A Spectral Analysis of Nonequiprobable Signalling

by

Hakan Altintepe, B.A.Sc.

A thesis submitted to the
School of Graduate Studies and Research
in partial fulfilment of the requirements
for the degree

Master of Applied Science

Ottawa-Carleton Institute for Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
University of Ottawa

ISBN 0-315-75062-6

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Abstract

Nonequiprobable signalling refers to techniques by which the possible transmitted signals (points in a constellation) do not occur with equal probability by intention. By selecting higher energy signals with lower probability, the power efficiency of signalling methods can be improved over equiprobable methods up to 1.53 dB. Two general methods to achieve such gains, referred to as shape gain, are the shaping codes of Calderbank & Ozarow and Forney's trellis shaping.

In this thesis, a review of the fundamentals of the shaping codes and trellis shaping methods is provided. A review of methods for the computation of power spectra for digitally coded signals relevant to nonequiprobable signalling is also presented. This background is used to provide an analysis of the spectral properties of the two-dimensional signalling associated with nonequiprobable signalling methods. Based on both analytic methods and simulation of specific examples, this analysis begins with a consideration of the spectral properties of memoryless modulation producing nonequiprobable signalling. Then, spectral properties of two examples of signals from shaping codes and sixteen examples of trellis shaping are numerically analyzed.

It was found that the statistics of the signal points affect the power spectrum of the coded output signal. Yet, if certain symmetry rules are obeyed in design of the transmission system, such that a 90 ° ratational symmetric constellation (180 ° symmetric for QAM signalling) is combined with any linear error correcting code, the power spectrum of the output signal sequence can be made invariant to the changes in the statistics of the signal points. These symmetry rules relate to the geometrical symmetry of the signal constellation and to the linearity of the codes used in communications system. Consequently, it has been verified that shaping gain can be achieved by nonequiprobable signalling without an expansion in the transmission bandwidth.

Acknowledgements

I would like to express my gratitude to Dr. Peter Galko for his continuous support and valuable suggestions throughout this thesis. I would like to acknowledge Dr. Abbas Yongacoğlu, and Dr. Mohsen Kavehrad for their contributions in providing me a broad understanding in coding and communications theory. Thanks to Langis Roy for patiently proof-reading the manuscript, and making comments which greatly improved chapter 1 and 3. Thanks to Sebastian Philippe-Auguste for his contributions to Chapter 4.

Specials thanks to my family who encouraged, and motivated me all the time despite the distance between us. I would also like to appreciate the initial support provided me by the Turkish Education Foundation.

Contents

List of Figures

List of Tables

Acronyms

BAWGN	Bandlimited Additive White Gaussian Channel
BCH	Bose, Chaudhuri, and Hocquenghem
BW	Bandwidth
CER	Constellation Expansion Ratio
CFM	Constellation Figure of Merit
CO	Calderbank & Ozarow
CPFSK	Continuous Phase, Frequency Shift Keying
CPU	Central Processing Unit
DFT	Discrete Fourier Transform
GF	Galois Field
NRZ	Non Return to Zero
PAM	Pulse Amplitude Modulation
PAR	Peak to Average Energy Ratio
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QASK	Quadrature Amplitude Shift Keying
SNR	Signal to Noise Ratio

List of Symbols

$\bar{a}$	An element of a lattice
C	Conventional code
$\bar{C}$	A signal constellation in N-dimensional signal space
$\bar{C}_B$	Baseline constellation in N-dimensional signal space
$\bar{C}^i$	ith subconstellation in $\bar{C}$
$\bar{C}_2$	A constituent two-dimensional constellation induced by $\bar{C}$
$\bar{C}_{B,2}$	Constituent two-dimensional constellation induced by $\bar{C}_B$
$\mathcal{C}(\Lambda/\Lambda', \mathrm{C})$	Coset code
$\overline{CER}(\bar{C})$	Constellation expansion of $\bar{C}$ due to nonequiprobable signalling
$\overline{CER}_2(\bar{C})$	Magnification factor of the constituent two-dimensional constellation of $\bar{C}$ due to nonequiprobable signalling
$CER_C(\mathrm{C}_c)$	Constellation expansion due to employment of (C)c
$CER_{c,2}(\Lambda)$	Constellation expansion ratio of $\bar{C}$ which is based on Λ in two-dimensions due to (lattice) coding
$CER_2(\bar{C})$	Constellation expansion ratio of constituent two-dimensional constellation induced by $\bar{C}$
$CER_n(\bar{C})$	Constellation expansion ratio of constituent n-dimensional constellation induced by $\bar{C}$
$CER_{s,2}(R_B)$	Constellation expansion ratio of $\bar{C}$ which is bounded by R_B due to the shape of R_B
$CFM(\bar{C})$	Constellation Figure of Merit of $\bar{C}$
$\bar{c}_k$	Coset representative of Λ'_k
$d(\mathrm{C})$	The minimum Euclidean distance of the L-level code C
$d^2_{\min}(\Lambda)$	Minimum squared separation of the elements of Λ
$d^2_{\min}(\bar{C})$	Minimum separation between the points in the constellation $\bar{C}$
$\frac{E_b}{N_0}$	Signal energy per bit to noise spectral density ratio
f_i	Frequency of the subconstellation $\bar{C}^i$
G	Number of subconstellations

$\mathbf{G}$	The generator matrix of code C
$\langle G, + \rangle$	Group
$g_{\Im,s}(r_0)$	The equivalence relation between the point r_0 in $\tilde{C}^0$ and the Imaginary coordinate of r_s in $\tilde{C}^s$
$g_{\Re,s}(r_0)$	The equivalence relation between the point r_0 in $\tilde{C}^0$ and the real coordinate of r_s in $\tilde{C}^s$
$g_s(r_0)$	The equivalence relation between the point r_0 in $\tilde{C}^0$ and r_s in $\tilde{C}^s$
$\mathcal{G}_{v,k}$	Greedy algorithm
$H(\tilde{C})$	Entropy of the equiprobable constellation $\tilde{C}$
$\overline{H}(\tilde{C})$	Entropy of the nonequiprobable constellation $\tilde{C}$
$\mathbf{H}^T$	The transpose of the syndrome former matrix of code C
$\Im\{\}$	Imaginary part of a complex number
$j(y, w, t)$	An equivalence class of sequences
$m(y, w, z)$	A mapping function from the L-ary words y, w and z to the signal points in the constellation
$N[coset]$	The norm of a coset
$\langle N, + \rangle$	Subgroup
$\mathbf{P}$	State transition matrix
$\mathbf{P}^{(m)}$	m-step state transition matrix
$PAR_2(\tilde{C})$	The peak to average energy ratio of $\tilde{C}$ in 2 dimensions
$PAR_N(\tilde{C})$	The peak to average energy ratio of equiprobable $\tilde{C}$ in N dimension
$\overline{PAR}_N(\tilde{C})$	The peak to average energy ratio of nonequiprobable $\tilde{C}$ in N dimensions
$P_{av}(\tilde{C})$	Average signal energy of the points contained in the equiprobable constellation $\tilde{C}$
$\overline{P}_{av}(\tilde{C})$	Average signal energy of the points contained in the nonequiprobable constellation $\tilde{C}$
P_i	Average normalized signal energy of the ith subconstellation $\tilde{C}^i$ in $\tilde{C}$
$P_{i,j}^{(n)}$	n-step transition probability from state i to state j.
$p(x)$	Continuous distribution on points in R_B
p_i	Discrete probability on points in $\tilde{C}$
$p(t)$	Time-limited pulse function
$\mathbf{R}$	Correlation matrix
$\Re\{\}$	Real part of a complex number
$\mathcal{R}$	Rotation operator
$r(C)$	The code rate of C
R_i	The boundary region enclosing the ith subconstellation $\tilde{C}^i$

$R_{B,\text{uncoded}}$	Boundary region of the constellation utilized with a uncoded transmission system
$R_{B,\text{coded}}$	Boundary region of the constellation combined with a trellis encoder
$R_{B,sq}$	Boundary region of square shaped constellation
$R_{B,cir}$	Boundary region of circular shaped constellation
R_B	Boundary region of $\bar{C}$
$R_{B,2}$	Boundary region of constituent two-dimensional constellation $\bar{C}_2$
$\mathbb{R}^N$	N-dimensional real vector space
$R(\Lambda)$	Fundamental region of Λ
$\lVert r \rVert$	The norm of a vector in real vector space $\mathbb{R}^N$
$\lVert r \rVert^2$	Energy of the point r
r_i	A signal point within the subconstellation $\bar{C}^i$
$r^2_{\max}(R_B)$	Peak energy of the signal points contained in R_B
$r_{x,i}$	The x coordinate of the signal point r_i
$r_{y,i}$	The y coordinate of the signal point r_i
$s(t)$	The syndrome sequence corresponding to the sequence t
$\mathbf{U}$	Information word
$\mathbf{V}$	Codeword
var[]	Variance of a random variable
$V(\Lambda)$	Fundamental volume of Λ
$w[n]$	Discrete window function
y_s	A legitimate sequence of code C_s
$\mathbb{Z}^N$	N-dimensional integer vector space
β	Normalized bit rate
$\mathbf{\Gamma}^i_x$	The ith trace of the correlation matrix $\mathbf{R}_x$
$\gamma_b(\bar{C})$	Biasing gain of the constellation $\bar{C}$ due to nonuniform distribution of points
$\gamma_c(\mathbb{Z}^N/\Lambda', \mathrm{C})$	Coding gain of a trellis code
$\gamma_{c,t}(\Lambda/\Lambda', \mathrm{C})$	Total coding gain of a trellis code
$\gamma_l(\Lambda)$	Lattice gain of Λ
$\gamma_l(\Lambda_1 : \Lambda_2)$	Lattice gain of Λ_1 over Λ_2
$\gamma_N(f_0, f_1, \cdots, f_{G-1})$	Biasing gain of nonequiprobable constellation constructed according to CO method
$\gamma_s(R_B)$	Shape gain of the boundary region R_B
$\gamma_t(\bar{C})$	Total shape gain
Λ	Lattice

Λ'	Sublattice of Λ
Λ'_k	Coset of Λ'
Λ/Λ'	Lattice partition
$[\Lambda/\Lambda']$	Coset representative set corresponding to the Λ/Λ' partition
$\|\Lambda/\Lambda'\|$	Order of Λ/Λ' partition
π	Stationary distribution matrix
$\Phi(f)$	Power density function
$\Omega_{\tilde{C}}$	Number of points in $\tilde{C}$
μ	The period of a irreducible Markov chain
μ_i	The period of the ith state in a Markov chain
$\otimes_N$	Spherical shaped boundary region R_B in N-dimensional signal space

Chapter 1

INTRODUCTION

1.1 Problem Statement

A digital communications system can be divided into three main components, the transmitter, the channel, and the receiver, as shown in Fig. 1.1. The transmitter and receiver are further divided into three subcomponents.

An information source produces messages, which are desired to be transmitted. Usually, these messages are converted into sequences of binary digits. During this conversion it is desired to represent the information messages by as few binary digits as possible. The process of efficiently converting the information messages into binary sequences is called *source coding*, and is implemented by the source encoder. The discrete channel encoder, placed after the source encoder, introduces redundancy into the information sequence to allow the system to better guard against noise disturbances and interference that can randomly corrupt the signal during transmission over the channel. At the final component of the transmitter, the coded digital sequence is translated into channel waveforms (signals) in a process known as modulation, and the resultant signal sequence is transmitted over the channel. Although source encoders and channel encoders are used to improve the performance of the communications over a noisy channel, we should note that there exists systems which do not utilize encoders, as the existing performance level of communications is already satisfactory.

At the receiving end of the communications system, the digital demodulator pro-

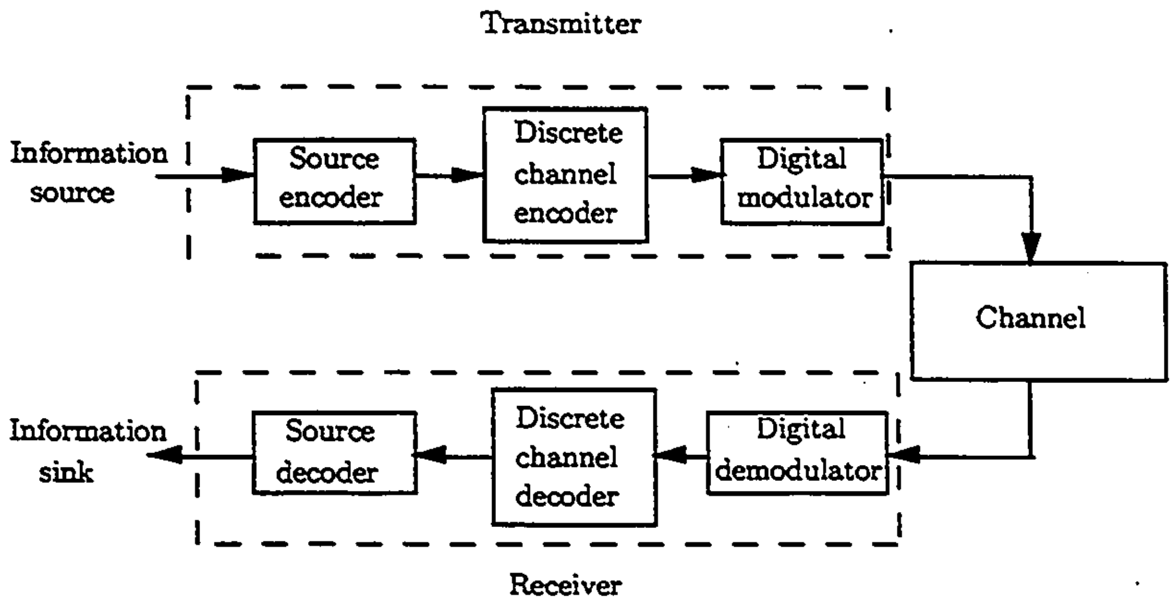


Figure 1.1: Illustration of a digital communications system.

cesses the channel-corrupted signals, and generates a digital sequence which consists of estimated data symbols that were available to the channel encoder in the transmitter. This estimated data sequence is decoded by the channel decoder. The source decoder accepts the output sequence of the channel decoder, and via the knowledge of the source encoding method used, attempts to provide to the information sink a reconstructed version of the original information message. There exist other types of receivers which utilize circuits that combine the demodulator and channel decoder, and receivers that employ complex demodulators to transfer extra information to the channel decoder about the possible errors which have occured in the demodulator (soft decision). However, for simplicity the system given in Fig. 1.1 is the general communications system model considered throughout this thesis.

Usually, the reconstructed message is identical to the original message, but sometimes due to the corruption of the transmitted signal sequence in the channel by external interferences, the digital demodulator or channel decoder makes decision errors which result in a difference between the original message and the reconstructed message. By introducing redundancy into the transmitted signal sequence, the communications system can be made more reliable against the external channel disturbances. Introducing redundancies inevitably produces an expansion in the transmission bandwidth and/or an increase in the transmitted signal power needed to obtain satisfactory performance. However, transmitted signal power may be very expensive; therefore in channels where added bandwidth is readily available, systems which trade-off bandwidth for power are preferred. Such systems are called *power limited channel* applications since the power limitation is the main constraint on the communications system. Use of conventional block codes and convolutional codes are two well-known methods applied to power limited channel applications to improve performance through expansion of the transmitted signal's bandwidth by a factor equal to the reciprocal of the code rate.

Because of the ever-increasing demand for communications, in many applications an unlimited allocation of the electromagnetic spectrum is not possible. In these cases, the application is called *bandlimited*, and an uncontrolled bandwidth expansion which would increase the communications system's immunity to channel disturbances is no longer acceptable. For bandlimited channels the system is designed to use bandefficient multi-level/phase modulation schemes such as PAM, PSK, or QAM (QASK). In this case, the performance gain can be obtained only by increasing the size of the alphabet (size of the signal constellation) of signals to be transmitted over that used in a corresponding uncoded system. This compensates for redundancy introduced by the code; thereby causing an expansion of the signal constellation, and consequently an increase in the transmitted signal power. Therefore, if coding is to provide a benefit, the performance gain of the code must be greater than the penalty paid for the constellation expansion.

Fig. 1.2 shows the channel capacity of a bandlimited additive white Gaussian noise channel (BAWGN), i.e., the maximum possible information rate in bits/s/Hz, and the

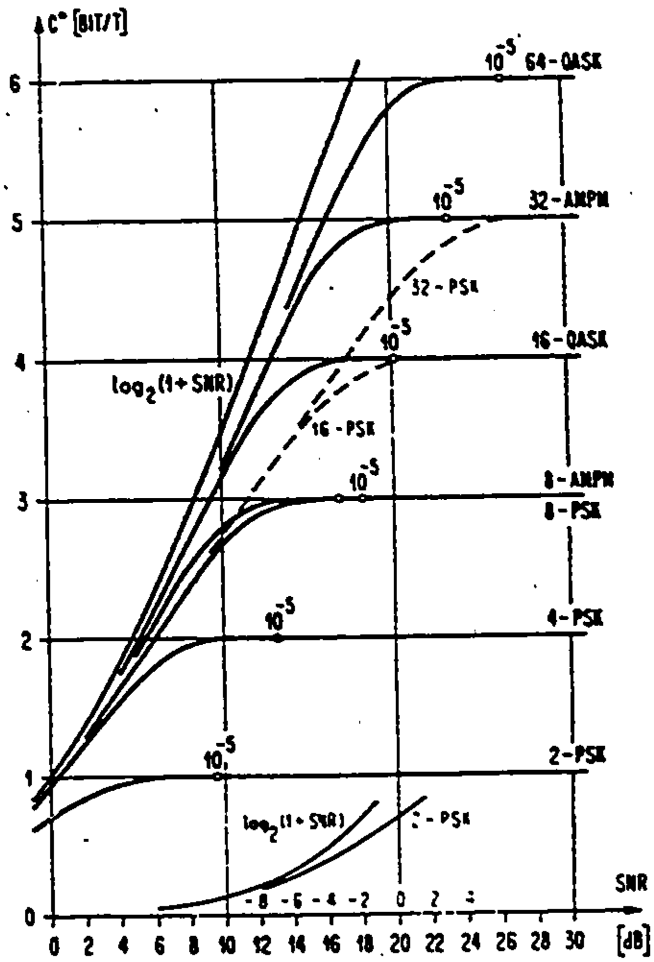


Figure 1.2: Channel capacity C* of BAWGN channels with discrete valued input and continuous valued output combined with two-dimensional modulation;(from [16]).

information rate achieved by some common uncoded multi-level/phase modulation systems as a function of signal to noise ratio (SNR). According to this figure, an uncoded 8-PSK modulation requires a SNR equal to 18 dB in order to transmit 3 bits/s/Hz, whereas a 16-QASK modulation requires only 9 dB SNR. Furthermore, the channel capacity sets 8.5 dB as the minimum SNR limit to obtain an information rate of 3 bits/s/Hz. Consequently, by doubling the number of channel signals from 8 to 16, almost all the gain in terms of channel capacity is obtained, and this is achieved by expanding the signal constellation and applying a well-designed code capable of extracting all the information from the channel-corrupted signals.

The most important class of codes that can be used with an expanded signal set in order to achieve some performance gain is the class of coset codes (they will be discussed in-depth in Chapter 2). The codes in this class achieve power gain, termed *coding gain*, by increasing the minimum separation (as seen by the decoder) between possible transmitted signal sequences so that the channel decoder can more readily discern the correct one from the channel-corrupted demodulated sequence. Increasing the minimum separation between the signal sequences can be done in two ways: either the individual signals are selected from a signal constellation, in which the minimum distance between the members of the constellation is maximized for a given transmitted signal power, or transmission of some signal sequences is prohibited so that the distances between the remaining sequences is maximized. The former method is implemented by lattice codes, and the latter by trellis codes. Since these codes are intended for use in bandlimited channel applications, no bandwidth expansion due to coding can be tolerated. Actually, according to the general assumption that the transmission bandwidth of digitally coded signal sequences which are generated by purely random input messages is proportional to the reciprocal of the symbol rate, trellis codes and lattice codes are believed to be not affect the bandwidth, since they do not alter the symbol rate over the uncoded transmission. Still, the redundancy introduced into the output symbol sequence by these codes could affect the power spectrum by varying the correlation between the elements of the output sequence.

Some researchers, e.g. [1], investigated special applications and derived conditions so that a bandwidth expansion would be prevented.

Recently, it has been noticed that the shape of the signal constellation, corresponding to the coset codes in bandlimited channel applications , also has an impact on the transmitted power. Coset codes that utilize spherically shaped signal constellations can yield up to 1.53 dB more power gain over uncoded systems with cubic shaped signal constellations. This additional gain is termed *shape gain.* The reason why performance is compared to cubic shaped constellation is that cubic constellations occur naturally in most uncoded schemes. Although shape gain does not require any special coding algorithm, and consequently is not expected to cause any bandwidth expansion, high gains can only be obtained with multidimensional signal constellations requiring extra effort in mapping the information-bearing symbol sequence into the channel signal sequence, possibly leading to an enormous increase in the complexity.

Shortly after the importance of shape gain was realized, Calderbank & Ozarow [2], and Forney [3] independently proposed two methods which enable communications systems to achieve high shape gain without directly using high-dimension constellations, and consequently avoiding high complexity in the signal mapping. Both of these algorithms are based on the nonuniform selection of the signal points within the constellation instead of selecting each signal point with equal probability as was believed to be best by Ungerboeck [4]. Because of the nonuniform distribution of points in the constellation, it is termed *nonequiprobable signalling.* The nonuniform selection of the points is accomplished by adding a new code, termed a shaping code, to the system which forces it to select low energy points more often than high energy points. Nonequiprobable signalling increases the complexity of the communications system, but after achieving a certain amount of coding gain using coset codes, it has been found wiser to invest in more complex shaping codes to achieve further power gain.

Since the shaping codes are intended to be used with communications systems for bandlimited channel applications, any bandwidth expansion due to nonequiprobable signalling is undesirable. In this research, the effect of nonequiprobable signalling on the power spectral density function of the channel signal sequence generated by coset codes and shaping codes is investigated, and the suitability of the shaping codes for bandlimited channel applications is verified. To achieve this goal, we first review the work done in calculating the power density function of digitally coded signal sequences, and assess the state of the art of nonequiprobable signalling. Finally, we combine the knowledge accumulated in both areas to calculate the power spectrum function of the signal sequences generated by modulators equipped with shaping codes, and determine the conditions which must be met in order to apply shaping codes without causing any bandwidth expansion, so that they can be used for bandlimited channel applications.

The following section is a literature survey of the calculation of power spectrum functions, and the developments of coding methods for bandlimited channel applications. Afterwards, in Chapter 2 we introduce the coset codes, the most widely used class of codes for bandlimited channel applications. Chapter 3 includes a general overview of the methods used to achieve shape gain such as multidimensional constellations and nonequiprobable signalling. Chapter 4 discusses the problem of the computation of the power spectral density function of digitally coded signal sequences. Finally, in Chapter 5, we simulate the communications systems equipped with shaping codes, and calculate the corresponding power spectral density function. At the end of this chapter we also determine the effects of nonequiprobable signalling on the power spectrum according to the results of the simulations, and reason our results with theoretical calculations. The last chapter is a summary of the results obtained during this research.

1.2 Literature Survey

Calculation of Power Spectrum Function of Digitally Coded Sequences

The first research papers on the calculation of power spectrums for various pulse sequences were published in the late 50's [6]. In the early 60's, W. R. Bennett and S. O. Rice [7] found the autocorrelation and the spectral density functions of the uncoded signal sequence generated by binary frequency-shift keying (BFSK) modulation schemes. In this research, both continuous- and discrete-phase BFSK signalling is examined. In 1971, B. S. Bosik [8] suggested that the spectral density function of coded digital signals in general could be calculated by evaluating the autocorrelation function $R(k)$, which represents the correlation between possible transmitted signals during time slots separated from each other by k symbol intervals. In this paper it was assumed that the statistics of the input symbols do not vary with time, i.e. they are stationary, and the system is ergodic. A limitation of this method is that it requires in principle the calculation of $R(k)$ for infinitely many different values of k, so that a calculation on a finite number of terms would introduce a certain amount of error. H. Yasuda and H. Inose proposed in 1970 a direct calculation method for power spectra of pulse sequences. They described a fairly simple solution for ergodic pulse sequences in [9] and for general sequences in [10]. Thereafter, P. Galko and S. Pasupathy [6] developed a Markov chain model to describe the signal sequence waveforms. According to them, one waveform is assigned to each state in the chain. Every symbol interval the chain makes a transition to some state j, generating the j^{th} waveform at the channel. Separate formulas are given in [6] for irreducible and reducible Markov models . Because of the simplicity of this method, it has been utilized for most power spectrum calculations in this thesis.

Besides the general methods for spectral density function calculations, some researchers have investigated special cases: Cariolaro and Tronca in [11] focused mainly on the spectra of general block codes combined with PAM signalling, with codewords of constant length N which are generated after channel encoding of binary information M-tuple word sequences. The authors represented the block encoder by a conventional finite-state synchronous sequential machine. The formulas derived in their paper are convenient to explain the spectral properties of block codes.

D. Divsalar and M. K. Simon [12] paid attention to the bandwidth expansion due to linear convolutional coding. Although it was widely believed that a purely random input gives a purely random output with a bandwidth expansion equal to the reciprocal of the code rate, the authors concluded that the effective bandwidth of the waveform produced by the correlated convolutional codes was narrower than that produced by uncorrelated (purely random) sequences.

One of the most practical results was put forward by E. Biglieri [1], who investigated the spectral properties of Ungerboeck codes. By simply using the method proposed in [11] he concluded that for all the encoder states, provided that the symbols available from a state of the encoder, or forcing the encoder to that state have zero mean, the spectrum of Ungerboeck coded signals is equal to that of uncoded signals.

Developments in Coding for Bandlimited Channel Applications

In the 50's and 60's mathematicians developed high density sphere packings based on lattices in multidimensional space. The algorithms based on these dense packings of signal points could achieve high power efficiency over an uncoded PAM system. In [13], D. Forney demonstrated convenient methods to construct multidimensional dense lattices using block codes. He also noted that with a 64-dimensional lattice, it is possible to obtain 9.4 dB power gain over an uncoded PAM system. These methods used to generate dense lattices are termed lattice codes.

After Ungerboeck's important publication in 1982 [16], introducing set partitioning methods, trellis codes gained a greater research momentum than lattice codes. The first trellis codes selected the signal points from within a one or two-dimensional constellation, whereas the subsequent generation of these codes, as investigated by L. Wei in [17], utilized multidimensional constellations in order to obtain a better tradeoff between complexity and coding gain.

Later, D. Forney and L. Wei were first to speak of shape gain for bandlimited channel applications in [18], and proposed opportunistic secondary channels, which then inspired some researchers to design nonequiprobable signalling schemes in order to obtain shape gain. The idea of opportunistic secondary channels mainly consists of assigning more than one signal point to some of the data words. If any of these words happens to be generated, there is an opportunity to transmit additional side information by choosing one of the two or more signal points associated with the data word. This causes an expansion in signal constellation and also increases the average signal power. However it was shown in [18] that the data rate of the opportunistic secondary channel may be higher than the increase in normalized bit rate for an ordinary system with increased average power.

In 1984, D. Forney published a research paper [19] on Voronoi constellations, which is another method of constructing multidimensional signal constellations. The power gain obtained with these constellations is a combination of coding gain and shape gain.

All the multidimensional coding techniques mentioned so far are found to achieve high coding gain; but they are also able to obtain some shape gain. Therefore multidimensional signalling is an indirect method of achieving shape gain. Calderbank & Ozarow [2] were the first to investigate nonequiprobable signalling techniques to obtain shape gain directly. They divided a low dimensional signal constellation into several subconstellations with increasing average power, where each of these subconstellations consists of an equal number of signal points. Different values were assigned to the probabilities of each subconstellation being selected; consequently they constructed a nonequiprobable signal constellation. The codes that they proposed are able to achieve 0.91 dB shape gain. Thereafter, Livingston [20] found that by dividing the signal constellation into nonequally-sized subconstellations, better shape gains could be achieved than those found in [2]. Finally, D. Forney developed a very powerful method, termed trellis shaping, to select the signal points with nonequal probabilities within a constellation and easily obtained 1 dB of shape gain.

Chapter 2

CHANNEL CODING BY MULTIAMPLITUDE SIGNALLING: COSET CODES

According to the channel capacity of the bandlimited additive white Gaussian channel (BAWGN), the performance of a transmission system employing optimally detected pulse amplitude modulation (PAM) for a memoryless source is approximately 9 dB below the performance of the best possible transmission system (i.e., for typical error rate goals [10^{-5} to 10^{-10} or so], a modulation method exists which would achieve these same error rates or better while transmitting signals with an average power level 9 dB less than that required by a PAM system). Viewed from signal space, the performance of any transmission system employing optimal detection depends primarily on the minimum distance between signal constellation points, while the average power of the transmitted signal is proportional to the average squared distance that the signal points are away from the origin. Thus one way of modifying PAM to improve its power efficiency would be to alter the signal constellation to move points towards the origin while maintaining the same minimal separation of signal points. Seeking such constellations can easily be seen to correspond to seeking the most dense packing of spheres of a given radius in signal space. Alternately, one can achieve an increase in the apparent distance between signal points (over time) by introducing some form of redundancy into the symbol stream given to the transmitter. This essentially alters the waveforms being used to represent different symbol sequences, increasing the distance between signal possibilities while preserving the average power.

Consider the first of these possibilities through the modulation scheme shown in Fig. 2.1 where we have a source symbol stream input to a mapping circuit which maps input source symbol sequences to signal points in a memoryless fashion. As an illustration, three possible signal constellations in signal spaces of different dimensions that we may consider are shown in Fig. 2.2. The constellation shown in Fig. 2.2.(a) represents the densest packing of spheres possible in a one-dimensional space. The

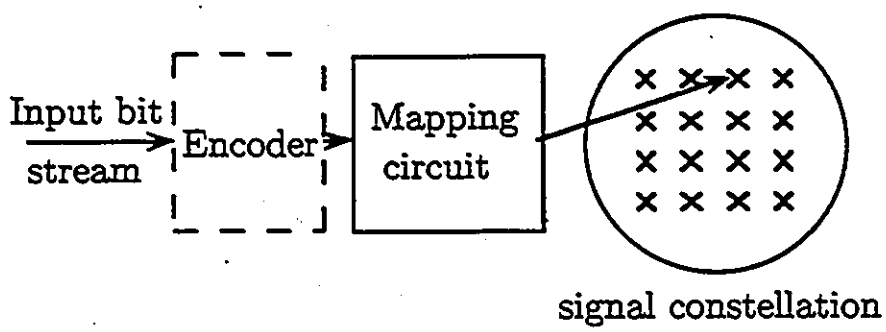


Figure 2.1: A general modulation system representation.

signal constellation in Fig. 2.2.(b) represents the two-dimensional constellation of points constructed by taking the Cartesian product of signals from the first example. It can readily be seen that constructing signal constellations in higher dimensions by taking Cartesian products results in a signal constellation containing the same number of signal points per unit volume for the same minimum separation distance as the original set and the same average symbol energy normalized to two-dimensions and thus does not increase the power efficiency of the higher dimensional constellation over that of the original constellation. Constellations with similar structures will have different average symbol energy, if they are constructed in signal spaces with different number of dimensions N. Average symbol energy is in this case proportional to N, but the information carried by each symbol is directly proportional to N. Therefore to make a fair comparison, the average symbol energy of each constellation concerned in this thesis is normalized into two dimensions by simply multiplying the average energy found in N dimensions by $N/2$. The two-dimensional constellation shown in Fig. 2.2.(c) represents the densest distribution of signal points possible in two-dimensions. For a large number of signal points, this constellation is found to have

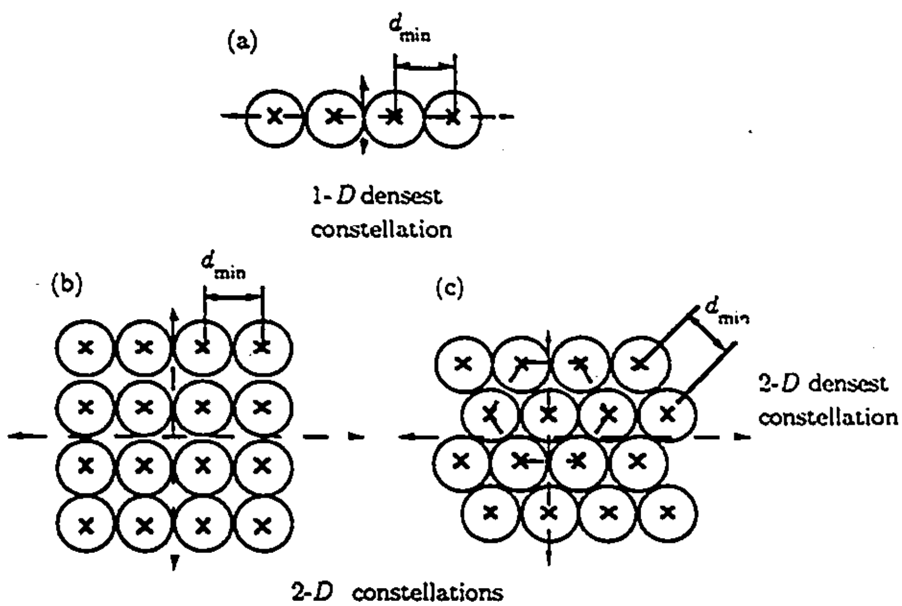


Figure 2.2: Examples for dense sphere packing in one and two-dimensional space. d_{min} denotes the minimal separation between the signal points.

an average symbol energy which is 0.6 dB less than the constellation in Fig. 2.2.(b) for the same minimum distance between points [13]. In general, it has been found that as the dimension of the signal space increases, more dense constellations are possible ultimately making up for the 9 dB penalty of simple PAM. For example, memoryless modulation with the densest 64-dimensional signal constellation has a 9.4 dB average power advantage over simple PAM. Such high dimensional dense constellations are however very difficult to implement.

Consider now the alternative of employing redundancy in the symbol stream. In the early 80's, Ungerboeck suggested a technique to achieve larger signal separations without bandwidth or power penalties [16]. This technique was referred to as *set partitioning* and consisted of dividing or partitioning a large signal constellation into subconstellations in which the minimum separation of points was larger than in the original complete signal constellation.

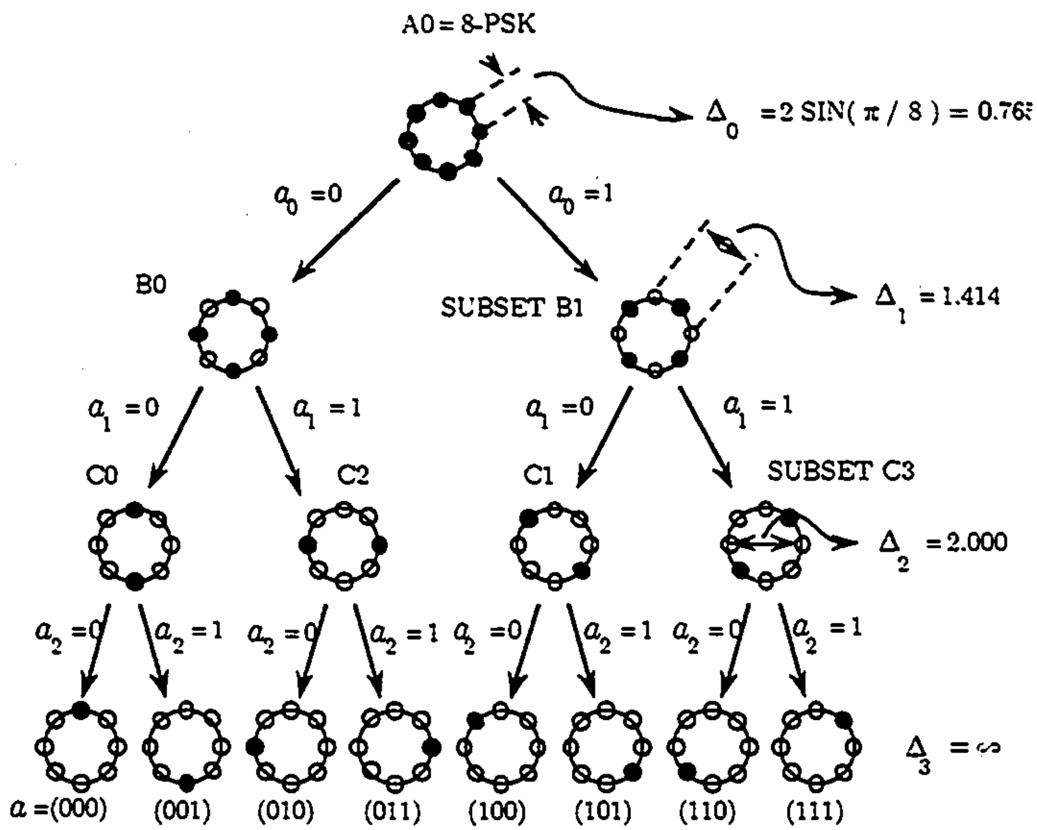


Figure 2.3: Partitioning of 8-PSK channel signals into subsets with increasing minimum subset distances.

Fig. 2.3 shows the partitioning of the 8-PSK signal constellation. Minimum distances between the points in the same subset increase in the downward direction along the partition tree. Every signalling interval, the collection of a certain number of the partitions at the final-end is used as the current signal constellation. The selection of these partitions is done by a finite state encoder in such a way that the minimum distance in the current constellation will be greater than in the unpartitioned original constellation. In this manner, Ungerboeck was able to demonstrate with a simple four-state code a 3 dB improvement in power efficiency over simple PAM, and that a 6 dB improvement could be attained with more complex encoders.

