and an overall power constraint of P, then the

total capacity of the parallel channels is given by

$$C = \frac{1}{2}\sum_{i=1}^{N} \log\left(1 + \frac{P_i}{N_i}\right)$$

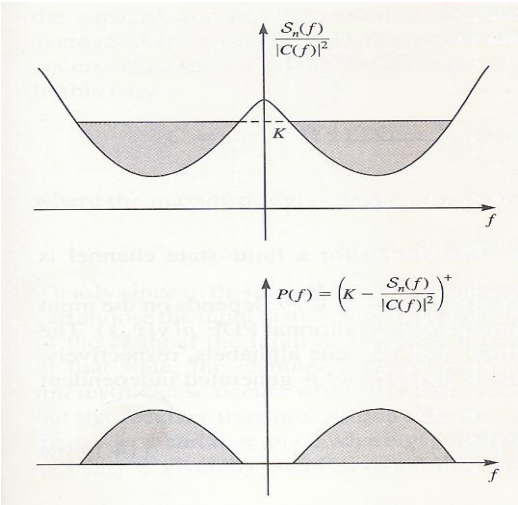
where P_i's are selected such that

$$P_i = (K - N_i)^+$$

subject to

$$\sum_{i=1}^{N} P_i = P$$

In addition to frequency selectivity which can be treated through water-filling arguments, a fading channel is characterized with time variations in channel characteristics,



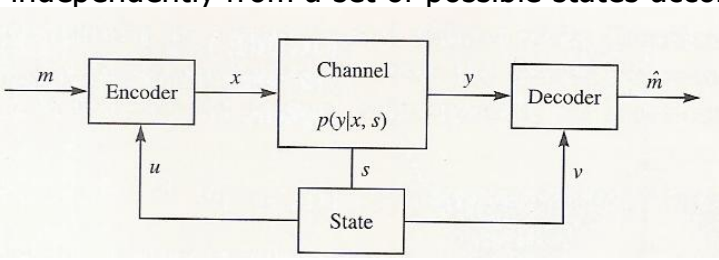
i.e., time selectivity. Since the capacity is defined in the limiting sense as the block length of the code tends to infinity, we can always argue that even in a slowly fading channel the block length can be selected large enough that in any block the channel experiences all possible states, and hence the time averages over one block are equal to the statistical averages. However, from a practical point of view, this would introduce a large delay which is not acceptable in many applications, for instance, speech communication on cellular phones. Therefore, for a delay-constrained system on a slowly fading channel, the ergodicity assumption is not valid.

A common practice to break the inherent memory in fading channels is to employ long interleavers that spread a code sequence across a long period of time, thus making individual symbols experience independent fading. However, employing long interleavers would also introduce unacceptable delay in many applications. These ob-

servations make it clear that the notion of capacity is more subtle in the study of fading channels, and depending on the coherence time of the channel and the maximum delay acceptable in the application under study, different channel models and different notions of channel capacity need to be considered. Since fading channels can be modeled as channels whose state changes, we first study the capacity of these channels.

3.2 Capacity of Finite-State Channels

A finite-state channel is a channel model for a communication environment that varies with time. We assume that in each transmission interval the state of the channel is selected independently from a set of possible states according to some probability



distribution on the space of channel states. The model for a finite-state channel is shown in Figure 14.1-2.

In this channel model, in each transmission the output y E 41 depends on the input x E Xand the state of the channel s E d5" through the conditional PDF p(yIx, s). The sets 'd; c_Y, and HY denote the input, the output, and the state alphabets, respectively, and are assumed to be discrete sets. The state of the channel is generated independent of the channel input according to

$$p(s) = \prod_{i=1}^{n} p(s_i)$$

and the channel is memoryless, i.e,

$$p(y|x, s) = \prod_{i=1}^{n} p(y_i|x_i, s_i)$$

The encoder and the decoder have access to noisy versions of the state denoted by u E W and v E h respectively. Based on an original idea of Shannon (1958), Salehi (1992), and Caire and Shamai (1999) have shown that the capacity of this channel can be given as

$$C = \max_{p(t)} I(T; Y|V)$$

In this expression the maximization is over p(t), the set of all probability mass functions on ~ where c7 denotes the set of all vectors of length 1,~l 1 with components from 0" The cardinality of the set c~is I X I and the set ~ is called the set of input strategies.

In the study of fading channels, certain cases of this channel model are of particular interest. The special case where C = S and V is a degenerate random variable corresponds to the case when complete channel state information (CSI) is available at the receiver and no channel state information is available at the transmitter. In this case the capacitv reduces to

$$C = \max_{p(x)} I(X; Y|S)$$

where

$$p(s, x, y) = p(s)p(x)p(y|x, s)$$

Note that since

$$I(X; Y|S) = \sum_s p(s)I(X; Y|S = s)$$

the capacity can be interpreted as the maximum over all input distributions of the average of the mutual information over all channel states. A second interesting case occurs when the state information is available at both the transmitter and the receiver. In this case

$$C = \max_{p(x|s)} I(X; Y|S) = \sum_s p(s) \max_{p(x|s)} I(X; Y|S = s)$$

where the maximization is on all joint probabilities of the form

$$p(s, x, y) = p(s)p(x|s)p(y|x, s)$$

Clearly since in this case the state information is available at the transmitter, the encoder can choose the input distribution based on the knowledge of the state. Since for each state of the channel the input distribution is selected to maximize the mutual information in that state, the channel capacity is the expected value of the capacities. A third interesting case occurs when complete channel information is available at the receiver but the receiver transmits only a deterministic function of it to the transmitter. In this case v = s and u = g(s), where g(.) denotes a deterministic function. In this case the capacity is given by [see Caire and Shamai (1999)]

$$C = \sum_u p(u) \max_{p(x|u)} I(X; Y|S, U = u)$$

This case corresponds to when the receiver can estimate the channel state but due to communication constraints over the feedback channel can transmit only a quantized version of the state information to the transmitter.

The underlying memoryless assumption in these cases makes these models appropriate for a fully interleaved fading channel.

4.0 Conclusion

Coding system techniques introduce redundancy through transmission of the party check codes, the extra transmissions provide diversity that improves the performance of coded system over fading channels.

5.0 Summary

This unit, we distinguish two different possibilities in dealing with capacity and coding for fading channels. In one case, the characteristics of the channels change fast enough with respect to the transmission duration of a block that a single block of information experiences all possible realizations of the channel frequently. The time average during the transmission duration of a single block are equal to the statistical (ensemble) averages over all possible channel realizations. Another possibility is that the block duration is short and each block experiences only a cross section of channel characteristics. Furthermore, the availability of state information at the receiver that is usually measured by transmitting tones over the channel at different frequencies help the receiver in increasing the channel capacity since the state of the channel can be interpreted as an auxiliary channel output.

6.0 Tutor Marked Assignment

1. Using Equation 14.1-19, determine the capacity of a finite-state channel in which state information is only available at the receiver.

2. Using equation 14.1-19, determine the capacity of a finite-sate channel in which the same state information is available at the transmitter and the receiver.

7.0 References/ Further Reading

Coding for Fading Channels by Biglieri (2005)

Unit 2: Ergodic and Outage Capacity

1.0 Introduction

In this first channel model, since all channel realizations one experienced during a block, an ergodic channel model is appropriate and ergodic capacity can be defined as the ensemble average of channel capacity overall possible channel realization. In the second channel model, where in each block different channel realization are experienced, for each block the capacity will be different. Thus, the capacity can best be modeled as a random variable. This is case, another notion of capacity known as outage capacity is more appropriate.

2.0 Objectives

At the end of this unit, you should be able to;

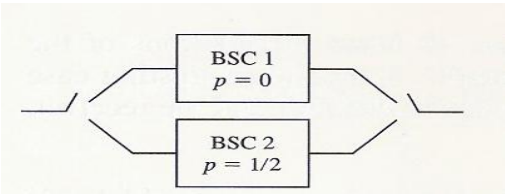
- Explain the Ergodic capacity of channel model and Rayleigh fading model.
- Discuss the outage capacity of Rayleigh fading models and the effect of diversity on outage capacity.

3.1 The Ergodic capacity of channel model

To study the difference between ergodic and outage capacity, consider the two-state channel shown in Figure 14.2-1. In this figure two binary symmetric channels, one with crossover probability p = 0 and one with crossover probability p = 1/2, are shown. We consider two different channel models based on this figure.

1. In channel model 1 the input and output switches choose the top channel (BSC 1) with probability 8 and the bottom channel (BSC 2) with probability 1 - 6, independently for each transmission. In this channel model each symbol is transmitted independently of the previous symbols, and the state of the channel is also selected independently for each symbol.

2. In channel model 2 the top and the bottom channels are selected at the beginning of the transmission with probabilities S and 1 - 6, respectively; but once a channel is selected, it will not change for the entire transmission period.



From Chapter 6 we know that the capacities of the top and bottom channels are Cl = 1 and CZ = 0 bits per transmission, respectively. To find the capacity of the first channel model, we note that since in this case for transmission of each symbol the channel is selected independently over a long block, the channel will experience both BSC component channels according to their corresponding probabilities. In this case time and_ ensemble averages can be interchanged, the notion of ergodic capacity, denoted by C, applies, and the results of the preceding section can be used. The capacity of this channel model depends on the availability of the state information. We distinguish three cases for the first channel model.