The easy implementation of "mapping by set partitioning" and its significant power gains without bandwidth expansion inspired further developments of similar encoders, most notably the family of *coset codes.* These correspond to signal encoders much like those proposed by Ungerboeck in which the signal constellations and partitions are finite portions of the geometric structures referred to as a lattice and cosets of a lattice [14]. (These terms will be explained in the next section.) In these encoders the output symbols are formed from the data source as follows:

1. The data source output stream is formed into data words of several symbols.

2. A conventional [block or convolutional] encoder operates on a certain number of the symbols of each data word to generate a new stream of redundant codewords.

3. The codewords so produced select one of the partitions of the signal constellation (i.e., one of the cosets).

4. The data symbols not used in the encoder for coset selection are used to select the signal in the coset that will be sent.

We can thus see that a coset code can be basically described by two entities, a lattice partition Λ/Λ' and a conventional code C. We shall denote coset codes with these components by $C(\Lambda/\Lambda', C)$. If the encoder for coset selection is a block code, the resulting coset code is termed a *lattice code,* while if a convolutional code is employed, the result is termed a *trellis code.* Fig. 2.4 shows the structure of an encoder for a binary coset code employing codewords of length n of which k bits are used to encode for coset selection among 2^{k+r} cosets.

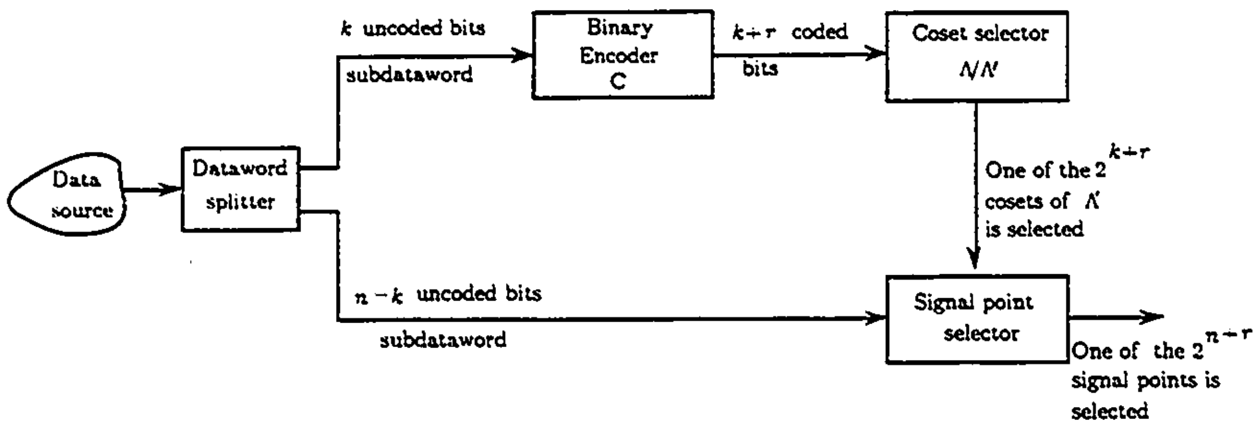


Figure 2.4: General structure of encoder for coset code.

2.1 Lattice Codes

To discuss lattice codes in any detail, we must make use of terminology from the branch of the mathematics known as abstract algebra. We therefore begin this discussion with the definitions of needed terms and concepts.

2.1.1 Lattices

Definition. A *group* $\langle G, +\rangle$ is a set G of objects together with a binary operation $+$ defined on these objects, which satisfy the following properties :

(i) The binary operation $+$ is associative:
if $a,\ b,\ c \in G$ then $(a+b)+c = a+(b+c)$.

(ii) There exists an element e, termed the *identity element* in G such that $e+a = a+e = a$ for all $a \in G$.

(iii) For each a in G, there exists an element a' in G with the property that $a'+a = a+a' = e$, where e represents the identity element. Such an element a' is termed an inverse of a with respect to $+$. ∎

Although a group $\langle G, +\rangle$ is composed of two entities – the set G and the operation $+$ on G – it is common to abuse notation and refer to the group as simply "G", where the operation on G is supposedly understood. For simplicity we shall make use of this notation as well when the binary operation involved is normally obvious (which will usually be some form of "addition" here).

Definition. A group is *Abelian* if its binary operation $+$ is commutative: if $a,\ b, \in G$ then $a + b = b + a$. ∎

Definition. If $\langle G, +\rangle$ is a group, and N is a subset of the set G such that $\langle N, +\rangle$ is itself a group then we say that $\langle N, +\rangle$ is a *subgroup* of $\langle G, +\rangle$ (or simply "N is a subgroup of G"). ∎

Definition. Let $\langle N, +\rangle$ be a subgroup of a group $\langle G, +\rangle$, and let $a \in G$. The set $N_a = \{a + h \mid h \in N\}$ is termed a coset of N, in particular the coset of N generated by the element a. The set of cosets is denoted by $\langle G, +\rangle / \langle N, +\rangle$ or simply G/N. We also denote N_a as well by "$a + N$". ∎

The number of distinct cosets is termed the order of G/N and is denoted by $|G/N|$. Any element that is the generator of a particular coset is a *coset representative* of that particular coset; furthermore the set of all coset representatives of each coset of N is termed the *coset representative set*, denoted by $[G/N]$. Since no restriction is imposed in selecting a representative within the elements of a coset, the coset representative set is not unique.

As an example, consider the set of integers $\mathbb{Z}$ forming a group under ordinary addition. Then, the set of even numbers $\mathbb{Z}_e$ forms a subgroup of $\mathbb{Z}$. The set $\{\cdots, -4+0, -2+0, 0+0, 2+0, 4+0, \cdots\}$ is a coset of $\mathbb{Z}_e$ generated by the element 0. Another set $\{\cdots, -4+1, -2+1, 0+1, 2+1, 4+1, \cdots\}$ forms another coset of $\mathbb{Z}_e$ generated by the element 1, which is the set of odd numbers, $\mathbb{Z}_o$. Furthermore, we see that all other cosets of $\mathbb{Z}_e$ generated by any element in $\mathbb{Z}$ will be identical to one of the above cosets, and so any element of $\mathbb{Z}$ will belong either to $\mathbb{Z}_e$ or $\mathbb{Z}_o$.

Lemma 2.1 Let N be a subgroup of G. A coset of N generated by an element a and that generated by another element b are either identical or have no elements in common. (It should be noted that a given coset can have many generators.)

Proof: Let's assume that c is a common element in two different cosets N_a and N_b of the subgroup N, generated by a and b, respectively such that:

$$c \in N_a \qquad \text{and} \qquad c \in N_b. \tag{2.1}$$

Because of the group properties ,

$$c - a = h_1 \in N, \tag{2.2}$$

$$c - b = h_2 \in N. \tag{2.3}$$

From equations (2.2) and (2.3), we can write

$$a = b + h_2 - h_1. \tag{2.4}$$

Now let $x = a + l$ be any element of N_a, where $l \in N$. Then

$$x = b + h_2 - h_1 + l. \tag{2.5}$$

Since N is a subgroup, $h_2 - h_1 + l \in N$, consequently $x \in N_b$. This shows that every $x \in N_a$ is also in N_b. By a similar argument we can show that for every x, $x \in N_b$, it is also true that $x \in N_a$. This completes the proof that two cosets of N are either disjoint, or identical.

Lemma 2.2 The union of all different cosets of N spans the group G.

Proof: If there exists an element a such that

$$a \in G \qquad \text{and} \qquad a \notin \bigcup_k N_k$$

then $a + N$ defines a new coset of N, such as N_a, and a becomes a member of the union of the cosets of N.

Since cosets are then either distinctly different or identical, and since the union of the cosets is G, the cosets of N in G, G/N, form a particularly natural partition of G with $|G/N|$ elements. Note that in the previous example, the set of integers $\mathbb{Z}$ is partitioned into two distinct cosets $\mathbb{Z}_e$ and $\mathbb{Z}_o$.

Definition. A real *lattice* Λ is a countably infinite set of points in $\mathbb{R}^N$ that forms an Abelian group under ordinary vector addition in $\mathbb{R}^N$. ▮

As an example, the set of all integers $\mathbb{Z}$ is a one-dimensional real lattice; in fact it is the prototype of all the other one dimensional real lattices since they can be generated from $\mathbb{Z}$ simply by scaling. The set $\mathbb{Z}^N$ of all N-tuples whose elements are integers (points in N dimensional space all with integer coordinates) is an N dimensional real lattice for any N. Fig. 2.5 illustrates two lattices in one and two dimensional space.

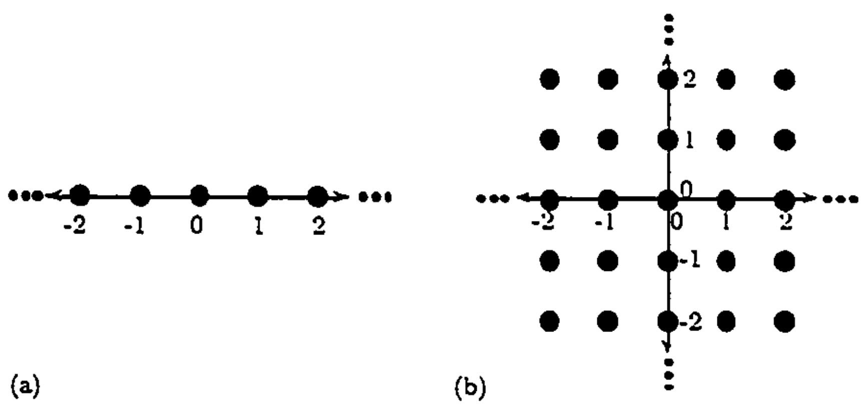


Figure 2.5: The lattices (a) $\mathbb{Z}$; (b) $\mathbb{Z}^2$.

Definition. A *sublattice* Λ' of a lattice Λ is a subcollection of points in Λ that forms a lattice. ▮

If Λ' is a sublattice of Λ, and a vector $\bar{c}_k$ is an element of Λ that is not contained in Λ', then the set formed by adding $\bar{c}_k$ to the elements of Λ' is a subset of Λ, denoted by Λ'_k, where

$$\Lambda'_k \triangleq \left\{ x \,\middle|\, x = \bar{\lambda} + \bar{c}_k \ \text{ for some } \ \bar{\lambda} \in \Lambda' \right\}.$$

By choosing different vectors $\bar{c}_k$, we can generate different subsets. Since each Λ'_k is the version of Λ' shifted by $\bar{c}_k$, the set Λ'_k is termed a *translate* of Λ'. Each of these translates is the coset of the Λ/Λ' partition generated by $\bar{c}_k$.

Every element $\bar{\lambda}$ of Λ can uniquely be written as a sum $\bar{\lambda} = \bar{\lambda}' + \bar{c}_k$, where $\bar{c}_k$ is the representative of the coset Λ'_k, in which $\bar{\lambda}$ lies, and $\bar{\lambda}'$ is an element of the sublattice Λ'. This is called a *coset decomposition* of Λ and is written as

$$\Lambda = \Lambda' +' [\Lambda/\Lambda'], \tag{2.6}$$

where $[\Lambda/\Lambda']$ is the coset representative set of the Λ/Λ' partition. In (2.6) "$+'$" denotes an addition operation defined between two sets such that if S_1 and S_2 represent two sets,

$$S_1 +' S_2 = \{s_1 + s_2 \mid s_1 \in S_1, \quad s_2 \in S_2\}.$$

In Fig. 2.6 a sample Λ/Λ' partitioning is illustrated, where a two dimensional integer lattice $\mathbb{Z}^2$ is partitioned into four cosets of the lattice $2\mathbb{Z}^2$. The latter lattice consists of all points in $\mathbb{R}^2$ whose elements are even integers, and can be seen to be simply the points in $\mathbb{Z}^2$ scaled by a factor 2 (the scaling operator for a real lattice is defined in the next section with other operators). The four coset representatives of the cosets of $2\mathbb{Z}^2$ can be selected as:

$$\bar{c}_0 = (0,0) \ , \bar{c}_1 = (1,0) \ , \ \bar{c}_2 = (0,1) \ , \ \bar{c}_3 = (1,1).$$

Definition. A *partition chain* $\Lambda/\Lambda'/\Lambda''/\cdots/\Lambda^{(n)}$ is a finite sequence of lattices such that each is a sublattice of the previous one, i.e., $\Lambda \supseteq \Lambda' \supseteq \Lambda'' \supseteq \cdots \supseteq \Lambda^{(n)}$. ∎

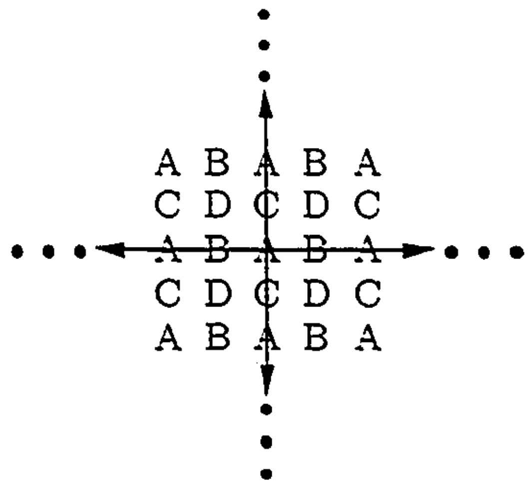


Figure 2.6: An example for $\mathbb{Z}^2/2\mathbb{Z}^2$ partition. Each letter represents a coset of $2\mathbb{Z}^2$.

A partition chain induces a multiterm coset decomposition chain. If $\Lambda/\Lambda'/\Lambda''$ is a partition chain, then the lattice Λ can be written in the form

$$\Lambda = \Lambda'' +' [\Lambda'/\Lambda''] +' [\Lambda/\Lambda']. \tag{2.7}$$

The usefulness of the partition chain derives from the fact that any Λ/Λ'' partitioning can be separated into two partitionings Λ/Λ' and Λ'/Λ'' if there exists a lattice Λ' such that Λ' is a sublattice of Λ and Λ'' is a sublattice of Λ'. Some examples of partition chains are given after the lattice operations are defined.

2.1.2 Lattice Operations

Definition. *Scaling*: A lattice Λ may be scaled by any real number r just by multiplying all the coordinates (N-tuples) of points contained in Λ by r. The resultant set is also a lattice, and is denoted by $r\Lambda$. ∎

As an example, if Λ contains all the points with integer coordinates in real Euclidean N-space, $\mathbb{R}^N$, such that

$$\Lambda = \mathbb{Z}^N,$$

then its scaled by 2 version, denoted by $2\mathbb{Z}^N$, is a lattice which comprises points within $\mathbb{R}^N$ with all even integer coordinates.

Definition. *Transformation:* If $\mathcal{T}$ is a transformation defined in $\mathbb{R}^N$, then $\mathcal{T}\Lambda$ is a lattice, which is generated by applying the vector transformation $\mathcal{T}$ to every element λ in Λ. ∎

One of the most useful transformations for our examples is the *rotation operator* $\mathcal{R}$, defined in $\mathbb{R}^2$ as

$$\mathcal{R} \triangleq \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{2.8}$$

$\mathcal{R}\Lambda$ contains vectors from Λ, that are rotated by 45°, and scaled by $\sqrt{2}$. For the case $\Lambda = \mathbb{Z}^2$, its rotated and scaled version $\mathcal{R}\Lambda$ is shown in Fig. 2.7 with black dots. The set of white dots in the same figure is simply a translated version of $\mathcal{R}\mathbb{Z}^2$ shifted by $\bar{c}_k = (1,0)$. Therefore this set is a coset of $\mathcal{R}\mathbb{Z}^2$ and the partition of Λ by $\mathcal{R}\Lambda$ divides Λ into two cosets. Interestingly, applying the $\mathcal{R}$ transformation once again to $\mathcal{R}\mathbb{Z}^2$, generates $\mathcal{R}^2\mathbb{Z}^2$, which is nothing more than $2\mathbb{Z}^2$. Therefore we conclude that $\mathcal{R}^2$ is equal to $2\mathcal{I}$, where $\mathcal{I}$ represents the identity transformation.

In the higher dimensional space of dimension $2N$, the rotation operator is defined [14] by the matrix

$$\mathcal{R}_{2N} \triangleq \begin{bmatrix} \mathcal{R} & 0 & \cdots & 0 \\ 0 & \mathcal{R} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathcal{R} \end{bmatrix} \tag{2.9}$$

Definition. *Cartesian Product:* The M-fold Cartesian product of the Λ in $\mathbb{R}^N$ with itself—i.e., the set of all MN-tuples $(\lambda_1, \lambda_2, \cdots, \lambda_M)$, where each λ_i is in Λ—is an MN-dimensional lattice denoted by Λ^M. ∎

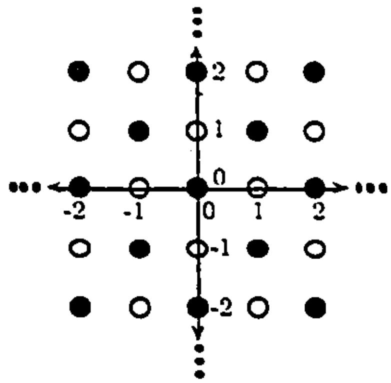


Figure 2.7: $\mathbb{Z}^2/\mathcal{R}\mathbb{Z}^2$ partitioning.

2.1.3 Geometric Properties of Lattices

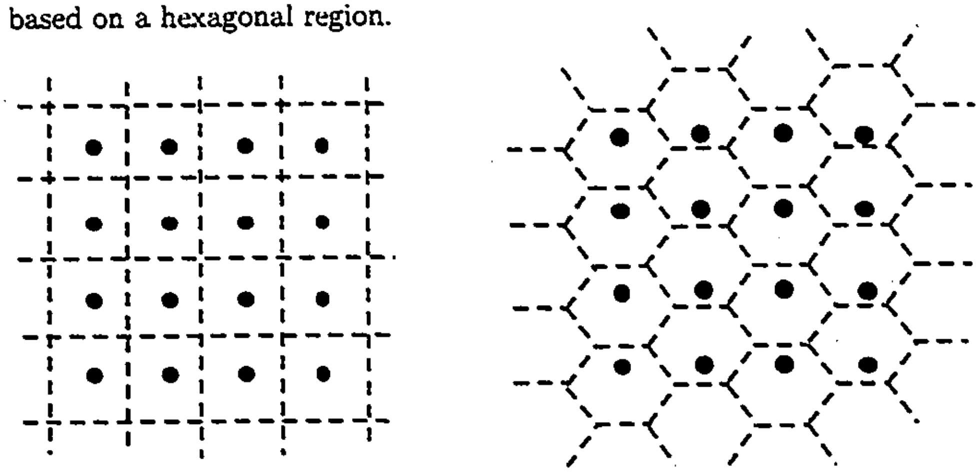
The geometric properties of a real lattice Λ arise from the geometry of the real N-dimensional Euclidean vector space $\mathbb{R}^N$. There are two important parameters of lattices that indicate the suitability of a lattice, Λ, for communications purposes: the *minimum squared distance* $d^2_{\min}(\Lambda)$ and the *fundamental volume* $V(\Lambda)$ of Λ. These two parameters convey enough information to determine the theoretical coding gain or lattice gain, $\gamma_l(\Lambda)$. In the following section, these parametersare defined along with other parameters that will be useful later in the design of good shaping codes.

Definition. For a lattice Λ, the *minimum squared distance* $d^2_{\min}(\Lambda)$ of the lattice is the least squared distance between two distinct points in the lattice:

$$d^2_{\min}(\Lambda) \triangleq \min_{\substack{c_i, c_j \in \Lambda \\ c_i \neq c_j}} \|\vec{c}_i - \vec{c}_j\|^2. \qquad \blacksquare \tag{2.10}$$

Associated with every lattice Λ in a real Euclidean N-space $\mathbb{R}^N$, we can define a region $R(\Lambda)$ with the properties that the union of regions $R(\Lambda) + \vec{x}$ for all $\vec{x} \in \Lambda$ is the entire vector space $\mathbb{R}^N$, and the regions $R(\Lambda) + \vec{x}$ for different $\vec{x}$ are disjoint. Such a region $R(\Lambda)$ is termed a *fundamental region.* Clearly, a fundamental region $R(\Lambda)$

contains only one element of Λ. Fundamental regions are not unique however. The regions $R(\Lambda) + \bar{x}$ for different $\bar{x}$ in Λ form a partition for $\mathbb{R}^N$. Fig. 2.8 shows two partitions of $\mathbb{R}^2$ for the lattice $\mathbb{Z}^2$, one based on a rectangular fundamental region, and one based on a hexagonal region.



(a) (b)

Figure 2.8: Two partitions of $\mathbb{R}^2$ for $\mathbb{Z}^2$ based on two different fundamental regions $R(\mathbb{Z}^2)$: (a) rectangular; (b) hexagonal. Both fundamental regions have a volume equal to 1.

Lemma 2.3 Let Λ' be a sublattice of Λ. The union of $R(\Lambda) + \bar{c}_k$, for all $\bar{c}_k$ in $[\Lambda/\Lambda']$ is a fundamental region $R(\Lambda')$ of Λ'.

Proof. From the definition of the fundamental region we can write

$$\bigcup_{\bar{\lambda}\in\Lambda} \{R(\Lambda) + \bar{\lambda}\} \quad = \quad \mathbb{R}^N. \tag{2.11}$$

Using the coset decomposition, we can state that also as

$$\bigcup_{\bar{\lambda}\in[\Lambda/\Lambda']+\Lambda'} \{R(\Lambda) + \lambda\} = \bigcup_{\bar{\lambda}'\in\Lambda'} \left(\bigcup_{\substack{\bar{\lambda}=\bar{\lambda}'+\bar{c}_k \\ \bar{c}_k\in[\Lambda/\Lambda']}} \{R(\Lambda) + \bar{\lambda}\} \right) \quad = \quad \mathbb{R}^N. \tag{2.12}$$

Here, it is not hard to verify that the region described in the large parenthesis is a fundamental region of Λ' translated by the vector $\bar{\lambda} \in \Lambda'$. Since $\lambda' + [\Lambda/\Lambda']$ is another coset representative set of Λ/Λ' partition we can terminate our proof:

$$R(\Lambda') = \bigcup_{\substack{\bar{\lambda} = \bar{c}_k \\ \bar{c}_k \in [\Lambda/\Lambda']}} \{R(\Lambda) + \bar{\lambda}\} = \bigcup_{\bar{c}_k \in [\Lambda/\Lambda']} \{R(\Lambda) + \bar{c}_k\}. \tag{2.13}$$

The volume of a fundamental region $R(\Lambda)$ of a lattice Λ is called the *fundamental volume* $V(\Lambda)$ of the lattice Λ. The reciprocal of $V(\Lambda)$ is termed the density of the lattice Λ or the volume per lattice point. Clearly the density of a lattice or the fundamental volume of the lattice does not depend on the choice of a fundamental region. (See Fig. 2.8).

Recalling the definition of the fundamental volume $V(\Lambda)$, and Lemma 2.3, we claim that if Λ' is a sublattice of Λ, inducing a Λ/Λ' partitioning, then

$$V(\Lambda') = |\Lambda/\Lambda'|\, V(\Lambda). \tag{2.14}$$

Also knowing that the fundamental volume of the integer lattice $\mathbb{Z}^N$ is always equal to 1, for Λ' a sublattice of $\mathbb{Z}^N$,

$$V(\Lambda') = |\mathbb{Z}^N/\Lambda'|. \tag{2.15}$$

From the communications engineering point of view, lattices are significant because almost all the signal constellations utilized with different modulation systems are based on a finite portion of some kind of lattice. Systems using the same symbol alphabet but different constellations based on different lattices result in different performances with each constellation. For two different lattices, Λ_1 and Λ_2, the ratio of the transmitted power to achieve a given error rate using a constellation based on Λ_1, to the transmitted power required to achieve the same error rate using a constellation based on Λ_2 (in the limit of lower and lower error rates) is defined as the theoretical coding gain or the *lattice gain* of the lattice Λ_1 over Λ_2. It can be shown that this is given by

$$\gamma_l(\Lambda_1 : \Lambda_2) = \frac{d^2_{\min}(\Lambda_1)}{d^2_{\min}(\Lambda_2)} \left(\frac{V(\Lambda_2)}{V(\Lambda_1)}\right)^{2/N}, \tag{2.16}$$

assuming that both constellations are bounded by regions with the same general shape, and they are large enough so that average power can be calculated by the continuous approximation method (Appendix A) accurately enough [22]. Usually the lattice gain of a lattice Λ is compared with a base constellation, which has traditionally been $\mathbb{Z}^N$. Therefore the lattice gain of a lattice Λ is defined as

$$\gamma_l(\Lambda) = \gamma_l(\Lambda : \mathbb{Z}^N) = \frac{d^2_{\min}(\Lambda)}{(V(\Lambda))^{2/N}}. \tag{2.17}$$

Equation (2.17) is derived from (2.16) using the fact that

$$V(\mathbb{Z}^N) = 1 \qquad \text{, and} \qquad d^2_{\min}(\mathbb{Z}^N) = 1. \tag{2.18}$$

Consider an example communications system where the signal points are selected within a lattice $2\mathcal{R}\mathbb{Z}^2$, where $\mathcal{R}$ denotes the rotation operator defined earlier. Considering the $\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partition, we can convert this partition into the $\mathbb{Z}^2/\mathcal{R}\mathbb{Z}^2/2\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partition chain, since every lattice on this chain is a sublattice of the previous one. We also notice that every step partition is a $\Lambda/\mathcal{R}\Lambda$ partition (since we showed earlier that $\mathcal{R}^2 = 2\mathcal{I}$). Furthermore, at every step Λ is partitioned in two cosets of Λ', therefore the order of $\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$, $|\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2| = 2 \times 2 \times 2 = 8$. Consequently, from (2.15) the fundamental volume $V(2\mathcal{R}\mathbb{Z}^2) = 8$. Since scaling a lattice Λ by a real number r increases the minimum squared distance $d^2_{\min}(\Lambda)$ by r^2, and rotation operator $\mathcal{R}$ consists of a scaling factor $r_{\mathcal{R}} = \sqrt{2}$, the minimum squared distance of the lattice $2\mathcal{R}\mathbb{Z}^2$

$$d^2_{\min}(2\mathcal{R}\mathbb{Z}^2) = (2 \times \sqrt{2})^2 = 8.$$

By using (2.17), the lattice gain is

$$\gamma_l(2\mathcal{R}\mathbb{Z}^2) = \gamma_l(2\mathcal{R}\mathbb{Z}^2 : \mathbb{Z}^2) = \frac{d^2_{\min}(2\mathcal{R}\mathbb{Z}^2)}{V(2\mathcal{R}\mathbb{Z}^2)^{2/2}} = \frac{8}{8} = 1$$

in $\mathbb{R}^2$. In other words no lattice gain is obtained. This result is expected since $2\mathcal{R}\mathbb{Z}^2$ is only a scaled and rotated version of $\mathbb{Z}^2$, and as proved in [14] lattice gain is invariant to scaling and applying transformations.

Another lattice referred to as the Schlafli lattice, D_4, is known to be a more efficient lattice than the integer lattice $\mathbb{Z}^4$ in $\mathbb{R}^4$. D_4 is the set of vectors in $\mathbb{Z}^4$ with an even number of odd valued coordinates. For example, a vector $\vec{a}$ with all integer coordinates in $\mathbb{R}^4$ such as $\vec{a} = (2,3,5,6)$ is an element of both $\mathbb{Z}^4$ and D_4 (there are two odd valued coordinates), but the vector $\vec{b} = \vec{a} + (1,0,0,0) = (3,3,5,6)$ is not contained in D_4. Since D_4 is a sublattice of $\mathbb{Z}^4$, it induces a partitioning on the integer lattice. The order of $\mathbb{Z}^4/D_4$ is 2 because all the points contained in $\mathbb{Z}^4$ have either an even number of odd valued coordinates or an odd number of odd valued coordinates. Furthermore the two simplest coset representatives can be selected as

$$\begin{aligned} c_0 &= (0,0,0,0), \\ c_1 &= (1,0,0,0). \end{aligned}$$

The minimum distance of D_4, $d^2_{\min}(D_4) = 2$, since the coordinates of two different points contained in D_4 must differ at least at two coordinates by 1 or by 2 at at least one coordinate. The fundamental volume is also calculated equal to 2 from (2.15). The lattice gain of D_4 is then,

$$\gamma_l(D_4) = \frac{d^2_{\min}(D_4)}{(V(D_4))^{2/4}} = \frac{2}{2^{2/4}} = 1.51\,\text{dB}. \tag{2.19}$$

If we scale D_4 to have the same minimum distance as $\mathbb{Z}^4$, (so that the performance of constellations in the same), we see that the scaled D_4 is a denser lattice then $\mathbb{Z}^4$. Thus selecting the signal points from D_4 represents an 1.51 dB power efficiency improvement over the ordinary uncoded PAM system. One point that should be kept in mind is that as the lattice gets denser, the number of nearest neighbours, i.e., the number of lattice points located a distance $d^2_{\min}(\Lambda)$ away from a lattice point, increases. This fact may cause significant degradation from the theoretical power gain calculations.

2.1.4 Binary Lattices and their Labelings of Partitions

Definition. Integer lattices (i.e., sublattices of $\mathbb{Z}^N$ in $\mathbb{R}^N$) that contain sublattices in the format of $2^m\mathbb{Z}^N$ in $\mathbb{R}^N$ for some integer m are termed *binary lattices*. The least value for m for which $2^m\mathbb{Z}^N$ is a sublattice of the binary lattice is termed the 2-depth of the binary lattice. ∎

Binary lattices are one of the most efficient and commonly used lattices in digital communications applications because of their high compatibility to the bit-oriented real world. The densest lattices known in 1, 4, 8, 16 and 24 dimensional space belong to the class of binary lattices [14].

The lattice $\mathcal{R}\mathbb{Z}^2$ serves as a good example of a binary lattice. As we discussed in the previous section, $\mathcal{R}\mathbb{Z}^2$ is a sublattice of $\mathbb{Z}^2$ and therefore an integer lattice, and it can be partitioned in the cosets of $2\mathbb{Z}^2$ such that the $\mathbb{Z}^2/\mathcal{R}\mathbb{Z}^2/2\mathbb{Z}^2$ partition chain exists. The existence of this chain implies that $\mathcal{R}\mathbb{Z}^2$ contains a sublattice in the format of $2^m\mathbb{Z}^N$, where $m = 1$. Therefore $\mathcal{R}\mathbb{Z}^2$ is a binary lattice. Clearly, since $2^0\mathbb{Z}^2 = \mathbb{Z}^2$ is not a sublattice of $\mathcal{R}\mathbb{Z}^2$ the 2-depth of $\mathcal{R}\mathbb{Z}^2$ is 1.

As a consequence of the definition of the binary lattices, we assert that if Λ a binary lattice, there exists a partition chain $\mathbb{Z}^N/\Lambda/2^m\mathbb{Z}^N$ for some integer m. The order of $\mathbb{Z}^N/2^m\mathbb{Z}^N$ is found to be 2^{mN}. Thus

$$|\mathbb{Z}^N/2^m\mathbb{Z}^N| = |\mathbb{Z}^N/\Lambda| \times |\Lambda/2^m\mathbb{Z}^N| = 2^{mN}. \tag{2.20}$$

Since the two factors in (2.20) are integers, they both have to be a power of 2. More generally, if Λ and Λ' are two binary lattices with 2-depths m and m', respectively; and Λ' is a sublattice of Λ, then the following partitionings,

$$\Lambda/\Lambda', \tag{2.21}$$

$$\mathbb{Z}^N/\Lambda/2^m\mathbb{Z}^N, \tag{2.22}$$

$$\mathbb{Z}^N/\Lambda'/2^{m'}\mathbb{Z}^N, \tag{2.23}$$

exist, from which we conclude that

$$\mathbb{Z}^N/\Lambda/\Lambda'/2^{mm'}\mathbb{Z}^N \tag{2.24}$$

also exists since $2^{mm'}\mathbb{Z}^N$ is a sublattice of $2^m\mathbb{Z}^N$ and of $2^{m'}\mathbb{Z}^N$. Since the order of the partitions at every step of this chain is a power of 2, the order of Λ/Λ',

$$|\Lambda/\Lambda'| = 2^K, \tag{2.25}$$

for some integer K.

Lemma 2.4 Let Λ and Λ' be binary lattices where $\Lambda' \subseteq \Lambda$, and $|\Lambda/\Lambda'| = 2^K$. There can be found a sequence of lattices $\Lambda_0 = \Lambda$, $\Lambda_1, \cdots, \Lambda_K = \Lambda'$ such that $\Lambda_0/\Lambda_1/\cdots/\Lambda_K$ is a lattice partition chain and at every step k of this chain the lattice Λ_k is divided into two cosets of Λ_{k+1}, $0 \leq k \leq K-1$. Then Λ has the coset decomposition

$$\Lambda = \Lambda' + \left\{\sum a_k g_k\right\}, \tag{2.26}$$

where $a_k \in \{0,1\}$ and $\{a_k g_k,\ a_k \in \{0,1\}\}$ is a set of coset representatives for $[\Lambda_k/\Lambda_{k+1}]$, $0 \leq k \leq K-1$.

The proof of Lemma 2.4 is given in [14], [15], and is not important to this thesis so is not repeated here.

According to Lemma 2.4 any defined partition of a binary lattice Λ into the cosets of another binary lattice Λ' can be broken up into K step two-way partitions Λ_k/Λ_{k+1}, $0 \leq k \leq K-1$; and there is a labelling of the coset representatives of the cosets of Λ such that

$$c(\bar{a}) = \sum_{0 \leq k \leq K-1} a_k g_k, \tag{2.27}$$

where a_k represents the k^{th} coordinate of the vector $\bar{a}$, and g_k is an element of Λ_k but not Λ_{k+1} such that $a_k g_k$ ($a_k \in \{0,1\}$) and g_k are the two elements of the coset representative set corresponding to Λ_k/Λ_{k+1} partitioning.

The described partition chain of binary lattices can be more easily understood by a tree diagram as shown in Fig. 2.9. At every level k, the coset representatives are actually the linear combinations of the $g_0, g_1, \cdots, g_k$ nonzero coset representatives of the $\Lambda_0/\Lambda_1, \Lambda_1/\Lambda_2, \cdots, \Lambda_k/\Lambda_{k+1}$ partitionings.

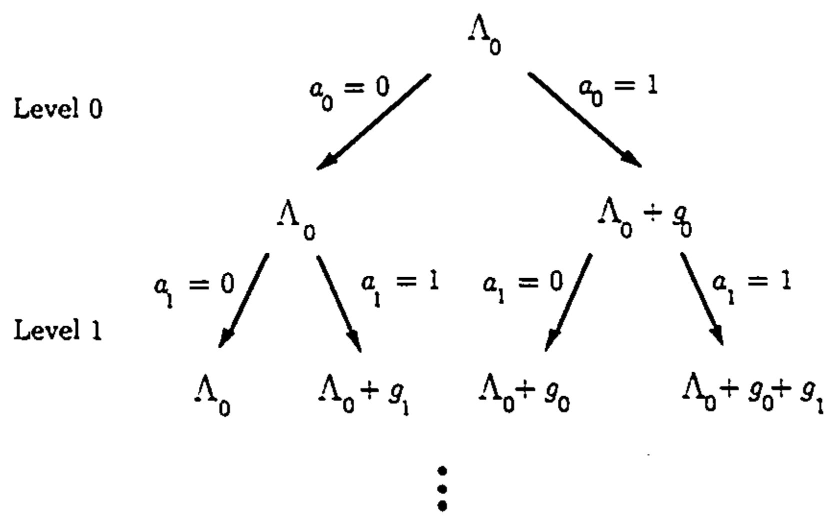


Figure 2.9: Ungerboeck labeling by partition tree.

This way of labeling of Λ/Λ' binary lattice partition has a unique result: If λ_1 and λ_2 are two points in Λ, which reside in two different cosets of Λ', and have labels which agree in their first l bits, but not in $(l+1)^{th}$ bit, then the squared distance between λ_1 and λ_2 is at least $d^2_{min}(\Lambda_L)$.

Suppose we are asked to partition $\mathbb{Z}^2$ into the cosets of $2\mathcal{R}\mathbb{Z}^2$. We can easily show that both lattices are binary lattices, which enables as to use the labeling method discussed earlier. We can break the $\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partition into $\mathbb{Z}^2/\mathcal{R}\mathbb{Z}^2, \mathcal{R}\mathbb{Z}^2/2\mathbb{Z}^2, 2\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partitions with nonzero coset representatives $g_0 = (1,0), g_1 = (1,1), g_2 = (-2,0)$, respectively. The 8 coset representatives of $\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ $c(\bar{a}) = c(a_0, a_1, a_2)$ are the linear combinations of $g_0, g_1, \cdots, g_k$ such that

$$\begin{aligned}
c(0,0,0) &= 0g_0 + 0g_1 + 0g_2 = (\ \ 0,0) \\
c(0,0,1) &= 0g_0 + 0g_1 + 1g_2 = (-2,0) \\
c(0,1,0) &= 0g_0 + 1g_1 + 0g_2 = (\ \ 1,1) \\
c(0,1,1) &= 0g_0 + 1g_1 + 1g_2 = (-1,1) \\
c(1,0,0) &= 1g_0 + 0g_1 + 0g_2 = (\ \ 1,0)
\end{aligned}$$

$$
\begin{aligned}
c(1,0,1) &= 1g_0 + 0g_1 + 1g_2 = (-1,0) \\
c(1,1,0) &= 1g_0 + 1g_1 + 0g_2 = (\ \ 2,1) \\
c(1,1,1) &= 1g_0 + 1g_1 + 1g_2 = (\ \ 0,1)
\end{aligned}
$$

2.1.5 Examples for Decomposition of Binary Lattices

In this section we will illustrate the relation between mod-2 binary lattices and binary block codes and show how to construct these lattices using binary block codes.