1. Case 1: No channel state information is available at the transmitter or receiver. In this case it is easy to verify that the average channel is a binary symmetric channel with crossover probability of 125, and hence the ergodic capacity is

$$\overline{C} = 1 - H_b\left(\frac{1-\delta}{2}\right)$$

2. Case 2: Channel state information available at the receiver. Using Equation 14.1-22, we observe that in this case we maximize the mutual-information with a fixed input distribution. But since regardless of the state of the channel a uniform input distribution maximizes the mutual information, the ergodic capacity of the channel is the average of the two capacities, i.e.,

3. Case 3: Channel state information is available at the transmitter and the receiver. Here we use Equation 14.1-23 to find the channel capacity. In this case we can maximize the mutual information individually for each state, and the capacity is the average of the capacities as $\overline{C} = \delta C_1 + (1 - \delta)C_2 = \delta$

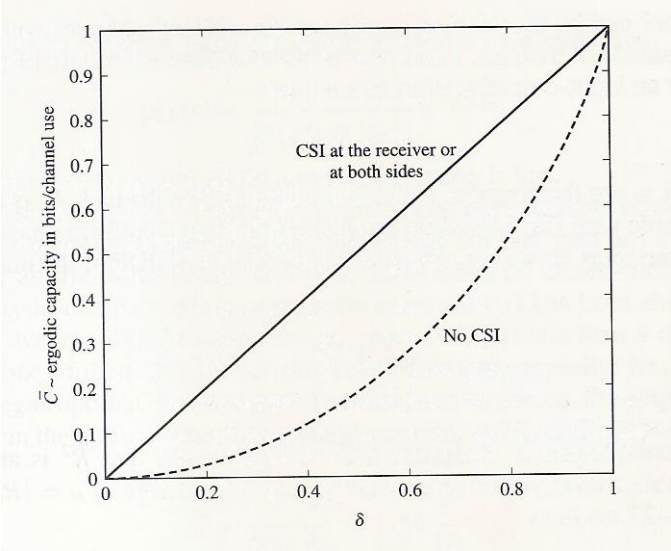
A plot of the two capacities as a function of 8 is given in Figure 14.2-2. Note that in this particular channel since the capacity achieving input distribution for the two channels states is the same, the results of cases 2 and 3 are the same. In general the capacities in these cases are different, as shown in Problem 14.7.

In the second channel model where one of the two channels BSC 1 or BSC 2 is selected only once and then used for the entire communication situation, the capacity in the Shannon sense is zero. In fact it is not possible to communicate reliably over this channel model at any positive rate. The reason is that if we transmit at a rate R > 0 and channel BSC 2 is selected, the error probability cannot be set arbitrarily small. Since channel BSC 2 is selected with a probability of 1 - 8 > 0, reliable communication at any rate R > 0 is impossible. In fact in this case the channel capacity is a binary random variable which takes values of 1 and 0 with probabilities 8 and 1 - 8, respectively. This is a case for which ergodic capacity is not applicable and a new notion of capacity called outage capacity is more appropriate (Ozarow et al. (1994)).

We note that since the channel capacity in this case is a random variable, if we transmit at a rate R > 0, there is a certain probability that the rate exceeds the capacity and the channel will be in outage. The probability of this event is called the outage probability and is given by

$P_{out}(R) = P\,[C < R] = Fc(R-)$

where FC(c) denotes the CDF of the random variable C and FC(R-) is the limit-fromleft of FC(c) at point c = R.



For any 0 <_ E < 1 we can define CE, the E-outage capacity of the channel, as the highest transmission rate that keeps the outage probability under E, i.e.,

$$C_\epsilon = \max\{R : \mathrm{P_{out}}(R) \le \epsilon\}$$

In the channel model 2, the ϵ-outage capacity of the channel is given by

$$C_\epsilon = \begin{cases} 0 & \text{for } 0 \le \epsilon < 1-\delta \\ 1 & \text{for } 1-\delta \le \epsilon < 1 \end{cases}$$

3.2 The Ergodic Capacity of the Rayleigh Fading Channel

In this section we study the ergodic capacity of the Rayleigh fading channel. The underlying assumption is that the channel coherence time and the delay restrictions of the channel are such that perfect interleaving is possible and the discrete-time equivalent of the channel can be modeled as a memoryless AWGN channel with independent Rayleigh channel coefficients. The lowpass discrete-time equivalent of this channel is described by an input-output relation of the form

yi = $\boldsymbol{R}_i x_i$ + ni

where xi and yi are the complex input and output of the channel, Ri is a complex iid random variable with Rayleigh distributed magnitude and uniform phase, and ni's are iid random variables drawn according to CN(0, No). The PDF of the magnitude of Ri is given by

$$p(r) = \begin{cases} \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} & r > 0 \\ 0 & r \le 0 \end{cases}$$

We know from Chapter 2, Equations 2.3-45 and 2.3-27, that R2 is an exponential random variable with expected value E[R2] = 202. Therefore, if p = IRi 12, then from Equation 2.3-27 we have

$$p(\rho) = \begin{cases} \frac{1}{2\sigma^2} e^{-\frac{\rho}{2\sigma^2}} & \rho > 0 \\ 0 & \rho \leq 0 \end{cases}$$

and since the received power is proportional to ρ, we have

$$P_r = 2\sigma^2 P_t$$

where Pt and Pr denote the transmitted and the received power, respectively. In the following discussion we assume that 202 = 1, thus Pt = Pr = P. The extension of the results to the general case is straightforward.

Depending on the availability of channel state information at the transmitter and receiver, we study the ergodic channel capacity in three cases.

No Channel State Information In this case the receiver knows neither the magnitude nor the phase of the fading coefficients Ri ; hence no information can be transmitted on the phase of the input signal. The input-output relation for the channel is given by

y = Rx + n

where R and n are independent circular complex Gaussian random variables drawn according to CAr(0, 202) and CA/(0, No), respectively. To determine the capacity of the channel in this case, we need to derive an expression for p(yIx) which can be written as

$$p(y|x) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} p(y|x, r, \theta) p(r)\, dr\, d\theta$$

where $p(r)$ is given by Equation 14.2–7 and

$$p(y|x, r, \theta) = \frac{1}{\pi N_0} e^{\frac{|y - re^{j\theta} x|^2}{N_0}}$$

It can be shown (see Problem 14.8) that Equation 14.2-11 simplifies to

$$p(y|x) = \frac{1}{\pi \left(N_0 + |x|^2\right)} e^{-\frac{|y|^2}{N_0 + |x|^2}}$$

This relation clearly shows that all the phase information is lost. It has been shown by Abou-Faycal et al. (2001) that when an input power constraint is imposed, the capacity achieving input distribution for this case has a discrete iid amplitude and an irrelevant phase. However, there exists no closed-form expression for the capacity in this case. Moreover, in the same work it has been shown that for relatively low average signal-to-

noise ratios, when PING is less than 8 dB, only two signal levels, one of them at zero, are sufficient to achieve capacity; i.e., in this case on-off signaling is optimal. As the signal-to-noise ratio decreases, the amplitude of the nonzero input in the optimal on-off signaling increases, and in the limit for P/No -->. 0 we obtain

$$\overline{C} = \frac{1}{\ln 2}\frac{P}{N_0} \approx 1.44\frac{P}{N_0}$$

By comparing this result with Equation" 14.1-8 it is seen that for low signal-to-noise ratios the capacity is equal to the capacity of an AWGN channel; but at high signal-to-noise ratios the capacity is much lower than the capacity of an AWGN channel. Although no closed form for the capacity exists, a parametric expression for the capacity is derived in Taricco and Elia (1997). The parametric form of the capacity is given by

$$P = \mu e^{-\gamma-\Psi(\mu)} - 1$$

$$\overline{C} = \frac{\mu - \gamma - \mu\Psi(\mu) - 1}{\ln 2} + \log_2 \Gamma(\mu)$$

where $\Psi(z)$ is the *digamma function* defined by

$$\Psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$$

and y = - ψ (1) ≈ 0.5772156 is Euler's constant.

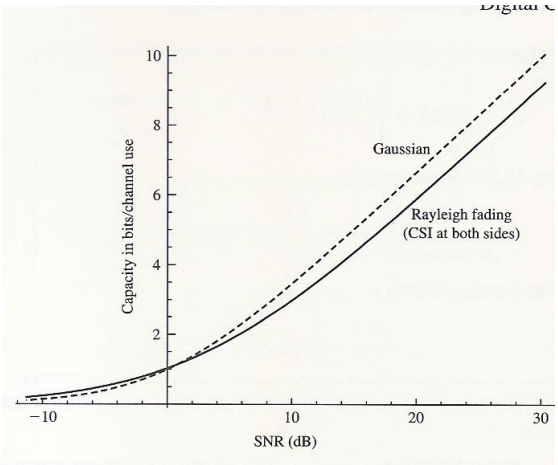
A plot of capacity in this case is shown in Figure 14.2-3. The capacity of AWGN is also given for reference. It is clearly seen that lack of information about the channel state is particularly harmful at high signal-to-noise ratios.

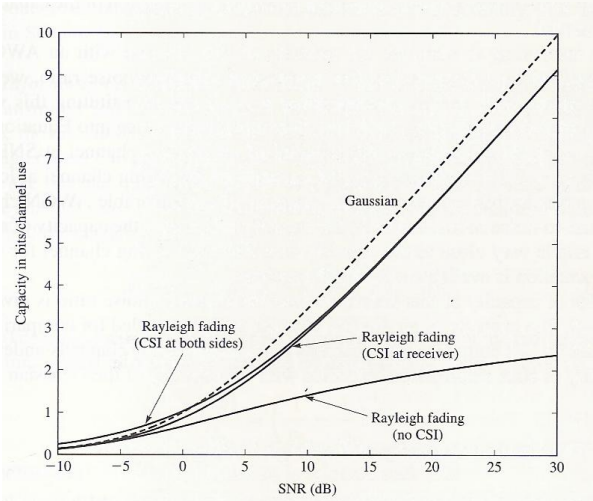
3.3 The outage capacity of Rayleigh fading channels

The outage capacity is considered when due to strict delay restrictions ideal inter-leaving is impossible and the channel capacity cannot be expressed as the average of the capacities for all possible channel realizations, as was done in the case of the

Capacity of Gaussians and Rayleigh fading channel with CSI at both sides.

ergodic capacity. We assume at rates less than capacity ideal coding is employed to make transmission effectively error-free. With this assumption, errors occur only when the rate exceeds capacity, i.e., when the channel is in outage.





Capacity of Gaussian and Rayleigh fading channel with different CSI.

For a Rayleigh fading channel the outage E-capacity is derived by using Equations 14.2-3 and 14.2-4 as

$$\begin{aligned} C_\epsilon &= \max\{R : \mathrm{P}_{\text{out}}(R) \le \epsilon\} \\ &= \max\{R : F_C(R^-) = \epsilon\} \\ &= F_C^{-1}(\epsilon) \end{aligned}$$

where FC(-) is the CDF of the random variable representing the channel capacity. For a Rayleigh fading channel with normalized channel gain, we have

C = log (1 + p SNR)

where p is an exponential random variable with expected value equal to l. The outage probability in this case is given by

$$\mathrm{P}_{\text{out}}(R) = \mathrm{P}\,[C < R]$$

which simplifies to

$$\begin{aligned} \mathrm{P}_{\text{out}}(R) &= \mathrm{P}\left[\rho < \frac{2^R - 1}{\mathrm{SNR}}\right] \\ &= 1 - e^{-\frac{2^R - 1}{\mathrm{SNR}}} \end{aligned}$$

Note that for high signal-to-noise ratios, i.e., for low outage probabilities, this expression can be approximated by

$$P_{out}(R) \approx \frac{2^R - 1}{SNR}$$

Solving for R from Equation 14.2–36 results in

$$R = \log[1 - SNR \ \ln(1 - P_{out})]$$

from which

$$C_\epsilon = \log[1 - SNR \ \ln(1 - \epsilon)]$$

We consider the cases of low and high signal-to-noise ratios separately. For low SNR values we have

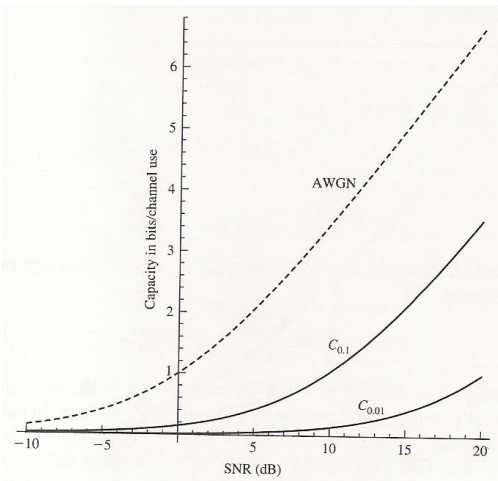
$$C_\epsilon \approx \frac{SNR}{\ln 2} \ln \frac{1}{1 - \epsilon}$$

Since the capacity of an AWGN at low SNR values is 112 SNR, we conclude that the outage capacity is a fraction of the capacity of an AWGN channel. In fact the capacity of an AWGN channel is scaled by a factor of In T-1-. For instance, for E = 0.1 this value is equal to 0.105, and the outage capacity of the Rayleigh fading channel is only one-tenth of the capacity of an AWGN channel with the same power. For very small E, this factor tends to E and we have

$$C_\epsilon \approx \epsilon C_{AWGN}$$

3.4 Effect of diversity on outage capacity.

If a communication system over a Raleigh fading channel employs L-order diversity, then the random variable p = /R/^{2} has a X^2 PDF with 2L degrees of freedom. In the special case of L = 1 we have a X^2 random variable with two degrees of freedom which is an exponential random variable studied so far. For L-order diversity we use



The CDF of a X2 random variable given by equation 2.3-24. we obtain

$$\mathrm{P}_{\text{out}}(R) = \mathrm{P}\left[\rho < \frac{2^R - 1}{\mathrm{SNR}}\right]$$
$$= 1 - e^{-\frac{2^R - 1}{\mathrm{SNR}}} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{2^R - 1}{\mathrm{SNR}}\right)^k$$

Equating $\mathbf{P}_{\text{out}}$ (R) to E and solving for R give the E-outage capacity CE for a channel with L-order diversity. The resulting CE is obtained by solving the equation.

$$e^{-\frac{2^{C_\epsilon} - 1}{\mathrm{SNR}}} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{2^{C_\epsilon} - 1}{\mathrm{SNR}}\right)^k = 1 - \epsilon$$

or equivalently

$$e^{-\frac{2^{C_\epsilon} - 1}{\mathrm{SNR}}} \sum_{k=L}^{\infty} \frac{1}{k!} \left(\frac{2^{C_\epsilon} - 1}{\mathrm{SNR}}\right)^k = \epsilon$$

No close-form solution for CE exists for arbitrary L/ Plots of C0.01 for different diversity orders as well as the capacity of an AWGN channel are given in figure 14.2-8. The noticeable improvement due to diversity is clear from this figure.

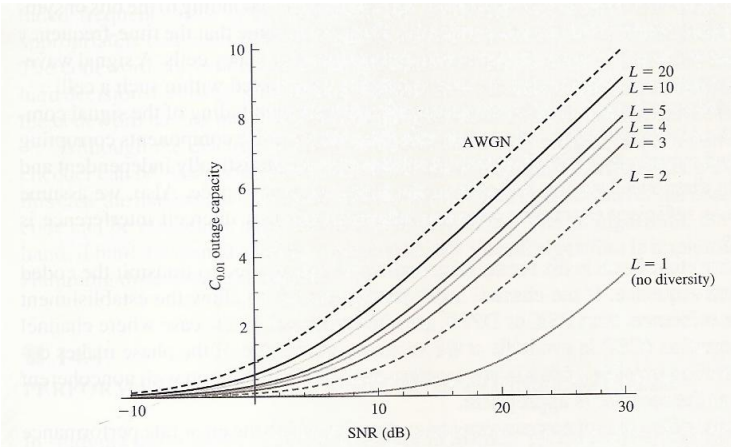


FIGURE 14.2–8
The outage capacity of fading channels with different diversity orders.

4.0 Conclusion

In this unit, we examined Ergodic capacity of channel model, Ergodic capacity of the Rayleigh fading channel. Beside, the outage capacity of the Rayleigh fading channels and the effect of diversity on outage capacity.

5.0 Summary

The channel capacity is a binary random variable which takes values of 1 and 0 with probabilities 8 and 1-8, respectively. This is a case for which Ergodic capacity is not applicable and a new notion of capacity called outage capacity is more appropriate.

6.0 Tutor Marked Assignment

1. Consider the BSC in which the channel can be in three states. In state S = 0 the output of the channel is always 0, regardless of the channel input, in state S = 2 the channel is noiseless, i.e the output is always equal to the input. We assume that P (S=0) = P (S=1) = P/2.

2. Determine the capacity of this channel, assuming no state information is available to the transmitter or the receiver.

3. Determine the capacity of the channel, assuming that channel state information S is available at both sides.

7.0 References/ Further Reading

The Importance of Coding for Digital Communication over a Fading Channel by Chase (1976).

Unit 3: Coding for and Performance of Coded Systems in Fading Channels

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Coding for Fading Channel

3.2 Performance of Coded Systems in Fading Channels

3.3 Performance of Fully Interleaved Fading Channels with CSI at the Receivers

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/ Further Reading

1.0 Introduction

In this unit, we will consider coding for fading channel and its performance. The capacity of a fading channel depends on the dynamics of the fading process.

2.0 Objectives

At the end of the unit, you should be able to;

- Explain coding for fading channel
- Explain performance of coded systems in fading channel
- Discuss performance by fully interleaved fading channels with CSI at the receivers.

3.1 Coding for fading channel

In Chapter 13 we have demonstrated that diversity techniques are very effective in overcoming the detrimental effects of fading caused by the time-variant dispersive characteristics of the channel. Time and/or frequency diversity techniques may be viewed as a form of repetition (block) coding of the information sequence. From this point of view, the combining techniques described in Chapter 13 represent soft decision decoding of the repetition code. Since a repetition code is a trivial form of coding, we now consider the additional benefits derived from more efficient types of codes. In particular, we demonstrate that coding provides an efficient means of obtaining diversity on a fading channel. The amount of diversity provided by a code is directly related to its minimum distance.

As explained in Section 13.4, time diversity is obtained by transmitting the signal components carrying the same information in multiple time intervals mutually separated by an amount equal to or exceeding the coherence time (At), of the channel. Similarly, frequency diversity is obtained by transmitting the signal components carrying the same

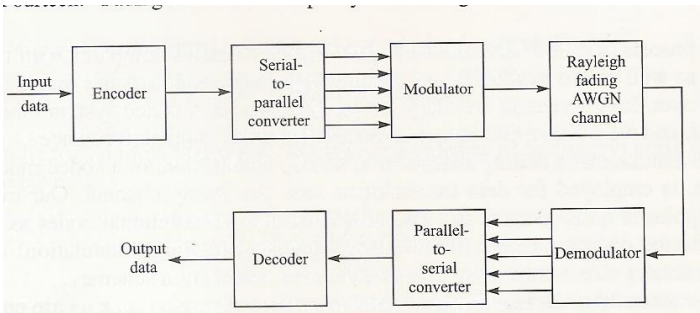
information in multiple frequency slots mutually separated by an amount at least equal to the coherence bandwidth (A f), of the channel. Thus, the signal components carrying the same information undergo statistically independent fading.

To extend these notions to a coded information sequence, we simply require that the signal waveform corresponding to a particular code bit or code symbol fade independently of the signal waveform corresponding to any other code bit or code symbol. This requirement may result in inefficient utilization of the available time-frequency space, with the existence of large unused portions in this two-dimensional signaling space. To reduce the inefficiency, a number of codewords may be interleaved in time or in frequency or both, in such a manner that the waveforms corresponding to the bits or symbols of a given codeword fade independently. Thus, we assume that the time-frequency signaling space is partitioned into nonoverlapping time-frequency cells. A signal waveform corresponding to a code bit or code symbol is transmitted within such a cell.

In addition to the assumption of statistically independent fading of the signal components of a given codeword, we assume that the additive noise components corrupting the received signals are white Gaussian processes that are statistically independent and identically distributed among the cells in the time-frequency space. Also, we assume that there is sufficient separation between adjacent cells that intercell interference is negligible.

An important issue is the modulation technique that is used to transmit the coded information sequence. If the channel fades slowly enough to allow the establishment of a phase reference, then PSK or DPSK may be employed. In the case where channel state information (CSI) is available at the receiver, knowledge of the phase makes coherent detection possible. If this is not possible, then FSK modulation with noncoherent detection at the receiver is appropriate.

A model of the digital communication system for which the error rate performance will be evaluated is shown in Figure 14.3-1. The encoder may be binary, nonbinary, or a concatenation of a nonbinary encoder with a binary encoder. Furthermore, the code



Mode of communications system with modulation/ demodulation and encoding/decoding.

generated by the encoder may be a block code a convolutional code, or, in the case of concatenation, a mixture of a block code and a convolutional code.