Definition. A *mod-2 lattice* is a binary lattice whose 2-depth is 1. This implies that $2\mathbb{Z}^N$ is a sublattice of all mod-2 binary lattices. ∎

Lemma 2.5 If Λ is a binary mod-2 lattice, then there exist a coset representative set corresponding to the $\mathbb{Z}^N/\Lambda$ partition, where each coset representative $\bar{c}_k$ of the coset of Λ is binary, i.e. N tuple of zeros and ones.

Proof. Consider the lattice $2\mathbb{Z}^N$ which is a sublattice of the binary mod-2 lattice Λ. $2\mathbb{Z}^N$ is the collection of all the N-tuples with all even valued coordinates. If $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are two vectors contained in the same coset of $2\mathbb{Z}^N$ then the coordinates of the difference vector $\bar{\lambda}_1 - \bar{\lambda}_2$ will also consists of all even valued coordinates; conversely, if the difference vector $\bar{\lambda}_1 - \bar{\lambda}_2$ of two vectors $\bar{\lambda}_1$ and $\bar{\lambda}_2$ in $\mathbb{Z}^N$ forms an N-tuple with all even valued numbers then $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are in the same coset of $2\mathbb{Z}^N$. For example $\bar{\lambda}_1 = (3,2,1,6)$ and $\bar{\lambda}_2 = (1,4,5,6)$ are in the same coset of $2\mathbb{Z}^N$ since $\bar{\lambda}_1 - \bar{\lambda}_2 = (2,-2,-4,0)$ consists of all even valued coordinates. Consequently a coset of $2\mathbb{Z}^N$ consists of points whose corresponding coordinates are congruent to each other modulo-2. The equivalence classes of modulo-2 arithmetic can be represented by "zero" or "one". If $\bar{\lambda}$ is any element of a coset of $\mathbb{Z}^N$ then simply by replacing every coordinate of $\bar{\lambda}$ by its equivalent "0" or "1" in modulo-2, we can find a new point with all binary coordinates in the same coset of $2\mathbb{Z}^N$ as $\bar{\lambda}$. By letting this binary vector represent the coset, and by repeating this procedure for all the cosets of $2\mathbb{Z}^N$ we can construct an all binary vector set for $[\mathbb{Z}^N/2\mathbb{Z}^N]$. Also, every coset of Λ will be a collection of some number of cosets of $2\mathbb{Z}^N$, since $2\mathbb{Z}^N$ is a sublattice of Λ. Therefore the cosets of Λ will contain some number of binary coset representatives of the cosets of $2\mathbb{Z}^2$. By selecting one of these binary elements as the representative of the coset of Λ, we can construct $[\mathbb{Z}^N/\Lambda]$ from all binary vectors.

Definition. An (n, k) *binary linear block code* is the set of 2^k binary code words of length n, which form a k-dimensional subspace of the vector space of all binary length n words over the field $GF(2)$. Every codeword $c(a)$ can be expressed as a linear combination of k linearly independent codewords g_l's such that

$$c(a) = \sum_{l=0}^{k-1} a_l g_l \tag{2.28}$$

where $a = (a_0, a_1, \cdots, a_{k-1})$ is the k tuple information word and $a_l \in \{0, 1\}$ for $0 \leq l \leq k-1$. ∎

Lemma 2.6 A N-dimensional real lattice Λ is a mod-2 binary lattice if and only if it is the set of all integer N-tuples that are congruent modulo-2 to one of the codewords $c(a)$ in a linear binary (n, k) block code C, where $N = n$. The minimum distance of Λ $d^2_{\min}(\Lambda) = \min\{4, d_H(\mathrm{C})\}$, where $d_H(\mathrm{C})$ is the minimum Hamming distance of the block code C.

Proof. Assuming that Λ is a binary mod-2 lattice,$|\Lambda/2\mathbb{Z}^2| = 2^K$ for some integer K while $|\mathbb{Z}^N/2\mathbb{Z}^N| = 2^N$. By Lemma 2.4, the $\Lambda/2\mathbb{Z}^N$ partition can be broken up into a partition chain, where every lattice in this chain is a binary mod-2 lattice. Again by Lemma 2.4 Λ has a coset decomposition

$$\Lambda = 2\mathbb{Z}^N +' \left\{\sum a_k g_k\right\} \tag{2.29}$$

where $a_k \in \{0, 1\}$, and g_k is the nonzero coset representative of every step partition of the partition chain as described in Lemma 2.4. Lemma 2.5 enables us to select all g_k's from within the N-tuple binary vector set. Since every vector in $2\mathbb{Z}^N$ is congruent to $\bar{0}$, according to (2.29) Λ is congruent to $\{\sum a_k g_k\}$ modulo-2, where it is a linear (N, K)

block code. Conversely, if $\{\sum a_k g_k\}$ is such a code, then the set of all integer N-tuples that are congruent modulo-2 to codewords of the block code is clearly a lattice, with $2\mathbb{Z}^N$ as a sublattice. The Hamming distance $d_H(\mathrm{C})$ of the code C reveals that every codeword is different at least at $d_H(\mathrm{C})$ digits from other codewords. Therefore every N-tuple in Λ differs at least at $d_H(\mathrm{C})$ digits by 1 or at one digit by 2 from any other point in Λ. Consequently, the minimum squared distance $d^2_{\min}(\Lambda) = \min\{d_H(\mathrm{C}), 4\}$.

A mod-2 binary lattice Λ is the union of 2^K cosets of $2\mathbb{Z}^N$ out of 2^N available such cosets in $\mathbb{Z}^N$. According to this fact and the discussions on partitioning and labelling of binary lattices, a binary mod-2 lattice can be constructed from the cosets of $2\mathbb{Z}^N$ as shown in Fig2.10. A K-tuple information word enters into a block code, whose 2^K codewords are the labels of the coset representatives of the 2^K cosets of $2\mathbb{Z}^N$,which are contained in the binary mod-2 lattice that we wish to construct.

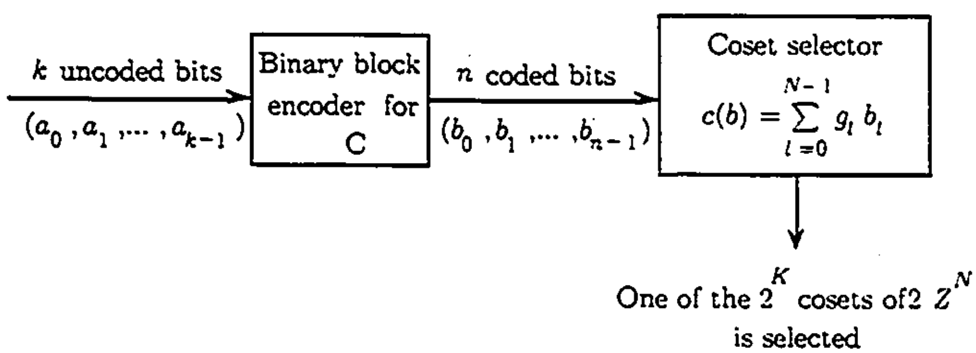


Figure 2.10: Illustration of modulo-2 binary lattice as union of 2^K cosets of $2\mathbb{Z}^N$.

As we see, binary mod-2 lattices and binary block codes with a minimum Hamming distance less than or equal to four are isomorphic. In fact, binary block codes with greater Hamming distances are isomorphic to other lattices such as Reed-Muller code to the Gosset lattice E_8 [22]. Considering the lattice codes such as in Fig. 2.10, the binary block encoder does not introduce redundancy between the signal points in the output signal sequence; therefore this nonredundant transmission is only reluctantly termed a "code" by some mathematicians [22]. Nevertheless, for engineers a lattice "code" is very useful for improving the power efficiency of transmission over uncoded ordinary PAM transmission on Gaussian channels.

2.2 Trellis Codes

Trellis codes form a subclass of coset codes $C(\Lambda/\Lambda', \mathrm{C})$, generated by the system shown in Fig. 2.4, where C represents a binary convolutional code [31]. Usually, Λ and Λ' are selected from the class of binary lattices because of the suitability in the bit oriented real world. Basically, a trellis code can be taken as the combination of a binary convolutional code and a mapping function defined from the output words of the convolutional encoder to the coset representatives of Λ' in Λ.

Convolutional codes are time-invariant codes, as are trellis codes. Trellis codes, however are not necessarily linear, while convolutional codes are always linear. The linearity of a trellis code depends on the linearity of the mapping function utilized. In the following sections of this thesis, only linear trellis codes will be considered. According to Fig. 2.4, at every signalling interval, for every k-tuple information word, the convolutional encoder generates a $(k+r)$-tuple binary codeword $\bar{a}$. Each of these 2^{k+r} codewords are mapped into one of the 2^{k+r} coset representatives $c(\bar{a})$ of the cosets of Λ' in Λ. Finally the uncoded $n-k$ bits determine the signal point within the coset represented by $c(\bar{a})$ to be transmitted over the channel. A trellis code $C(\Lambda/\Lambda', \mathrm{C})$ therefore consists of the sequences of elements of Λ that are congruent to some coset representative sequence $(c(\bar{a}_{t+1}), c(\bar{a}_{t+2}), \cdots)$ modulo Λ', where $a_{t+1}, a_{t+2}, \cdots$ is a *legitimate code sequence* of C, i.e., a binary word sequence that can be generated at the output of a convolutional encoder.

A trellis diagram for a 2^{ν}-state, rate-$k/(k+r)$ convolutional code is an extended state transition diagram for the encoder that generates C. For each symbol interval the diagram has 2^{ν} nodes representing each possible state of the encoder. The nodes of the consecutive time intervals are connected to each other by branches if the encoder may change its state from that corresponding to the first node to state corresponding to the second node. Since there are 2^k possible inputs and hence 2^k possible preceding or following states, there are 2^k branches leaving and entering each node and each branch can be labeled with the $(k + r)$-tuple $\bar{a}$ that represents the encoder output associated with the state transition. By replacing the labels $\bar{a}$ of the branches of the trellis diagram of the convolutional code C with the corresponding coset representative $c(\bar{a})$, the trellis diagram for the trellis code is obtained.

For example, a trellis code can be designed by using a convolutional encoder such that the generator matrix of the code is given as $G = [1 + D^2, 1 + D + D^2]$. Together with this code, the $\mathbb{Z}^2/2\mathbb{Z}^2$ partition characterizes the trellis code whose block diagram is given in Fig. 2.11. As seen before, the $\mathbb{Z}^2/2\mathbb{Z}^2$ partition may be composed into two way partitions such as $\mathbb{Z}^2/\mathcal{R}\mathbb{Z}^2$ and $\mathcal{R}\mathbb{Z}^2/2\mathbb{Z}^2$, which allows one to utilize a binary labelling method. Let $g_0 = (1, 0)$ and $g_1 = (1, 1)$ be the nonzero coset

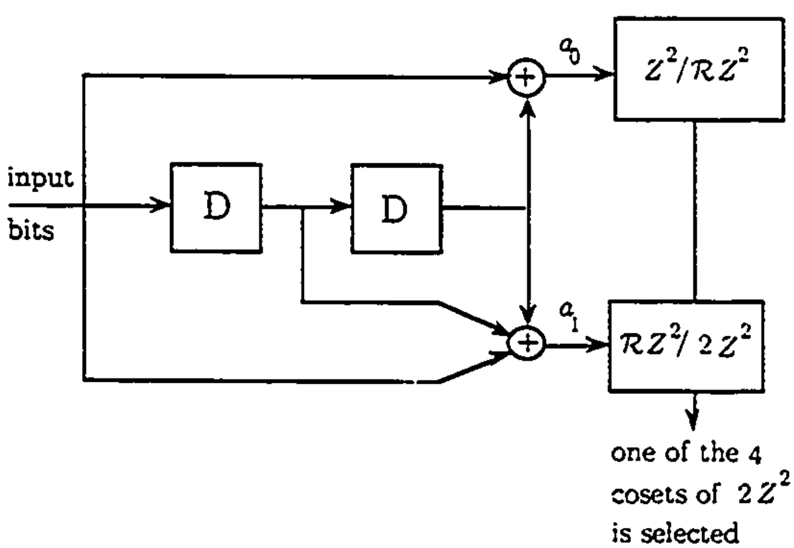


Figure 2.11: An encoder for $C(\mathbb{Z}^2/2\mathbb{Z}^2, C)$ trellis code, where C is characterized by its generator matrix $G = [1 + D^2, 1 + D + D^2]$.

representatives of $\mathbb{Z}^2/\mathcal{R}\mathbb{Z}^2$ and $\mathcal{R}\mathbb{Z}^2/2\mathbb{Z}^2$ partitions, respectively. The resultant four coset representatives $c(\vec{a}) = c(a_0, a_1)$ can be written as

$$c(a_0, a_1) = a_0 g_0 + a_1 g_1. \tag{2.30}$$

Finally, Fig. 2.12 illustrates the trellis diagram of the designed trellis code along with the trellis diagram of the convolutional code used used in the trellis code.

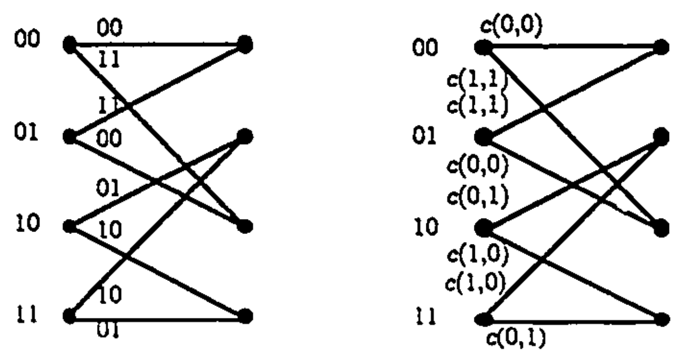


Figure 2.12: Trellis diagram of the convolutional code C and that of the trellis code $\mathcal{C}(\mathbb{Z}^2/2\mathbb{Z}^2, \mathrm{C})$.

In a trellis code $\mathcal{C}(\Lambda/\Lambda', \mathrm{C})$, Λ can simply be the integer lattice $\mathbb{Z}^N$. In this case the reduction in power requirements for transmission compared to uncoded ordinary PAM transmission is obtained by introducing redundancy between successive symbols in the output symbol sequence. This gain is the coding gain $\gamma_c(\mathbb{Z}^N/\Lambda', C)$ of the trellis code. If any lattice other then $\mathbb{Z}^N$ is utilized as Λ, an additional gain is obtained due to selecting the signal points within a denser structure. In the previous section, we defined this additional gain as lattice gain $\gamma_l(\Lambda)$. Consequently, the total gain obtained by the trellis code $\mathcal{C}(\Lambda/\Lambda', \mathrm{C})$ is the product of these two separate gains:

$$\gamma_{c,t}(\Lambda/\Lambda', \mathrm{C}) = \gamma_c(\mathbb{Z}^N/\Lambda', \mathrm{C})\gamma_l(\Lambda), \tag{2.31}$$

where the coding gain of the trellis code is

$$\gamma_c(\mathbb{Z}^N/\Lambda', \mathrm{C}) = \frac{d^2_{\min}(\text{coded})}{d^2_{\min}(\text{uncoded})} \left(\frac{V(R_{B,\text{uncoded}})}{V(R_{B,\text{coded}})}\right)^{2/N}, \tag{2.32}$$

assuming that for both coded and uncoded transmission (the coded transmission is a *coset* trellis code, and the uncoded transmission is the case where the convolutional encoder in Fig. 2.4 is absent so that the information bits are directly mapped to the coset representatives of the 2^k cosets of Λ'), signal constellations (i.e., the domain of the channel symbols) are bounded by the regions $R_{B,\text{coded}}$ and $R_{B,\text{uncoded}}$ with the same shape. Coding introduces redundancy into the transmission, and therefore causes a decrease in the amount of actual information transmitted. The decrease in the information transmitted can be compensatedfor by an expansion in signal constellation. The ratio $V(R_{B,\text{coded}})/V(R_{B,\text{uncoded}})$ represents the constellation expansion ratio due to coding, and is approximately equal to 2^r, where r is the number of redundant bits introduced into the system by the convolutional encoder. The minimum squared distance $d^2_{\min}$(uncoded) is simply the separation of the signal points in the lattice Λ, $d^2_{\min}(\text{uncoded}) = d^2_{\min}(\Lambda)$; $d^2_{\min}$(coded) represents the minimum squared Euclidean distance between the output symbol sequences of the trellis code, and determined by the minimum squared distance $d^2_{\min}(\Lambda')$, and by the method of selecting the cosets. Before we give its formal definition we introduce two new concepts:

Definition. The *norm* of a coset is the minimum norm of the points contained in that coset. ∎

Definition. The *minimum distance between sequences* is the summation of the differences of the norms of the two corresponding elements in each sequence. ∎

The minimum squared distance of the output sequences of a trellis code

$$d^2_{\min}(\text{coded}) \triangleq \min\left[d^2_{\min}(\Lambda'), \|\lambda_1(D) - \lambda_2(D)\|^2\right], \tag{2.33}$$

where $\|\lambda_1(D) - \lambda_2(D)\|^2$ is the minimum squared distance between two distinct paths $\lambda_1(D)$ and $\lambda_2(D)$, starting and ending in a common state. The elements of $\lambda_1(D)$ and $\lambda_2(D)$ are the cosets of Λ'.

Let's calculate the minimum distance $d^2_{\min}$(coded) of the trellis code designed in the previous example. The norms of the four cosets of $2\mathbb{Z}^2$ in $\mathbb{Z}^2$, the coset representatives and the assignment of the output words of the convolutional encoder to each coset are presented in Table 2.1.

Table 2.1: Coset representatives names and norms of the four cosets of $2\mathbb{Z}^2$ in $\mathbb{Z}^2$.

a_0, a_1	coset rep. $c(a)$	sym. of cosets	$\mathcal{N}$[coset]
$0,0$	$(0,0)$	A	0
$0,1$	$(1,1)$	B	2
$1,0$	$(1,0)$	C	1
$1,1$	$(2,1)$	D	1

Fig. 2.13 illustrates two typical paths with a distance between them equal to $d^2_{\min}$(coded), and also shows the elements of the coset representatives by illustrating the four cosets of $2\mathbb{Z}^2$. The minimum distance of the coded transmission

$$\|\lambda_1(D) - \lambda_2(D)\|^2 = \mathcal{N}[2,1] - \mathcal{N}[0,0] + \mathcal{N}[1,1] - \mathcal{N}[0,0] + \mathcal{N}[2,1] - \mathcal{N}[0,0] = 4.$$

Also the minimum squared distance of lattice $2\mathbb{Z}^2$ is $d^2_{\min}(2\mathbb{Z}^2) = 4$. Therefore the minimum distance of the trellis code is

$$d^2_{\min}(\text{coded}) = \min[4,4] = 4$$

In this example one of the paths shown in Fig. 2.13 just happened to be the all zero path. As for convolutional codes, if the distribution of the distances from a given code sequence to all other code sequences does not depend on the given code sequence, then $d^2_{\min}$(coded) is the minimum squared distance from the all-zero code sequence to any other code sequence. Such codes are called *distance invariant codes* [14].

Having defined the components of the total coding gain, from (2.31),(2.32) and

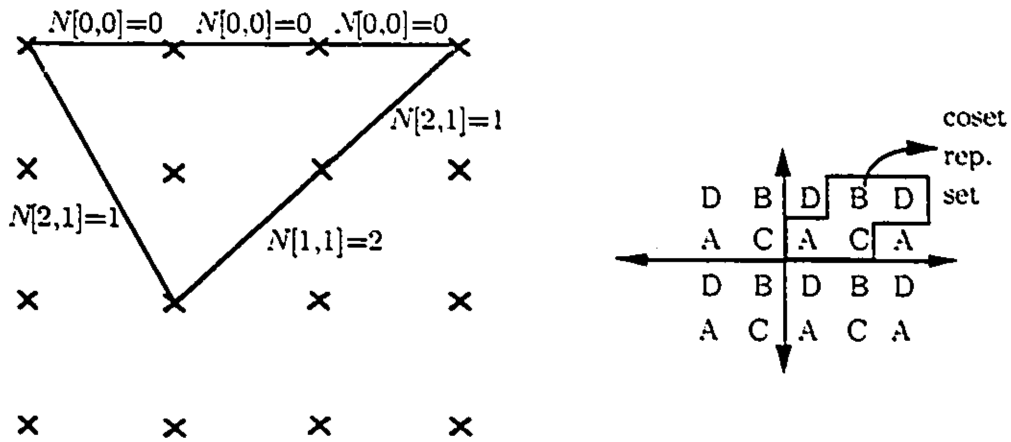


Figure 2.13: Paths apart at $d^2_{\min}$(coded) and the coset representatives of the cosets of $2\mathbb{Z}^2$, where each letter denotes different coset.

(2.17), it is found that

$$\gamma_{c,t}(\Lambda/\Lambda', \mathrm{C}) = \gamma_t(\Lambda)\gamma_c(\mathbb{Z}^N/\Lambda', C) = \frac{d^2_{\min}(\Lambda)}{V(\Lambda)^{2/N}} \frac{d^2_{\min}(\text{coded})}{d^2_{\min}(\text{uncoded})} \left(\frac{V(R_{B,\text{uncoded}})}{V(R_{B,\text{coded}})}\right)^{2/N} \tag{2.34}$$

since

$$d^2_{\min}(\text{uncoded}) = d^2_{\min}(\Lambda) \qquad \text{and} \qquad \frac{V(R_{B,\text{coded}})}{V(R_{B,\text{uncoded}})} = 2^r, \tag{2.35}$$

$$\gamma_{c,t}(\Lambda/\Lambda', \mathrm{C}) = \frac{d^2_{\min}(\text{coded})}{V(\Lambda)^{2/N}} \left(2^{-r}\right)^{2/N}. \tag{2.36}$$

In the previous example the signal points were originally selected within the integer lattice; therefore $V(\Lambda) = V(\mathbb{Z}^2) = 1$, and the minimum distance due to coding is $d^2_{\min}(\text{coded}) = 4$; the corresponding coding gain yields

$$\gamma_{c,t}(\mathbb{Z}^2/2\mathbb{Z}^2, \mathrm{C}) = \frac{4}{1^{2/2}} \left(2^{-1}\right)^{2/2} = 2 = 3\,\mathrm{dB}.$$

Chapter 3

THE NOTION OF SHAPING

Many coding schemes have been introduced throughout the evolution of telecommunications which allow better information transmission over bandlimited channels. Coset codes, i.e., lattice codes and trellis codes, attracted much interest as they could provide high code rates coupled with good performance. As mentioned earlier, it was

Table 3.1: The coding gain of the best known lattice codes in $\mathbb{R}^N$ for various dimensions .

N (dimension)	$\gamma_l(\Lambda)$	N (dimension)	$\gamma_l(\Lambda)$ (dB)
1	0.00	16	4.52
2	0.62	17	4.60
3	1.00	18	4.75
4	1.51	19	4.91
5	1.81	20	5.11
6	2.21	21	5.30
7	2.58	22	5.53
8	3.01	23	5.76
9	3.01	24	6.02
10	3.16	32	6.02
11	3.38	36	6.19
12	3.63	40	6.62
13	3.78	48	7.78
14	3.96	64	8.09
15	4.21		

proven that lattice codes could bridge the gap between the performance of ordinary uncoded PAM and the channel capacity, as shown in Table 3.1. This table lists the lattice gain, as given by (2.17), of the most power efficient lattice codes discovered so far in $\mathbb{R}^N$ for different dimensions [13]. A modest 3 dB gain requires a signal constellation based on a lattice in 8 dimensional space. In order to obtain 6 dB gain, the number of dimensions of the vector space, where the lattice is constructed increases rapidly to 24. Beyond 5 dB of gain, obtained by using a 20 dimensional signal constellation, a doubling of the number of dimensions is necessary for every 1.5 dB increase in power gain. Unfortunately, as the number of dimensions increases, the decoding task of the lattice becomes so complicated that it negates the attractiveness of the lattice codes. Similarly, lattice based trellis codes can achieve up to 6.22 dB total power gain [14] over ordinary uncoded PAM systems. However, it is shown in [14] that taking complexity into account at a fixed error rate, after the initial 3–4 dB of gain, the effective coding gain increases only 0.4 dB, if the complexity is doubled. (By effective coding gain we mean the total coding gain decreased by an error coefficient factor which is proportional to the number of nearest neighbors in the lattice on which the trellis code is based, since an increase in the number of nearest neighbors increases the possibility of error events). In other words, finding denser constellations in $\mathbb{R}^N$ (as for lattice codes), or adding more redundancy between the following signal points in the output symbol sequence (as for trellis codes), is of little help after a certain gain is achieved.

Still, there is another factor that has an impact on the power gain of the transmission system. This is the shape of the boundary region R_B that covers the finite portion of the lattice used as the signal constellation. For example, Fig. 3.1 illustrates two signal constellations which are bounded by circular and square shaped regions. Each constellation is assumed to be large enough so that we can use the continuous approximation method to calculate the normalized average power of these two signal constellations (in fact all the calculations stated in this chapter will be based on large constellations unless otherwise stated, where the normalized average power is calcu-

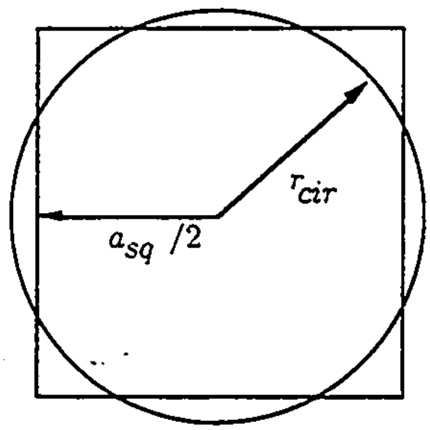


Figure 3.1: Illustration of a square shaped and a circular shaped constellation in $\mathbb{R}^2$ with equal sizes.

lated by the continuous approximation method). Clearly, both regions have the same volume so that the number of signal points in both constellations is equal, resulting in an equal amount of information carried by each symbol from both constellations. According to the calculations performed in Appendix D, it has been found that the circular constellation requires 0.2 dB less power than the square shaped constellation.

The improvement in power efficiency of the boundary region with any shape, over that of the cubic shaped boundary region in N dimensional space is termed the *shape gain* $\gamma_s(R_B)$ of R_B. The best boundary regions in terms of shape gain are always spherically shaped. As the number of dimensions N increases, the maximum shape gain achievable with any boundary region R_B in $\mathbb{R}^N$ increases, reaching 1.53 dB as N becomes arbitrarily large (see Appendix B). Therefore, after achieving the first 3–4 dB gain by using coset codes it may become more attractive to use additional computational complexity in shaping rather than in complex coding when seeking to achieve the best possible performance.

Multidimensional and nonequiprobable signalling are the two principal methods that are used to achieve a significant shape gain for a given constellation. In the remainder of this chapter. we will discuss both of these methods in detail.

3.1 Multidimensional Constellations

Transmission systems utilizing signal constellations based on multidimensional lattices can achieve higher power efficiency than ordinary uncoded PAM. This is partly due to the fact that dense signal point packings can be found in multidimensional space, and partly due to the increase in potential shape gain $\gamma_s(R_B)$ of the boundary region R_B as the number of dimensions increases (Appendix B). Table 3.2 summarizes the maximum potential shape gain obtainable in N-dimensional space. Regardless of the number of dimensions, the maximum shape gain is achieved only with spherically shaped signal constellations.

Table 3.2: Energy savings of N-dimensional spherical constellations compared to cubic constellations.

N (dimension)	$\gamma_s(R_B)$
2	0.20
4	0.45
8	0.73
16	0.98
24	1.10
32	1.17
48	1.26
64	1.31

Since large coding and shape gain is achievable in high multidimensional vector space, multidimensional constellations attracted significant attention among communications engineers. In an N-dimensional space, a signal point is represented by an N-tuple vector $\vec{x} = (x_1, x_2, \cdots, x_N)$. This vector can be expressed equivalently by N/n n-tuple vectors

$$\vec{x}_1 = (x_1, x_2, \cdots, x_n)$$

$$
\begin{aligned}
\bar{x}_2 &= (x_{n+1}, x_{n+2}, \cdots, x_{2n}) \\
&\vdots \\
\bar{x}_{N/n} &= (x_{N-n+1}, x_{N-n+2}, \cdots, x_N)
\end{aligned} \tag{3.1}
$$

and $\bar{x} = (\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{N/n})$, assuming that N is a multiple of n. Considering a signal constellation $\tilde{C}$ in $\mathbb{R}^N$; by expressing each point in $\tilde{C}$ by corresponding N/n n-tuples, we construct N/n n-dimensional signal constellation. Every such constellation is termed a constituent n-dimensional constellation induced by N-dimensional $\tilde{C}$. Geometrically, every n-dimensional constituent constellation is a projection of $\tilde{C}$ into that n-dimensional space. Depending on the constellation, $\tilde{C}$, N/n constituent constellations can be different from each other. In reality, multidimensional signal space can be realized by using two space orthogonal electric field polarizations to communicate on the same carrier frequency. Therefore an N-dimensional signal constellation $\tilde{C}$ is usually divided into its $N/2$ constituent two dimensional constellations, and every symbol in $\tilde{C}$ is regarded as $N/2$ consecutive symbols from each of the $N/2$ constituent constellation induced by $\tilde{C}$. For practical reasons (in order to utilize the same constellation for consecutive $N/2$ signalling interval), all of the $N/2$ constituent constellations are usually assumed to be identical. Throughout this thesis, we will thus denote all the constituent two dimensional constellations simply by $\tilde{C}_2$. For example, in four dimensions a signal point is marked as (x_1, x_2, x_3, x_4). The coordinates of the corresponding points in the $2D$ constituent constellation are then (x_1, x_2) and (x_3, x_4). Another important modulation scheme corresponds to dividing $\tilde{C}$ in $\mathbb{R}^N$, into N 1-dimensional constituent constellations, denoted by $\tilde{C}_1$, and induced by $\tilde{C}$.

Returning to 2-dimensional signalling, the $N/2$ 2-dimensional symbols are transmitted one after another over the channel in $N/2$ consecutive symbol intervals. Clearly, each 2-dimensional symbol conveys $\beta = 2b/N$ bits of information, where β is termed the *normalized bit rate.* An N-dimensional signal constellation $\tilde{C}$ which is designed to transmit b bits/symbol information must contain $\Omega_{\tilde{C}} = 2^b$ signal points, and $\tilde{C}$ must be a subset of the Cartesian product of its constituent 2-dimensional constellation $\tilde{C}_2$

with itself $N/2$ times. Hence the number of points in $\tilde{C}$ should be

$$\Omega_{\tilde{C}} \leq \left(\Omega_{\tilde{C}_2}\right)^{N/2}. \tag{3.2}$$

Equivalently, for a given $\Omega_{\tilde{C}} = 2^b = 2^{\beta N/2}$, (3.2) yields

$$2^{\beta N/2} \leq \left(\Omega_{\tilde{C}_2}\right)^{N/2} \quad\Longrightarrow\quad 2^{\beta} \leq \Omega_{\tilde{C}_2}. \tag{3.3}$$

In general it is desired that the constituent 2-dimensional constellation be as small as possible, yielding an equality in (3.2); but this is normally not the case and $\Omega_{\tilde{C}_2}$ is strictly greater than 2^{β}. This is termed the *constellation expansion* of the 2-dimensional constituent constellations induced by $\tilde{C}$, and is characterized by the *constellation expansion ratio* (CER) defined as

$$CER_2(\tilde{C}) \triangleq \frac{\Omega_{\tilde{C}_2}}{\Omega_{\tilde{C}}^{2/N}} = \frac{\Omega_{\tilde{C}_2}}{2^{\beta}}. \tag{3.4}$$

Clearly, the lower bound for $CER_2(\tilde{C})$ is 1. Furthermore, the CER of $\tilde{C}_n$ induced by $\tilde{C}$ is defined as

$$CER_n(\tilde{C}) \triangleq \frac{\Omega_{\tilde{C}_n}}{\Omega_{\tilde{C}}^{n/N}}. \tag{3.5}$$

To demonstrate the constellation expansion due to dividing the multidimensional constellation into its constituent constellations in fewer dimensions, we select a two dimensional constellation $\tilde{C}$ from $\mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$, a translate of the integer lattice $\mathbb{Z}^2$. Now, determine its 1-dimensional constituent constellation $\tilde{C}_1$. Fig. 3.2 illustrates $\tilde{C}$ and $\tilde{C}_1$.

There are 12 signal points in $\tilde{C}$, whereas $\tilde{C}_1$ contains 4 points. By using (3.5), the constellation expansion ratio of $\tilde{C}_1$ is found as

$$CER_1(\tilde{C}) = \frac{4}{12^{1/2}} = 1.15. \tag{3.6}$$

The suitability of a signal constellation for information transmission mainly depends on two factors. The first one is the minimum separation between the signal

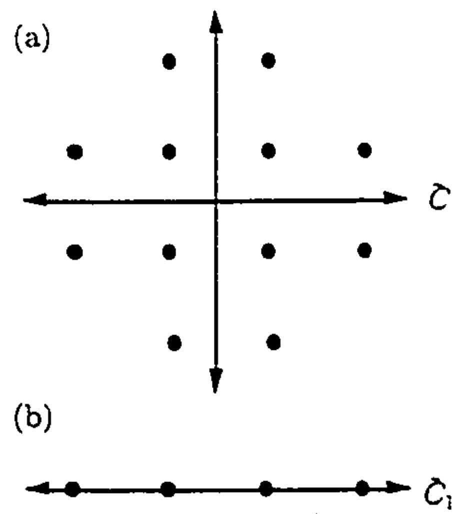


Figure 3.2: (a): A 12-point two-dimensional constellation $\tilde{C}$; (b): corresponding constituent one-dimensional constellation $\tilde{C}_1$.

points contained in the constellation, and denoted by $d^2_{\min}(\tilde{C})$. The second factor is the average signal energy $P_{av}(\tilde{C})$, which is defined as the average energy of signal points in $\tilde{C}$ in N-dimensional signal space, normalized to two dimensions so that

$$P_{av}(\tilde{C}) \triangleq \frac{2}{N} E\left\{\|x\|^2\right\}. \tag{3.7}$$

For a given normalized bit rate β, the efficiency of the constellation $\tilde{C}$ is increases as $d^2_{\min}(\tilde{C})$ increases and $P_{av}(\tilde{C})$ decreases. Therefore a general constellation figure of merit (CFM) can be defined as:

$$CFM(\tilde{C}) \triangleq \frac{d^2_{\min}(\tilde{C})}{P_{av}(\tilde{C})}. \tag{3.8}$$

Note that if all points in a signal constellation $\tilde{C}$ were scaled by some factor r to produce a new constellation $\tilde{C}_r$, then we would find $d^2_{\min}(\tilde{C}_r) = r^2 d^2_{\min}(\tilde{C})$ and $P_{av}(\tilde{C}_r) = r^2 P_{av}(\tilde{C})$ so that $CFM(\tilde{C}_r) = CFM(\tilde{C})$. Thus to evaluate $CFM(\tilde{C})$, we could just as easily evaluate it for a scaled version of $\tilde{C}$. It is customary to always consider the scaled version of a constellation for which $d^2_{\min}(\tilde{C}_r) = 1$, referred to as a normalized version of $\tilde{C}$.

It is usual to state the value of $CFM(\tilde{C})$ in comparison to that of the cubic constellation from the integer lattice $\mathbb{Z}^N$, which we label $\tilde{C}_B$. Recall that $d^2_{\min}(\tilde{C}_B) = d^2_{\min}(\mathbb{Z}^N) = 1$. The $CFM(\tilde{C}_B)$ of $\tilde{C}_B$ in $\mathbb{Z}^N$ for any N has been shown in [18] to be given by

$$CFM(\tilde{C}_B) = \frac{6}{2^\beta}, \tag{3.9}$$

where β is the normalized bit rate defined previously.

The ratio of normalized peak signal energy in the constituent 2-dimensional constellation $\tilde{C}_2$ to the normalized average signal energy $P_{av}(\tilde{C})$ of $\tilde{C}$ is defined as *peak to average energy ratio* of $\tilde{C}$ and is denoted by $PAR_2(\tilde{C})$. PAR can also be measured as the ratio of the normalized peak signal energy in $\tilde{C}$ to the normalized average signal energy $P_{av}(\tilde{C})$. In this case it would be denoted by $PAR_N(\tilde{C})$.

Both $CER_2(\tilde{C})$ and $PAR_2(\tilde{C})$ are very important characteristics of a signal constellation, and we would like both of them to be as small as possible. This is particularly true when impairments other than additive Gaussian noise (such as linear or nonlinear distortion, or phase jitter) are also present. Such impairments are important in voiceband data transmission, [17].

For example, any N-dimensional cubic constellation $\tilde{C}$, where N is assumed to be even, can be considered as an $N/2$ fold Cartesian product of two-dimensional square constellation. Since the normalized average power of the Cartesian product of one set of points with itself is the same as the normalized average power of the set itself, we define the normalized average power of an N-dimensional cubic constellation to be equal to that of a 2-dimensional square constellation, which for large constellations is calculated to be $P_{av}(\tilde{C}) = a^2/6$ (Appendix D), where a denotes the length of one side of the square. The peak signal energy in the constituent constellation, which is also a square constellation. induced by $\tilde{C}$ is found as $a^2/2$, yielding

$$PAR_2(\tilde{C}) = \frac{\frac{a^2}{2}}{\frac{a^2}{6}} = 3. \tag{3.10}$$

A general method of constructing an N-dimensional signal constellation $\tilde{C}$ is to take the signal points from some N-dimensional lattice Λ (or from a translate of Λ which is $\Lambda + \bar{c}$) that lie within some region R_B, termed the boundary region, in N-dimensional signal space. The region R_B should be large enough to enclose the desired number of signal points; we require $\Omega_{\tilde{C}} = 2^b$ points to transmit b bits per N dimensional symbol.