To explain the modulation, demodulation, and decoding, consider a linear binary block code in which k information bits are encoded into a block of n bits. For simplicity and without loss of generality, let us assume that all n bits of a codeword are transmitted simultaneously over the channel on multiple frequency/time cells. A codeword ci having bits {cij} is mapped into signal waveforms and interleaved in time and/or frequency and transmitted. The dimensionality of the signal space depends on the modulation system. For instance, if FSK modulation is employed, each transmitted symbol is a point in the two-dimensional space, hence the dimensionality of the encoded/modulated signal is 2n. Since each codeword conveys k bits of information, the bandwidth expansion factor for FSK is Be = 2n/ k.

The demodulator demodulates the signal components transmitted in independently faded frequency/ time cells, providing the sufficient statistics to the decoder which appropriately combines them for each codeword to form the M = 2k decision variables. The codeword corresponding to the maximum of the decision variables is selected. If hard decision decoding is employed, the optimum maximum-likelihood decoder selects the codeword having the smallest Hamming distance relative to the received codeword.

Although the discussion above assumed the use of a block code, a convolutional encoder can be easily accommodated in the block diagram shown in Figure 14.3-1. For this case the maximum-likelihood soft decision decoding criterion for the convolutional code can be efficiently implemented by means of the Viterbi algorithm. On the other hand, if hard decision decoding is employed, the Viterbi algorithm is implemented with Hamming distance as the metric.

3.2 Performance of coded systems in fading channels

In studying the capacity of fading channels in Section 14.2 we noted that the notion of capacity in fading channels is more involved that the notion of capacity for a standard memoryless channel. The capacity of a fading channel depends on the dynamics of the fading process and how the coherence time of the channels compares with the code length as well as the availability of channel state information at the transmitter and the receiver. In this section we study the performance of a coded system on a fading channel, and we observe that the same factors affect the code performance.

We assume that a coding scheme followed by modulation, or a coded modulation scheme, is employed for data transmission over the fading channel. Our treatment at this point is quite general and includes block and convolutional codes as well as concatenated coding schemes followed by a general signaling (modulation) scheme. This treatment also includes block or trellis-coded modulation schemes.

We assume that M signal space coded sequences fx1, x2, . . . ,xM} are employed to transmit one of the equiprobable messages 1 _< m <_ M. Each codeword xi is a sequence of n symbols of the form

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$$

where each xii is a point in the signal constellation. We assume that the signal constellation is two-dimensional, hence xij's are complex numbers.

Depending on the dynamics of fading and availability of channel state information, we can study the effect of fading and derive bounds on the performance of the coding scheme just described.

Coding for Fully Interleaved Channel Model

In this model we assume a very long interleaver is employed and the codeword components are spread over a long interval, much longer than the channel coherence time. As a result, we can assume that the components of the transmitted codeword undergo independent fading. The channel output for this model, when xi is sent, is given by

$$y_j = R_j x_{ij} + n_j, \qquad 1 \le j \le n$$

where the Rj represents the fading effect of the channel and the nj is the noise. In this model due to the interleaving, Rd's are independent and no's are iid samples drawn according to CV(0, No). The vector input-output relation for this channel is given by

$$y = Rx + n$$

where R is an $n \times n$ diagonal matrix

$$R = \text{diag}(R_1, R_2, \ldots, R_n) = \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ 0 & R_2 & 0 & \cdots & 0 \\ 0 & 0 & R_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & R_n \end{bmatrix}$$

and n is a vector with independent no's as its components. The Rd's are in general complex, denoting the magnitude and the phase of the fading process.

The maximum-likelihood decoder, having received y, uses the rule

$$\hat{m} = \arg\max_{1 \le m \le M} p(\boldsymbol{y}|\boldsymbol{x}_m)$$

to detect the transmitted message rn. By the independence of fading and noise components we have

$$p(\boldsymbol{y}|\boldsymbol{x}_m) = \prod_{i=1}^{n} p(y_j|x_{mj})$$

The value of p(yj/x,nj) depends on the availability of channel state information at the receiver.

CSI Available at the Receiver In this case the output of the channel consists of the output vector y and the channel state sequence (ri, r2, . . . , rn) which are realizations of random variables R1, R2, .. ., Rn, or equivalently the realization of matrix R. Therefore, the maximum-likelihood rule, P[observed Iinput], becomes

$$\prod_{i=1}^{n} p(y_j, r_j|x_{mj}) = \prod_{i=1}^{n} p(r_j)p(y_j|x_{mj}, r_j)$$

Substituting Equation 14.4-7 into 14.4-5 and dropping the common positive factor rjn=1 p(rj) result in

$$\hat{m} = \arg\max_{1 \le m \le M} \prod_{i=1}^{n} p(y_j | x_{mj}, r_j)$$

No CSI Available at the Receiver In this case the ML rule is

$$\hat{m} = \arg\max_{1 \le m \le M} \prod_{i=1}^{n} p(y_j | x_{mj})$$

where

$$p(y_j | x_{mj}) = \int p(r_i) p(y_j | x_{mj}, r_j)\, dr_i$$

3.3 Performance of fully interleaved fading channels with CSI at the receivers.

A bound on error probability can be obtained by using an approach similar to the one used in Section 6.8-1. Using Equation 6.8-2, we have

$$\begin{aligned} P_{e|m} &\le \sum_{\substack{m'=1 \\ m' \ne m}}^{M} \mathrm{P}[\boldsymbol{y} \in D_{mm'} | \boldsymbol{x}_m \text{ sent}] \\ &= \sum_{\substack{m'=1 \\ m' \ne m}}^{M} P_{m \to m'} \end{aligned}$$

where $P_{m \to m}$, is the pairwise error probability (PEP), i.e., the probability of error in a binary communication system consisting of two signals x,n and xm, when xm is transmitted. Here we derive an upper bound on the pairwise error probability by using the Chernov bounding technique. For other methods of studying the pairwise error probability, the reader is referred to Biglieri et al. (1995, 1996, 1998a).

A Bound on the Pairwise Error Probability To compute a bound on the PEP, we note that since in this case CSI is available at the receiver, according to Equation 14.4-8, the channel conditional probabilities are p(y j I xml , r j) and hence

$$P_{m \to m'} = \int \mathrm{P}[\boldsymbol{x}_m \to \boldsymbol{x}_{m'} | \boldsymbol{R} = \boldsymbol{r}] p(\boldsymbol{r})\, d\boldsymbol{r}$$

where

$$\begin{aligned} \mathrm{P}[\boldsymbol{x}_m \to \boldsymbol{x}_{m'} | \boldsymbol{R} = \boldsymbol{r}] &= \mathrm{P}\left[\ln \frac{p(\boldsymbol{y} | \boldsymbol{x}_{m'}, \boldsymbol{r})}{p(\boldsymbol{y} | \boldsymbol{x}_m, \boldsymbol{r})} > 0\right] \\ &= \mathrm{P}[Z_{mm'}(\boldsymbol{r}) > 0] \end{aligned}$$

and the likelihood ratio $Z_{mm'}(\boldsymbol{r})$ becomes

$$\begin{aligned} Z_{mm'}(\boldsymbol{r}) &= \ln \frac{p(\boldsymbol{y} | \boldsymbol{x}_{m'}, \boldsymbol{r})}{p(\boldsymbol{y} | \boldsymbol{x}_m, \boldsymbol{r})} \\ &= \frac{1}{N_0} \left(\| \boldsymbol{y} - \boldsymbol{r}\boldsymbol{x}_m \|^2 - \| \boldsymbol{y} - \boldsymbol{r}\boldsymbol{x}_{m'} \|^2 \right) \\ &= \frac{1}{N_0} \sum_{j=1}^{n} Z_{mm'j}(r_j) \end{aligned}$$

with

$$\begin{aligned} Z_{mm'j}(r_j) &= |y_j - r_j x_{mj}|^2 - |y_j - r_j x_{m'j}|^2 \\ &= |r_j|^2 \left(|x_{mj}|^2 - |x_{m'j}|^2 \right) + 2\mathrm{Re}\left[y_j^* r_j (x_{m'j} - x_{mj}) \right] \end{aligned}$$

Since we are assuming xn is transmitted, we have yj = rjxmj + n j. Substituting this into Equation 14.4-15 and simplifying yield

$$Z_{mm'j}(r_j) = -|r_j|^2|x_{mj} - x_{m'j}|^2 - 2\mathrm{Re}\left[r_j n_j^*(x_{mj} - x_{m'j})\right]$$
$$= -|r_j|^2 d_{mm'j}^2 - N_j$$

where Nj is a real zero-mean Gaussian random variable with variance 2 1 r j I2dm,n, j No and d,n,n, j is the Euclidean distance between the constellation points representing the j th components of x_m and x_m.

Substituting Equation 14.4-16 into Equation 14.4-13 yields 1 n

$$Z_{mm'}(\boldsymbol{r}) = \frac{1}{N_0}\sum_{j=1}^{n}\left(-|r_j|^2 d_{mm'j}^2 - N_j\right)$$

Using this result, equation 14.4-13.

$$\mathrm{P}\left[\boldsymbol{x}_m \to \boldsymbol{x}_{m'} \,|\, \boldsymbol{R} = \boldsymbol{r}\right] = \mathrm{P}\left[\sum_{j=1}^{n}\left(|R_j|^2 d_{mm'j}^2 + N_j\right) < 0 \,\middle|\, \boldsymbol{R} = \boldsymbol{r}\right]$$

Apply the Chernov bounding technique discussed in section 2.4 give

$$\mathrm{P}\left[\sum_{j=1}^{n}\left(|R_j|^2 d_{mm'j}^2 + N_j\right) < 0 \,\middle|\, \boldsymbol{R} = \boldsymbol{r}\right] = \mathrm{E}\left[e^{\nu\sum_{j=1}^{n}\left(|R_j|^2 d_{mm'j}^2 + N_j\right) < 0} \,\middle|\, \boldsymbol{R} = \boldsymbol{r}\right]$$
$$\leq \min_{\nu<0}\prod_{j=1}^{n}\mathrm{E}\left[e^{\nu\left(|R_j|^2 d_{mm'j}^2 + N_j\right)} \,\middle|\, R_j = r_j\right]$$