For large constellations, the volume of the boundary region $V(R_B)$ is approximately equal to the product of the number of points enclosed by R_B, and the fundamental volume of the lattice Λ from which the signal points are selected. Consequently, for larger and larger constellations,

$$\Omega_{\tilde{C}} = \frac{V(R_B)}{V(\Lambda)}. \tag{3.11}$$

Since $d^2_{\min}(\tilde{C}) = d^2_{\min}(\Lambda)$, we found that

$$CFM(\tilde{C}) = \frac{d^2_{\min}(\tilde{C})}{P_{av}(\tilde{C})} = \frac{d^2_{\min}(\Lambda)}{P_{av}(\tilde{C})}. \tag{3.12}$$

Thus the ratio of the $CFM(\tilde{C})$ to the $CFM(\tilde{C}_B)$ is

$$\frac{CFM(\tilde{C})}{CFM(\tilde{C}_B)} = \frac{d^2_{\min}(\Lambda)}{P_{av}(\tilde{C})}\frac{2^{\beta}}{6}. \tag{3.13}$$

Since $\Omega_{\tilde{C}} = 2^b = 2^{\beta N/2}$, when we transmit b bits per N dimensional symbol we have from (3.11)

$$2^{\beta} = \left(\frac{V(R_B)}{V(\Lambda)}\right)^{2/N}. \tag{3.14}$$

Replacing 2^{β} in (3.13) then yields

$$\frac{CFM(\tilde{C})}{CFM(\tilde{C}_B)} = \left(\frac{d^2_{\min}(\Lambda)}{(V(\Lambda))^{2/N}}\right)\left(\frac{(V(R_B))^{2/N}}{6P_{av}(\tilde{C})}\right). \tag{3.15}$$

The ratio above gives the power efficiency of the constellation $\tilde{C}$ relative to the baseline constellation $\tilde{C}_B$. The first factor of the product in (3.15) is the lattice gain defined in Chapter 2:

$$\gamma_l(\Lambda) = \frac{d^2_{\min}(\Lambda)}{(V(\Lambda))^{2/N}}. \tag{3.16}$$

The second factor of (3.15) is the *shape gain:*

$$\gamma_s(R_B) = \frac{(V(R_B))^{2/N}}{6P_{av}(\tilde{C})}. \tag{3.17}$$

Thus the power efficiency of a constellation $\tilde{C}$ is separated into two parts, the lattice gain of the lattice on which $\tilde{C}$ is based, and the shape gain of R_B which encloses $\tilde{C}$.

Components of the Constellation Expansion Ratio $CER_2(\tilde{C})$

Assuming that $\tilde{C}$ is a signal constellation in $\mathbb{R}^N$ and $\tilde{C}_2$ is the constituent 2-dimensional constellation induced by $\tilde{C}$, $\tilde{C}_2$ will be a finite portion of a two dimensional lattice Λ_2, where Λ_2 is a projected version of the lattice Λ, on which $\tilde{C}$ is based, onto $\mathbb{R}^2$. Let $R_{B,2}$ denote the boundary region of the 2-dimensional constituent constellation $\tilde{C}_2$, then the number of signal points in $\tilde{C}_2$ can be written as

$$\Omega_{\tilde{C}_2} = \frac{V(R_{B,2})}{V(\Lambda_2)}, \tag{3.18}$$

and the number of signal points in $\tilde{C}$ is already known to be given by

$$\Omega_{\tilde{C}} = \frac{V(R_B)}{V(\Lambda)}. \tag{3.19}$$

By using (3.4), we find

$$CER_2(\tilde{C}) = \left(\frac{V(R_{B,2})}{V(R_B)^{2/N}}\right) \times \left(\frac{(V(\Lambda))^{2/N}}{V(\Lambda_2)}\right). \tag{3.20}$$

Here again, we were able to separate the effect of shaping and coding on the constellation expansion. where the constellation expansion ratio of Λ due to coding is defined as

$$CER_{c,2}(\Lambda) \triangleq \frac{V(\Lambda)^{2/N}}{V(\Lambda_2)}. \tag{3.21}$$

and the constellation expansion ratio of R_B due to shaping is

$$CER_{s,2}(R_B) = \frac{V(R_{B,2})}{V(R_B)^{2/N}}. \tag{3.22}$$

Components of the Peak to Average Energy Ratio $PAR_2(\tilde{C})$

The definition of $PAR_2(\tilde{C})$ given before can be formulated as

$$PAR_2(\tilde{C}) = \frac{r^2_{\max}(R_{B,2})}{P_{av}(\tilde{C})}, \tag{3.23}$$

where $r^2_{\max}(R_{B,2})$ denotes the peak energy of the points in the constituent $2D$ constellation $\tilde{C}_2$ bounded by $R_{B,2}$. Similarly, the peak to average energy ratio of the constituent $2D$ constellation is

$$PAR_2(\tilde{C}_2) = \frac{r^2_{\max}(R_{B,2})}{P_{av}(\tilde{C}_2)}. \tag{3.24}$$

From (3.17), the shape gain of $R_{B,2}$ is

$$\gamma_s(R_{B,2}) = \frac{(V(R_{B,2}))^{2/N}}{6P_{av}(\tilde{C}_2)}. \tag{3.25}$$

Substituting (3.17), (3.24) and (3.25) into (3.23) yields

$$\begin{aligned} PAR_2(\tilde{C}) &= PAR_2(\tilde{C}_2)\frac{V(R_{B,2})}{6\gamma_s(R_{B,2})}\frac{\gamma_s(R_B)6}{(V(R_B))^{2/N}} \\ &= PAR_2(\tilde{C}_2)\frac{V(R_{B,2})}{(V(R_B))^{2/N}}\frac{\gamma_s(R_B)}{\gamma_s(R_{B,2})} \\ &= PAR_2(\tilde{C}_2)CER_{s,2}(R_B)\frac{\gamma_s(R_B)}{\gamma_s(R_{B,2})}. \end{aligned} \tag{3.26}$$

Equation (3.26) reveals that the peak to average power ratio in two dimensions is a product of the peak to average ratio of the constituent $2D$ constellation, the constellation expansion ratio of the boundary region R_B and the improvement in shape gain resulting from the use of higher dimensional constellations. Since the latter factor is the factor we would like to increase, (3.26) suggests that we choose a boundary region R_B with a small constellation expansion ratio and with a good peak to average power ratio for its constituent $2D$ constellation.

Comparison of Shape Gain with PAR and CER

The following table shows the shape gain, constellation expansion ratio and peak to average power ratio of N dimensional constellations $\tilde{C}$ with spherical boundary region $\otimes_N$. The results here are based on the calculations given in Appendix B.

Table 3.3: Shape gains, constellation expansion ratios and peak to average energy ratios of N spheres.

N	$\gamma_s(\otimes_N)$ (dB)	$CER_{s,2}(\otimes_N)$	$PAR_2(\tilde{C})$
2	0.2	1.000	2
4	0.46	1.414	3
8	0.73	2.213	5
12	0.88	2.994	7
16	0.98	3.764	9
24	1.10	5.289	13
32	1.17	6.800	17

Note that $PAR_2(\tilde{C})$ increases rapidly; only when $N = 4$ is the PAR of the spherical constellation the same as the PAR of the cubic constellation. The second column in Table 3.3 shows the maximum obtainable shaping gain in N dimensional space by equiprobable signalling, i.e., when every signal point in the signal constellation is equally likely to be selected. A circular shaped constellation has 0.2 dB power efficiency over a square shaped constellation in 2 dimensional space, and by increasing the number of dimensions of the signal space to 16, we can obtain almost 1 dB of shape gain.

3.1.1 Voronoi Constellations

A *Voronoi constellation* is a multidimensional constellation based on the lattice partition Λ/Λ', where Λ' is a sublattice of Λ. Basically, this constellation is the set of minimum norm coset representatives of $|\Lambda/\Lambda'|$ cosets of Λ'. In [19] these constellations are thoroughly investigated, and the methods used to construct them are described.

This discussion will not be repeated here. Voronoi constellations have most of the desirable attributes, and they can achieve significantly higher shape gains than any other multidimensional constellations considered earlier, but at the cost of somewhat larger constellation expansion and peak to average energy ratios. Another problem with this method is that the complexity of addressing the signal points is governed by the complexity of decoding the lattice. Decoding such a lattice can be quite complex.

3.2 Nonequiprobable Signalling

In the previous section it is shown that equiprobable multidimensional constellations can achieve a shape gain of up to 1.53 dB as the number of dimensions becomes arbitrarily large. In practice, to construct a spherical shaped constellation in multidimensional signal space is a very difficult task since the addressing of the output symbols of the encoder to the individual signal points in the constellation becomes arduous when the number of dimensions of the signal space increases. Fortunately, nonequiprobable signalling, as discussed shortly, is a method to approximate the shape of a multidimensional constellation in relatively lower dimensions to achieve high shape gain without suffering the complexity of working in multidimensional signal spaces.

Previously, it was shown that an N-dimensional constellation could be realized as a subset of the Cartesian product of its $N/2$ 2-dimensional constituent constellations, assuming that N is even. Equally, we might reverse this idea and construct a $2M$-dimensional constellation from considering linking 2-dimensional constellations. The question in considering such an approach is how to select or construct such 2-dimensional constellations $\tilde{C}_2$ so that the resulting N-dimensional constellation $\tilde{C}$ would have the properties of a spherically shaped constellation.

So far, it has been assumed that the signal points in a multidimensional constellation $\tilde{C}$ are uniformly distributed, meaning that every point is equally likely to be transmitted. However the probability distribution of the points in $\tilde{C}_2$ will not be uniform unless $\tilde{C}$ is exactly the Cartesian product of $N/2$ times of $\tilde{C}_2$ with itself. Assuming that $\tilde{C}_2$, induced by spherically shaped $\tilde{C}$ is sufficiently large that the selection a point in $\tilde{C}_2$ by the encoder is adequately approximated as the selection of a point in $R_{B,2}$ according to a continuous 2-dimensional density function $p(\bar{x})$, it has been shown [18] that as N becomes arbitrarily large, the density function $p(\bar{x})$ approaches that of a truncated Gaussian distribution:

$$p(\bar{x}) \sim \frac{k}{(2\pi\sigma^2)} \exp\{-\|\bar{x}\|^2/2\sigma^2\} \qquad \text{for } \bar{x} \in R_{B,2}, \tag{3.27}$$

for some scaling factor k, slightly larger than 1, where σ^2 corresponds to the variance of each dimension of the constellation. Also the boundary region $R_{B,2}$ has been found to be spherically shaped. The entropy per two dimensions is calculated by using continuous distribution function as :

$$H(\tilde{C}_2) = \int_{\bar{x}\in R_{B,2}} p(\bar{x}) \log_2 p(\bar{x})\, d\bar{x} \;\sim\; \log_2 2\pi e\sigma^2. \tag{3.28}$$

Then the average signal energy of $\tilde{C}_2$ is calculated to be

$$P_{av}(\tilde{C}_2) = \int_{\bar{x}\in R_{B,2}} p(\bar{x}) \|\bar{x}\|^2\, d\bar{x} \;\sim\; 2\sigma^2. \tag{3.29}$$

A cubic shaped equiprobable signal constellation $\tilde{C}_B$ in N dimensional space would have a square shaped constituent 2-dimensional constellation $\tilde{C}_{B,2}$, where the distribution probability on points is uniform; therefore comparing $\tilde{C}_2$ with a square shaped equiprobable constellation with the same entropy as $H(\tilde{C}_2)$ is equivalent to comparing $\tilde{C}$ with the cubic shaped constellation in N-dimensional space in terms of power efficiency. The average signal energy of $\tilde{C}_{B,2}$ with an entropy equal to $H(\tilde{C}_2)$ is found as

$$P_{av}(\tilde{C}_{B,2}) = \frac{2^{\beta}}{6} = \frac{2\pi e\sigma^2}{6}. \tag{3.30}$$

The improvement in the power efficiency of the nonequiprobable constituent 2-dimensional constellation $\tilde{C}_2$ over equiprobable baseline constellation $\tilde{C}_{B,2}$ is termed the total shape gain of $\tilde{C}_2$ which is found to be

$$\gamma_b(\tilde{C}_2) = \frac{P_{av}(\tilde{C}_{B,2})}{P_{av}(\tilde{C}_2)} = \frac{2\pi e\sigma^2/6}{2\sigma^2} = \frac{e\pi}{6} = 1.53 \text{ dB}. \tag{3.31}$$

The total shape gain partially depends on the nonuniform distribution of points in $\tilde{C}_2$ and partially depends on the boundary region $R_{B,2}$. The first component of the total shape gain is termed as the biasing gain. The latter component was already defined as the shape gain $\gamma_s(R_{B,2})$. Consequently, the total shape gain can be expressed in general as

$$\gamma_t(\tilde{C}) = \gamma_b(\tilde{C}) \times \gamma_s(R_B), \tag{3.32}$$

where $\tilde{C}$ denotes a nonequiprobable signal constellation, R_B is the region that encloses $\tilde{C}$, and $\gamma_b(\tilde{C})$ denotes the biasing gain of $\tilde{C}$ due to nonuniform distribution of points. The concept of biasing gain will be discussed more deeply in the following section.

As we see in (3.31) the same shape gain which is obtained by spherical constellation in very large dimensional signal space can also be obtained by two dimensional signal constellations where the distribution of points approaches that of a truncated Gaussian distribution. Note that the ultimate shape gain 1.53 dB can never be obtained either by spherical constellations (construction of infinite dimensional signal constellations is impossible.), or by nonequiprobable 2-dimensional constellations since the actual discrete distribution function p_i on points $i \in \tilde{C}_2$ can never be truly approximated by the continuous distribution $p(x)$ no matter how large the constellation is.

Still, many algorithms have been suggested in the literature to construct nonequiprobable constellations in order to obtain high shape gains in low dimensional signal spaces, without being bothered by the hard addressing problem of the signal points in high dimensional spaces. One of these methods was suggested by Calderbank & Ozarow [2], and is called the Calderbank & Ozarow (CO) method. Another was suggested by D. Forney [3], and called as trellis shaping. In the rest of this chapter, we will describe these methods.

3.2.1 The Calderbank & Ozarow Method

The Calderbank & Ozarow method of generating nonequiprobable signalling is based on partitioning an N dimensional signal space into G subconstellations $\tilde{C}^i$, with increasing average power as i increases for $0 \leq i \leq G-1$. Calderbank and Ozarow [21] proposed subconstellations with an equal number of points without claiming that this is necessarily optimum. Each subconstellation is selected with different frequency f_i, while the signal points in the same subconstellation are equally likely to be selected. The case where all f_i are equal corresponds to equiprobable signalling where all points in the constellation are equally probable to be selected. Livingston [20] demonstrated later that by using subconstellations with different sizes but with equal frequencies such that $f_0 = f_1 = \cdots = f_{G-1}$, also leads to nonequiprobable signalling, and yields the same gain as the CO method with even fewer subregions.

The partitioning of the signal constellation $\tilde{C}$ is done as follows: A region R_0 is selected within $\mathbb{R}^N$; thereafter by scaling R_0 the following nested sequence is obtained:

$$\begin{aligned} R_0 &= \alpha_0 R_0 \\ R_1 &= \alpha_1 R_0 \\ &\vdots \\ R_{G-1} &= \alpha_{G-1} R_0, \end{aligned} \tag{3.33}$$

where R_{G-1} is equal to the boundary region R_B of $\tilde{C}$. The sequence of R_i's allows us to divide $\tilde{C}$ into subconstellations $\tilde{C}^0, \tilde{C}^1, \cdots, \tilde{C}^{G-1}$ according to the rule

$$\begin{aligned} \tilde{C}^0 &= \tilde{C} \cap R_0 \quad \text{and} \\ \tilde{C}^i &= \tilde{C} \cap (R_i - R_{i-1}) \quad \text{for} \quad 1 \leq i \leq G-1. \end{aligned} \tag{3.34}$$

Since each boundary region is a scaled version of the center boundary region R_0, the volume contained in these regions are also scaled such that

$$V(R_i) = k_i R_0. \tag{3.35}$$

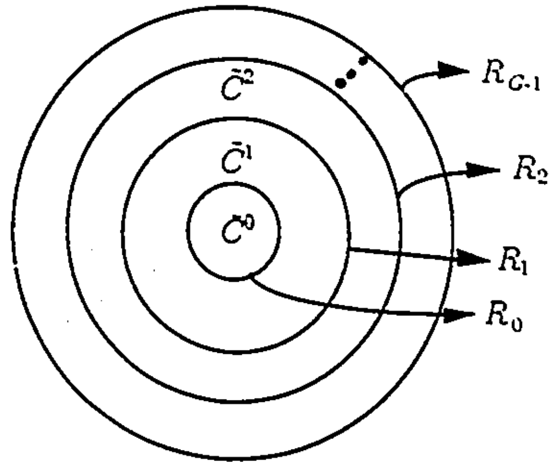


Figure 3.3: Decomposition of a two-dimensional constellation $\tilde{C}$ into its subconstellations $\tilde{C}^0, \tilde{C}^1, \cdots, \tilde{C}^{G-1}$.

Making a continuous approximation we see that

$$\Omega_{R_i} = k_i \Omega_{R_0}, \tag{3.36}$$

$$P_{av}(R_i) = k_i^{2/N} P_0, \tag{3.37}$$

where Ω_{R_i} denotes the number of signal points enclosed by the region R_i, $P_{av}(R_i)$ is the normalized average signal energy of the points contained in R_i, and P_0 denotes the average energy of the points within the subconstellation $\tilde{C}^0$ such that $P_0 = P_{av}(R_0)$.

The normalized average signal energy P_i of $\tilde{C}^i$ is the calculated as

$$\begin{aligned} P_i &= \frac{P_{av}(R_i)\Omega_{R_i} - P_{av}(R_{i-1})\Omega_{R_{i-1}}}{\Omega_{R_i} - \Omega_{R_{i-1}}} \\ &= \frac{k_i^{2/N} P_0 k_i \Omega_{R_0} - k_{i-1}^{2/N} P_0 k_{i-1} \Omega_{R_0}}{k_i \Omega_{R_0} - k_i \Omega_{R_0}} \\ &= \frac{k_i^{(2/N)+1} - k_{i-1}^{(2/N)+1}}{k_i - k_{i-1}} P_0. \end{aligned} \tag{3.38}$$

If all the subconstellations contain an equal number of signal points, then $k_i = i + 1$

for $0 \leq i \leq G-1$, and (3.38) becomes

$$P_i = \left[(i+1)^{(2/N)+1} - i^{(2/N)+1}\right] P_0. \tag{3.39}$$

The normalized average signal energy of the overall constellation $\tilde{C}$ depends on the frequency of each subconstellation $\tilde{C}^i$

$$\begin{aligned} \overline{P}_{av}(\tilde{C}) &= \sum_{i=0}^{G-1} P_i f_i \\ &= P_0 \sum_{i=0}^{G-1} \frac{\left(k_i^{2/N+1} - k_{i-1}^{2/N+1}\right)}{k_i - k_{i-1}} f_i. \end{aligned} \tag{3.40}$$

The bar in the formulas above denotes nonequiprobable signalling, and throughout this thesis whenever we use a bar it will be understood that the symbol under the bar corresponds to nonequiprobable signalling. The probability of an individual signal point in $\tilde{C}^i$ being selected can be found as

$$P(\ x \text{ is selected},\ x \in \tilde{C}^i\ ) = \frac{f_i}{\Omega_{\tilde{C}^i}}. \tag{3.41}$$

The entropy of the resultant nonequiprobable constellation can be calculated as follows:

$$\begin{aligned} \overline{H}(\tilde{C}) &= -\sum_{i=0}^{G-1} \sum_{j=1}^{\Omega_{\tilde{C}^i}} \left(\frac{f_i}{\Omega_{\tilde{C}^i}} \log_2 \frac{f_i}{\Omega_{\tilde{C}^i}}\right) \\ &= -\sum_{i=0}^{G-1} f_i \log_2 \frac{f_i}{\Omega_{\tilde{C}^i}} \\ &= -\sum_{i=0}^{G-1} f_i \log_2 f_i + \sum_{i=0}^{G-1} f_i \log_2 \Omega_{\tilde{C}^i} \\ &= H(f_0, f_1, \cdots, f_{G-1}) + \sum_{i=0}^{G-1} f_i \log_2 \left[(k_i - k_{i-1})\Omega_{\tilde{C}^0}\right] && (3.42) \\ &= H(f_0, f_1, \cdots, f_{G-1}) + \log_2 \Omega_{\tilde{C}^0} + \sum_{i=0}^{G-1} f_i \log_2 (k_i - k_{i-1}). && (3.43) \end{aligned}$$

A nonequiprobable constellation with entropy $\overline{H}(\tilde{C})$ is compared with an equiprobable constellation with the same entropy (and so $2^{\overline{H}(\tilde{C})}$ points) in order to investigate the consequences of nonequiprobable signalling (both constellations are based on the same lattice Λ). Such an equiprobable constellation would occupy a region R_B in the signal space with a volume

$$V(R_B) = V(R_0)\frac{2^{\overline{H}(\tilde{C})}}{2^{\log_2 \Omega_{\tilde{C}}^0}} = V(R_0)2^{\overline{H}(\tilde{C})-\log_2 \Omega_{\tilde{C}0}}. \tag{3.44}$$

The constellation expansion ratio due to using nonequiprobable signal constellations is defined as the magnification factor of the signal constellation due to nonequiprobable signalling, and in the case of the CO method

$$\begin{aligned}\overline{CER}(\tilde{C}) &\triangleq \frac{V(R_{G-1})}{V(R_B)} = \frac{V(R_0)k_{G-1}}{V(R_0)2^{\overline{H}(\tilde{C})-\log_2 \Omega_{\tilde{C}0}}} \\ &= \frac{k_{G-1}}{2^{\left[H(f_0,f_1,\cdots,f_{G-1})+\sum_{i=0}^{G-1} f_i \log_2(k_i-k_{i-1})\right]}}. \end{aligned}\tag{3.45}$$

Here, $\tilde{C}$ denotes the signal constellation in N-dimensional signal space, where nonequiprobable signal points reside. For almost all the discussion of nonequiprobable signalling will concentrate on a two-dimensional signal space. If all the constituent 2-dimensional constellations of $\tilde{C}$ are identical, the magnification factor $CER_2(\tilde{C})$ for the size of the constituent two-dimensional constellation is given by

$$\overline{CER}_2(\tilde{C}) = \left(\overline{CER}(\tilde{C})\right)^{2/N}. \tag{3.46}$$

To calculate the power efficiency of nonequiprobable signalling, we need to find the normalized average signal energy of the corresponding equiprobable constellation:

$$P_{av}(\tilde{C}) = P_0\left(\frac{V(R_B)}{V(R_0)}\right)^{2/N} = P_0 2^{\frac{2}{N}\left[H(\tilde{C})-\log_2 \Omega_{\tilde{C}0}\right]}. \tag{3.47}$$

Consequently, the gain obtained by nonequiprobable signalling in N dimensional signal space. termed the *biasing gain*. is given as

$$\gamma_N(f_0,f_1,\cdots,f_{G-1}) = \frac{P_{av}(\tilde{C})}{\overline{P}_{av}(\tilde{C})} = \frac{P_0 2^{\frac{2}{N}\left[H(\tilde{C})-\log_2 \Omega_{\tilde{C}0}\right]}}{P_0 \sum_{i=0}^{G-1} \frac{\left(k_i^{2/N+1}-k_{i-1}^{2/N+1}\right)}{k_i-k_{i-1}} f_i} \tag{3.48}$$

$$= \frac{2^{\frac{2}{N}\left[H(f_0,f_1,\cdots,f_{G-1})+\sum_{i=0}^{G-1} f_i \log_2(k_i-k_{i-1})\right]}}{\sum_{i=0}^{G-1} \frac{\left(k_i^{2/N+1}-k_{i-1}^{2/N+1}\right)}{k_i-k_{i-1}} f_i}. \tag{3.49}$$

Equation (3.49) can be simplified for equally sized subconstellations such that $k_i = i + 1$ for $0 \leq i \leq G - 1$:

$$\gamma_N(f_0, f_1, \cdots, f_{G-1}) = \frac{2^{\frac{2}{N}[H(f_0, f_1, \cdots, f_{G-1})]}}{\sum_{i=0}^{G-1}\left[(i+1)^{2/N+1} - i^{2/N+1}\right] f_i}. \tag{3.50}$$

The peak to average energy ratio for a nonequiprobable constellation can be calculated first by determining the peak signal energy:

$$r_{\max}^2(R_{G-1}) = (k_{G-1})^{2/N} r_{\max}^2(R_0), \tag{3.51}$$

and then using (3.40) to give us that

$$\begin{aligned}
\overline{PAR}_N(\bar{C}) &= \frac{k_{G-1}^{2/N}}{\sum_{i=0}^{G-1} \frac{\left(k_i^{2/N+1} - k_{i-1}^{2/N+1}\right)}{k_i - k_{i-1}} f_i} \frac{r_{\max}^2(R_0)}{P_0} \\
&= \frac{k_{G-1}^{2/N}}{\sum_{i=0}^{P-1} \frac{\left(k_i^{2/N+1} - k_{i-1}^{2/N+1}\right)}{k_i - k_{i-1}} f_i} PAR_2(\bar{C}^0) \\
&= \gamma_N(f_0, f_1, \cdots, f_{G-1}) PAR_N(\bar{C}^0) \left(CER(\bar{C})^{2/N}\right) \\
&= \gamma_N(f_0, f_1, \cdots, f_{G-1}) PAR_N(\bar{C}^0) CER_2(\bar{C}).
\end{aligned} \tag{3.52}$$

Components of the Total Shape Gain

The biasing gain $\gamma_N(f_0, f_1, \cdots, f_{G-1})$, calculated in (3.49) is only one of the components of the total shape gain of the N dimensional nonequiprobable constellation, as noted previously. The biasing gain is directly obtained by approximating the probability distribution on points in the constellation to a truncated Gaussian distribution. The other component of the total shape gain is the shape gain of the boundary region R_B that encloses the nonequiprobable constellation, and called as shape gain of the region R_B as we defined at the beginning of Chapter 3. (The reader is cautioned not to confuse the notions of shape gain and the total shape gain). The total shape gain is the product of the shape gain and the biasing gain, and for the CO method, is given by

$$\gamma_t = \gamma_N(f_0, f_1, \cdots, f_{G-1}) \times \gamma_s(R_B) \tag{3.53}$$

Earlier we concluded that the total shape gain is absolutely limited to 1.53 dB in any signal space. Therefore the biasing gain should be limited to 1.53 dB, the maximum shape gain obtainable in that N dimensional space, hence it yields 1.33 dB in two dimensions (see Table 3.3), and decreases as the number of dimensions increases. Consequently, nonequiprobable signalling is only significant in a low dimensional signal space.

On the other hand as the number of subconstellations increases, the obtained biasing gain approaches the above set limit. Fig. 3.5 illustrates the increase in potential biasing gain with respect to an increase in the number of subconstellations in two dimensional space. As $G \rightarrow \infty$, the potential biasing gain approaches $e/2 \doteq 1.33$ dB.

Therefore, we conclude that ultimate shape gain can be obtained in low dimensional spaces by partitioning an existing signal constellation into an infinite number of subconstellations, applying a spherical boundary region to the partitioned constellation, and implementing a method to activate the subconstellations with low average signal energy more frequently than those with higher average signal energy. In the first sight this method may seem to be very complicated, but even with 4 subconstellations and a trivial square shaped boundary region, it is possible to obtain more than 1 dB total shape gain. In the next section we will see how to apply the CO method to real systems employing what we refer to as a shaping code.

Construction of Shaping Codes

Because of the separability of shaping and coding, shape gain is usually achieved by a separate code, which is employed in selecting the signal points with different probabilities so that the overall probability distribution on points looks like a truncated Gaussian distribution. There is no restriction on selecting shaping codes. If a block code is used, then the result is nothing other than an equiprobable signalling method in multidimensional signal space. Nonetheless the CO method is more attractive than any other multidimensional signalling such as employing Voronoi constellations, where the addressing of points requires complicated algorithms. In [23], different codes have been investigated as shaping codes, but in this thesis we will employ only block codes for shaping purposes. No matter which code is selected, there is a basic procedure to be followed.

First, a mapping function from the output symbols of the shaping code to the subcostellations must be properly defined, and the frequencies f_i of these subconstellations must be arranged in such a way that within the overall signal constellation, subconstellations with higher average signal energy are selected with a lower probability than those with lower average signal energy.

If a $r(C)$ rate code is used as a shaping code, then the biasing gain $\overline{\gamma}_N(\mathrm{C})$, constellation expansion ratio $\overline{CER}(\mathrm{C})$ and peak to average power $\overline{PAR}_2(\mathrm{C})$ become

$$\overline{\gamma}_N(\mathrm{C}) = \frac{2^{\frac{2}{N}\left[r(C)+\sum_{i=0}^{C-1} f_i \log_2(k_i-k_{i-1})\right]}}{\sum_{i=0}^{C-1} \frac{\left(k_i^{2/N+1}-k_{i-1}^{2/N+1}\right)}{k_i-k_{i-1}} f_i}, \tag{3.54}$$

$$\overline{CER}(\mathrm{C}) = \frac{k_{C-1}}{2^{\left[r(C)+\sum_{i=0}^{C-1} f_i \log_2(k_i-k_{i-1})\right]}}, \tag{3.55}$$

$$\overline{PAR}_N(\mathrm{C}) = \overline{PAR}_N(\tilde{C}). \tag{3.56}$$

In the above equations, "C" denotes the shaping code, where $\tilde{C}$ denotes the nonequiprobable signal constellation.

A simple code suitable for shaping is provided by the *greedy algorithm* [23], which produces a set of binary vectors in order of increasing Hamming weight

$$\mathcal{G}_{v,k} = \{c \text{ is binary } v\text{-tuple} \mid w(c) \leq k\}, \tag{3.57}$$

where $w(c)$ denotes the Hamming weight of the v tuple c. $\mathcal{G}_{v,k}$ includes all binary codewords c of length v, with k or fewer nonzero digits. For example the codewords in $\mathcal{G}_{4,2}$ are given in Table 3.4. The mapping from the digits of the codewords of the

Table 3.4: Codewords contained in $\mathcal{G}_{4,2}$.

$\mathcal{G}_{4,2}$ code		
0 0 0 0	1 0 0 0	0 1 0 1
0 0 0 1	1 0 0 1	0 1 1 0
0 0 1 0	1 0 1 0	0 0 1 1
0 1 0 0	1 1 0 0	

shaping code to the subconstellations can be done as follows: Suppose the constellation $\bar{C}$ is divided into 2^L equally sized subconstellations. Then an L level binary shaping code $C = [C_1, C_2, \cdots, C_L]$ is utilized, where the codes at all levels are selected from the family of greedy algorithms. To each subconstellation an L tuple label $c(c_1, c_2, \cdots, c_L)$ is assigned. This is done as follows: Every subconstellation is put into order of increasing average signal energy. To every label a decimal priority number $\bar{p}$ is then assigned where

$$\bar{p} = \sum_{i=1}^{L} c_i 2^{L-i}, \tag{3.58}$$

with c_i being the i^{th} digit of the L-tuple label, corresponding to the i^{th} level code C_i. The labels with smaller $\bar{p}$ come first in the order of the labels then those with larger $\bar{p}$. After all the labels are ordered a one to one match is done with the order of the subconstellations. If p_i is the frequency of having a zero within the codewords of the component codeword c_i for $1 \leq i \leq L$, then

$$\begin{aligned} f_{(0,0,\cdots,0,0)} &= p_1 p_2 p_3 \cdots p_{L-1} p_L \\ f_{(0,0,\cdots,0,1)} &= p_1 p_2 p_3 \cdots p_{L-1}(1-p_L) \\ f_{(0,0,\cdots,1,0)} &= p_1 p_2 p_3 \cdots (1-p_{L-1}) p_L. \\ &\vdots \end{aligned} \tag{3.59}$$

In this equation set, $f_{\bar{c}} = f_{(c_1,c_2,\cdots,c_L)}$ denotes the frequency of the subconstellation labeled by $\bar{c} = (c_1, c_2, \cdots, c_L)$. If $p_1 > p_2 > \cdots > p_L > 1/2$, we find that

$$f_{(0,0,\ldots,0,0,0)} > f_{(0,0,\ldots,0,0,1)} > f_{(0,0,\ldots,0,1,0)} > \cdots > f_{(1,1,\ldots,1,1,1)}. \tag{3.60}$$

which shows that subconstellations with a lower average signal energy will be more frequently selected than those with higher average signal energy as required for CO method.

Table 3.5 lists the biasing gain, CER and PAR obtained as a result of utilizing a two level shaping code $[C_1, C_2]$, where $C_1 = \mathcal{G}_{m,k1}$ and $C_2 = \mathcal{G}_{m,k2}$.

Table 3.5: Parameters of two-level shaping codes $[C_1, C_2]$ where $C_1 = \mathcal{G}_{m,k1}$ and $C_2 = \mathcal{G}_{m,k2}$.

m	k_1	k_2	$\bar{\gamma}_N$(C) (dB)	CER_2(C)	PAR_2(C)
6	1	3	0.43	1.55	1.71
8	1	3	0.50	1.72	1.93
8	2	4	0.49	1.45	1.51
10	2	4	0.55	1.47	1.67
12	2	5	0.60	1.50	1.73
15	2	5	0.65	1.65	1.91
16	3	6	0.66	1.46	1.70
19	3	7	0.70	1.51	1.77
20	3	7	0.71	1.54	1.82

Example: Consider a one level code $C_1 = \mathcal{G}_{8,2}$ as a shaping code. We divide the signal constellation into two subconstellations for nonequiprobable signalling. For simplicity, each constellation is assigned an equal number of signal points. The subconstellation with lower average signal energy will be called "light", labeled $\tilde{C}^0$, and the other one "heavy", labeled $\tilde{C}^1$. The codewords selected from $\mathcal{G}_{8,2}$ follow one after another in series to generate a binary sequence. Every element of this sequence determines which one of the subconstellations the transmitted signal point is from in every signalling interval. A "0" activates the light subconstellation $\tilde{C}^0$, whereas a "1" activates the heavy one $\tilde{C}^1$. A block diagram of the transmitter is given in Fig. 3.4, where the subconstellation selector includes a lookup table for the code $\mathcal{G}_{8,2}$. Each light point r_0 in $\tilde{C}^0$ has a corresponding heavy point r_1 in $\tilde{C}^1$. The conventional encoder selects one of the light points within $\tilde{C}^0$, and the signal point selector decides whether to transmit the light point selected, or its corresponding heavy point.

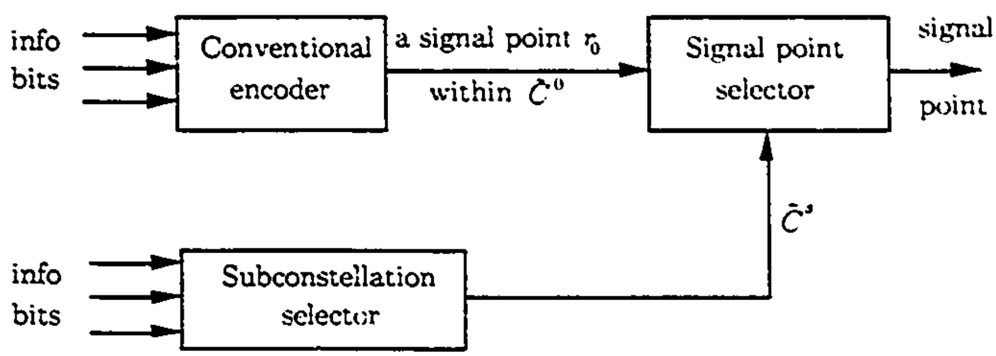


Figure 3.4: Block diagram of a modulation system for nonequiprobable signalling as suggested by Calderbank & Ozarow.