(14.4-1

Where/Rj/ denotes the envelop of the fading process. Substituting this result into equation. 14.4-12 gives

$$P_{m\to m'} \leq \min_{\nu<0}\prod_{j=1}^{n}\int \mathrm{E}\left[e^{\nu\left(|R_j|^2 d_{mm'j}^2 + N_j\right)} \,\middle|\, R_j = r_j\right] p(r_j)\, dr_j$$

Ricean Fading Here we assume that /Rj/, the envelope of the fading process, has a Ricean PDF as given by equation 2.3-56. we can directly apply the result of example 2.4-2 in section 2.4 and in particular equation 2.4-25 to obtained

$$P_{m\to m'} \le \prod_{j=1}^{n} \frac{1}{1+\frac{d^2_{mm'j}}{2N_0}\sigma^2} \exp\left[-\frac{\frac{d^2_{mm'j}}{4N_0}s^2}{1+\frac{d^2_{mm'j}}{2N_0}\sigma^2}\right]$$

and finally, from Equation 14.4–11 we have

$$P_e \le \frac{1}{M}\sum_{m=1}^{M}\sum_{\substack{m'=1\\ m'\neq m}}^{M}\prod_{j=1}^{n} \frac{1}{1+\frac{d^2_{mm'j}}{2N_0}\sigma^2} \exp\left[-\frac{\frac{d^2_{mm'j}}{4N_0}s^2}{1+\frac{d^2_{mm'j}}{2N_0}\sigma^2}\right]$$

In equation 14.4-21 and 14.4-22, σ^2 and s are the parameters of the Ricean random variable determining the envelope of the fading process. The pairwise error probability can also be expressed in terms of the Rice factor K as (see equation 2.4-26)

$$P_{m\to m'} \le \prod_{j=1}^{n} \frac{K+1}{K+1+\frac{Ad^2_{mm'j}}{4N_0}} \exp\left[-\frac{\frac{AKd^2_{mm'j}}{4N_0}}{K+1+\frac{Ad^2_{mm'j}}{4N_0}}\right]$$

4.0 Conclusion

In this unit, it is a crystal clear that the factors affecting the performance of a coded system on a Rayleigh fading channel are quite different from the factors affecting the performance on Gaussian channels.

5.0 Summary

We study the performance of a coded system on a fading channel, and observe that the same factors affect the code performance.

6.0 Tutor Marked Assignment

A fading channel model that is flat in both time and frequency can be modeled as y = Rx + n, where then fading factor R remains constant for the entire duration of the transmission of the codeword. Determine the optional decision rule for this channel for Ricean fading when the slate information is available at the receiver and when it is not available.

7.0 References/ Further Reading

Coding for Fading Channel by Biglieri (2005)

UNIT 4: TRELLIS-CODED MODULATION FOR FADING CHANNEL

1.0 Introduction

For code design on Gaussian channels, when soft decision decoding is employed, two parameters determine the performance of the code. These parameters are the minimum Euclidean distance of the code and the multiplicity of the code.

2.0 Objectives

At the end of this unit, you should be able to;

- Describe TCM systems for fading channels.
- Explain Multiple Trellis-coded Modulation

3.1 TCM Systems for Fading Channel

Trellis-coded modulation was described in Section 8.12 as a means for achieving a coding gain on bandwidth-constrained channels, where we wish to transmit at a bit

rate-to-bandwidth ratio R/ W > 1. For such channels, the digital communication system is designed to use bandwidth-efficient multilevel or multiphase modulation (PAM, PSK, DPSK, or QAM), which allows us to achieve an RI W > 1. When coding is applied in signal design for a bandwidth-constrained channel, a coding gain is desired without expanding the signal bandwidth. This goal can be achieved, as described in Section 8.12, by increasing the number of signal points in the constellation over the corresponding uncoded system, to compensate for the redundancy introduced by the code, and designing the trellis code so that the Euclidean distance in a sequence of transmitted symbols corresponding to paths that merge at any node in the trellis is larger than the Euclidean distance per symbol in an uncoded system. In contrast, traditional coding schemes used on fading channels in conjunction with FSK or PSK modulation expand the bandwidth of the modulated signal for the purpose of achieving signal diversity.

In designing trellis-coded signal waveforms for fading channels, we may use the same basic principles that we have learned and applied in the design of conventional coding schemes. In particular, the most important objective in any coded signal design for fading channels is to achieve as large a diversity order as possible.

As indicated above, the candidate modulation methods that achieve high bandwidth efficiency are M-ary PSK, DPSK, QAM, and PAM. The choice depends to a large extent on the channel characteristics. If there are rapid amplitude variations in the received signal, QAM and PAM may be particularly vulnerable, because a wideband automatic gain control (AGC) must be used to compensate for the channel variations. In such a case, PSK or DPSK is more suitable, since the information is conveyed by the signal phase and not by the signal amplitude. DPSK provides the additional benefit that carrier phase coherence is required only over two successive symbols. However, there is an SNR degradation in DPSK relative to PSK.

The discussion and the design criteria provided in Section 14.5 show that a good TCM code for the Gaussian channel is not necessarily a good code for the fading channel. It is quite possible that a trellis code has a large Euclidean distance but has a low effective code length or product distance. In particular some of the good codes designed by Ungerboeck for the Gaussian channel (Ungerboeck (1983)) have parallel branches in their trellises. The existence of parallel branches in TCM codes is due to the existence of

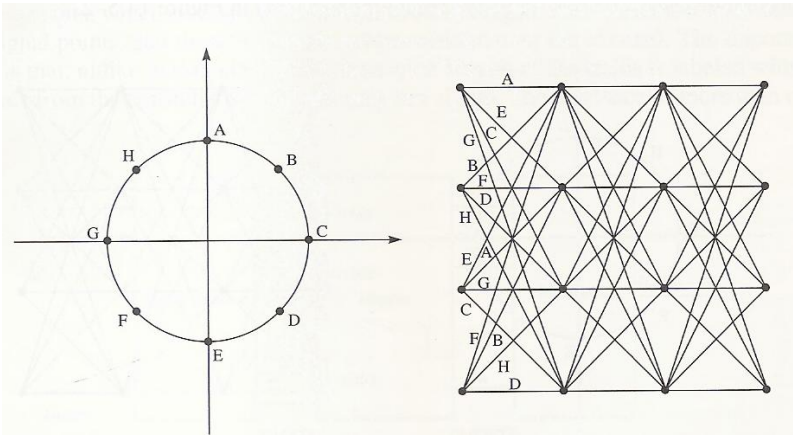
uncoded bits, as explained in Chapter 8. Obviously, two paths in the trellis that are similar on all branches but correspond to different branches on a parallel branch have a minimum distance of 1 and provide a diversity order of unity. Such codes are not desirable for transmission over fading channels due to their low diversity order and should be avoided. This is not, however, a problem with the Gaussian channel, and in fact many good TCM schemes that work satisfactorily on Gaussian channels have parallel branches in their trellis representation.

To design TCM schemes with high diversity order, we have to make sure that the paths in the trellis corresponding to different code sequences have long runs of different branches, and the branches are labeled by different symbols from the code constellation. In order for two code sequences to have a diversity order of L, the corresponding paths in the code trellis must remerge at least L branches after diverging, and the two paths on these L branches must have different labels. This clearly indicates that for L > 1 parallel transitions have to be excluded.

Let us consider an (n, k, K) convolutional code as shown in Figure 8.1-1. The number of memory elements in this code is Kk, the number of states in the trellis representing this code is 2k(x_i), and 2k branches enter and leave each state of the trellis. Without loss of generality we consider the all-zero path and a path diverging from it. The diverging path from the all-zero path corresponds to an input of k bits that contains at least one l. Since the number of memory elements of the code is Kk, it takes K sequences of k-bit inputs, all equal to zero, to move the 1 (or Is) out of the kK memory units, thus bringing back the code to the all-zero state and remerging the path with the all-zero path. This shows that the two paths that have emerged from one state can remerge after at least K branches, and hence this code can potentially provide a diversity order of K. Therefore, the diversity order that a convolutional code can provide is equal to K, the constraint length of the convolutional code. To employ this potential diversity order, we need to have enough points in the signal constellation to assign different signal points to different branches of the trellis.

Let us consider the following trellis code studied by Wilson and Leung (1987). The trellis diagram and the constellation for this TCM scheme are shown in Figure 14.5-1 As seen in the figure, the trellis corresponding to this code is a fully connected trellis, and there are no parallel branches on it, i.e., each branch of the trellis corresponds to a single

point in the constellation. The diversity order for this trellis is 2; therefore the error probability is inversely proportional to the square of the signal-to-noise-ratio. The product distance provided by this code is 1.172. It can be easily verified that the squared free Euclidean distance for this code is $d^2_{free} = 2.586$; therefore the coding



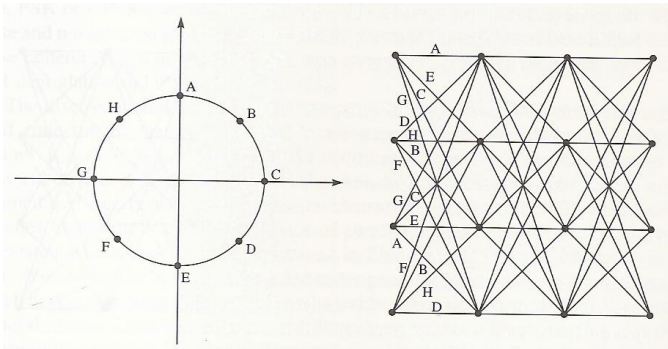
gain of the TCM scheme in Figure 14.5-1, when used for transmission over an AWGN channel, is 1.1 dB which is 1.9 dB inferior to the coding gain of the Ungerboeck code of comparable complexity given in Section 8.12.

In Schlegel and Costello (1989) a class of 8-PSK rate 2/3 TCM codes for various constraint lengths is introduced. The search for good codes in this work is done among all codes that can be designed by employing a systematic convolutional code followed by mapping to the 8-PSK signal constellation. It turns out that the advantage of this design procedure is more noticeable at higher constraint lengths. In particular, this design approach results in the same codes obtained by Ungerboeck (1983) when the constraint length is small. At high constraint lengths these codes are capable of providing both higher diversity orders and higher product distances compared to the codes designed by Ungerboeck. For example, for a trellis with 1024 states, these codes can provide a diversity order of 5 and a (normalized) product distance of 128. For comparison, the Ungerboeck code with the same complexity can provide a diversity order of 4 and a product distance of 32.

In Du and Vucetic (1990), Gray coding is employed in the mapping from a convolutional code output to the signal constellation. An exhaustive search is performed on 8-PSK TCM schemes, and it is shown that, particularly at lower constraint lengths, these codes have a better performance compared to those designed in Schlegel and Costello (1989). As the number of states increases, the performance of the codes designed in Schlegel

and Costello (1989) is better. As an example for a 32-state trellis code, the approach of Du and Vucetic (1990) results in a diversity order of 3 and a normalized product distance of 32, whereas the corresponding figures for the code designed in Schlegel and Costello (1989) are 3 and 16, respectively.

In Jamali and Le-Ngoc (1991), not only is the design problem of good 4-state 8-PSK trellis codes addressed, but also general design rules are formulated for the Rayleigh fading channel. These design principles can be viewed as the generalization of the design rules formulated in Ungerboeck (1983) for the Gaussian channel. Application of these rules results in improved performance. As an example, by applying these rules one obtains the signal constellation and the trellis shown in Figure 14.5-2.



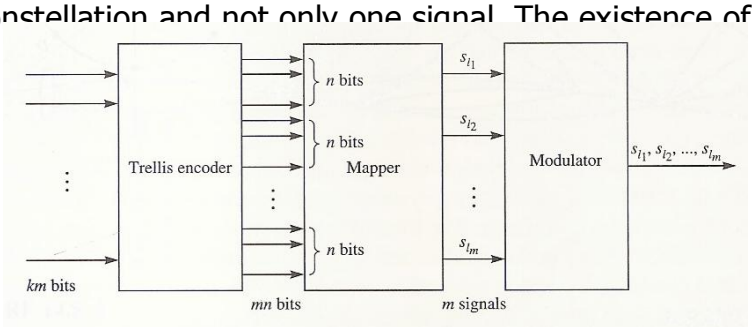
It is easy to verify that the coding gain of this code over an AWGN channel (as expressed by the free Euclidean distance) is 2 dB, which is 0.9 dB superior to the code designed in Wilson and Leung (1987) and shown in Figure 14.5-1, and only 1 dB inferior to the Ungerboeck code with a comparable complexity. It is also easy to see that the product distance of this code is twice the product distance of the code shown in Figure 14.5-1, and therefore the performance of this code over a fading channel is superior to the performance of the code designed in Wilson and Leung (1987). Since the squared product distance of this code can be shown to be twice the squared product distance of the code shown in Figure 14.5-1, the asymptotic performance improvement of this code compared to the one designed in Wilson and Leung (1987), when used over fading channels, is 10 log /2- = 1.5 dB. The encoder for this code can be realized by a convolutional encoder followed by a natural mapping to the 8-PSK signal set.

3.2 Multiple Trellis-Coded Modulations

We have seen that the performance of trellis code modulation schemes on fading channels is primarily determined by their diversity order and product distance. In particular, we saw that trellises with parallel branches are to be avoided in transmission over fading channels due to their low (unity) diversity order. In cases where high bit rates are to be transmitted under severe bandwidth restrictions, the signal constellation consists of many signal points. In such cases, to avoid parallel paths in the code trellis, the number of trellis states should be very large, resulting in a very complex decoding scheme.

An innovative approach to avoid parallel branches and at the same time to avoid a very large number of states is to employ multiple trellis-coded modulation (MTCM) as first formulated in Divsalar and Simon (1988c). The block diagram for a multiple trellis-coded modulation is shown in Figure 14.5-3.

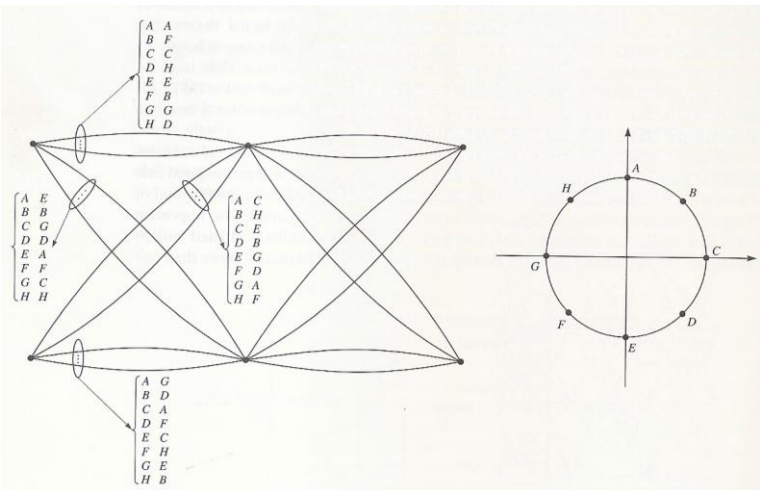
In the multiple trellis-coded modulation depicted in Figure 14.5-3, at each instance of time K = km information bits enter the trellis encoder and are mapped into N = nm bits, which correspond to m signals from a signal constellation with a total of 2' signal points, and these m signals are transmitted over the channel. The important fact is that, unlike the standard TCM, here each branch of the trellis is labeled with m signals from the constellation and not only one signal. The existence of more than one



signal corresponding to each trellis branch results in higher diversity order and therefore improved performance when used over fading channels. In fact, MTCM schemes can have a relatively small number of states and at the same time avoid a reduced diversity order. The throughput (or spectral bit rate, defined as the ratio of the bit rate to the bandwidth) for this system is k, which is equivalent to an uncoded (and a conventional TCM) system. In most implementations of MTCM, the value of n

is selected to be k + 1. Note that with this choice, the case m = 1 is equivalent to conventional TCM. The rate of the MTCM code is R = KIN = k/n.

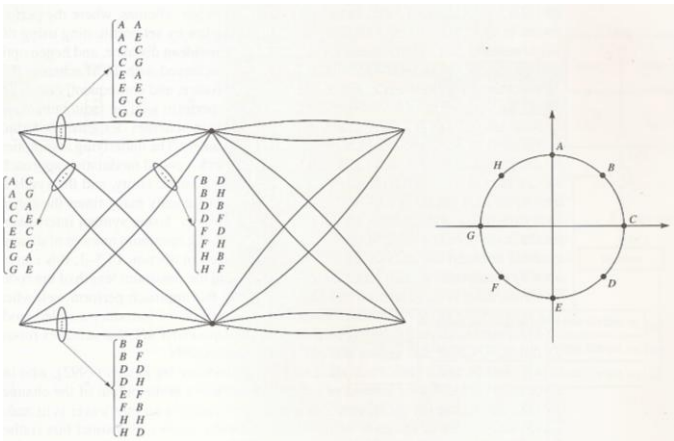
In the following example we give a specific TCM scheme and discuss its performance in a fading environment. The signal constellation and the trellis for this example are shown in Figure 14.5-4. For this code we assume m = 2, k = 2, and n = 3. Therefore, the rate of this code is 2/3, and the trellis selected for the code is a two-state trellis. At each instant of time K = km = 4 information bits enter the encoder. This means that there are 2K = 16 branches leaving each state of the trellis. Due to the symmetry in the structure of the trellis, there exist eight parallel branches connecting any two states of the trellis. The difference, however, with conventional trellis-coded modulation is that here we assign two signals in the signal space to each branch of the trellis. In fact, corresponding to the K = 4 information bits that enter the encoder, N = nm = 6 binary symbols leave the encoder. These six binary symbols are used to select two signals from the 8-PSK constellation shown in Figure 14.5-4 (each signal



requires three binary symbols). The mappings of the branches to the binary symbols are also shown in Figure 14.5-4. Close examination of the mappings suggested in this figure shows that although there exist parallel branches in the trellis for this code, the diversity order provided by this code is equal to 2.

It is seen from the above example that multiple trellis-coded modulation can achieve good diversity, which is essential for transmission through the fading channel, without requiring complex trellises with a large number of states. It can also be shown (see Divsalar and Simon (1988c)), that this same technique can provide all the benefits of using the asymmetric signal sets, as described in Divsalar et al. (1987), without the dif-

ficulties encountered with time jitter and catastrophic trellis codes. Optimum set partitioning rules for multiple trellis-coded modulation schemes are investigated in Divsalar and Simon (1988b) (see also Biglieri et al. (1991)). It is important to note that the signal set assignments to the trellis branches shown in Figure 14.5-4 are not the best possible signal assignments if this code is to be used over an AWGN channel. In fact, the signal set assignment shown in Figure 14.5-5 provides a performance 1.315 dB superior to the signal set assignment of Figure 14.5-4 when used over an AWGN channel. However, obviously the signal assignment of Figure 14.5-5 can only provide a diversity order equal to unity as opposed to the diversity order of 2 provided by the signal assignment of Figure 14.5-4. This means that on fading channels the performance of the code shown in Figure 14.5-4 is superior to the performance of the code shown in Figure 14.5-5.



4.0 Conclusion

For fading channels the code parameters with highest impact on code performance are the code diversity or effective length, given by the minimum Hamming distance of the code and lastly, the product distance of the code. This parameter results in a shift in the error probability plot of the code and has the same effect at all signal-to-noise ratio. Besides, the multiplicity of the mode N min is another parameter.

5.0 Summary

Trellis-coded modulation for fading channels has been considered in this unit, we also examine the performance of Trellis code modulation which primarily determined by their diversity order and product distance.

6.0 Tutor Marked Assignment

Explain briefly why multiple Trellis coded modulation can achieve god diversity.

7.0 References/ Further Reading

Trellis-Coded Modulation for Fading Channels by Biglieri et al (1991) and LeNgoc (1994)

UNIT 5: BIT-INTERLEAVED CODED MODULATION

1.0 Introduction

2.0 Objectives

3.0 Main Content

3.1 Bit-Interleaved Modulation

3.2 Coding in the Frequency Domain

3.3 Use of Constant-Weights Codes and Concatenated Codes for a Fading Channel

4.0 Conclusion

5.0 Summary

6.0 Tutor Marked Assignment

7.0 References/ Further Reading

1.0 Introduction

In the unit, we consider Bit-interleaved coded modulation which makes the diversity order of the code equal to the minimum number of distinct bits (rather than channel symbols) by which two Trellis paths differ.