Now, consider the properties of the $\mathcal{G}_{8,2}$ code. This code contains all the binary codewords of length 8 with a Hamming weight less or equal to 2. Since there are $\binom{8}{0} = 1$ codewords with Hamming weight 0, $\binom{8}{1} = 8$ codewords of unit weight and $\binom{8}{2} = 28$ codewords of weight 2, the total number of codewords in $\mathcal{G}_{8,2}$ is 37. Since the codewords are selected equiprobably, the codewords would each convey $\log_2 37 = 5.209$ bits of information, or $(\log_2 37)/8 = 0.651$ bits of information for transmitted bit. Thus the code rate $r(C)$ of the shaping code $\mathcal{G}_{8,2}$ is

$$r(\mathcal{G}_{8,2}) = 0.651 \text{ bits/symbol interval.}$$

Every time a "1" appears at the output of the subconstellation selector, one of the heavy signals will be transmitted, and for every "0", a light signal will be transmitted. Since each codeword within $\mathcal{G}_{8,2}$ is selected with equal probability, the frequency of having a "0" or "1" at the output of the subconstellation selector is equal to how often a "0" or "1" is used within the codewords of $\mathcal{G}_{8,2}$, respectively. Therefore

$$\begin{aligned} f_0 &= \frac{\binom{8}{0}8+\binom{8}{1}7+\binom{8}{2}6}{8\left[\binom{8}{0}+\binom{8}{1}+\binom{8}{2}\right]} = 0.784. \\ f_1 &= 1-f_0 = 0.216. \end{aligned} \tag{3.61}$$

For the above values of f_0 and f_1, the entropy $\overline{H}(f_0, f_1)$ is calculated to be

$$\overline{H}(f_0, f_1) = -\sum_{i=0}^{1} f_i \log_2 f_i = 0.753 \text{ bits/symbol interval.}$$

By (3.54), the biasing gain obtainable with the $\mathcal{G}_{8,2}$ code is

$$\gamma_2(\mathcal{G}_{8,2}) = \frac{2^{0.651}}{\sum_{i=0}^{1}(2i+1)f_i} = 0.40 \text{ dB.} \tag{3.62}$$

On the other hand, (3.50) gives the potential biasing gain for two signal point sets with frequencies f_0 and f_1 is

$$\gamma_2(f_0, f_1) = \frac{2^{0.753}}{\sum_{i=0}^{1}(2i+1)f_i} = 0.71 \text{ dB.}$$

The difference between the potential and the actual biasing gain is due to the low code rate $r(\mathcal{G}_{8,2})$. More complex codes can achieve a code rate closer to $H(f_0, f_1)$, which increases the biasing gain. Fig. 3.5 shows the theoretical biasing gain vs "x", a parameter used to find the optimum frequency combinations of subconstellations, and the actual biasing gain obtained by $\mathcal{G}_{8,2}$. The relation between f_0 and "x" is given as

$$f_0 = \frac{1-x}{1-x^2}. \tag{3.63}$$

Summarizing the results of the previous example, if a greedy algorithm $\mathcal{G}_{v,k}$ is used for shaping purposes, and the original signal constellation is partitioned into

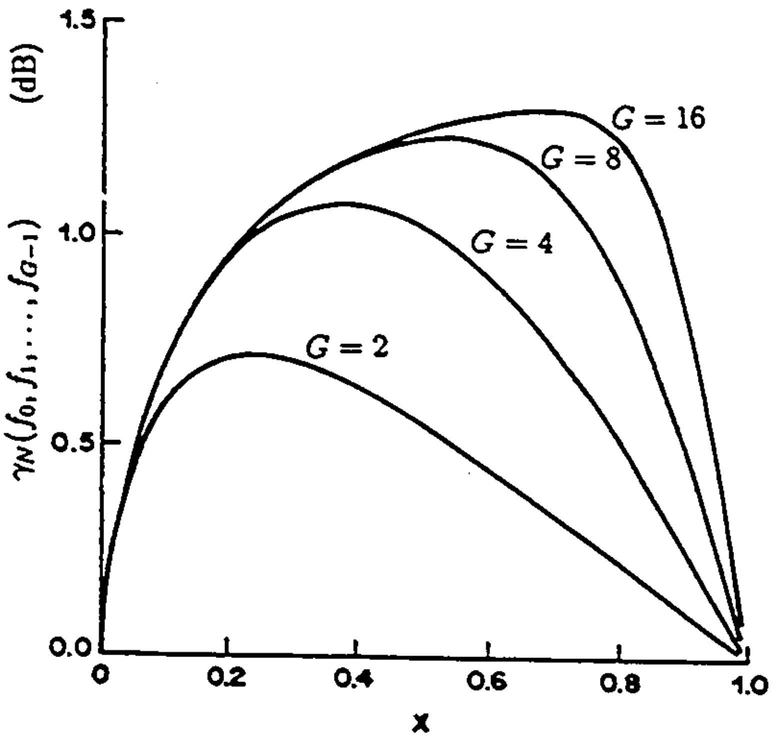


Figure 3.5: The biasing gain $\gamma_2(f_0, f_1, \cdots, f_{G-1})$ as a function of x, where $f_i = (1 - x)x^i/(1 - x^G)$; (from [2]).

two subconstellations, then the code rate and the frequencies of each subconstellation will be

$$R(\mathcal{G}_{v,\kappa}) = \frac{1}{v}\log_2\left(\sum_{j=0}^{k}\binom{v}{j}\right) \tag{3.64}$$

$$f_0 = \frac{\sum_{j=0}^{k}\binom{v}{j}(v-j)}{\left(\sum_{j=0}^{k}\binom{v}{j}\right)v} \tag{3.65}$$

$$f_1 = 1 - f_0 \tag{3.66}$$

Coset Codes Combined with the Calderbank & Ozarow Method

As described in Chapter 2, modulation schemes based on coset codes employ a partitioning of the signal constellation into subsets. While part of the information code bits select one legitimate subset, the rest of the information bits select a signal point within the previously activated subset.

Calderbank & Ozarow showed [2] that it is possible to integrate coding and shaping within a common multi-level framework. At the top of the partition tree, the least significant bits are employed for coding, while the most significant bits are used for shaping at the bottom.

Consider an L-level partition of a constellation $\tilde{C}$. $\tilde{C}$ is partitioned into t_1 level 1 subconstellations, denoted by $\tilde{C}^1$, meaning that these subconstellations are obtained by the first level code. If this corresponds to a lattice partitioning, then all $\tilde{C}^1$'s will be based on the cosets of the same lattice. As the partitioning continues each $\tilde{C}^1$ is partitioned into t_2 level 2 subconstellations, $\tilde{C}^2$, and so on. Fig. 3.6 illustrates a typical partition tree. Every branch from a subset at level i to t_{i+1}, corresponding to subsets at level $i+1$, is labeled as $0, 1, \cdots, t_{i+1}-1$. The subsets of the last level partition, $\tilde{C}^L$'s can be labeled by paths $(a_1, a_2, \cdots, a_L)$, such that each a_j, $0 \leq a_j \leq t_j - 1$. The separation between the points in the same subset is termed the *intra subset* distances

δ_i, for $0 \leq i \leq L$, where

$$\delta_i = \min_{\substack{a,b \in \tilde{C}^i \\ a \neq b}} \{d(a,b)\}, \tag{3.67}$$

in which $d(a,b)$ denotes the Euclidean distance between the two points a and b. If the Hamming distance of C_i, the code used for partitioning $\tilde{C}^{i-1}$ into $\tilde{C}^i$, is equal to d_i, then the minimum Euclidean distance of the L-level code $C = [C_1, C_2, \cdots, C_L]$ is given by

$$d(C) = \min\{d_1\delta_0, d_2\delta_1, \cdots, d_L\delta_{L-1}, \delta_L\}. \tag{3.68}$$

This partitioning is easier to explain through an example: The following problem is

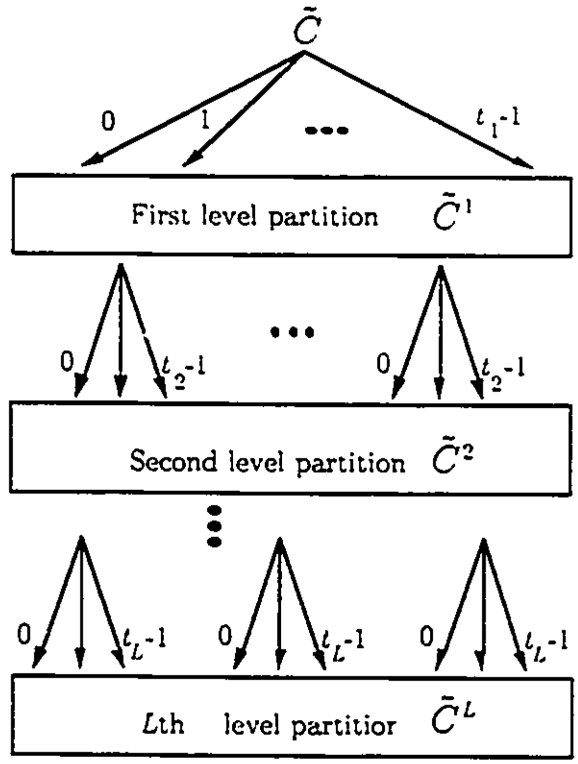


Figure 3.6: Multilevel partitioning of $\tilde{C}$.

illustrated in [2], and repeated here to give to the reader a better understanding of this concept. Fig. 3.7 shows a three level partition of a two dimensional constellation. The first two levels correspond to the lattice partition chain such as $Z^2/2Z^2/4Z^2$. If x and y denote the coordinates of the individual signal points in the constellation $\bar{C}$, then the mapping function from the output bits $(a_1, a_2, a_3, a_4, a_5)$ of the three level code to the individual signal points is given as

$$\begin{aligned} 2x &\equiv 2a_1 + 4a_3 - 1 \quad (\text{mod } 8) \\ 2y &\equiv 2a_2 + 4a_4 - 1 \quad (\text{mod } 8) \end{aligned} \tag{3.69}$$

The last symbol a_5 distinguishes light signal points ($a_5 = 0$) from the heavy ones ($a_5 = 1$). The intra subset distances are $\delta_0 = 1$, $\delta_1 = 4$, $\delta_2 = 16$, and $\delta_3 = \infty$.

In this example, the $\mathcal{G}_{8,2}$ code is used as C_3 for shaping purposes. The code C_1 is a rate 1/2 binary convolutional code with free distance d_1, and it is assumed that $d_1 < 8$. The component code C_2 is a single parity check code. The minimum distance of this three level code is calculated by (3.68) as

$$d(C) = \min\{d_1, 4d_2, 16d_3, \infty\} = d_1 \tag{3.70}$$

since the Hamming distance of a parity check code is 2 and the greedy code has a Hamming distance of 1.

C_1 and C_2 are used to obtain the coding gain where C_3 is used to obtain the shape gain. The multi-level method completely decouples the shaping bits and their mapping complexity from the remaining system. It is also important that the shaping code does not affect the minimum distance, coding gain or path multiplicity of the conventional code [2].

3.2.2 Trellis Shaping

After Calderbank & Ozarow asserted that shaping in high dimensional signal spaces is equivalent to nonequiprobable signalling in low dimensions, a new area of interest was created for researchers, namely developing efficient methods of mapping a sequence of

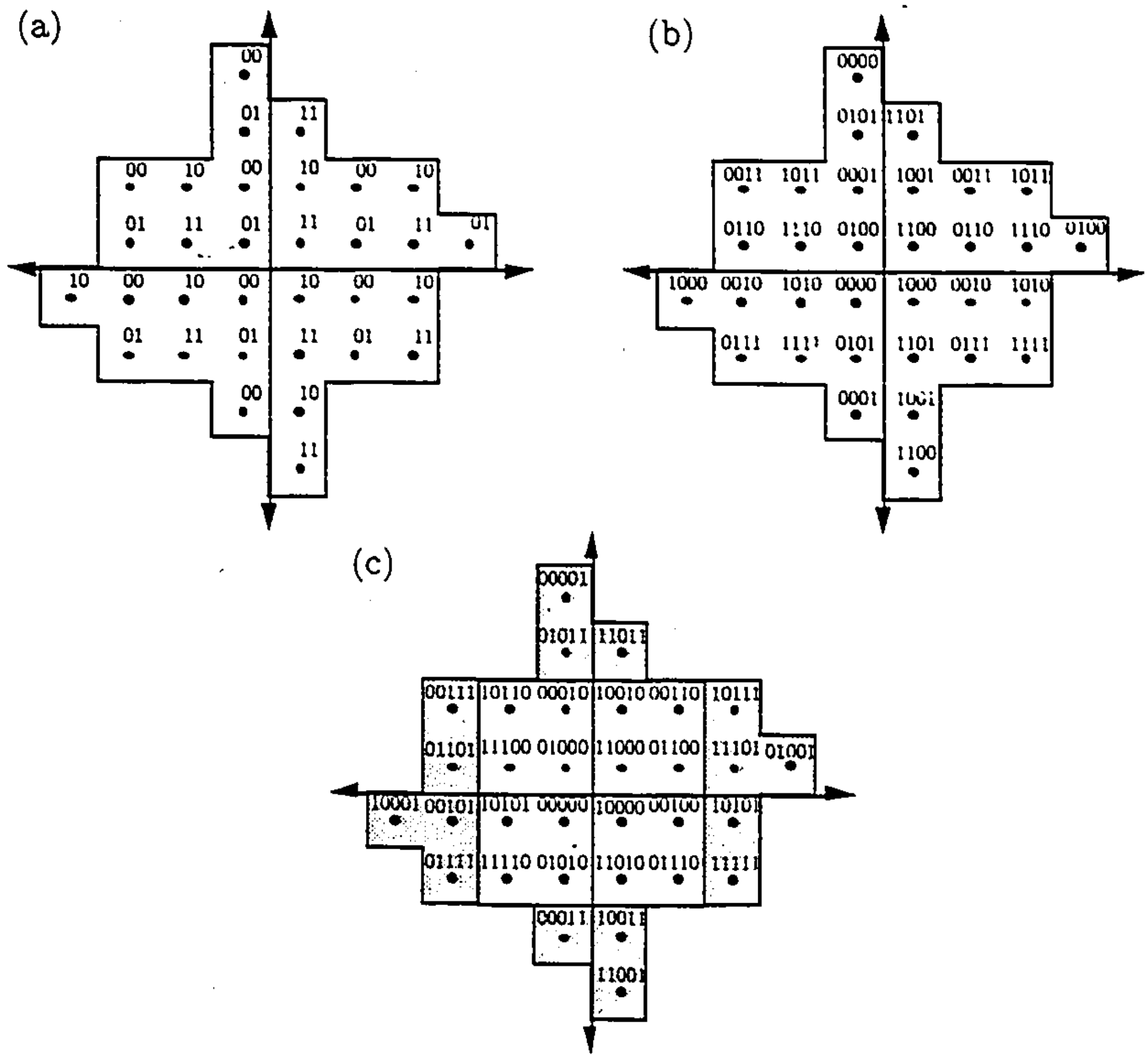


Figure 3.7: A three-level partition of a 32-point two-dimensional constellation: (a): $\mathbb{Z}^2/2\mathbb{Z}^2$; (b): $2\mathbb{Z}^2/4\mathbb{Z}^2$; (c): $4\mathbb{Z}^2/4\mathcal{R}\mathbb{Z}^2$ (the coordinates of the signal points are 1/2-integers between -4 and 4).

random input bits to the channel signals, so that signals with higher energies would be less likely to be selected than those with lower energies.

In the previous section, the CO method was described, which is based on partitioning a low dimensional constellation into G subconstellations of increasing average signal energy. Although they have suggested that no restrictions would be applied to the type of code used for shaping, they mainly investigated block codes as shaping codes. In [3], Forney introduced a new method termed *trellis shaping*. In this method, a signal constellation, $\tilde{C}$, is divided into subregions, and a convolutional code or a trellis code is used to select subregions within the constellation. As it is known, the output sequence of a convolutional or trellis code contains dependencies, among its elements, extending throughout the sequence. Hence, this sequence-oriented method can only be described in an infinite dimensional sequence space. Trellis shaping can easily achieve 1 dB of shape gain with two dimensional constellations by using 4-state encoders, as will be explained shortly. Another advantage of this method is its compatibility with coded modulation systems. The constellation expansion ratio and peak to average energy ratio can be limited to a reasonable factor by using *peak constraint* method. With trellis precoding, a method combined with coding, trellis shaping and channel equalization, the channel capacity of non-ideal Gaussian channels can be approached in principle [3]. Because of all these properties, it is worthwhile examining trellis shaping closely.

Principles of Trellis Shaping

The concept of trellis shaping is based on the coding system shown in Fig 3.8 which encodes a data source consisting of a stream of binary b-tuples. This system divides the input data word on the jth interval into three subwords $\bar{x}_j$ of k_c bits, $\bar{s}_j$ of r_s bits, and $\bar{w}_j$ of $n_u = b - (k_c + r_s)$ bits. The sequences of partial data words $\{\bar{x}_j\}$ and $\{\bar{s}_j\}$ are separately encoded to produce codeword streams $\{\bar{y}_j\}$ and $\{\bar{t}_j\}$, respectively, where codewords $\bar{y}_j$ has n_c bits, and $\bar{t}_j$ has n_s bits. The codewords in the jth signal interval are combined with the uncoded source bits in $\bar{w}_j$ to form an overall codeword

$(\bar{y}_j, \bar{w}_j, \bar{t}_j)$, which is then used to select a point a_j from the signal constellation $\bar{C}$ corresponding to the signal to be sent in the jth interval according to the mapping $m(\bar{y}_j, \bar{w}_j, \bar{t}_j) = a_j$.

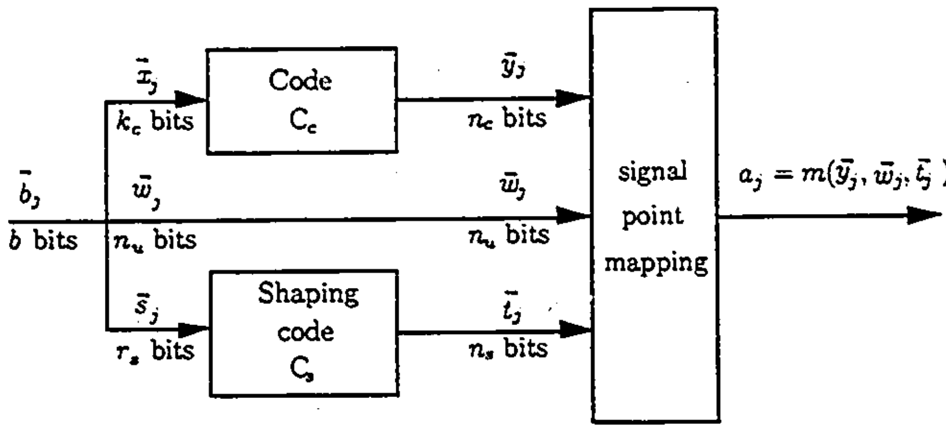


Figure 3.8: A basic modulation system with trellis shaping.

This mapping will divide the signal constellation into 2^{n_s} subregions $R(t)$ consisting of the signal points $m(\bar{y}, \bar{w}, \bar{t})$ for some n_c-tuple $\bar{y}$, n_u-tuple $\bar{w}$ and for some fixed n_s-tuple $\bar{t}$, or equally well into $2^{n_c+n_u}$ subsets $A(y, w)$, each consisting of signal points $m(\bar{y}, \bar{w}, \bar{t})$ for some fixed $\bar{y}$, $\bar{w}$ and some $\bar{t}$.

Now suppose we construct a receiver for this system. Such a receiver in the absence of noise environment would see at its input a sequence of signal points $\{a_j\} = m(\{\bar{y}_j\}, \{\bar{w}_j\}, \{\bar{t}_j\})$ and produce at its output that data word sequence $\{\bar{b}_j\} = (\{\bar{x}_j\}, \{\bar{w}_j\}, \{\bar{s}_j\})$. We recognize that due to the redundancy present in the $\{\bar{t}_j\}$ sequence it might be possible to alter the transmitted signal sequence $\{a_j\}$ by varying the sequence $\{\bar{t}\}$ of partial codewords, and still have the receiver correctly produce the transmitted data word sequence $\{\bar{b}_j\}$. If we choose such an altered sequence $\{\bar{z}_j\}$ from amongst the codewords on the basis of finding the one which produces a signal point sequence with lowest average signal energy then we may be able to lower the average power of the transmitted signal, i.e., achieve a shaping gain. The fact that the redundancy in the coded sequence $\{\bar{y}_j\}$ is unaltered may allow the value of d^2_{min} in the output sequence to be unaffected by such an adjustment in the transmitted signals.

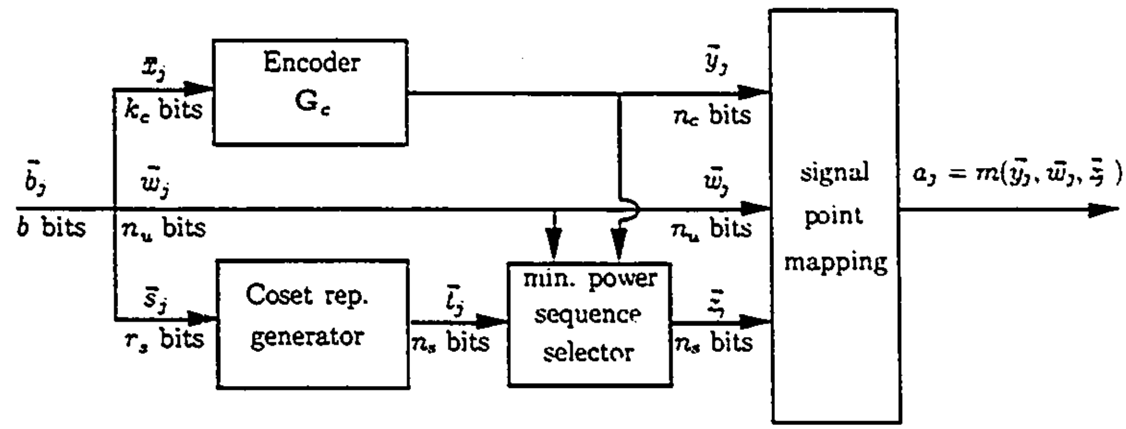


Figure 3.9: Illustration of coded modulation system with trellis shaping.

The algorithm sketched above turns out to be achievable with convolutional codes and leads to the system shown in Fig. 3.9 where $\bar{z}_j$ represents the optimal choice for replacing of $\bar{t}_j$, G_c is the generator matrix of the code C_c, G_s is the generator matrix of the code C_s, and the shaping coder is termed the "*coset representative generator*" as it will be explained shortly.

The code C_c used for coding purposes adds $r_c = n_c - k_c$ redundant bits to each codeword resulting in a constellation expansion ratio,

$$CER_c(C_c) = 2^{r_c}, \tag{3.71}$$

due to C_c only. If no trellis shaping were implemented, the signal constellation would contain $2^{n_c+n_u+r_s} = 2^{b+r_c}$ signal points , on the contrary, with shaping, there are $2^{n_c+n_u+n_s}$ points which yields a constellation expansion ratio,

$$\overline{CER}_N(\tilde{C}) = 2^{n_s-r_s} = 2^{k_s}. \tag{3.72}$$

Labelling and Partitioning of Signal Set with Trellis Shaping

Whatever method is implemented, all communications systems require " a mapping function from the input bit sequence to the signal point sequence which is completely invertible at the receiver, and does not cause any catastrophic error propagation".

In the previous figure, the modulation system with trellis shaping generates a signal sequence $\{a\} = m(\{\bar{y}\}, \{\bar{w}\}, \{\bar{z}\})$ from the input bit sequence $(\{\bar{x}\}, \{\bar{w}\}, \{\bar{s}\})$. In the channel, due to external effects the signal sequence $\{a\}$ is corrupted and some noisy sequence $\{a'\}$ is received at the receiver. The receiver tries to recover an estimate $(\{\bar{\hat{x}}\}, \{\bar{\hat{w}}\}, \{\bar{\hat{s}}\})$ of the original input sequence from the transmitter. For this procedure, a suitable partitioning of the signal set and a mapping function from the input bits $(\{x\}, \{y\}, \{s\})$ to the signal points must be defined.

First, the signal constellation $\tilde{C}$ is partitioned into 2^{n_c} subsets of $\tilde{C}$, denoted by $A_Y(\bar{y})$. Throughout this section, the capital letters Y, W, Z will denote the set of possible binary words that can be assigned to $\bar{y}, \bar{w}, \bar{z}$, respectively. Consequently, $Y \times W \times Z$ denotes the label alphabet. If the points in $\tilde{C}$ are selected within a lattice Λ, and the Λ/Λ' lattice partition is applied, then every subset $A_Y(\bar{y})$ is actually a coset of Λ' such that $A_Y(\bar{y}) = \Lambda' + \bar{c}_Y(\bar{y})$, where $\bar{c}_Y(\bar{y})$ is the coset representative selected by the n_c-tuple $\bar{y} \in Y$. If this partitioning is implemented by a convolutional code C, then the resultant output sequence $m(\{\bar{y}\}, \{*\}, \{*\})$ is a legitimate code sequence of the trellis code $C(\Lambda/\Lambda', \mathrm{C})$ ("$*$" represents any possibly valid codeword.)

The constellation $\tilde{C}$, or equivalently the finite portion of the lattice Λ, is divided into 2^{n_s} subregions labeled $R_Z(\bar{z})$. There are many ways of choosing such subregions, but the performance of the trellis shaping is not sensitive to this choice [3]. The only condition for defining $R_Z(\bar{z})$ is that every subregion $R_Z(\bar{z})$ must contain 2^{n_u} points from every subset $A_Y(\bar{y})$ of $\tilde{C}$. A new subset $A_{YZ}(\bar{y}, \bar{z})$ of $\tilde{C}$must now be defined as the intersection of the subset $A_Y(\bar{y})$ and the subregion $R_Z(\bar{z})$ such that

$$A_{YZ}(\bar{y}, \bar{z}) = A_Y(\bar{y}) \cap R_Z(\bar{z}). \tag{3.73}$$

The mapping function $m(\bar{y}, \bar{w}, \bar{z})$ selects one of the subsets $A_Y(\bar{y})$ using the n_c-tuple

$\bar{y} \in Y$ and one subregion $R_Z(\bar{z})$ using the n_s-tuple $\bar{z} \in Z$. The intersection of these two, $A_{YZ}(\bar{y}, \bar{z})$, contains 2^{n_u} points, and one of these is selected using the n_u-tuple $\bar{w} \in W$.

The partition of $\tilde{C}$ in the above described way is powerful but not practical to implement as the signal set gets larger. One systematic way of partitioning $\tilde{C}$ is the lattice partitioning.

Trellis Shaping via Lattice Partitions

The signal constellation $\tilde{C}$ and the mapping of codewords to signal points cannot be arbitrarily selected if we are to obtain shaping gain without degradation of performance in noise. One way to ensure such a degradation does not occur is to base the signal constellation on a binary lattice Λ_c and a sublattice $\Lambda_{c'}$ where $|\Lambda_c/\Lambda_{c'}| = 2^{n_c}$. The signal constellation $\tilde{C}$ is chosen to consist of points from Λ_c selected so that there are $2^{n_u+n_s}$ points from each coset of $\Lambda_{c'}$. The mapping of codewords to the signal points may then be made by having the codeword $\bar{y}_j$ select the coset from which we use the codewords $(\bar{w}_j, \bar{z}_j)$ to determine which element in the coset is sent. Regardless of how $(\bar{w}_j, \bar{z}_j)$ map to the coset elements, the coding gain provided by the code C_c is preserved. Note that this discussion would hold equally well if all points in the constellation were translated by a fixed vector $\bar{c}$.

In order to simplify an implementation, the mapping of $(\bar{w}_j, \bar{z}_j)$ to codewords needs to be further structured (especially when considering large signal sets). This is accomplished by employing a partition chain $\Lambda_c/\Lambda_{c'}/\Lambda_s/\Lambda_{s'}$ where $|\Lambda_c/\Lambda_{c'}| = |Y| = 2^{n_c}$, $|\Lambda_{c'}/\Lambda_s| = |W| = 2^{n_u}$ and $|\Lambda_s/\Lambda_{s'}| = |Z| = 2^{n_s}$ so that $|\Lambda_c/\Lambda_{s'}| = 2^{n_c+n_u+n_s}$; the constellation $\tilde{C}$ is chosen to consist of one element from each of the cosets of $\Lambda_{s'}$ in Λ_c. Thus there exists a boundary region $R_B(\tilde{C})$ which is actually a fundamental region $R(\Lambda_{s'})$ of $\Lambda_{s'}$. Furthermore all the cosets of $\Lambda_{s'}$ are possibly translated by some fixed vector $\bar{c}$, so that the constellation is in fact drawn from $\Lambda_c + \bar{c}$.

The fundamental region $R(\Lambda_{s'})$ is divided into $|\Lambda_s/\Lambda_{s'}| = 2^{n_s}$ fundamental regions $R(\Lambda_s)$ of Λ_s by assigning each of the 2^{n_s} elements of each cosets of Λ_s within $R(\Lambda_{s'})$ to a different fundamental region $R(\Lambda_s)$, consequently each $R(\Lambda_s)$ will contain $|W| = 2^{n_w}$ points from each of the $|Y| = 2^{n_c}$ cosets of $\Lambda_{c'}$. Let every such fundamental region $R(\Lambda_s)$ of Λ_s labeled by $R_Z(z)$ correspond to one of the n_s-tuple $z \in Z$. Then any coset of Λ_s in Λ_c in the lattice chain $\Lambda_c/\Lambda_{c'}/\Lambda_s$ can be expressed as the lattice Λ_s translated by some coset representative of a coset from $\Lambda_c/\Lambda_{c'}$ and a coset representative from $\Lambda_{c'}/\Lambda_s$. If $\{c_Y(\bar{y}),\ \bar{y} \in Y\}$, and $\{c_W(\bar{w}),\ \bar{w} \in W\}$ denote the coset representatives of $\Lambda_{c'}$ and Λ_s, respectively, then the cosets of $\Lambda_c/\Lambda_{c'}/\Lambda_s$ partition chain can be expressed as

$$s_{YW}(\bar{y},\bar{w}) = \Lambda_s + c_Y(\bar{y}) + c_W(\bar{w}) = \Lambda_s c_{YW}(\bar{y},\bar{w}), \tag{3.74}$$

where $c_{YW}(\bar{y},\bar{w})$ represents the coset representatives of the Λ_c/Λ_s partition. Since a fundamental region $R(\Lambda_s)$, consequently every $R_Z(\bar{z})$, contains only one point from each coset of Λ_s, the intersection of some coset $s_{YW}(\bar{y},\bar{w})$ of Λ_s and some subregion $R_Z(z)$ uniquely identifies every point within the region $R_B(\bar{C})$ such that

$$m(\bar{y},\bar{w},\bar{z}) = s_{YW}(\bar{y},\bar{w}) \cap R_Z(\bar{z}). \tag{3.75}$$

Consider the signal constellation $\bar{C}$ in Fig. 3.10, which is based on the $\mathbb{Z}^2 + (\frac{1}{2} + \frac{1}{2})$ translate of the 2-dimensional integer lattice, so that $\Lambda_c = \mathbb{Z}^2$. $\bar{C}$ contains 32 points, and is partitioned in subsets via $\mathbb{Z}^2/2\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2/4\mathcal{R}\mathbb{Z}^2$ partition chain. As expected, the boundary region of $4\mathcal{R}\mathbb{Z}^2$ (the outer rotated square) is also a fundamental region of $\bar{C}$. The numbers in the figure represent the points that are elements of the constellation and the value of these numbers represent to which coset of $|\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2| = 8$ cosets of $2\mathcal{R}\mathbb{Z}^2$ the point belongs. The signal constellation is divided into $|2\mathcal{R}\mathbb{Z}^2/4\mathcal{R}\mathbb{Z}^2| = 4$ subregions $R_Z(z)$, where each is an element of a fundamental region $R(2\mathcal{R}\mathbb{Z}^2)$ of the lattice $2\mathcal{R}\mathbb{Z}^2$ since it contains only one element from each of the 8 cosets of $2\mathcal{R}\mathbb{Z}^2$. Every point is labeled by $a = m(\bar{y},\bar{w},\bar{z})$. The 2-tuple $\bar{y}$ and 1-tuple $\bar{w}$ selects one of the 8 cosets of $2\mathcal{R}\mathbb{Z}^2$, while the 2-tuple z selects the subregion $R_Z(\bar{z})$ containing only one point from the preselected coset of $2\mathcal{R}\mathbb{Z}^2$. Fig. 3.11 illustrates $\mathbb{Z}^2/2\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partition chain which corresponds to $\Lambda_c/\Lambda_{c'}/\Lambda_s$.

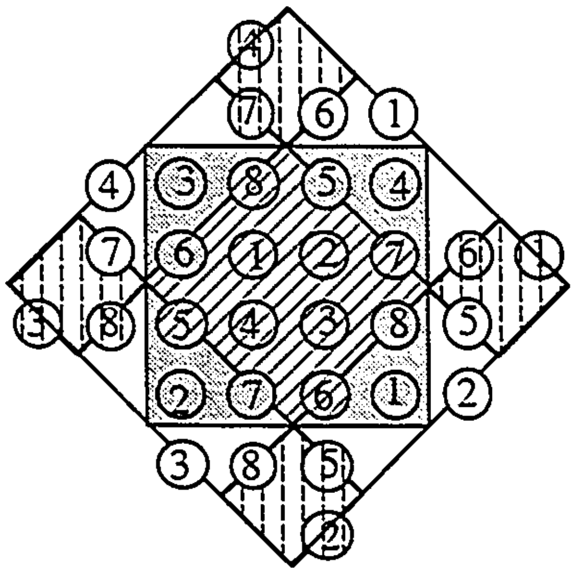


Figure 3.10: 32-point signal constellation divided into four subregions.

According to Fig. 3.11, the coset representatives of the cosets of $2\mathbb{Z}^2$ are given as

$$c_Y(y_1, y_2) = y_1(1,0) + y_2(0,1),$$

and the coset representatives of $2\mathcal{R}\mathbb{Z}^2$ in $2\mathbb{Z}^2$ are

$$c_W(\bar{w}) = w_1(0,2).$$

Consequently,

$$\begin{aligned} c_{YW}(\bar{y}, \bar{w}) &= y_1(1,0) + y_2(0,1) + w_1(0,2) \\ &= \begin{vmatrix} y_1 & y_2 & w_1 \end{vmatrix} \begin{vmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{vmatrix}, \end{aligned}$$

and according to (3.75),

$$m(\bar{y}, \bar{w}, \bar{z}) = \begin{vmatrix} y_1 & y_2 & w_1 \end{vmatrix} \begin{vmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{vmatrix} \cap R_Z(\bar{z}) + (\frac{1}{2}, \frac{1}{2}).$$

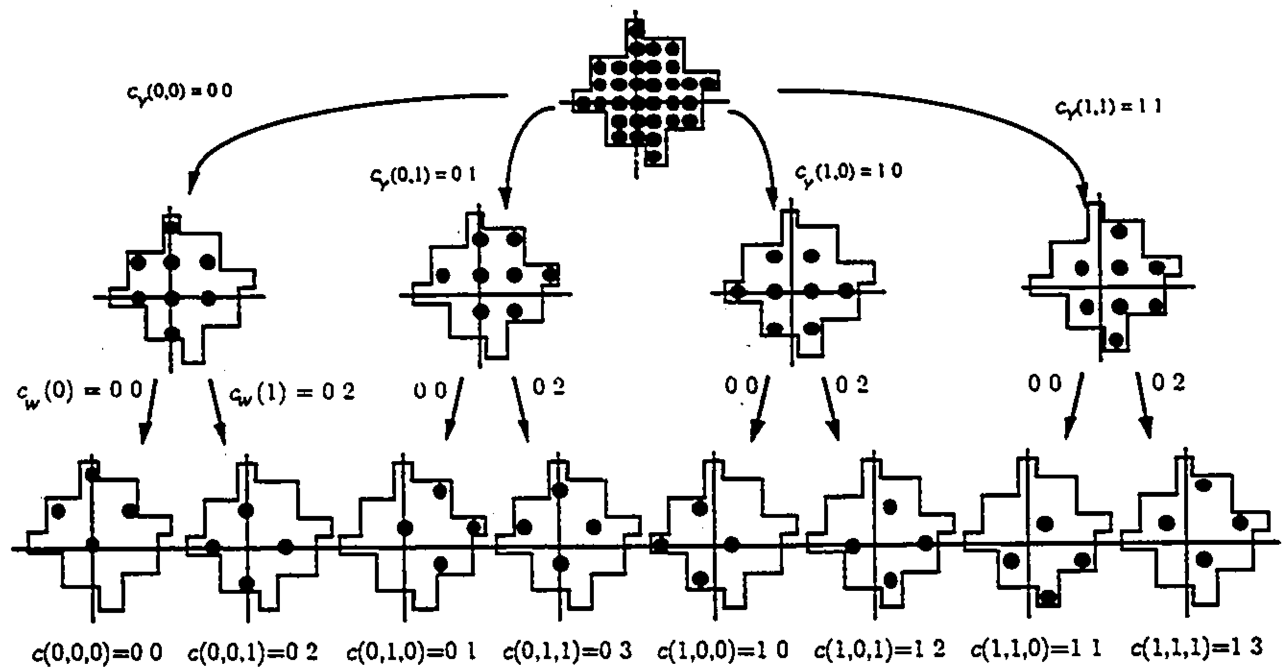


Figure 3.11: The $\mathbb{Z}^2/2\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partition chain.

The Mathematics of Trellis Shaping

The trellis shaping method employs a convolutional code $\mathbf{C}_s$ for shaping. In order to discuss this aspect of the scheme we shall first briefly review some of the relevant basic concepts and notations we shall require.

Let $\mathbf{C}_s$ denote a rate k_s/n_s binary linear convolution code. An encoder for $\mathbf{C}_s$ has a sequence of binary k_s-tuples $\{\bar{x}_s\}$ as input, and produces a sequence of binary n_s-tuples $\{\bar{y}_s\}$ as output. If we let $\bar{X}_s(D)$ denote the D-transform of $\{\bar{x}_s\}$

$$\bar{X}_s(D) \triangleq \sum_i (\{\bar{x}_s\})_i D^i, \tag{3.76}$$

where $(\{\bar{x}_s\})_i$ represents the ith elements of the sequence $\{\bar{x}_s\}$, and $\bar{Y}_s(D)$ denote the D-transform of $\{\bar{y}_s\}$, then we express the relationship between $\{\bar{x}_s\}$ and $\{\bar{y}_s\}$ as

$$\bar{Y}_s(D) = X_s(D)\mathbf{G}_s(D), \tag{3.77}$$

where $\mathbf{G}_s(D)$ is a $k_s \times n_s$ matrix termed the generator matrix of the code. The elements of $\mathbf{G}_s(D)$ are in general the ratio of two polynomials in D which both have all coefficients as either 0 or 1 (i.e., are binary polynomials).