2.0 Objectives

At the end of this unit, we should be able to;

- Explain bit interleaved modulation
- Discuss coding in the frequency domain
- Understand the use of constant-weight code and concatenated codes for a fading channel

3.1 Bit-Interleaved Modulation

In Section 8.12 we have seen that a coded modulation system in which coding and modulation are jointly designed as a single entity provides good coding gain over Gaussian channels with no expansion in bandwidth. These codes employ labeling by set partitioning on the code trellis rather than common labeling techniques such as Gray labeling, and these codes achieve their good performance over Gaussian channels by providing large Euclidian distance between trellis paths corresponding to different coded sequences. On the other hand, a code has good performance on a fading channel if it can provide high diversity order, which depends on the minimum Hamming distance of the code, as was seen in Section 14.4-1. For a code to have good performance under both channel models, it has to provide high Euclidean and high Hamming distances. We have previously seen in Chapter 7 that for BPSK and BFSK modulation schemes the relation between Euclidean and Hamming distances is a simple relation given by Equations 7.2-15 and 7.2-17, respectively. These equations indicate that for these modulation schemes Euclidean and Hamming distances are optimized simultaneously.

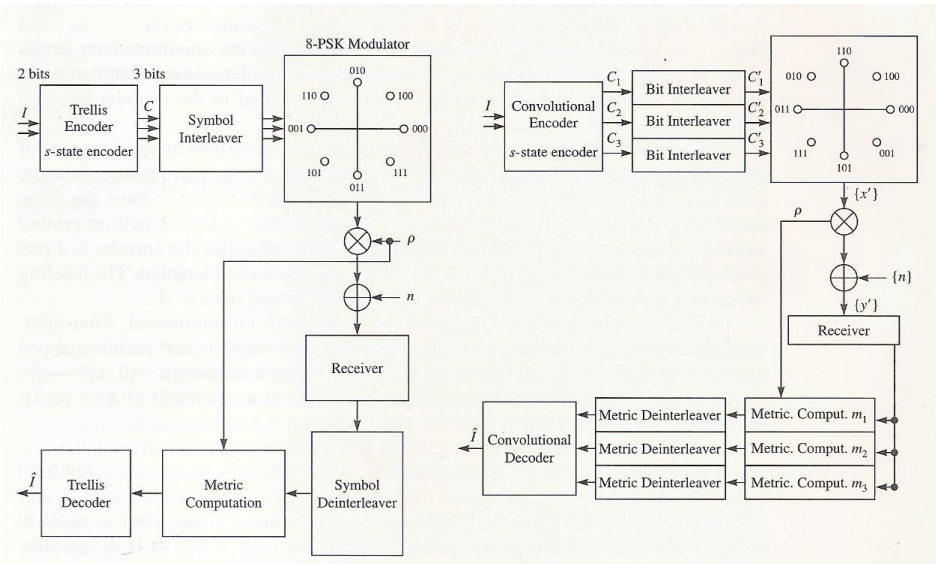
For coded modulation where expanded signal sets are employed, the relation between Euclidean and Hamming distances is not as simple as the corresponding relations for BPSK and BFSK. In fact, in many coded modulation schemes, where the performance is

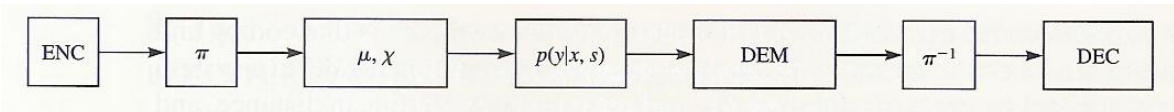
optimized through labeling the trellis branches by set partitioning using the Ungerboeck's rules (Ungerboeck (1983)), optimal Euclidean distance, and hence optimal performance on the AWGN channels model, is achieved with TCM schemes that have parallel branches and thus have a Hamming distance, and consequently diversity order, equal to unity. These codes obviously cannot perform well on fading channels. In Section 14.5 we gave examples of coded modulation schemes designed for fading channels that achieve good diversity gain on these channels. The underlying assumption in designing these codes was that similar to Ungerboeck's coded modulation approach, the modulation and coding have to be considered as a single entity, and the symbols have to be interleaved by a symbol interleaver of depth usually many times the coherence time of the channel to guarantee maximum diversity. Using symbol interleavers results in the diversity order of the code being equal to the minimum number of distinct symbols between the codewords; and as we have seen in Section 14.5-1, this can be done by eliminating parallel transitions and increasing the constraint length of the code. However, there is no guarantee that the codes using this approach perform well when transmitted over an AWGN channel model. In this section we introduce a coded modulation scheme, called bit-interleaved coded modulation (BICM), that achieves robust performance under both fading and AWGN channel models.

Bit-interleaved coded modulation was first introduced by Zehavi (1992), who introduced a bit interleaver instead of a symbol interleaver at the output of the channel encoder and before the modulator. The idea of introducing a bit interleaver is to make the diversity order of the code equal to the minimum number of distinct bits (rather than channel symbols) by which two trellis paths differ. Using this scheme results in a new soft decision decoding metric for optimal decoding that is different from the metric used in standard coded modulation. A consequence of this approach is that coding and modulation can be done separately. Separate coding and modulation results in a system that is not optimal in terms of achieving the highest minimum Euclidean distance, and therefore the resulting code is not optimal when used on an AWGN channel. However, the diversity order provided by these codes is generally higher than the diversity order of codes obtained by set partitioned labeling and thus provides improved performance over fading channels. A block diagram of a standard TCM system and a bit-interleaved coded modulation system are shown in Figure 14.6-1. In both systems a rate 2/3 convolutional code with an 8-PSK constellation is employed. In the TCM system, the symbol

outputs of the encoder are interleaved and then modulated using the 8-PSK constellation and transmitted over the fading channel, in which p and n denote the fading and noise processes. In the BICM system, instead of the symbol interleaver we are using three independent bit interleavers that individually interleave the three bit streams. In both systems deinterleavers (at symbol and bit level, respectively) are used at the receiver to undo the effect of interleaving. Note that the fading process (CSI) is available at the receiver in both systems.

Bit-interleaved coded modulation was extensively studied in Caire et al. (1998). This comprehensive study generalized the system introduced by Zehavi (1992), which used multiple bit interleavers at the output of the encoder, and instead used a single bit





interleaver that operates on the entire encoder output. The block diagram of the system studied in Caire et al. (1998) is shown in Figure 14.6-2. The encoder output is applied to to an interleaver denoted by π. The output of the interleaver is modulated by the modulator consisting of a label map $/, t$ followed by a signal set X. The channel model is a state channel with state s which is assumed to be a stationary, finite-memory vector channel whose input and output symbols x and y are N-tuples of complex numbers. The state s is independent of the channel input x, and conditioned on s, the channel is memoryless, i.e.,

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{s}) = \prod_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{s}_i)$$

The state sequence s is assumed to be a stationary finite-memory random process; i.e., there exists some integer v >_ 0 such that for all integers r and s and all integers *v < kl < k2 < ... < k,* and *jl < j2 < ... < js* < 0, the sequences (sk..... *sk,)* and (s i..... $s_{j,}$) are independent. The integer v represents the maximum memory length of the state process. The output of the channel enters the demodulator that computes the branch metrics which after deinterleaving are supplied to the decoder for final decision.

Both coded modulation and BICM systems can be described as special cases of the block diagram of Figure 14.6-2. A coded modulation system results when the encoder is defined over the label alphabet A and A and X $\subset C^N$ have the same cardinality, i.e., when IAI = I X I = M. The labeling map *Et :* A --> X acts on symbol interleaved encoder outputs individually. For Ungerboeck codes the encoder is a rate *k/n* convolutional code, and A is the set of binary sequences of length *n.* The labeling function *it* is obtained through applying the set partitioning rules to X.

In BICM, a binary code is employed and its output is bit-interleaved. After interleaving the bit sequence is broken into subsequences of length *n,* and each is mapped onto a constellation X C *CN* of size I X I = M = 2^n using a mapping u: $\{0,1\}^n \rightarrow$ X.

Let 'x $\in$ X and let e^i (x) denote the ith bit of the label x; obviously e^i (x) *E {0, 1].* We define

$$\mathcal{X}_b^i = \{x \in \mathcal{X} : \ell^i(x) = b\}$$

where Xb denotes the set of all points in the constellation whose label is equal to *b E {0,* 1} at position i. It can be easily seen that if P *[b = 0] = P[b*= 1] = 1/2, then

$$p(y|\ell^i(x) = b, s) = 2^{-(m-1)} \sum_{x \in \mathcal{X}_b^i} p(y|x, s)$$

The computation of the bit metrics at the demodulator depends on the availability of the channel state information. If CSI is available at the receiver, then the bit metric for the ith bit of the symbol at time k is given by the log-likelihood

$$\lambda^i(y_k, b) = \log \sum_{x \in \mathcal{X}_b^i} p(y_k|x, s)$$

and for the case with no CSI we have

$$\lambda^i(y_k, b) = \log \sum_{x \in \mathcal{X}_b^i} p(y_k|x)$$

where b E (0,1) and 1 ≤ i ≤ n. In the bit metric calculation for the no CSI case, we have

$$p(y_k|x) = \int p(y_k|x, s)p(s)\,ds$$

Finally, the decoder uses the ML bit metrics to decode the codeword c E C according to

$$\hat{c} = \arg\max_{c \in \mathcal{C}} \sum_{i=1}^{N} \lambda^i(y_k, c_k)$$

which can be implemented using the Viterbi algorithm.

A simpler version of bit metrics can be found using the approximation

$$\log \sum_i a_i \approx \max_i \log a_i$$

which is similar to Equation 8.8-33. With this approximation we have the approximate bit metric

$$\tilde{\lambda}^i(y_k, b) = \begin{cases} \max\limits_{x \in \mathcal{X}_b^i} \log p(y_k|x, s) & \text{CSI available} \\ \max\limits_{x \in \mathcal{X}_b^i} \log p(y_k|x) & \text{no CSI} \end{cases}$$

It turns out that BICM performs better when it is used with Gray labeling as opposed to labeling induced by the set partitioning rules. The Gray and set partitioning labeling for 16-QAM constellation is shown in Figure 14.6-3. Gray labeling is possible for certain constellations. For instance, Gray labeling is not possible for a 32-QAM constellation. In such cases a quasi-Gray labeling achieves good performance.