If the elements of $\mathbf{G}_s(D)$ are simply binary polynomials, we say the code is *feedback free*. Similar to the notion of a parity check matrix for a block code, for each linear code there exists a matrix $\mathbf{H}_s^T(D)$ with binary rational polynomial elements which has the property that for any $\bar{Y}_s(D)$ from the code,

$$\bar{Y}_s(D)\mathbf{H}_s^T(D) = \bar{0}. \tag{3.78}$$

If $\dot{\bar{Y}}_s(D)$ is the D-transform of some sequence which is not an output of the code $\mathbf{C}_s$, then

$$\dot{\bar{Y}}_s(D)\mathbf{H}_s^T(D) \neq \bar{0}. \tag{3.79}$$

The matrix $\mathbf{H}_s(D)$ is referred to as a syndrome former matrix of $\mathbf{C}_s$. The matrix $\mathbf{H}_s^T(D)$ can be considered as the generator matrix of another linear convolutional code with rate r_s/n_s referred to as the dual code C_s^T of C_s. It can be shown [3] that the dual code C_s^T has the same number of states as an encoder for C_s with the fewest possible states. The matrix $\mathbf{H}_s^T(D)$ is known to have a left inverse $(\mathbf{H}_s^{-1})^T(D)$, where

$$(\mathbf{H}_s^{-1}(D))^T\,\mathbf{H}_s^T(D) = (\mathbf{H}_s(D)\,\mathbf{H}_s^{-1}(D))^T = \mathbf{I}. \tag{3.80}$$

Since the code C_s is linear, the set of possible code sequences forms a group with respect to the operation of binary sequence addition denoted by $\ominus$. The set of all code sequences characterizes the code C_s, therefore, such set is also denoted by C_s. Suppose $\{\bar{t}\}$ is some sequence of n_s-tuples which is not a possible output for the code. Consider the set of sequences formed by adding this sequence to the elements of C_s

$$C_s \ominus' \{\bar{t}\} \triangleq \{\{\bar{y}_s\} \ominus \{\bar{t}\} \mid \{\bar{y}_s\} \in C_s\}. \tag{3.81}$$

If we consider any sequence $\{\bar{k}\}$ from $C_s \ominus' \{\bar{t}\}$, i.e., $\{\bar{k}\} = \{\bar{y}_s\} \ominus \{\bar{t}\}$ for some $\{\bar{y}_s\} \in C_s$, and its D-transform $\bar{K}(D)$, we may form the syndrome of $\{\bar{k}\}$ as

$$\bar{S}(D) = \bar{K}(D)\mathbf{H}_s^T(D) = \bar{T}(D)\mathbf{H}_s^T(D), \tag{3.82}$$

since $\bar{Y}_s(D)\mathbf{H}_s^T(D) = \bar{0}$. The quantity $\bar{S}(D)$ is the D-transform of a sequence $\{\bar{s}\}$ referred to as the syndrome sequence of $\{\bar{k}\}$. From (3.80) and (3.82) it is possible to find a sequence $\{\bar{t}\}$ corresponding to a given syndrome sequence $S(D)$ through

$$\bar{T}(D) = S(D)(\mathbf{H}_s^T(D))^{-1} = S(D)(\mathbf{H}_s^{-1}(D))^T. \tag{3.83}$$

Lemma 3.7 If a sequence $\{\bar{t}\}$ is not an element of the set of convolutional code sequences C_s, then $C_s \ominus' \{\bar{t}\}$ is a coset of C_s in the space of all sequences of binary n_s-tuples. The sequence $\{\bar{k}\}$ is an element of $C_s \ominus' \{\bar{t}\}$ if and only if

$$\bar{K}(D)\mathbf{H}_s^T(D) = \bar{S}(D) = \bar{T}(D)\mathbf{H}_s^T(D).$$

Proof: It is clear from the earlier discussion that $C_s \ominus' \{\bar{t}\}$ is a coset of C_s, and that if $\{\bar{k}\} \in C_s \ominus' \{\bar{t}\}$ then $\bar{K}(D)\mathbf{H}_s^T(D) = \bar{S}(D)$. All that remains to show is that if $\bar{K}(D)\mathbf{H}_s^T(D) = \bar{S}(D)$, then $\{\bar{k}\} \in C_s \ominus' \{\bar{t}\}$. We can decompose the sequence $\{\bar{k}\}$ into sequences such that $\{\bar{k}\} = \{\bar{q}\} \ominus \{\bar{t}\}$, consequently,

$$\begin{aligned} \bar{K}(D)\mathbf{H}_s^T(D) &= \bar{T}(D)\mathbf{H}_s^T(D) + \bar{Q}(D)\mathbf{H}_s^T(D) = \bar{S}(D), \\ &= \bar{S}(D) + \bar{Q}(D)\mathbf{H}_s^T(D) = \bar{S}(D). \end{aligned} \tag{3.84}$$

Equation (3.84) is true only if $\{\bar{q}\} \in C_s$ so that $\{\bar{k}\} \in C_s \ominus' \{\bar{t}\}$.

Because of Lemma 3.7, we see that the relation

$$\bar{K}(D) = \bar{S}(D)(\mathbf{H}_s^{-1}(D))^T,$$

represents a means of assigning a coset representative to each of the cosets of C_s, hence, $(\mathbf{H}_s^{-1}(D))^T$ corresponds to a convolutional code we may term a coset representative generator.

Now consider a receiver which decodes the received signal by first determining an estimate of which signal point $\{a\} = m(\{\bar{y}\}, \{\bar{w}\}, \{\bar{z}\})$ was sent, which is then used to determine an estimate $(\{\hat{\bar{y}}\}, \{\hat{\bar{w}}\}, \{\hat{\bar{z}}\})$ of the transmitted sequence of code words $(\{\bar{y}\}, \{\bar{w}\}, \{\bar{z}\})$. The elements of these codewords are then used to estimate the transmitted data words $(\{\bar{x}\}, \{\bar{w}\}, \{\bar{s}\})$ as follows:

$$\hat{\bar{X}}(D) = \hat{\bar{Y}}(D)\mathbf{G}_c^{-1}(D), \tag{3.85}$$

$$\hat{\bar{W}}(D) = \hat{\bar{W}}(D), \tag{3.86}$$

$$\hat{\bar{S}}(D) = \hat{\bar{Z}}(D)\mathbf{H}_s^T(D). \tag{3.87}$$

If $\{\hat{\bar{y}}\} = \{\bar{y}\}$, then

$$\hat{\bar{X}}(D) = \bar{Y}(D)\mathbf{G}_c^{-1}(D) = \bar{X}(D), \tag{3.88}$$

resulting in $\{\hat{\bar{x}}\} = \{\bar{x}\}$. Clearly, if $\{\hat{\bar{w}}\} = \{\bar{w}\}$ then there would be no error in decoding the $\{\bar{w}\}$ codeword sequence. We observe from (3.87), that if $\{\hat{\bar{z}}\} = \{\bar{z}\}$, then $\{\hat{\bar{s}}\} = \{\bar{s}\}$, provided that

$$\bar{Z}(D) = \bar{S}(D)(\mathbf{H}_s^{-1}(D))^T + \bar{Y}_s(D), \tag{3.89}$$

where $\bar{Y}_s(D)$ corresponds to the D-transform of any sequence $\{\bar{y}_s\}$ with $\{\bar{y}_s\} \in C_s$. Thus, the set of codeword sequences for $\{\bar{z}\}$ that would not cause a decoding failure are precisely a coset of the code C_s, for which the sequence that corresponds to $\bar{S}(D)(\mathbf{H}_s^{-1}(D))^T$ is a coset generator. The selection of a sequence to minimize the transmitted power via trellis shaping is thus a question of choosing the best sequence from the coset of C_s selected by the encoder $(\mathbf{H}_s^{-1}(D))^T$. As a result of this, the trellis shaping system takes on the form shown in Fig. 3.12.

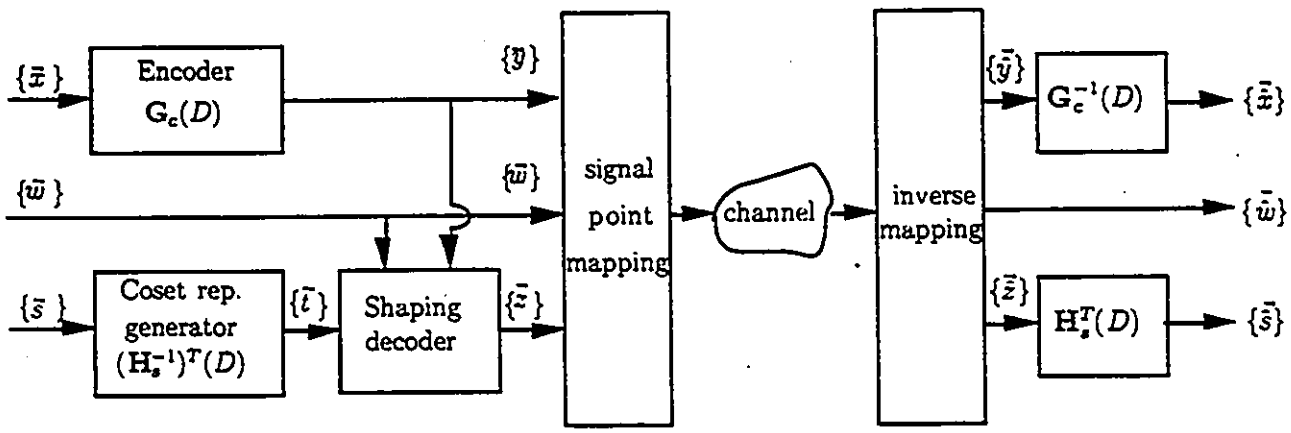


Figure 3.12: A complete communications system with trellis shaping.

It is important to note that if $G_c^{-1}(D)$ and $H_s^T(D)$ are feedback free, then the effects of errors in selecting estimates $(\{\hat{\bar{y}}\}, \{\hat{\bar{w}}\}, \{\hat{\bar{z}}\})$ are limited in duration in the output, and this would prevent the possibility of catastrophic errors. The remaining issue to be discussed is how to determine the appropriate elements of the coset.

Recall that the objective is to minimize the transmitted average power. This is equivalent to finding the sequence in the coset with minimal energy as we proceed. Since the energy of any sequence of signal points is just the sum of energies of the component points, the sequence to send can be found by applying the principles of dynamic programming, i.e., by using the Viterbi algorithm with appropriate path metrics. This task is accomplished by the circuit termed *shaping decoder*.

The Shaping Decoder

The shaping decoder is used to select the most appropriate sequence within the sequence coset, selected by the input sequence, according to any criterion, set by the designer. In this thesis, as mentioned previously, the criterion is to minimize the average signal energy of the transmitted symbol sequence. Based on the introduction given in the previous section, the set of the sequences

$$j(\{\bar{y}\},\{\bar{w}\},\{\bar{t}\}) = \{\{a\} = m(\{\bar{y}\},\{\bar{w}\},\{\bar{z}\}),\ \ \{\bar{z}\} \in \mathrm{C}_s \ominus' \{\bar{t}\}\} \tag{3.90}$$

can be regarded as an equivalence class of possible transmitted sequences, each corresponding to the same input sequence $(\{\bar{x}\},\{\bar{w}\},\{\bar{s}\})$.

Actually, $j(\{\bar{y}\},\{\bar{w}\},\{\bar{t}\})$ is a set of sequences whose elements $a = m(\bar{y},\bar{w},\bar{z})$ are modified versions of the elements of the legitimate code sequences of C_s, where $a = (\bar{y},\bar{w},\bar{y}_s \ominus \bar{t})$ corresponds to $\bar{y}_s$. Therefore the trellis diagram of C_s can also be used to represent the sequences of the equivalence class $j(\{\bar{y}\},\{\bar{w}\},\{\bar{t}\})$, by modifying the signals assigned to the branches by

$$a_j = m(\bar{y}_j,\bar{w}_j,\bar{y}_{s,j} \ominus \bar{t}_j).$$

By utilizing a Viterbi decoder, the minimum energy path through the trellis diagram can be found by assigning the norms, $\|a_j\|^2$, of the branch signals as the cost of the corresponding branches. Since this trellis diagram will be the same as that of the code C_s, the complexity of the system will be equal to the complexity of the decoder of C_s.To clarify this process consider the trellis diagram of the shaping decoder shown in Fig. 3.15.

Fig. 3.13 demonstrates a modulation system equipped with trellis shaping. A two-state, rate 1/2 convolutional code with generator matrix $\mathrm{G}_c(D) = [1 + D^2, D]$ is used for error control coding purposes. The elements of the bit sequences $\{\bar{y}\}$ and $\{\bar{w}\}$ select one of the cosets of $\mathbb{Z}^2/2\mathbb{Z}^2$ and $2\mathbb{Z}^2/2\mathcal{R}\mathbb{Z}^2$ partitions, respectively. The 32 point signal constellation shown in Fig 3.10 is used to select the signal points which form the output symbol sequence of the modulation system. The shaping code has the

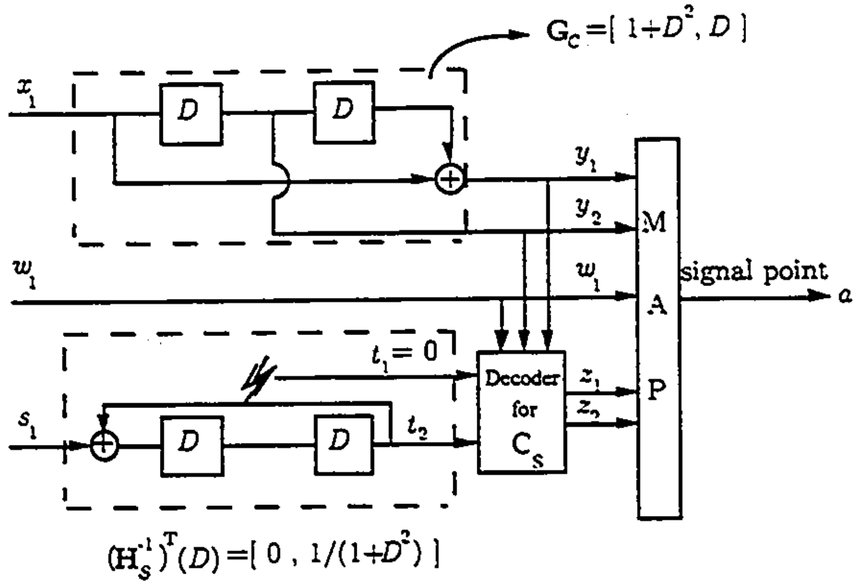


Figure 3.13: An example modulation system with trellis shaping.

same generator matrix $G_s = G_c$ for which a syndrome former matrix is found to be $H_c(D) = [D, 1+D^2]$, which reveals the following as corresponding coset representative generator circuits:

$$
\begin{aligned}
(H_s^{-1})^T &= [D, 1], \\
(H_s^{-1})^T &= [1/D, 0], \\
(H_s^{-1})^T &= [0, 1/(1+D^2)].
\end{aligned}
$$

The first circuit represents a feedback free coset representative generator, while the latter two are not and must be checked to ensure they do not cause any catastrophic error propagation. These two circuits generate a sequence of 2-tuples where one bit of each element of these sequences is always zero. In Fig. 3.13, the third coset representative is applied, which will be used in the remainder of this example.

In Fig 3.10, the signal constellation is divided into four subregions $R_Z(z)$, each containing one point from each of the 8 cosets of $2\mathcal{R}\mathbb{Z}^2$, and y_1, y_2 and w_1 select one of these cosets at every symbol interval. At the same time, the coset representative generator selects one of the subregions: The intersection point of the selected coset and the selected subregion correspond to a candidate signal point. The sequence formed by these candidate signals enters the decoder for C_s, where an equivalent sequence to the candidate sequence is selected with minimized average signal energy. Fig 3.14

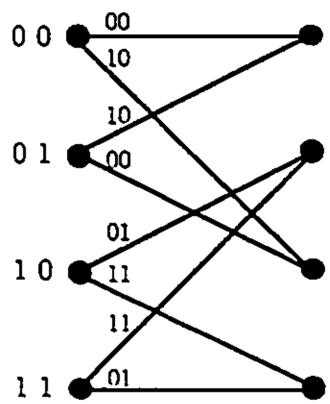


Figure 3.14: Trellis diagram of the convolutional code C_s.

shows the trellis diagram of C_s and Fig 3.15 illustrates the modified trellis diagram for the decoder for C_s. The difference between both of these diagrams lies in the branch labels. In the first diagram, the branches are n_s-tuple binary words corresponding to y_s. In the second one, they are the norms of the signals corresponding to $a(y, w, t \ominus y_s)$.

A simulation of these systems revealed that the required average signal energy was 2.197. Without employing trellis shaping, the system would require a 16-point constellation with an average signal energy equal to 2.5; therefore, the total shape gain obtained is

$$\gamma_t(C_s) = \frac{2.5}{2.197} = 0.56 \text{ dB}.$$

Fig. 3.16 shows the probability distribution of points in a one-dimensional constituent constellation induced by the two-dimensional constellation shown in Fig.

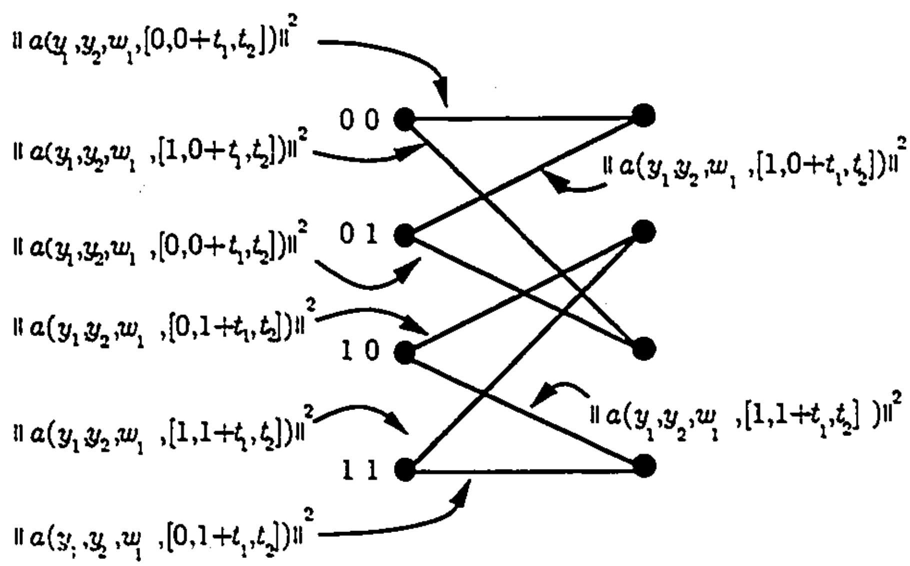


Figure 3.15: Modified trellis diagram of the shaping code for C_s decoder.

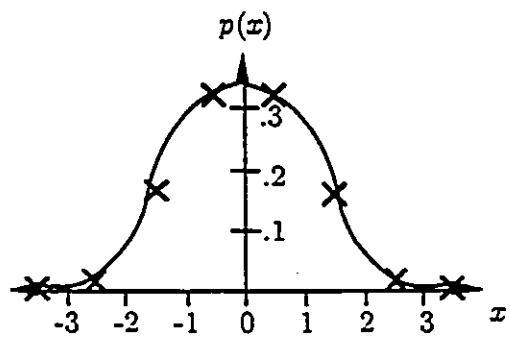


Figure 3.16: Probability distribution $p(x)$ on points in the constituent one dimensional constellation induced by 2-dimensional nonequiprobable signalling by the encoder given in the previous example.

3.10. Clearly this distribution resembles a truncated Gaussian distribution. The simulation results revealed that the nonequiprobable signal constellation, due to trellis shaping, has an entropy $\overline{H}(\tilde{C}) = 4.15$, although each symbol actually conveys only 4 bits of information. This 0.15 bits per two dimensions of unutilized constellation expansion equates to a loss of

$$\left(2^{0.15}\right)^{2/N} = 2^{0.15} = 0.45 \text{ dB}$$

in average signal energy. This accounts for part of the difference between the actual shape gain of 0.56 dB and the ultimate shape gain of 1.53 dB. The reason for the remaining difference is the small size of the signal constellation. If we examine Fig. 3.16, we see that we only have four different probability levels; thus our nonequiprobable constellation corresponds to that suggested by Calderbank & Ozarow, with equally-sized four subconstellations. From Fig 3.5, such a constellation can theoretically achieve 1.06 dB total shape gain, which is approximately equal to the summation (in dB) of the gain we obtained by trellis shaping and the loss of gain caused by the unutilized constellation expansion.

The choice of the shaping code has some effects on shaping gain as well. The same simulation was repeated with a different $\mathbf{C}_s$ characterized by

$$\begin{aligned} \mathbf{G}_s &= [1 + D^2, 1 + D + D^2], \\ \mathbf{H}_s^T &= [1 + D + D^2, 1 + D^2], \\ \left(\mathbf{H}_s^{-1}\right)^T &= [1 + D + D^2, D^2]. \end{aligned}$$

This produced a shaping gain of 0.64 dB and an entropy of the signal constellation of 4.14 bits/symbol. The excess amount of entropy corresponds to 0.42 dB in average signal energy. Again, the sum of 0.42 dB and 0.64 dB gives 1.06 dB, in accordance with our expectations.

The simulation was repeated with larger signal constellations, and by utilizing a 256-point constellation, 0.95 dB of shape gain was obtained.

Earlier we stated that PAR is an important criteria in the performance of the transmission systems. The peak to average energy ratio for the system in Fig. 3.13 with $\mathbf{G}_s = [1 + D^2, 1 + D + D^2]$, is calculated as

$$\overline{PAR}_N(\bar{C}) = \frac{50/4}{2.15} = 5.81.$$

This ratio is above the acceptable limits (square shaped base constellation has a PAR value of 3). In order to decrease PAR, we can use the peak constraint method, which suggests introducing additional constraints on the signal selection in order to decrease PAR with a small penalty in shape gain. Prohibiting the selection of some high energy signals decreases PAR, whereas it affects the shape gain very little since such signals are very rarely selected by the system. Consequently, if we avoid using the four highest energy points for the system in Fig. 3.13 with $\mathbf{G}_s = [1 + D^2, 1 + D + D^2]$, we see that

$$\gamma_t(\mathrm{C}_s) = \frac{2.5}{2.157} = 0.64 \text{ dB},$$
$$\overline{PAR}_2(\mathrm{C}_s) = \frac{34/4}{2.157} = 3.94,$$

and if we avoid using the eight highest energy signals,

$$\gamma_t(\mathrm{C}_s) = \frac{2.5}{2.165} = 0.62 \text{ dB}$$
$$\overline{PAR}_2(\mathrm{C}_s) = \frac{26/4}{2.165} = 3.00.$$

Chapter 4

SPECTRAL ANALYSIS OF CODED DIGITAL SIGNALS

In the preceding chapters techniques which attempt to improve the performance of communication systems in the presence of noise were discussed. Performance in noise is, however, not the only factor of importance. As discussed in Chapter 1, the spectral occupancy of signals being transmitted in a communications system in comparison to the information rate is also a concern. In this chapter methods for determining the spectral occupancy of signals associated with the signalling techniques we have being discussing, based on the power spectral density function of the signals, will be discussed. The power spectral density function provides the distribution in frequency of the signals time-average mean power. These techniques will be employed in Chapter 5 to determine the spectral occupancy of signals associated with coset coding.

4.1 Power Spectral Density Function of Markov Chain Driven Signals

A *Markov chain model* is a model which describes the characteristics of a random process that obeys Markovian probabilistic laws. Markovian statistics require that given the values of a random process in the present and in the past, the statistical description of the possible future behavior of this process depends only on the most recent value of the process available. This can be interpreted as the statement that all the influence of the processes past values on its future values is contained in the most recent available state of the process [30]. Such a model is able to describe most of the output signal sequences of a telecommunications system generated by a random input bit sequence, and can be used as a framework to compute the power spectral density function.

4.1.1 Properties of Markov Chain

First, some terms associated with the discussion of a Markov chain modelmust be defined.

Definition. The *period* μ_i of a state i of a Markov chain is defined to be:

$$\mu_i \triangleq g.c.d.\left\{n \,|\, P_{i,i}^{(n)} > 0 \quad n \geq 1\right\}, \tag{4.1}$$

where $g.c.d.$ refers to the greatest common divisor of the set and $P_{i,i}^{(n)}$ refers to the n-step transition probability from state i to itself, i.e., starting from state i, the probability of returning to the same state i after n transitions. If the period of a state i is equal to 1, then the state is termed *aperiodic*; otherwise it is said to be *periodic*. ∎

Definition. A Markov chain is said to be *irreducible*, if for all states i and j, there is some n such that $P_{i,j}^{(n)} > 0$, i.e., every state can be reached from every other state. ∎

If a chain is irreducible, all the states in the chain have the same period μ, and the number of transitions in any path from one given state to another given state is unique modulo μ. if we let $l_{i,j}$ denote the number of transitions modulo μ, in any path from state i to the state j, then we find that

$$\begin{aligned} l_{i,i} &= 0, \\ l_{i,j} + l_{j,k} &= l_{i,k}, \qquad \text{(modulo } \mu). \end{aligned} \tag{4.2}$$

The set of states can be partitioned into μ subsets $B_1, B_2, \cdots, B_\mu$, such that within any subset $l_{i,j} = 0$, where i and j belong to the same subset. Also, any path from state i, $i \in B_x$, to any state j, $j \in B_y$, requires $l_{i,j} = y - x$ transitions.

Example 1: The chain illustrated in Fig. 4.1 is irreducible, because every state can be reached from every other state. It is also aperiodic since diverging from any state, after three or four transitions, it is possible to return to the initial state, yielding

$$\mu_i = g.c.d.\,\{3, 4, \cdots\} = 1. \qquad \text{for every } i$$

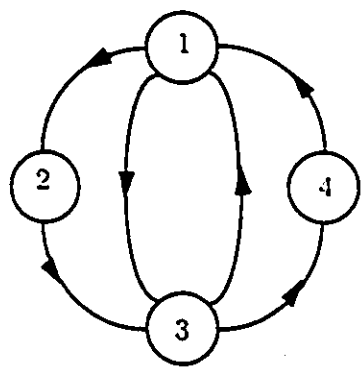


Figure 4.1: A four state Markov chain.

Example 2: Fig. 4.2 shows an irreducible chain with $\mu = 2$. The letters p and q assigned to the individual branches denote the transition probability in the direction indicated by the branch arrow. Since $l_{1,3} \equiv 0 \bmod 2$ and $l_{2,4} \equiv 0 \bmod 2$, the states $1, 3$ are in the same subset B_1 and the states $2, 4$ are in B_2. The state transition matrix, i.e., the matrix whose i, j-th element is $P_{i,j}^{(1)}$ is given by

$$\mathbf{P} = \begin{vmatrix} 0 & q & 0 & p \\ q & 0 & p & 0 \\ 0 & p & 0 & q \\ p & 0 & q & 0 \end{vmatrix}$$

The m-step transition matrix $\mathbf{P}^{(m)}$, i.e., the matrix whose i, j-th element is $P_{i,j}^{(m)}$

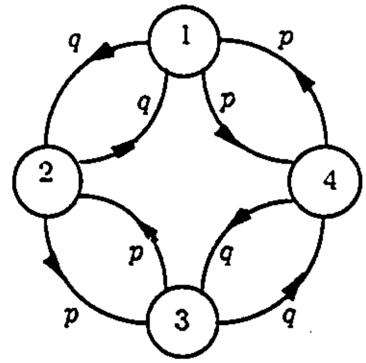


Figure 4.2: Example 2, state diagram.

is calculated by taking the m^{th} power of $\mathbf{P}$. For this example $\mathbf{P}^{(m)}$ is calculated as [6]

$$\mathbf{P}^{(m)} = \begin{cases} \frac{1}{2}\begin{bmatrix} 1+(p-q)^m & 0 & 0 & 1-(p-q)^m \\ 0 & 1+(p-q)^m & 1-(p-q)^m & 0 \\ 0 & 1-(p-q)^m & 1+(p-q)^m & 0 \\ 1-(p-q)^m & 0 & 0 & 1+(p-q)^m \end{bmatrix}, & \text{for } m \text{ even,} \\ \frac{1}{2}\begin{bmatrix} 0 & 1-(p-q)^m & 1+(p-q)^m & 0 \\ 1-(p-q)^m & 0 & 0 & 1+(p-q)^m \\ 1+(p-q)^m & 0 & 0 & 1-(p-q)^m \\ 0 & 1+(p-q)^m & 1-(p-q)^m & 0 \end{bmatrix}, & \text{for } m \text{ odd.} \end{cases} \tag{4.3}$$

Definition. For every K-state irreducible Markov chain (K is a finite integer) there exists a unique row vector $\pi = [\pi_1, \pi_2, \cdots, \pi_K]$ said to be a *stationary distribution matrix* which satisfies:

(i) $\pi\mathbf{P} = \pi$,

(ii) for all i, $0 \leq \pi_i \leq 1$, and

(iii) $\sum_{i=1}^{K} \pi_i = 1$. ∎

In general, it can also be shown [30] that the m-step transition probabilities $P_{i,j}^{(m)}$ of a irreducible Markov chain can be written in the form

$$P_{i,j}^{(m)} = \pi_j + \Delta_{i,j}^{(m)} + \epsilon_{i,j}^{(m)}, \tag{4.4}$$

where $\pi_j + \Delta_{i,j}^{(m)}$ is a periodic function of m whose value is $\mu\pi_j$ if there is a m-step transition path between the states i and j, i.e., $l_{i,j} \equiv m$, (modulo μ), but otherwise is zero. The term $\epsilon_{i,j}^{(m)}$ is one which is bounded by a geometrically decaying term (as a function of m), and can be nonzero only when the sum of the first two terms is nonzero.

In other words, the m-step transition probability $P_{i,j}^{(m)}$ converges to $\mu\pi_j$ as m becomes arbitrarily large; $\epsilon_{i,j}^{(m)}$ denotes the difference between $P_{i,j}^{(m)}$ and this asymptotic value at a given m provided that $l_{i,j} \equiv m$, (modulo μ).

4.1.2 Spectral Power Density Function of Digital Coded Signals

Let's consider a Markov chain of K states, where K is a finite number. The elements $P_{i,j}$ of the state transition matrix $\mathbf{P} = \|P_{i,j}\|$ denote the probability of a transition from state i to state j. To every state i, one of the channel signals $S_i(t)$ is assigned. At every symbol interval T, the Markov chain makes a transition from current state i to some state j, generating the j^{th} channel signal at the output of the transmitter.

The superposition of the transmitted channel signals is a sample sequence $s(t)$ in the ensemble of the random process, such that

$$s(t) = \sum_{n=0}^{\infty} s_{M_n}(t - nT), \tag{4.5}$$

where M_n is the random variable describing the state of the chain after the n^{th} transition. The spectral power density function of such a generated signal has been shown to be given by [6]:

$$\Phi(f) = \frac{1}{T}\left\{\sum_{k=1}^{K} \pi_k \left|S_k(f)\right|^2 - \mu \sum_{i=1}^{\mu} \left| \sum_{k\in B_i} \pi_k S_k(f) \right|^2 + \right.$$

$$2\Re \sum_{u,v=1}^{K} \sum_{l=1}^{\infty} \pi_u \epsilon_{u,v}^{(l)} \exp\{\jmath 2\pi f l T\} S_u(f) S_v^*(f)$$

$$+ \left| \sum_{u=1}^{K} \pi_u \exp\{\jmath 2\pi f l_{m,u} T\} S_u(f) \right|^2 \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T\mu}\right) \Bigg\} \tag{4.6}$$

or, similarly

$$\Phi(f) = \frac{1}{T}\left\{ I - \mu II + 2III + IV \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T\mu}\right) \right\}, \tag{4.7}$$

where

$$I = \sum_{k=1}^{K} \pi_k |S_k(f)|^2, \tag{4.8}$$

$$II = \sum_{i=1}^{\mu} \left| \sum_{k \in B_i} \pi_{k_i} S_k(f) \right|^2 \tag{4.9}$$

$$III = \Re \sum_{u,v=1}^{K} \sum_{l=1}^{\infty} \pi_u \epsilon_{u,v}^{(l)} \exp\{\jmath 2\pi f l T\} S_u(f) S_v^*(f), \tag{4.10}$$

$$IV = \left| \sum_{u=1}^{K} \pi_u \exp\{\jmath 2\pi f l_{m,u} T\} S_u(f) \right|^2. \tag{4.11}$$

In these equations

(i) K denotes the number of the states.

(ii) T is the symbol interval, i.e., the duration of the channel occupation by the signal $s_i(t)$ if the next state of the chain is state i.

(iii) $S_k(f)$ is the Fourier transform of the signal $s_k(t)$, that is a time limited signal, i.e. $s_k(t) = 0$ when $t \notin [0, T)$.

(iv) μ denotes the period of the chain and π_i is the i^{th} element of the stationary distribution matrix π.

(v) m is some arbitrary selected state on which the result does not depend.

(vi) B_i denotes a subclass of the states.

Let us show how equation (4.6) can be used for power density function calculations of some well-known modulation systems.

Example 3: Binary Antipodal NRZ-PAM signalling: In this case there are two channel signals: $s_1(t)$ and $s_2(t)$, which can be selected by the encoder every symbol interval T, with the property that $s_2(t) = -s_1(t)$. This system can be characterized by a two state Markov chain as shown in Fig. 4.3.

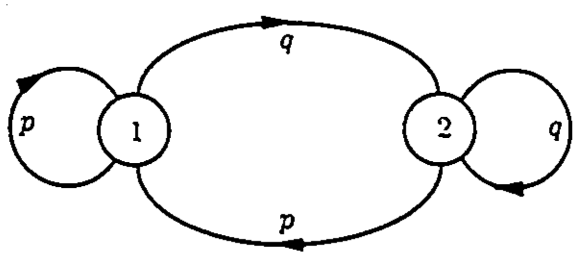


Figure 4.3: Markov chain model of Polar NRZ-PAM signalling.

Let $s_1(t)$ be the channel signal corresponding to symbol "0" in the output symbol sequence of the encoder, and $s_2(t)$ to symbol "1". The letter p denotes the probability of transmitting the symbol "1", and $q = 1 - p$ the probability that a "0" will be transmitted. Unless p or q is zero, the Markov chain representing this system is an irreducible aperiodic chain, i.e., $\mu = 1$. The probability transition matrix is given by:

$$\mathbf{P} = \begin{vmatrix} p & q \\ p & q \end{vmatrix}.$$

The m-step transition matrix is:

$$\mathbf{P}^{(m)} = \begin{vmatrix} p & q \\ p & q \end{vmatrix}^m = \begin{vmatrix} p & q \\ p & q \end{vmatrix}$$

with $\pi = [p, q]$, and

$$\epsilon_{i,j}^{(m)} = 0 \qquad \text{for all } i, j, m.$$

The power spectrum of the individual channel signals are found as

$$\begin{aligned}
|S_1(f)|^2 &= |\mathcal{F}\{s_1(t)\}|^2, \\
|S_2(f)|^2 &= |\mathcal{F}\{s_2(t)\}|^2 = |\mathcal{F}\{-s_1(t)\}|^2 = |-S_1(f)|^2 = |S_1(f)|^2,
\end{aligned}$$

and the cross multiplication of the Fourier transforms of the channel signals is

$$S_1(f)S_2^*(f) = S_1^*(f)S_2(f) = -|S_1(f)|^2;$$

inserting these equations into (4.8) through (4.11) yields

$$\begin{aligned}
I &= (p+q)|S_1(f)|^2 = |S_1(f)|^2, \\
II &= |pS_1(f) + qS_2(f)|^2 = (p-q)^2|S_1(f)|^2, \\
III &= 0, \\
IV &= II = (p-q)^2|S_1(f)|^2.
\end{aligned}$$

Finally,

$$\Phi(f) = \frac{1}{T}\left\{|S_1(f)|^2 - (p-q)^2|S_1(f)|^2 \frac{1}{T}(p-q)^2|S_1(f)|^2 \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right)\right\}.$$

Since $1-(p-q)^2 = (1-p+q)(1+p-q) = 2p2q = 4pq$, then

$$\Phi(f) = \frac{1}{T}\left\{\left[4pq + \frac{1}{T}(p-q)^2 \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right)\right] |S_1(f)|^2\right\}.$$

Coded NRZ-PAM signalling: Fig. 4.4 illustrates the block diagram of the encoder used in this example. All the symbols are selected within the $\{-1, 1\}$ alphabet. The probability of $a_n = -1$ is denoted by q, whereas the probability of $a_n = 1$ is p. There are four channel signals denoted by $s_1(t)$, $s_2(t)$, $s_3(t)$ and $s_4(t)$. The following table illustrates the relationship between the input symbols, the output symbols and the channel signals.

If $y_{n,1}$ and $y_{n,2}$ are connected to the circuit for an equal time interval , i.e., for $T/2$ seconds each time, and if the time limited signals $s(t)$ and $-s(t)$,more readily

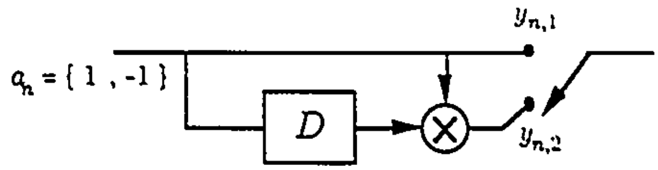


Figure 4.4: NRZ-PAM encoder.