The channel model for BICM, when ideal interleaving is employed, is a set of n independent memoryless parallel channels with binary inputs that are connected via a random switch to the encoder output. Each channel corresponds to one particular bit position from the total n bits. The capacity and the cutoff rate for this channel model under the assumption of full CSI at the receiver and no CSI are computed in Caire et al. (1998). Figure 14.6-4 shows the cutoff rate for different BICM systems for different QAM signaling schemes over AWGN and Rayleigh fading channels.

3.2 Coding in the Frequency Domain

Instead of bitwise or symbolwise interleaving in the time domain to increase diversity of a coded system and improve the performance over a fading channel, we can achieve

similar diversity order by spreading the transmitted signal components in the frequency domain. A candidate modulation scheme for this case is FSK which can be demodulated noncoherently when tracking the channel phase is not possible.

A model for this communication scheme is shown in Figure 14.3-1 where each bit {cij} is mapped into FSK signal waveforms in the following way. If cij = 0, the tone foi is transmitted; and if cii = l, the tone fly is transmitted. This means that 2n tones or cells are available to transmit the n bits of the codeword, but only n tones are transmitted in any signaling interval.

The demodulator for the received signal separates the signal into 2n spectral components corresponding to the available tone frequencies at the transmitter. Thus, the demodulator can be realized as a bank of 2n filters, where each filter is matched to one of the possible transmitted tones. The outputs of the 2n filters are detected noncoherently. Since the Rayleigh fading and the additive white Gaussian noises in the 2n frequency cells are mutually statistically independent and identically distributed random processes, the optimum maximum-likelihood soft decision decoding criterion requires that these filter responses be square-law-detected and appropriately combined for each codeword to form the M = 2k decision variables. The codeword corresponding to the maximum of the decision variables is selected. If hard decision decoding is employed, the optimum maximum-likelihood decoder selects the codeword having the smallest Hamming distance relative to the received codeword. Either a block or a convolutional code can be employed as the underlying code in this system.

Probability of Error for Soft Decision Decoding of Linear Binary Block Codes

Consider the decoding of a linear binary (n, k) code transmitted over a Rayleigh fading channel, as described above. The optimum soft-decision decoder, based on the maximum-likelihood criterion, forms the M = 2k decision variables.

$$
\begin{aligned}
U_i &= \sum_{j=1}^{n} \left[(1 - c_{ij})|\, y_{0j}|^2 + c_{ij}|\, y_{1j}|^2\right] \\
&= \sum_{j=1}^{n} \left[|\, y_{0j}|^2 + c_{ij}\left(|\, y_{1j}|^2 - |\, y_{0j}|^2\right)\right], \qquad i = 1, 2, \ldots, 2^k
\end{aligned}
$$

where $/y_{rj}/^2$, j = 1, 2, . . . , n, and r = 0, 1 represent the squared envelopes at the outputs of the 2n filters that are tuned to the 2n possible transmitted tones. A decision is made in favor of the code word corresponding to the largest decision variable of the set {Ui}

Our objective in this section is the determination of the error rate performance of the soft-decision decoder. Toward this end, let us assume that the all-zero code word cl is transmitted. The average received signal-to-noise ratio per tone (cell) is denoted by p~. The total received SNR for the n tones in np, and, hence, the average SNR per bit is

$$\bar{\gamma}_b = \frac{n}{k}\bar{\gamma}_c = \frac{\bar{\gamma}_c}{R_c}$$

where R, is the code rate.

The decision variable Ul corresponding to the code word cl is given by Equation 14.7-1 with clj = 0 for all j. The probability that a decision is made in favor of the mth code word is just

$$\begin{aligned} P_2(m) &= P(U_m > U_1) = P(U_1 - U_m < 0) \\ &= P\left[\sum_{j=1}^{n}(c_{1j} - c_{mj})\left(|y_{1j}|^2 - |y_{0j}|^2\right) < 0\right] \\ &= P\left[\sum_{j=1}^{w_m}\left(|y_{0j}|^2 - |y_{1j}|^2\right) < 0\right] \end{aligned}$$

where w_n is the weight of the mth code word. But the probability in Equation 14.7-3 is just the probability of error for square-law combining of binary orthogonal FSK with w,nth-order diversity. That is,

$$\begin{aligned} P_2(m) &= p^{w_m}\sum_{k=0}^{w_m-1}\binom{w_m - 1 + k}{k}(1-p)^k \\ &\le p^{w_m}\sum_{k=0}^{w_m-1}\binom{w_m - 1 + k}{k} = \binom{2w_m - 1}{w_m}p^{w_m} \end{aligned}$$

where

$$p = \frac{1}{2 + \bar{\gamma}_c} = \frac{1}{2 + R_c\bar{\gamma}_b}$$

As an alternative, we may use the Chernov upper bound derived in Section 13.4, which in the present notation is

$$P_2(m) \leq [4p(1-p)]^{w_m}$$

The sum of the binary error events over the M - 1 nonzero-weight code words gives an upper bound on the probability of error. Thus,

$$P_e \leq \sum_{m=2}^{M} P_2(m)$$

Since the minimum distance of the linear code is equal to the minimum weight, it follows that

$$(2 + R_c\bar{\gamma}_b)^{-w_m} \leq (2 + R_c\bar{\gamma}_b)^{-d_{\min}}$$

The use of this relation is conjunction with Equations 14.7-5 and 14.7-8 yields a simple, albeit looser, upper bound that may be expressed in the form

$$P_e < \frac{\sum_{m=2}^{M} \binom{2w_m - 1}{w_m}}{(2 + R_c\bar{\gamma}_b)^{d_{\min}}}$$

This simple bound indicates that the code provides an effective order of diversity equal to dmin. An even simpler bound is the union bound

$$P_e < (M-1)[4p(1-p)]^{d_{\min}}$$

which is obtained from the Chernov bound given in Equation 14.7-7. As an example serving to illustrate the benefits of coding for a Rayleigh fading channel, we have plotted in Figure 14.7-1 the performance obtained with the extended Golay (24,12) code and the performance of binary FSK and quaternary FSK each with dual diversity. Since the extended Golay code requires a total of 48 cells and k = 12, the bandwidth expansion factor Be = 4. This is also the bandwidth expansion factor for binary and quaternary FSK with L = 2. Thus, the three types of waveforms are compared on the basis of the same bandwidth expansion factor. Note that at Pb = 10-4, the Golay code

outperforms quaternary FSK by more than 6 dB, and at Pb = 10-5, the difference is approximately 10 dB.

The reason for the superior performanc of the Golay code is its large minimum distance (dtnin = 8), which translates into an equivalent eighth-order (L = 8) diversity. In contrast, the binary and quaternary FSK signals have only second-order diversity. Hence, the code makes more efficient use of the available channel bandwidth. The price that we must pay for the superior performance of the code is the increase in decoding complexity.

3.3 Use of Constant-weight Codes and Concatenated Codes for a Fading Channel

Our treatment of coding for a Rayleigh channel to this point was based on the use of binary FSK as the modulation technique for transmitting each of the binary digits in a code word. For this modulation technique, all the 2k code words in the (n, k) code have identical transmitted energy. Furthermore, under the condition that the fading on the n transmitted tones is mutually statistically independent and identically distributed, the average received signal energy for the M = 2k possible code words is also identical. Consequently, in a soft-decision decoder, the decision is made in favor of the code word having the largest decision variable.

The condition that the received code words have identical average SNR has an important ramification in the implementation of the receiver. If the received code words do not have identical average SNR, the receiver must provide bias compensation for each received code word so as to render it equal energy. In general, the determination of the appropriate bias terms is difficult to implement because it requires the estimation of the average received signal power; hence, the equal-energy condition on the received code words considerably simplifies the receiver processing.

There is an alternative modulation method for generating equal-energy waveforms from code words when the code is constant-weight, i.e., when every code word has the same number of Is. Note that such a code is non-linear. Nevertheless, suppose we assign a single tone or cell to each bit position of the 2k code words. Thus, an (n, k) binary block code has n tones assigned. Waveforms are constructed by transmitting the tone corresponding to a particular bit in a code word if that bit is a 1; otherwise, that tone is

not transmitted for the duration of the interval. This modulation technique for transmitting the coded bits is called on-off keying (OOK). Since the code is constant-weight, say, w, every coded waveform consists of w transmitted tones that depend on the positions of the is in each of the code words.

As in FSK, all tones in the OOK signal that are transmitted over the channel are assumed to fade independently across the frequency band and in time from one code word to another. The received signal envelope for each tone is described statistically by the Rayleigh distribution. Statistically independent additive white Gaussian noise is assumed to be present in each frequency cell.

The receiver employs maximum-likelihood (soft-decision) decoding to map the received waveform into one of the M possible transmitted code words. For this purpose, n matched filters are employed, each matched to one of the n frequency tones. For the assumed statistical independence of the signal fading for the n frequency cells and additive white Gaussian noise, the envelopes of the matched filter outputs are squared and combined to form the M decision variables

$$U_i = \sum_{j=1}^{n} c_{ij} |y_j|^2, \qquad i = 1, 2, \ldots, 2^k$$

where $/y/^2$ corresponds to the squared envelope of the filter corresponding to the jth frequency, where j = l, 2, . . . , n.

4.0 Conclusion

The performance gains due to code concatenation are still significant, but not impressive because the hadamard codes alone yield a large diversity, so that the increased diversity arising from concatenation does not result in as large as gain performance for the range of error rates covered.

5.0 Summary

In this unit, we summarized that the important parameter in the selection of both the outer and the inner codes is the minimum distance of the resultant concatenate code

required to achieve a specified level of performance. The ultimate choice is made on the basis of decoding complexity and bandwidth requirements.

6.0 Tutor Marked Assignment

Show that the pariwise error probability for a fully interleaved Rayleigh fading channel with fading process Ri can be bounded by;

Where the expectation is taken with respect to Ri's. From above conclude the following bound in the pairwise error probability.

7.0 References/Further Reading

Bit-Interleaved Coded Modulation by Zehari (1992).