Table 4.1: Mapping of information symbols to the channel symbols.

a_{n-1}	a_n	$y_{n,1}$	$y_{n,2}$	signal
1	1	1	-1	$s_1(t)$
1	-1	-1	1	$s_2(t)$
-1	1	1	1	$s_3(t)$
-1	-1	-1	-1	$s_4(t)$

$s(t) = 0$ for $t \notin [0, T/2)$, represent the channel signals corresponding to the symbols "-1" and "1", respectively, then the possible output signals can be written as:

$$\begin{aligned} s_1(t) &= s(t) - s(t - T/2), \\ s_2(t) &= -s(t) + s(t - T/2), \\ s_3(t) &= s(t) + s(t - T/2), \\ s_4(t) &= -s(t) - s(t - T/2); \end{aligned}$$

and the corresponding Fourier transforms are:

$$\begin{aligned} S_1(f) &= (1 - E)S(f), \\ S_2(f) &= (-1 + E)S(f), \\ S_3(f) &= (1 + E)S(f), \\ S_4(f) &= (-1 - E)S(f); \end{aligned}$$

where $E = \exp\{-j2\pi fT/2\}$. Fig. 4.5 shows the state diagram of the system given above. The corresponding matrices for this Markov chain are given below:

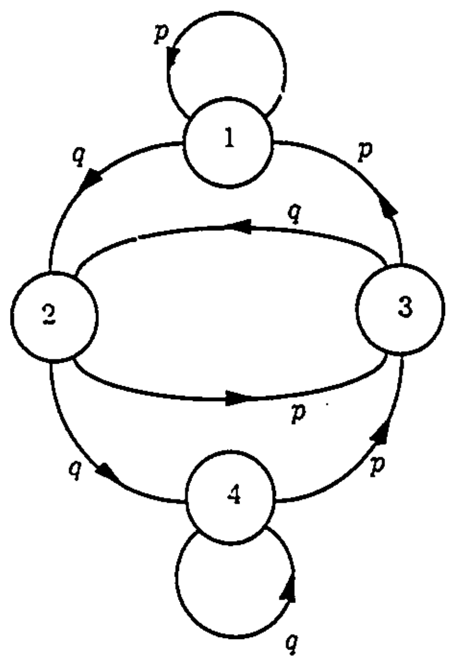


Figure 4.5: Markov chain model of Polar coded NRZ-PAM signalling.

$$\mathbf{P} = \begin{bmatrix} p & q & 0 & 0 \\ 0 & 0 & p & q \\ p & q & 0 & 0 \\ 0 & 0 & p & q \end{bmatrix} \qquad \mathbf{P}^{(n)} = \begin{bmatrix} p^2 & pq & pq & q^2 \\ p^2 & pq & pq & q^2 \\ p^2 & pq & pq & q^2 \\ p^2 & pq & pq & q^2 \end{bmatrix} \quad \text{for } n \geq 2$$

$$\epsilon^{(l)} = \begin{bmatrix} pq & q^2 & 0 & 0 \\ 0 & 0 & p^2 & pq \\ pq & q^2 & 0 & 0 \\ 0 & 0 & p^2 & pq \end{bmatrix} \qquad \epsilon^{(l)} = \|0\| \quad \text{for } l \geq 2$$

yielding

$$\pi = \left[\begin{array}{cccc} p^2 & pq & pq & q^2 \end{array} \right], \tag{4.12}$$

and, $\mu = 1$. Consequently,

$$\begin{aligned}
I &= \Big\{p^2(1-E)(1-E^*) + pq\left[(-1+E)(-1+E^*) + (1+E)(1+E^*)\right] \\
&+ q^2(-1-E)(-1-E^*)\Big\}\,|S(f)|^2 \\
&= \Big\{p^2\left[2-(E+E^*)\right] pq\left[2-(E+E^*)+2+(E+E^*)\right] + q^2\left[2+(E+E^*)\right]\Big\}\,|S(f)|^2 \\
&= \Big\{p^2\left[2-(E+E^*)\right] + 4pq + q^2\left[2+(E+E^*)\right]\Big\}\,|S(f)|^2 \\
&= \left[2+(q-p)(E+E^*)\right]|S(f)|^2 \\
II &= \Big|\Big\{p^2(1-E) + pq(-1+E+1+E) + q^2(-1-E)\Big\}\,S(f)\Big|^2 \\
&= \Big|\Big\{(p-q) - E(p-q)^2\Big\}\,S(f)\Big|^2 \\
&= (p-q)^2\left[1-E(p-q)\right]\left[1-E^*(p-q)\right]|S(f)|^2 \\
&= (p-q)^2\Big[1+(p-q)^2-(p-q)(E+E^*)\Big]\,|S(f)|^2 \\
&= (p-q)^2\Big[2(p^2+q^2)-(p-q)(E+E^*)\Big]\,|S(f)|^2 \\
III &= \Re\Big\{\Big[2pq(p-q)^2 + 2pq(q-p)E^*\Big]\exp \jmath 2\pi fT|S(f)|^2\Big\} \\
&= \Big\{2pq(p-q)^2\cos 1\pi fT + 2pq(q-p)\Re\{E^*\}\cos 2\pi fT \\
&\quad -2pq(q-p)\Im\{E^*\}\sin 2\pi fT\Big\}\,|S(f)|^2
\end{aligned}$$

Since $\Im\{E^*\} = -\sin \pi fT$ and $\Re\{E^*\} = \cos \pi fT$,

$$\Phi(f) = \frac{1}{T}\Big\{2+(q-p)(E+E^*) - (p-q)^2\Big[2(p^2+q^2)-(p-q)(E+E^*)\Big]$$

$$
\begin{aligned}
&+ \quad 4pq(p-q)\left[(p-q)\cos 2\pi fT - \cos\pi fT\right]\}\,|S(f)|^2 \\
&+ \quad \frac{1}{T}\sum_{n=-\infty}^{\infty}\delta\left(f-\frac{n}{T}\right)\left[2(p^2+q^2)-(p-q)2\cos\pi fT\right](p-q)^2|S(f)|^2 \\
&= \quad \frac{4}{T}\left\{pq(p^2+q^2)+pq(q-p)\cos\pi fT+4pq(p-q)^2\cos 2\pi fT\right\}|S(f)|^2 \\
&+ \quad 2(p-q)^2\left[p^2+q^2-(p-q)\cos\pi fT\right]|S(f)|^2\frac{1}{T}\sum_{n=-\infty}^{\infty}\delta\left(f-\frac{n}{T}\right)
\end{aligned}
$$

is the final power spectrum function of the coded NRZ-PAM modulator.

Equation (4.6) provides a very useful tool for calculating the power spectral density function of a digitally coded signal sequence. A well-written computer program can compute a significant portion of the whole spectrum in a few seconds.

4.2 Numerical Estimation of the Power Spectrum

In some cases the Markov chain model needed to describe a transmitted signal can be very complicated if the modulation memory spans several symbol periods, or the output symbol alphabet is very large. The Markov chain describing the signal consists of many states, so that drawing the state diagram and determining the transition probabilities is not feasible. In these cases a numerical estimate of the power spectrum is sought. One of the most important methods for numerically estimating the power spectrum is known as periodogram analysis. This method is based on the discrete Fourier transform of finite-length segments of samples of the signal being analysed.

4.2.1 The Periodogram

Let us consider a continuous-time signal $s(t)$, which is in fact the superposition of randomly selected channel signals generated by random data at the output of the communications systems. This random signal first passes through an anti-aliasing low pass filter, producing a new stationary random signal $x(t)$ with bandlimited power spectrum. According to the sampling theorem, every bandlimited signal can be uniquely represented by its samples taken at a rate greater than twice its bandwidth.

This $x(t)$ is sampled with a sampling rate $1/T_s$, where π/T_s corresponds to the cutoff frequency (in radians) of the anti-aliasing lowpass filter, and the sampled version of $x(t)$ is denoted by $x[n]$. Then, $x[n]$ is a discrete-time random signal whose power spectrum $P_{xx}(\Omega)$ is proportional to the power spectrum $P_{ss}(\omega)$ of the continuous signal $s(t)$, over the bandwidth of the anti-aliasing filter as given in [33]

$$P_{xx}(\Omega) = \frac{1}{T_s} P_{ss}\left(\frac{\Omega}{T_s}\right), \qquad \text{for} \qquad |\Omega| < \pi. \tag{4.13}$$

As a consequence of (4.13), a reasonable estimate of $P_{xx}(\Omega)$ will provide a reasonable estimate of $P_{ss}(\omega)$. In order to allow the power spectrum of a discrete signal to be computed, the signal must be of finite duration. The random sequence $x[n]$ is truncated into finite length segments (L samples), using a window function $w[n]$ of length L. The Fourier transform of the truncated segment $v[n] = x[n]w[n]$ is

$$V(e^{j\Omega}) = \sum_{n=0}^{L-1} w[n]x[n]e^{-j\Omega n}, \qquad \text{for} \qquad |\Omega| < \pi, \tag{4.14}$$

and an estimate of the power spectrum of $v[n]$ is given [33] as

$$I(\Omega) = \frac{1}{LU}|V(e^{j\Omega})|^2, \tag{4.15}$$

where U is a scaling factor arising from the nature of the window function $w[n]$. When the window $w[n]$ is a rectangular window, then the estimate in (4.15) is called the *periodogram* of the sequence. For other types of window it is called the *modified periodogram* [33]. In this thesis only rectangular windows are employed. In this case the scaling factor U in (4.15) is 1 and vanishes.

The N-point DFT, $V(e^{j\Omega})$, of the sequence $v[n]$ will generate a length N sequence of numbers, where each number corresponds to the Fourier transform value at a particular DFT frequency, namely a multiple of $2\pi/N$. Hence, a periodogram can estimate the power spectrum $I(\Omega)$ only at these discrete frequencies. This fact can be represented as

$$I(\Omega_k) = \frac{1}{L}|V[k]|^2, \qquad \text{where} \qquad \Omega_k = \frac{2\pi}{N}k. \tag{4.16}$$

for the rectangular window $w[n]$. In periodogram analysis the length of the window and the number of DFT points N play a very important role. As L increases the biasing factor in the estimate of the periodogram vanishes, i.e., the periodogram is asymptotically unbiased [33]. This statement suggests that L should be choosen as large as possible. On the other hand, in 1968 Jenkins and Watts ([33],pp.734) showed that as the window length L increases

$$\mathrm{var}[I(\Omega)] \simeq P_{xx}^2(\Omega). \tag{4.17}$$

Since the variance in the estimate of the power spectrum does vanish as the window length increases, the periodogram is not a consistent estimator. In other words, by increasing L and N, the consistency of the periodogram estimator does not improve. As N increases, the difference $2\pi/N$ between two adjacent DFT frequencies decreases, yielding a finer resolution but not a greater accuracy. In order to reduce $\mathrm{var}[I(\Omega)]$ to acceptable levels, a number of periodograms of different segments of the signal may be averaged.

Periodogram Averaging

A data sequence $x[n]$ which is the sampled version of the filtered channel signal $x(t)$, of length Q such that $0 \leq n \leq Q-1$, is divided into segments of length L, where overlapping of these segments is allowed:

$$x_r[n] = x[rR+n]w[n], \tag{4.18}$$

with $0 \leq n \leq L-1$. If $R < L$, the segments overlap, and if $R = L$, the segments are contiguous. The number of available segments K is the largest integer that satisfies

$$R(K-1)+(L-1) \leq (Q-1). \tag{4.19}$$

The periodogram of the r^{th} segment is given as

$$I_r(\Omega) = \frac{1}{L}|V_r(e^{j\Omega})|^2, \tag{4.20}$$

where

$$V_r(e^{j\Omega}) = \sum_{n=0}^{L-1} w[n]x_r[n]e^{-j\Omega n}. \tag{4.21}$$

Periodogram averaging consists of averaging the K periodogram estimates $I_r(\Omega)$'s, which is formulated as

$$\bar{I}(\Omega) = \frac{1}{K}\sum_{r=0}^{K-1} I_r(\Omega). \tag{4.22}$$

The discussion in [33] reveals that the average periodogram is still a biased estimate of the power spectrum where the bias vanishes as the length of the window size L increases. Interestingly, the variance of the average periodogram, provided that overlapping of the data segments does not exceed 50% of the segment length, is

$$\mathrm{var}[\bar{I}(\Omega)] = \frac{1}{K}\mathrm{var}[I_r(\Omega)] = \frac{1}{K}P_{xx}^2(\Omega). \tag{4.23}$$

Consequently, as K increases, the variance of the average periodogram approaches zero, as desired, since the power spectrum of the signal $x(t)$ is bounded. However, to increase K one either has to increase the length of the available data sequence, or (assuming $R = L$ and Q is fixed) decrease L, which reduces the resolution. Letting the segments overlap greatly can also be considered, but experiments show that selecting R less then $L/2$ does not cause a further decrease in the variance since the individual periodograms become less and less independent as the overlap increases [33].

4.2.2 Computation of the Power Spectrum of Digital Coded Signals by Periodogram Averaging

In this thesis, the interest is in the power spectrum of different modulation systems which are employed with some kind of shaping feature. For many of these systems it is possible to represent the system by its state diagram; but if trellis shaping is applied it usually is not feasible to find a practical Markov chain model. As a result, numerical estimation of the power spectrum must be done using the periodogram approach.

First, assume that the spectrum of the output sequence resides almost totally in the range $10/T$ Hz, where T is the symbol interval of the modulator. We also assume that the time-limited $-\sin\omega_c t$ and $\cos\omega_c t$ signal pair is used as the orthogonal basis signals to represent a 2-dimensional signal constellation, where w_c denotes the carrier frequency.

The continuous channel waveform corresponding to the output symbol sequence of a QAM modulator is formulated as

$$s(t) = \sum_{k=0}^{\infty} \sqrt{\frac{2\varepsilon}{T}} \left(A_{x,k}\cos\omega_c t - A_{y,k}\sin\omega_c t\right) p(t-kT), \tag{4.24}$$

where $p(t)$ is a time-limited pulse function, and ε is a constant corresponding to the signal energy of the base signals. According to [34], pp. 149, $s(t)$ can be modified as

$$s(t) = x(t)\cos\omega_c t - y(t)\sin\omega_c t, \tag{4.25}$$

where

$$\begin{aligned} x(t) &= \sqrt{\frac{2\varepsilon}{T}} \sum_{k=0}^{\infty} A_{x,k} p(t-kT), \\ y(t) &= \sqrt{\frac{2\varepsilon}{T}} \sum_{k=0}^{\infty} A_{y,k} p(t-kT); \end{aligned} \tag{4.26}$$

furthermore

$$s(t) = \Re\left[s_b(t)e^{\jmath\omega_c t}\right], \tag{4.27}$$

where $s_b(t)$ denotes the base-band representation of $s(t)$, and

$$s_b(t) = x(t) + \jmath y(t) = \sqrt{\frac{2\varepsilon}{T}} \sum_{k=0}^{\infty} (A_{x,k} + \jmath A_{y,k}) p(t-kT). \tag{4.28}$$

Due to the earlier assumption that $s_b(t)$ is bandlimited,

$$S_b(f) = 0, \qquad |f| > \frac{5}{T} = BW. \tag{4.29}$$

where $S_b(f)$ denotes the Fourier transform of $s_b(t)$ and BW is the short-form for bandwidth. According to the sampling theorem the minimum sampling rate is $2\,BW =$

$10/T$; consequently, this sampling rate reveals 10 samples in each symbol interval. For the case when $p(t)$ is a square pulse,

$$p(t) = \begin{cases} 1, & 0 \leq t < T; \\ 0, & \text{otherwise}, \end{cases} \tag{4.30}$$

the 10 samples of $x[n]$ have the same value:

$$x[n] = \sqrt{\frac{2\varepsilon}{T}}(A_{x,k} + \jmath A_{y,k}) \qquad \text{for} \qquad 10k \leq n < (k+1)10, \tag{4.31}$$

or equivalently

$$x[10r + i] = x[10r] \qquad \text{for} \qquad 0 \leq i < 10 \qquad \text{and} \qquad r \geq 0. \tag{4.32}$$

Lemma 4.8 If a discrete sequence $x[n]$ satisfies the following equation,

$$x[rp + i] = x[rp] \qquad \text{for} \qquad 0 \leq i < p \qquad \text{and} \qquad r \geq 0,$$

then the N-point DFT $\tilde{X}[n]$ of the sequence $x[n]$ is given by:

$$\tilde{X}[n] = \left(\sum_{s=0}^{h-1} x_p[s] \exp\{-\jmath\frac{2\pi}{h}s\breve{u}\}\right) C_n, \tag{4.33}$$

assuming that N is a multiple of p such that $N = h \times p$ where

$$\begin{aligned} C_n &= \sum_{i=0}^{p-1} \exp\{-\jmath\frac{2\pi}{N}ni\} \\ \breve{u} &= \text{the modulo } h \text{ reduction of } n \\ x_p[s] &= x[sp]. \end{aligned}$$

Note that the first multiplier in (4.33) is just an h-point DFT of the sequence $x_p[s]$.

Proof: The N-point DFT $\tilde{X}[n]$ of the sequence $x[n]$ is defined as

$$\tilde{X}[n] = X(e^{-\jmath\Omega_n}) = \sum_{k=0}^{N-1} x[k]e^{-\jmath\Omega_n k} \tag{4.34}$$

where $w_n = \frac{2\pi}{N}n$. Since $N = hp$, we can modify (4.34) as

$$\begin{aligned}\tilde{X}[n] &= \sum_{s=0}^{h-1}\sum_{i=0}^{p-1} x[sp+1]\exp\{-\jmath\Omega_n(sp+i)\} \\ &= \sum_{s=0}^{h-1} x[sp]\exp\{-\jmath\Omega_n sp\}\sum_{i=0}^{p-1}\exp\{-\jmath\Omega_n i\} \\ &= \sum_{s=0}^{h-1} x[sp]\exp\{-\jmath\frac{2\pi}{h}ns\}\sum_{i=0}^{p-1}\exp\{-\jmath\Omega_n i\} \qquad (4.35)\end{aligned}$$

by denoting $x[sp] = x_p[s]$,

$$\tilde{X}[n] = \sum_{s=0}^{h-1} x_p[s]\exp\{-\jmath\Omega'_n s\}\sum_{i=0}^{p-1}\exp -\jmath\Omega_n i, \qquad (4.36)$$

where $\Omega'_n = 2\pi n/h$; but also

$$\Omega'_{n+h} = 2\pi(n+h)/h = 2\pi n/h = \Omega'_n, \qquad (4.37)$$

and by replacing n by its modulo h reduced equivalent $\check{u}$, (4.35) can be rewritten as

$$\tilde{X}[n] = \sum_{s=0}^{h-1} x_p[s]\exp\{-\jmath\frac{2\pi}{h}su\}\sum_{i=0}^{p-1}\exp\{-\jmath\frac{2\pi}{N}ni\} \qquad (4.38)$$

as required.

According to Lemma 4.8, by calculating the h-point DFT of the $(A_{x,k} + \jmath A_{y,k})$ sequence, the $10h$-point DFT of $x[n]$ in (4.32) can also be calculated. For rectangular windowing the periodogram of $x[n]$ is

$$I_r(\frac{2\pi n}{N}) = \frac{1}{L}|\tilde{X}_r[n]|^2, \qquad (4.39)$$

and the average periodogram is

$$\bar{I}\left(\frac{2\pi n}{N}\right) = \frac{1}{K}\sum_{r=0}^{K-1} I_r\left(\frac{2\pi n}{N}\right) \qquad (4.40)$$

for $0 \leq n < N$.

After some experiments, the following values were found to be the most suitable for the calculations in this research:

$$\begin{aligned} h &= 1024 \\ N &= 10h = 10240 \\ L &= N = 10240 \\ K &= 100 \end{aligned}$$

Finally, all the numerical power spectrum calculations in the next chapter, are done according to the flowchart given in Fig. 4.6.

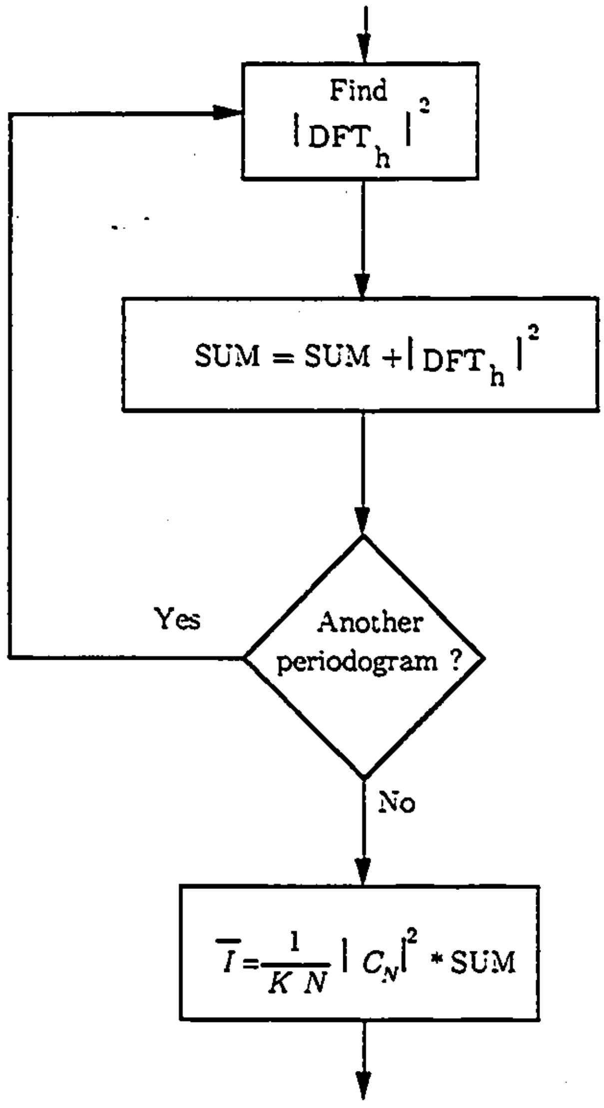


Figure 4.6: Flowchart of periodogram averaging algorithm.

Chapter 5

SPECTRAL ANALYSIS OF NONEQUIPROBABLE SIGNALLING

In previous chapters it was shown that the power efficiency of a communications system operating on a BAWGN channel could be improved by up to 1.53 dB over that of an ordinary PAM system by applying nonequiprobable signalling methods. In this chapter, the spectral characteristics of the output signal sequences generated by nonequiprobable signalling is investigated. The goal is to determine whether the shape gain achieved as a consequence of the nonuniform distribution of the signal points within a constellation causes any expansion in the transmission bandwidth.

The previously described methods introduce additional statistical dependencies between the elements of the output signal while achieving a truncated-Gaussian like distribution of signal points in a constellation. Shape gain is in fact independent of all these dependencies. If it were possible to impose such a distribution on signal points without introducing any correlation between the elements of the signal sequence, the shape gain could be perfectly achieved. Although such ideal systems do not exist in reality, the spectral analysis of their signals is much simpler than those generated by real systems, and they still provide a great deal of information that is valid for systems employing the CO method or trellis shaping. Therefore it is assumed that there exist memoryless systems that can impose a truncated Gaussian distribution on the signal points without introducing any statistical dependencies between the signals of different symbol intervals, and investigate the effect on the power spectrum caused only by the nonuniform distribution of points. Thereafter, we generate nonequiprobable signals by applying the CO method or trellis shaping, and analyse the resultant power spectrum in order to determine the effects on the spectrum due to nonuniform distributions in real systems.

5.1 Spectral Analysis of Memoryless Signalling with Nonequiprobable Constellations

In this section nonequiprobable signal constellations produced by the shaping methods suggested by Calderbank & Ozarow [2] and Livingston [20] are considered. The power spectrum of the signals generated by this manner are found by constructing a Markov chain model of the signals and then using (4.6).

The situations considered are those associated with two dimensional signal constellations. Thus an introduction of this format is given followed by a description of the examples which will be considered.

5.1.1 Signalling with Two Dimensional Signal Constellations

Any two-dimensional signal constellation may be expressed as a set of linear combinations of two orthogonal signals $p_1(t)$ and $p_2(t)$. For a signal point $s_m(t)$ on $[o, T)$, there are two constants A_{mc} and A_{ms} for which

$$s_m(t) = \sqrt{\frac{2\varepsilon}{T}}\left[A_{mc}p_1(t) + A_{ms}p_2(t)\right], \qquad \text{for } t \in [0, T). \tag{5.1}$$

These signals are identified by the ordered pair (A_{mc}, A_{ms}). In QAM, the transmitted signals have this form in which case

$$p_1(t) = p(t)\cos\omega_c t,$$

$$p_2(t) = -p(t)\sin \omega_c t, \tag{5.2}$$

where $p(t)$ is some lowpass waveform which is zero outside of the symbol interval. In its simplest form, $p(t)$ is a rectangular pulse

$$p(t) = \begin{cases} 1, & 0 \leq t < T; \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

In this case, (5.1) becomes

$$s_m(t) = \sqrt{\frac{2\varepsilon}{T}}\,(A_{mc}\cos w_c t - A_{ms}\sin w_c t)\,p(t). \tag{5.4}$$

The Fourier transform $S_m(f)$ of the signal $s_m(t)$ is calculated as:

$$\begin{aligned} S_m(f) &= \sqrt{\frac{T\varepsilon}{2}}\,\{(A_{mc}\cos[\pi(f-f_c)T] + A_{ms}\sin[\pi(f-f_c)T])\,\mathrm{sinc}[(f-f_c)T] \\ &+ (A_{mc}\cos[\pi(f+f_c)T] - A_{ms}\sin[\pi(f+f_c)T])\,\mathrm{sinc}[(f+f_c)T]\} \\ &- j\sqrt{\frac{T\varepsilon}{2}}\,\{(A_{mc}\sin[\pi(f-f_c)T] - A_{ms}\cos[\pi(f-f_c)T])\,\mathrm{sinc}[(f-f_c)T] \\ &+ (A_{mc}\sin[\pi(f+f_c)T] + A_{ms}\cos[\pi(f+f_c)T])\,\mathrm{sinc}[(f+f_c)T]\}. \end{aligned} \tag{5.5}$$

Fig. 5.1 approximates the magnitude of the Fourier transform obtained by (5.5). As

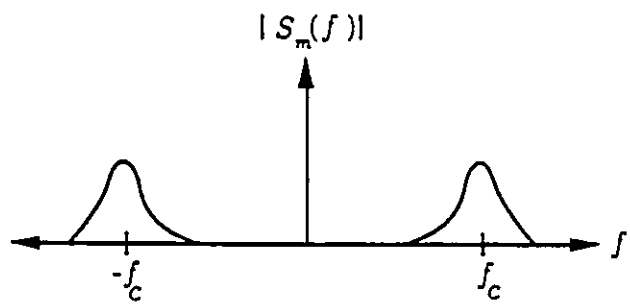


Figure 5.1: Approximate power spectrum of a modulated signal.

seen in the figure, this function is symmetric with respect to the origin of the frequency axis. Equation (5.5) can be divided into two functions of f, centred around f_c and $-f_c$, such that

$$S_m(f) = S_{m^-}(f) + S_{m^-}(f), \tag{5.6}$$

where

$$
\begin{aligned}
S_{m^+}(f) &= \sqrt{\frac{T\varepsilon}{2}}\,\{(A_{mc}\cos[\pi(f-f_c)T] + A_{ms}\sin[\pi(f-f_c)T]) \\
&\quad - \jmath\,(A_{mc}\sin[\pi(f-f_c)T] - A_{ms}\cos[\pi(f-f_c)T])\} \\
&\quad \times \operatorname{sinc}[(f-f_c)T]
\end{aligned} \tag{5.7}
$$

and

$$
\begin{aligned}
S_{m^-}(f) &= \sqrt{\frac{T\varepsilon}{2}}\,\{(A_{mc}\cos[\pi(f+f_c)T] - A_{ms}\sin[\pi(f+f_c)T]) \\
&\quad - \jmath\,(A_{mc}\sin[\pi(f+f_c)T] + A_{ms}\cos[\pi(f+f_c)T])\} \\
&\quad \times \operatorname{sinc}[(f+f_c)T]
\end{aligned} \tag{5.8}
$$

Note that $S_{m^-}(f) = S^*_{m^-}(-f)$ since $s_m(t)$ is a real signal, yielding

$$
S_m(f) = S_{m^-}(f) + S^*_{m^-}(-f). \tag{5.9}
$$

Note that for sufficiently large values of the carrier frequency, $f_cT \gg 1$, $S_{m^+}(f)$ and $S_{m^-}(f)$ do not overlap; therefore $S_{m^+}(f)$ dominates the calculations where $f > 0$, and $S_{m^-}(f)$ where $f < 0$. Inserting (5.9) for $1 \leq m \leq \Omega_{\hat{C}}$ ($\Omega_{\hat{C}}$ represents the number of signal points) into (4.6) would yield a power spectral function consisting of three main components: summation of the products in the form of $S_{m^+}(f)S^*_{n^+}(f)$, summation of products of $S_{m^-}(-f)S^*_{n^-}(-f)$, and summation of $S_{m^+}(f)S^*_{n^-}(-f)$ for $1 \leq m, n \leq \Omega_{\hat{C}}$. Clearly, the last component always yields zero since $S_{m^-}(f)$ and $S_{n^+}(-f)$ do not overlap. The first component constitutes the power spectrum for $f > 0$. and the second component for $f < 0$. Actually these two components are equal since the resultant power spectrum must be an even function of f. Therefore, the power spectrum need only be calculated for positive values of f, which depends only on $S_{m^+}(f)$. This allows $S_{m^+}(f)$ to be inserted into (4.6) as the Fourier transform of the channel signal associated with the mth state for computation of the spectral power density function (later in this chapter).

Nonequiprobable Signal Constellations

As examples, consider three nonequiprobable 2-dimensional signal constellations that are divided into two subconstellations $\tilde{C}^0$ and $\tilde{C}^1$. The two constellations $\tilde{C}^0$ and $\tilde{C}^1$ are selected with frequencies f_0 and f_1, respectively. These constellations will be used later in comparisons of nonequiprobable constellations with equiprobable ones. Before describing each constellation, two new concepts are introduced which have significant importance in later discussions:

(i) 90° rotational symmetric constellation:

If for a given signal point a_0 within the 2-dimensional constellation $\tilde{C}$, with a probability π_{a_0} of being selected, the 90° rotated version of a_0, labeled as a_1, is also a signal point within $\tilde{C}$ with probability π_{a_1} of being selected, and

$$\pi_{a_0} = \pi_{a_1}, \tag{5.10}$$

then the signal constellation $\tilde{C}$ is called a 90° rotational symmetric constellation.

(ii) 180° rotational symmetric constellation:

If for a given signal point a_0 within the 2-dimensional constellation $\tilde{C}$, with a probability π_{a_0} of being selected, the 180° rotated version of a_0, labeled as a_1, is also a signal point in $\tilde{C}$ with

$$\pi_{a_0} = \pi_{a_1}, \tag{5.11}$$

then $\tilde{C}$ has 180° rotational symmetry.

<u>64-Point $2D$ Nonequiprobable Constellation with Two Subregions:</u>

Fig 5.2 illustrates a constellation $\tilde{C}$ whose points are drawn from a translate of the integer lattice $\mathbb{Z}^2$, namely $\mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$. The 16 points with least energy form the inner

subconstellation $\tilde{C}^0$, and the remaining 48 points form the outer subconstellation $\tilde{C}^1$. Since every 90 ° rotated version of each subconstellation is identical to itself, $\tilde{C}$ has 90 ° rotational symmetry. There are three times more points in the outer constellation

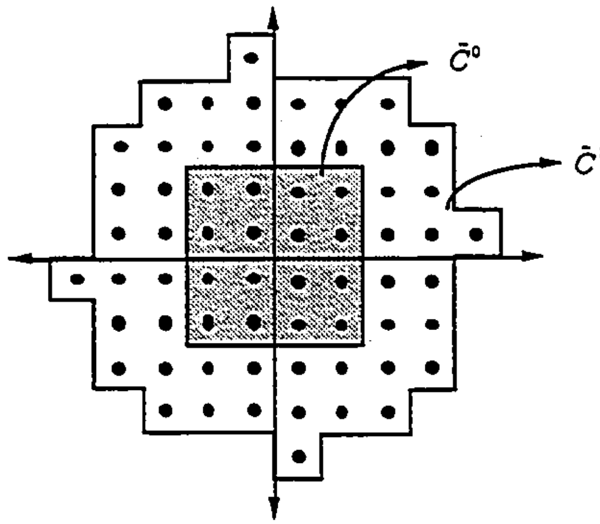


Figure 5.2: 64-point nonequiprobable signal constellation.

than the inner one; this ratio is indicated in [20] as optimum if each subconstellation is selected with equal probability such that

$$f_0 = f_1 = \frac{1}{2}. \tag{5.12}$$

The normalized average signal energy of each constellation is calculated to be

$$P_0 = \frac{5}{2} = 2.5, \tag{5.13}$$

$$P_1 = \frac{77}{6} = 12.8. \tag{5.14}$$

The normalized average signal energy of the overall constellation is (3.40):

$$\overline{P}_{av}(\tilde{C}) = P_0 f_0 + P_1 f_1 = \frac{1}{2}2.5 + \frac{1}{2}\frac{77}{6} = 7.667. \tag{5.15}$$

The entropy of the partitioned constellation was given in (3.43), and consequently,

$$\begin{aligned}\overline{H}(\tilde{C}) &= H(\frac{1}{2},\frac{1}{2}) + \log_2 16 + \frac{1}{2}\log_2 1 + \frac{1}{2}\log_2 3, \\ &= 1 + 4 + 0.792 = 5.792 \text{ bits/symbol.} \end{aligned} \tag{5.16}$$

This entropy corresponds to a $2^{\overline{H}(\tilde{C})} \approx 55$ point equiprobable constellation. Such a constellation $\tilde{C}_{eq}$ has average power $P_{av}(\tilde{C}_{eq}) = 8.755$. Consequently the total shape gain obtained by using a 64-point nonequiprobable constellation is

$$\gamma_t(\tilde{C}) = \frac{8.755}{7.667} = 1.14 = 0.58 \text{ dB}. \tag{5.17}$$

The constellation expansion ratio due to nonequiprobable signalling can be calculated via (3.45) to be

$$\overline{CER}_2(\tilde{C}) = \frac{4}{2^{1.792}} = 1.16, \tag{5.18}$$

or equivalently by taking the ratio of the number of points in the nonequiprobable constellation to that in the corresponding equiprobable constellation such that $\overline{CER}_2(\tilde{C}) = \frac{64}{55} = 1.16$. The peak to average energy ratio,

$$\overline{PAR}_2(\tilde{C}) = \frac{21}{7.667} = 2.74. \tag{5.19}$$

<u>32-Point $2D$ Nonequiprobable Constellation with Two Subregions:</u>

Fig. 5.3 shows the subconstellations $\tilde{C}^0$ and $\tilde{C}^1$. The inner constellation $\tilde{C}^0$ contains 8 points while the outer one $\tilde{C}^1$ has 24 points. As in the previous case each subconstellation is equally likely to be selected, so $f_0 = f_1 = \frac{1}{2}$. The remaining parameters of $\tilde{C}$ are as follows:

$$P_0 = \frac{3}{2}, \quad P_1 = \frac{37}{6}, \quad \overline{P}_{av}(\tilde{C}) = \frac{23}{6}, \quad \overline{H}(\tilde{C}) = 4.792. \tag{5.20}$$

The equivalent equiprobable constellation $\tilde{C}_{eq}$ contains $2^{4.792} = 28$ points, and has an average signal energy of $P_{av}(\tilde{C}_{eq}) = \frac{9}{2} = 4.5$; yielding

$$\gamma_t = \frac{9/2}{23/6} = 0.70 \text{ dB}, \tag{5.21}$$

$$\overline{CER}_2(\tilde{C}) = \frac{32}{28} = 1.14, \tag{5.22}$$

$$\overline{PAR}_2(\tilde{C}) = \frac{34/4}{23/6} = 2.22. \tag{5.23}$$

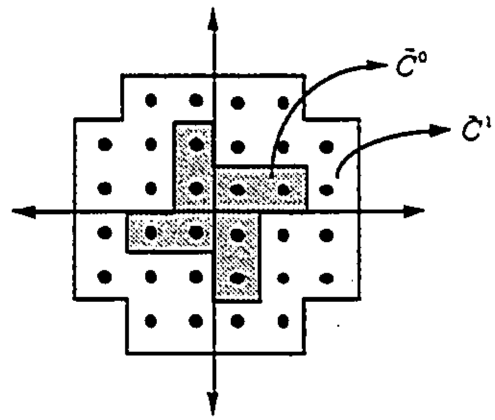


Figure 5.3: 32-points signal constellation.

16-Point $2D$ Nonequiprobable Constellation with two Subregions:
$\tilde{C}$ is illustrated in Fig. 5.4. As in the previous cases, each subconstellation $\tilde{C}^0$ and $\tilde{C}^1$ is selected with equal probability so that $f_0 = f_1 = \frac{1}{2}$. The calculation of the parameters is similar to that for the previous constellations, and this parameters are given in Table 5.1.

Equiprobable Signal Constellations

For comparison of nonequiprobable constellations with equiprobable constellations, a variety of equiprobable constellations were constructed based on the $\mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$ translate of the integer lattice $\mathbb{Z}^2$. Fig. 5.5 illustrates each of these constellations, and their parameters are summarized in Table 5.1.

5.1.2 Computation of Spectral Density Function

A memoryless transmission system combined with one of the previously described equiprobable or nonequiprobable constellations can be represented by a memoryless Markov chain model. and their power spectrums can be calculated using (4.6).

For the case of an equiprobable constellation, the Markov chain has $\Omega_{\tilde{C}}$ states

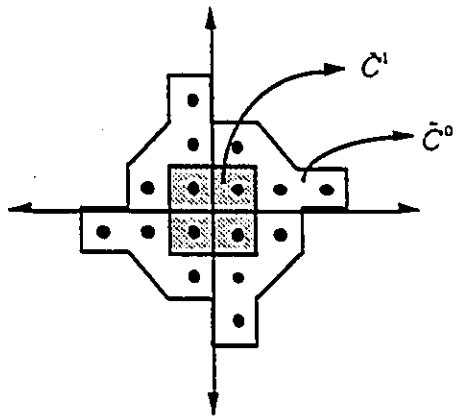


Figure 5.4: 16-point nonequiprobable signal constellation.

Table 5.1: Parameters of constructed equiprobable and nonequiprobable constellations.

Constellation	Entropy bits/symbol	Average Power $P_{av}(\tilde{C})^a$	$PAR_2(C)$
4-point EQ.	2.000	600.0	1.00
16-point EQ.	4.000	1500.0	1.80
32-pointEQ.	5.000	2400.0	1.70
55-point EQ.	5.781	3634.3	2.11
64-pointEQ.	6.000	4200.0	2.33
16-point NEQ.	3.793	1371.1	2.08
32-point NEQ.	4.793	1919.7	2.22
64-point NEQ.	5.793	3176.5	2.74

[a] $P_{av} = \frac{\varepsilon}{T}\left\{\langle A_{mc}^2\rangle + \langle A_{ms}^2\rangle\right\}$, and $\varepsilon = 1/4$.

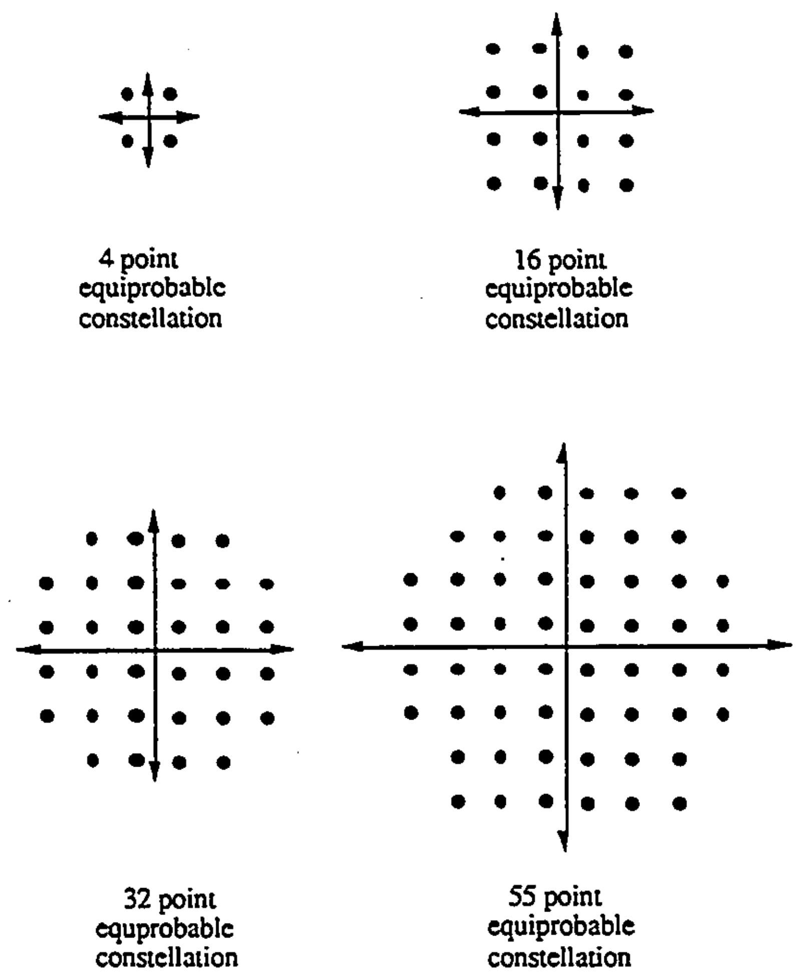


Figure 5.5: Various equiprobable signal constellations.

(one for each of the $\Omega_{\tilde{C}}$signal points) and has branches from every state to every other state which correspond to the transition probabilities $P_{i,j} = 1/\Omega_{\tilde{C}}$. This is an aperiodic ($\mu = 1$) irreducible-Markov chain with a constant stationary distribution whose elements are

$$\pi_i = \frac{1}{\Omega_{\tilde{C}}}, \qquad \text{for} \qquad 1 \leq i \leq \Omega_{\tilde{C}}. \tag{5.24}$$

Since the chain is memoryless, in (4.10), $\epsilon_{u,v}^{(l)} = 0$ for all u, v, and $l \geq 1$. Consequently (4.10) yields zero. The channel waveform assigned to the ith signal point within the constellation $\tilde{C}$ is

$$s_i(t) = \alpha_i p_1(t) + \beta_i p_2(t), \qquad \text{where} \qquad i = 1, 2, \ldots, \Omega_{\tilde{C}}. \tag{5.25}$$

Consequently, the Fourier transform of the ith channel waveform is

$$\begin{aligned} S_i(f) &= \mathcal{F}\{s_i(t)\} = \mathcal{F}\{\alpha_i p_1(t) + \beta_i p_2(t)\} \\ &= \alpha_i P_1(f) + \beta_i P_2(f), \end{aligned} \tag{5.26}$$

where

$$P_1(f) = \mathcal{F}\{p_1(t)\}, \qquad \text{and} \qquad P_2(f) = \mathcal{F}\{p_2(t)\}. \tag{5.27}$$

Thus,

$$\begin{aligned} |S_i(f)|^2 &= (\alpha_i P_1(f) + \beta_i P_2(f))(\alpha_i P_1^*(f) + \beta_i P_2^*(f)) \\ &= \alpha_i^2 |P_1(f)|^2 + \beta_i^2 |P_2(f)| \\ &\quad + \alpha_i \beta_i (P_1(f) P_2^*(f) + P_1^*(f) P_2(f)). \end{aligned} \tag{5.28}$$

Furthermore, (4.9) can be expressed as

$$\begin{aligned} II &= \left| \sum_{i=1}^{\Omega_{\tilde{C}}} \pi_i S_i(f) \right|^2 = \left| \sum_{i=1}^{\Omega_{\tilde{C}}} \pi_i \alpha_i P_1(f) + \pi_i \beta_i P_2(f) \right|^2 \\ &= \left(\sum_{i=1}^{\Omega_{\tilde{C}}} \pi_i \alpha_i P_1(f) + \pi_i \beta_i P_2(f) \right) \left(\sum_{i=1}^{\Omega_{\tilde{C}}} \pi_i \alpha_i P_1^*(f) + \pi_i \beta_i P_2^*(f) \right) \\ &= \sum_{i,j=1}^{\Omega_{\tilde{C}}} (\pi_i \alpha_i P_1(f) + \pi_i \beta_i P_2(f))(\pi_j \alpha_j P_1^*(f) + \pi_j \beta_j P_2^*(f)), \end{aligned} \tag{5.29}$$

or equivalently,

$$
\begin{aligned}
II &= \left(\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \alpha_i \alpha_j\right) |P_1(f)|^2 + \left(\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \beta_i \beta_j\right) |P_2(f)|^2 \\
&\quad + \left(\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \alpha_i \beta_j\right) \left(P_1(f) P_2^*(f) + P_1^*(f) P_2(f)\right). \qquad (5.30)
\end{aligned}
$$

Equation (4.8) can also be simplified using (5.28) yielding

$$
\begin{aligned}
I &= \sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i |S_i(f)|^2 = \left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \alpha_i^2\right) |P_1(f)|^2 + \left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \beta_i^2\right) |P_2(f)|^2 \\
&\quad + \left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \alpha_i \beta_i\right) \left(P_1(f) P_2^*(f) + P_1^*(f) P_2(f)\right). \qquad (5.31)
\end{aligned}
$$

For $\mu = 1$, $II = IV$ [see (4.9] and (4.11)), so that the final power spectrum function can be expressible as

$$
\begin{aligned}
\Phi(f) &= \frac{1}{T}\Bigg\{\left[\left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \alpha_i^2\right) - \left(\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \alpha_i \alpha_j\right)\right] |P_1(f)|^2 \\
&\quad + \left[\left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \beta_i^2\right) - \left(\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \beta_i \beta_j\right)\right] |P_2(f)|^2 \\
&\quad + \left[\left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \beta_i \alpha_i\right) - \left(\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \alpha_i \beta_j\right)\right] \left(P_1(f) P_2^*(f) + P_1^*(f) P_2(f)\right)\Bigg\} \\
&\quad + \left(\frac{1}{T^2}\right)\left\{\sum_{i,j=1}^{\Omega_{\mathcal{C}}} \left(\pi_i \alpha_i P_1(f) + \pi_i \beta_i P_2(f)\right)\left(\pi_j \alpha_j P_1^*(f) + \pi_j \beta_j P_2^*(f)\right)\right\} \\
&\quad \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right). \qquad (5.32)
\end{aligned}
$$

If it is desired to decouple the effect of a change in the statistics of the signal points on the shape of the spectral power density function, a sufficient condition would be to separate $\Phi(f)$ into two factors, one containing only the statistical information of the signal points, namely the π_i's, and the other containing the frequency component f.

If the signal points are selected within a 90° rotational symmetric constellation, for every signal point $\bar{a}_0$ in $\tilde{C}$ such that

$$\bar{a}_0 = \alpha_0 + j\beta_0, \tag{5.33}$$

the following points will be also contained in $\tilde{C}$:

$$\bar{a}_1 = \alpha_1 + j\beta_1 \ , \quad \bar{a}_2 = \alpha_2 + j\beta_2 \ , \quad \bar{a}_3 = \alpha_3 + j\beta_3 \tag{5.34}$$

yielding

$$\sum_{i,j=1}^{K} \pi_i \pi_j \alpha_i \beta_j = \sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \beta_i \beta_j = \sum_{i,j=1}^{\Omega_{\mathcal{C}}} \pi_i \pi_j \alpha_i \alpha_j = 0. \tag{5.35}$$

also,

$$\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \alpha_i \beta_i = 0 \ , \quad \left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \alpha_i^2 \right) = \left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \beta_i^2 \right). \tag{5.36}$$

From (5.32), $\Phi(f)$ can thus be simplified to

$$\Phi(f) = \frac{1}{T} \left\{ \left(\sum_{i=1}^{\Omega_{\mathcal{C}}} \pi_i \alpha_i^2 \right) \left[|P_1(f)|^2 + |P_2(f)|^2 \right] \right\}. \tag{5.37}$$

Equation (5.37) reveals that, if a 90° rotational symmetric constellation is utilized, the statistics of the signal points are separable from the frequency components in the power density function, and a change in the probabilities of selecting the signal points from the constellation does not shape the spectrum of the output signal sequence.

If a 180° rotational symmetric constellation is utilized, then

$$\bar{a}_0 = \alpha_0 + j\beta_0 \tag{5.38}$$

and

$$\bar{a}_1 = \alpha_1 + j\beta_1 \tag{5.39}$$

are elements of $\tilde{C}$, the set of points in $\tilde{C}$ satisfies (5.35), and therefore the power spectrum function becomes

$$\begin{aligned}\Phi(f) &= \frac{1}{T}\left\{\left(\sum_{i=1}^{\Omega_{\tilde{C}}}\pi_i\alpha_i^2\right)|P_1(f)|^2+\left(\sum_{i=1}^{\Omega_{\tilde{C}}}\pi_i\beta_i^2\right)|P_2(f)|^2\right.\\ &+ \left.\left(\sum_{i=1}^{\Omega_{\tilde{C}}}\pi_i\alpha_i\beta_i\right)\left[P_1(f)P_2^*(f)+P_1^*(f)P_2(f)\right]\right\}. \qquad (5.40)\end{aligned}$$

In (5.40), the frequency component and the statistics of the signal points can only be separated if

$$|P_1(f)|^2 = |P_2(f)|^2 \qquad (5.41)$$

and

$$P_1(f)P_2^*(f) + P_1^*(f)P_2(f) = 0. \qquad (5.42)$$

As a special case, if the signal pair given in (5.2) are used as orthogonal base signals, then (5.41), and (5.42) are satisfied, and the power spectrum function simplifies to

$$\Phi(f) = \frac{1}{T}\left\{\sum_{i=1}^{\Omega_{\tilde{C}}}\pi_i\left(\alpha_i^2+\beta_i^2\right)|P_1(f)|^2\right\}. \qquad (5.43)$$

Clearly the statistics of the signal points are separated from the frequency component f. Hence it is concluded that 180° rotational symmetry is sufficient to ensure that a change in the probabilities of selecting signals within $\tilde{C}$ does not shape the spectrum, if the constellation is combined with QAM.

In order to verify these results numerically, these signalling systems were simulated to compare the spectrum of memoryless transmission systems with various equiprobable and nonequiprobable constellations. In making any such comparison, both constellations should transmit information at the same rate, otherwise the comparison would not be fair. Since each constellation has a different entropy expressed in information symbols/transmitted symbols, their information rate can be equilized by adjusting the symbol interval T. In the simulations, the symbol interval was adjusted to provide a true information rate of 2400 bits/second. The minimal separation between the signal points in each constellation was also equalized since this has a strong impact on the reliability of the transmitted information. The constellations are also scaled so that $d^2_{\min}(\bar{C})$ is normalized to 1 for each $\bar{C}$ by selecting $\varepsilon = 0.25$. Finally, symbols in every constellation are assumed to be transmitted by QAM.

Table 5.2: Results of the first simulation.

Constellation	Entropy bits/sym.	Average Power $P_{av}(\bar{C})$[a]	$PAR_2(C)$	Symbol Interval T (ms)	90% P.BW. (Hz)	$BW. \times T$ $\times 10^3$
4-point EQ.	2.000	600.0	1.00	0.833	1018.23	848.5
16-point EQ.	4.000	1500.0	1.80	1.667	509.12	848.6
32-point EQ.	5.000	2400.0	1.70	2.083	407.29	848.6
55-point EQ.	5.781	3634.3	2.11	2.409	352.66	849.5
64-point EQ.	6.000	4200.0	2.33	2.500	339.41	848.5
16-point NEQ.	3.793	1371.1	2.08	1.580	536.63	848.0
32-point NEQ.	4.793	1919.7	2.22	1.997	423.50	845.7
64-point NEQ.	5.793	3176.5	2.74	2.414	351.50	848.3

[a] $P_{av} = \frac{\varepsilon}{T}\left\{\langle A^2_{mc}\rangle + \langle A^2_{ms}\rangle\right\}$, and $\varepsilon = 1/4$.

From (4.6), the power spectrum function of the signals is calculated, and then integrated numerically in order to determine the 90% fractional power bandwidth of the transmitted signal, i.e., the width of the frequency band where 90% of the total signal power is allocated. The results are summarized in Table 5.2 along with the characteristic parameters of the corresponding constellations. According to this result all constellations generate the same time-bandwidth product. This can be interpreted as the symbol interval being the only variable affecting the signal bandwidth BW. This nonequiprobable signalling does not result in any bandwidth expansion. These results are in accord with our previous derivations, because all the signal constellations listed in Table 5.2 have 90° rotational symmetry. We also see that in Table 5.2

there is no major increase in peak to average energy ratio due to nonequiprobable signalling. Consequently, it seems that shape gain can truly be obtained with little cost by transmission systems combined with nonequiprobable signal constellations in low dimensional signal space. Fig. 5.6 illustrates that theoretically, there is always a better nonequiprobable constellation than an equiprobable constellation either for limited average power case or limited available bandwidth case.

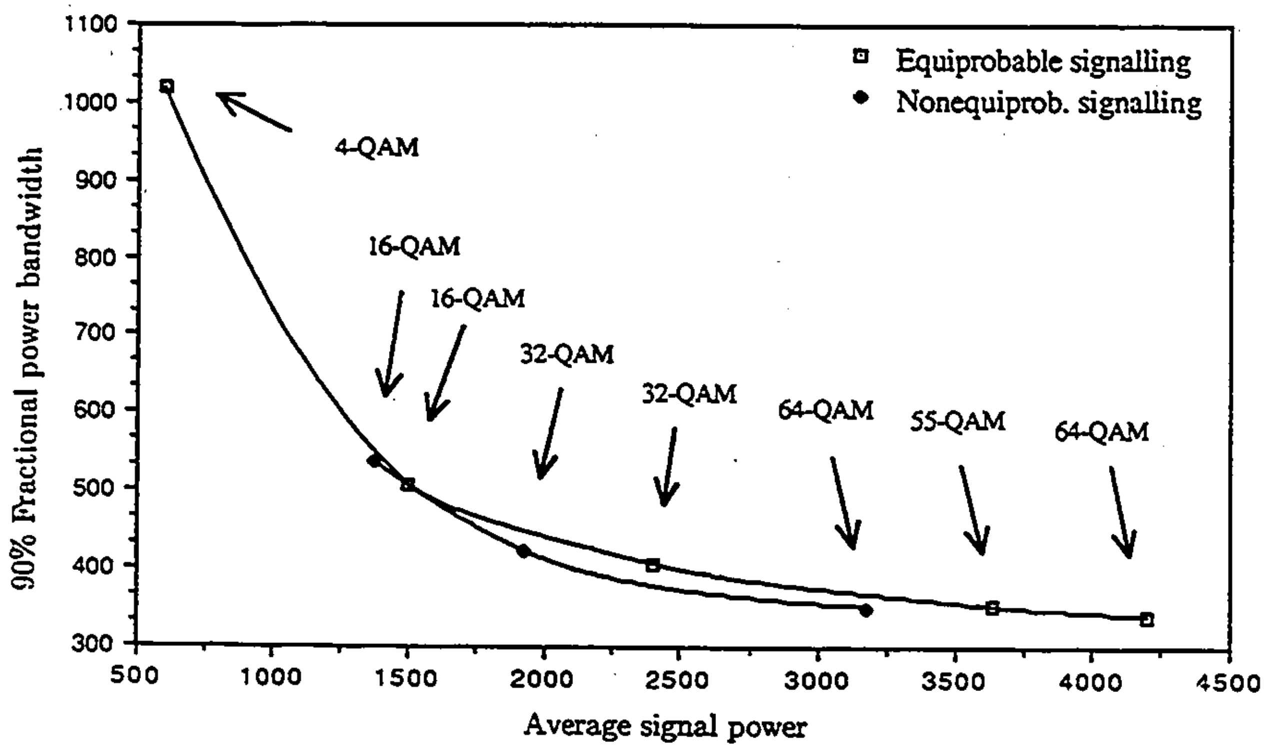


Figure 5.6: Comparison of nonequiprobable constellations with equiprobable constellations.

A second example was considered to test the previous derivations. The 64-point nonequiprobable signal constellation with the two subconstellations shown in Fig. 5.2 is utilized for nonequiprobable signalling. For a different set of frequencies f_0 and f_1, the 90% fractional power bandwidth was found to be the same for all relative frequencies. Therefore has been no effect is observed on the bandwidth due to a change in the statistics of the points in the constellation.

A third example was also considered. Three different two-dimensional 16 point signal constellations, $\bar{C}_a, \bar{C}_b$ and $\bar{C}_c$, are divided into two subconstellations as shown in Fig. 5.7, and combined with QAM. For different sets of frequencies of occurrence of the subconstellations, i.e., f_0 and f_1, the average signal energy and 90% fractional power bandwidth is calculated. Each constellation is designed to transmit 2400 bits per second. The results are summarized in Table 5.3-5.5 and plotted on a bandwidth vs. average signal energy graph as illustrated in Fig. 5.8. When we compare the time-bandwidth product of each constellation, we notice that it remaines constant for $\bar{C}_b$ and $\bar{C}_c$ but not for $\bar{C}_a$, which means the statistics of the subconstellations have affected the bandwidth for the constellation $\bar{C}_a$. In this example only $\bar{C}_a$ did not preserve its

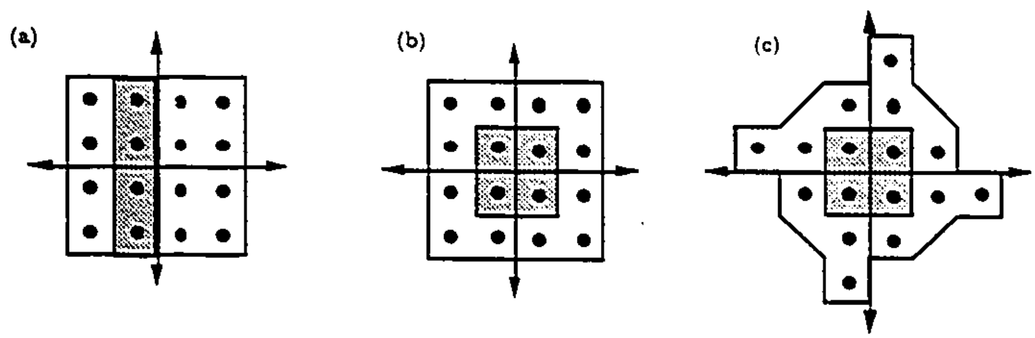


Figure 5.7: Three different 16-point nonequiprobable constellation: (a) $\bar{C}_a$; (b) $\bar{C}_b$: (c) $\bar{C}c$. Shaded region represents the subconstellation $\bar{C}^0$, and the remaining dots are within $\bar{C}^1$.

180° rotational symmetry. According to our earlier derivations, this non-symmetry may influence the statistics of the points of the power spectrum. Furthermore if we compare the bandwidth efficiency of $\bar{C}_a$ with that of $\bar{C}_b$ in Fig. 5.8, we see that the effect of statistics of points on the bandwidth is never in a favorable direction.

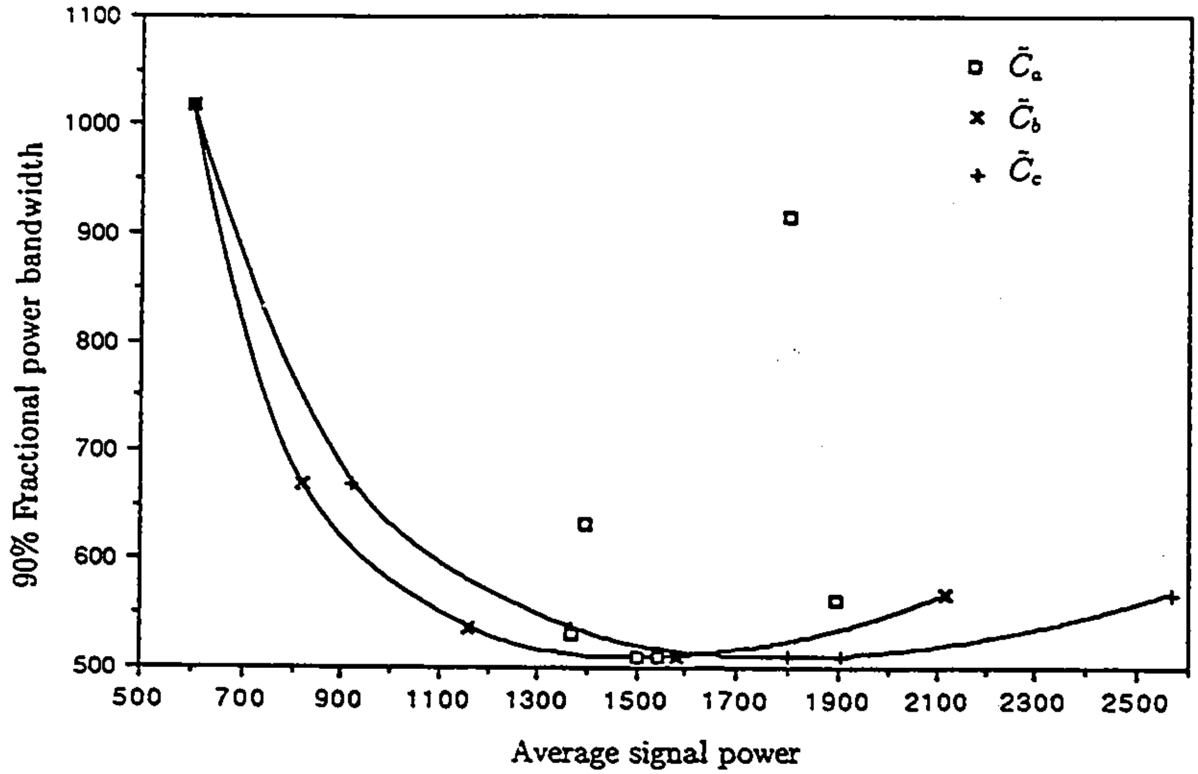


Figure 5.8: Average transmit power (for $d^2_{\min}(\Lambda) = 1$) vs. 90% fractional power bandwidth curves corresponding to the previous experiment.

Table 5.3: Simulation results with $\bar{C}_a$.

Symbol Interval T (ms)	f_0	f_1	Average Power P_{av} [a]	%90 *Power* BW. (Hz)	BW. $\times T$ $\times 10^3$
0.833	1.0	0.0	1800.0	914.65	762.2
1.266	0.8	0.2	1395.2	631.54	800.0
1.580	0.5	0.5	1371.1	530.24	837.9
1.667	0.25	0.75	1500.0	508.80	848.6
1.662	0.2	0.8	1543.9	510.47	848.6
1.494	0.0	1.0	1896.8	561.89	839.3

Table 5.4: Simulation results with $\tilde{C}_b$.

Symbol Interval T (ms)	f_0	f_1	Average Power P_{av} [a]	%90 *Power* BW. (Hz)	BW. $\times T$ $\times 10^3$
0.833	1.0	0.0	600.0	1018.23	845.1
1.266	0.8	0.2	816.2	669.80	848.0
1.580	0.5	0.5	1160.2	536.63	848.0
1.667	0.25	0.75	1500.0	508.80	848.0
1.662	0.2	0.8	1584.0	510.07	847.9
1.494	0.0	1.0	2120.0	567.70	848.0

Table 5.5: Simulation results with $\tilde{C}_c$.

Symbol Interval T (ms)	f_0	f_1	Average Power P_{av} [a]	%90 *Power* BW. (Hz)	BW. $\times T$ $\times 10^3$
0.833	1.0	0.0	600.0	1018.23	845.1
1.266	0.8	0.2	921.5	669.80	848.0
1.580	0.5	0.5	1371.1	536.63	848.0
1.667	0.25	0.75	1800.0	508.80	848.0
1.662	0.2	0.8	1904.9	510.07	847.9
1.494	0.0	1.0	2566.3	567.70	848.0

[a] $P_{av} = \frac{\epsilon}{T}\left\{\langle A_{mc}^2\rangle + \langle A_{ms}^2\rangle\right\}$, and $\epsilon = 1/4$.

5.2 Spectral Analysis of Nonequiprobable Signalling

In Chapter 3, two methods were proposed to attain shape gain by nonequiprobable signalling:

(i) The Calderbank & Ozarow method using mainly block coding,

(ii) and trellis shaping using convolutional codes due to Forney.

The exact computation of the spectrum of signals generated using these methods is extremely difficult, so to determine their spectral density the simulation and numerically estimation methods discussed in Chapter 4 are used. Three example codes are considered. In the first two the CO method is employed based on various greedy algorithms. In the last example trellis shaping is used.

The systems utilized in the first two examples can be taken as block encoders since they consist of combinations of two block encoders: the first one is used for conventional coding purposes, and the second one for shaping purposes. The power spectrum of a block encoder combined with two-dimensional signal constellations can be calculated by further simplifying (4.6).

Power Spectrum of Block Codes with Two-dimensional Constellations

An (M, N) memoryless block code is defined as the ensemble of length N codewords utilizing symbols selected from a L-ary alphabet, and an assignment of the datawords to the codewords. The Galois field $GF(2^n)$ ([31], pp. 24), for some n, is the most widely used L-ary alphabet ($L = 2^n$) in digital data transmission, because information is usually coded in binary form for practical reasons. Every length M information word, with symbols from $GF(2^m)$, is uniquely assigned to one of the codewords by the encoder, so that

$$L^N = (2^n)^N \geq (2^m)^M . \tag{5.44}$$

After the block encoder selects a codeword to be transmitted according to the current input word available to the encoder, the selected codeword enters a mapping circuit, where each of the N digits is mapped to one of the signal points within a constellation $\tilde{C}$. An L point constellation is necessary to assign every element of the L-ary alphabet to one of the signal points. If T seconds is required to transmit one digit, then $NT = T_N$ is the time required to transmit one codeword. Equation (5.25) gives the representation of a signal point (α_i, β_i) in terms of channel signals. The channel waveform corresponding to the uth codeword of the block code can be represented as

$$s_u(t) = \sum_{i=0}^{N-1} \alpha_{u,i} p_1(t - iT) + \beta_{u,i} p_2(t - iT), \tag{5.45}$$

where $p_1(t) = p_2(t) = 0$ for $t \in [0, T)$. Clearly, $s_u(t) = 0$ for $t \notin [0, T_N)$.

By assigning each channel waveform to one state, we can construct a Markov chain model for the block code that consists of $(2^m)^M$ codewords and consequently the chain contains $\Omega_{\tilde{C}} = (2^m)^M$ states. Since there is no memory involved from one codeword to the next

$$\epsilon_{i,j}^{(l)} = 0 \qquad \text{for all } \; i, j, \;\text{ and } \; l \geq 1, \tag{5.46}$$

and the period of the chain μ is equal to 1, hence, (4.10) is always zero. The Fourier transform of $s_u(t)$ is given by

$$S_u(f) = P_1(f) \sum_{i=0}^{N-1} \alpha_{u,i} e^{-j2\pi f T_i} + P_2(f) \sum_{j=0}^{N-1} \beta_{u,j} e^{-j2\pi f T_j}. \tag{5.47}$$

The energy density function of each waveform is then

$$\begin{aligned} |S_u(f)|^2 &= |P_1|^2 \sum_{i,i'}^{N-1} e^{j2\pi f T i'} \alpha_{u,i'} \alpha_{u,i} e^{-j2\pi f T i} \\ &+ |P_2|^2 \sum_{j,j'}^{N-1} e^{j2\pi f T j'} \alpha_{u,j'} \alpha_{u,j} e^{-j2\pi f T j} \\ &+ P_1(f) P_2^*(f) \sum_{i,j'}^{N-1} e^{j2\pi f T j'} \beta_{u,j'} \alpha_{u,i} e^{-j2\pi f T i} \end{aligned}$$

$$+ \quad P_1^*(f)P_2(f)\sum_{j,i'}^{N-1} e^{j2\pi fTi'}\alpha_{u,i'}\beta_{u,j}e^{-j2\pi fTj}. \qquad (5.48)$$

If π_i denotes the stationary distribution of state i then (4.8) yields

$$\begin{aligned} I = \sum_{u=1}^{2^M}\pi_u\,|S_u(f)|^2 &= \sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\left\{|P_1(f)|^2\sum_{i,j=0}^{N-1} e^{j2\pi fTj}\alpha_{u,i}\alpha_{u,j}e^{-j2\pi fTi}\right.\\ &+ |P_2(f)|^2\sum_{i,j=0}^{N-1} e^{j2\pi fTj}\beta_{u,i}\beta_{u,j}e^{-j2\pi fTi}\\ &+ P_1(f)P_2^*(f)\sum_{i,j=0}^{N-1} e^{j2\pi fTj}\beta_{u,i}\alpha_{u,j}e^{-j2\pi fTi}\\ &+ \left.P_1^*(f)P_2(f)\sum_{i,j=0}^{N-1} e^{j2\pi fTj}\alpha_{u,i}\beta_{u,j}e^{-j2\pi fTi}\right\} \qquad (5.49)\end{aligned}$$

Equation (4.9) yields

$$\begin{aligned} II &= \left|\sum_{u=1}^{\Omega_{\mathcal{C}}} S_u(f)\right|^2\\ &= \left|P_1(f)\sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\sum_{i=0}^{N-1}\alpha_{u,i}e^{-j2\pi fTi} + P_2(f)\sum_{v=1}^{\Omega_{\mathcal{C}}}\pi_v\sum_{j=0}^{N-1}\beta_{v,j}e^{-j2\pi fTj}\right|^2\\ &= |P_1(f)|^2\sum_{i,i'=0}^{N-1} e^{j2\pi fTi'}\left(\sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\alpha_{u,i}\right)\left(\sum_{v=1}^{\Omega_{\mathcal{C}}}\pi_v\alpha_{v,i'}\right)e^{-j2\pi fTi}\\ &+ |P_2(f)|^2\sum_{j,j'=0}^{N-1} e^{j2\pi fTj'}\left(\sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\beta_{u,j}\right)\left(\sum_{v=1}^{\Omega_{\mathcal{C}}}\pi_v\beta_{v,j'}\right)e^{-j2\pi fTj}\\ &+ P_1(f)P_2^*(f)\sum_{i,j'=0}^{N-1} e^{j2\pi fTj'}\left(\sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\alpha_{u,i}\right)\left(\sum_{v=1}^{\Omega_{\mathcal{C}}}\pi_v\beta_{v,j'}\right)e^{-j2\pi fTi}\\ &+ P_1^*(f)P_2(f)\sum_{j,i'=0}^{N-1} e^{j2\pi fTi'}\left(\sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\beta_{u,j}\right)\left(\sum_{v=1}^{\Omega_{\mathcal{C}}}\pi_v\alpha_{v,i'}\right)e^{-j2\pi fTj}. \qquad (5.50)\end{aligned}$$

The summations in (5.50) can also be written in matrix multiplication form

$$II = \left|\sum_{u=1}^{\Omega_{\mathcal{C}}} S_u(f)\right|^2 = |P_1(f)|^2\left(\mathbf{V}\mathbf{R}_{\alpha\alpha\infty}\mathbf{V}^{T*}\right)$$

$$
\begin{aligned}
&+ \quad |P_2(f)|^2 \left(\mathbf{V}\mathbf{R}_{\beta\beta\infty}\mathbf{V}^{T*}\right) \\
&+ \quad P_1(f)P_2^*(f) \left(\mathbf{V}\mathbf{R}_{\alpha\beta\infty}\mathbf{V}^{T*}\right) \\
&+ \quad P_1^*(f)P_2(f) \left(\mathbf{V}\mathbf{R}_{\beta\alpha\infty}\mathbf{V}^{T*}\right),
\end{aligned} \tag{5.51}
$$

with

$$
\mathbf{V} = \left|1\,,\, e^{j2\pi fT}\,,\, \cdots\,,\, e^{j2\pi fT(N-1)}\right|, \tag{5.52}
$$

and

$$
\begin{aligned}
\mathbf{R}^{i,j}_{\alpha\alpha\infty} &= \left(\sum_{u=1}^{\Omega_C} \pi_u \alpha_{u,i}\right)\left(\sum_{v=1}^{\Omega_C} \pi_v \alpha_{v,j}\right), \\
\mathbf{R}^{i,j}_{\beta\beta\infty} &= \left(\sum_{u=1}^{\Omega_C} \pi_u \beta_{u,i}\right)\left(\sum_{v=1}^{\Omega_C} \pi_v \beta_{v,j}\right), \\
\mathbf{R}^{i,j}_{\alpha\beta\infty} &= \left(\sum_{u=1}^{\Omega_C} \pi_u \alpha_{u,i}\right)\left(\sum_{v=1}^{\Omega_C} \pi_v \beta_{v,j}\right), \\
\mathbf{R}^{i,j}_{\beta\alpha\infty} &= \left(\sum_{u=1}^{\Omega_C} \pi_u \beta_{u,i}\right)\left(\sum_{v=1}^{\Omega_C} \pi_v \alpha_{v,j}\right).
\end{aligned} \tag{5.53}
$$

In these equations $\mathbf{R}^{i,j}$ denotes the element of the matrix $\mathbf{R}$ at the intersection of the ith row and the jth column. From equation (5.53) we conclude that

$$
\mathbf{R}_{\alpha\beta\infty} = \mathbf{R}^T_{\beta\alpha\infty}, \tag{5.54}
$$

and consequently,

$$
\mathbf{V}\mathbf{R}_{\alpha\beta\infty}\mathbf{V}^{T*} = \mathbf{V}\mathbf{R}^T_{\beta\alpha\infty}\mathbf{V}^{T*} = \left(\mathbf{V}^*\mathbf{R}_{\beta\alpha\infty}\mathbf{V}^T\right)^T = \left(\mathbf{V}\mathbf{R}_{\beta\alpha\infty}\mathbf{V}^{T*}\right)^{T*} \tag{5.55}
$$

Since $\mathbf{V}\mathbf{R}_{\beta\alpha\infty}\mathbf{V}^{T*}$ is a 1×1 matrix, its transpose is equal to itself, yielding

$$
\mathbf{V}\mathbf{R}_{\alpha\beta\infty}\mathbf{V}^{T*} = \left(\mathbf{V}\mathbf{R}_{\beta\alpha\infty}\mathbf{V}^{T*}\right)^*. \tag{5.56}
$$

Inserting (5.56) into (5.51), reveals that

$$
\begin{aligned}
H &= |P_1(f)|^2 \left(\mathbf{V}\mathbf{R}_{\alpha\alpha\infty}\mathbf{V}^{T*}\right) + |P_2(f)|^2 \left(\mathbf{V}\mathbf{R}_{\beta\beta\infty}\mathbf{V}^{T*}\right) \\
&+ 2\Re\left\{P_1(f)P_2^*(f)\left(\mathbf{V}\mathbf{R}_{\alpha\beta\infty}\mathbf{V}^{T*}\right)\right\}.
\end{aligned} \tag{5.57}
$$

Similarly from (5.49)

$$I=\sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\left|S_u(f)\right|^2 = \left|P_1(f)\right|^2\left(\mathbf{V}\mathbf{R}_{\alpha\alpha0}\mathbf{V}^{T*}\right) + \left|P_2(f)\right|^2\left(\mathbf{V}\mathbf{R}_{\beta\beta0}\mathbf{V}^{T*}\right) + 2\Re\left\{P_1(f)P_2^*(f)\left(\mathbf{V}\mathbf{R}_{\beta\alpha0}\mathbf{V}^{T*}\right)\right\}, \quad (5.58)$$

where

$$\begin{aligned}\mathbf{R}_{\alpha\alpha0}^{i,j} &= \sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\alpha_{u,i}\alpha_{u,j}\\ \mathbf{R}_{\beta\beta0}^{i,j} &= \sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\beta_{u,i}\beta_{u,j}\\ \mathbf{R}_{\alpha\beta0}^{i,j} &= \sum_{u=1}^{\Omega_{\mathcal{C}}}\pi_u\alpha_{u,i}\beta_{u,j}.\end{aligned} \quad (5.59)$$

Consequently, (4.7) yields

$$\begin{aligned}\Phi(f) &= \frac{1}{T_N}\left|P_1(f)\right|^2\left[\mathbf{V}\left(\mathbf{R}_{\alpha\alpha0}-\mathbf{R}_{\alpha\alpha\infty}\right)\mathbf{V}^{T*}+\frac{1}{T_N}\left(\mathbf{V}\mathbf{R}_{\alpha\alpha\infty}\mathbf{V}^{T*}\right)\sum_{k=-\infty}^{\infty}\delta\left(f-\frac{k}{T_N}\right)\right]\\ &+ \frac{1}{T_N}\left|P_2(f)\right|^2\left[\mathbf{V}\left(\mathbf{R}_{\beta\beta0}-\mathbf{R}_{\beta\beta\infty}\right)\mathbf{V}^{T*}+\frac{1}{T_N}\left(\mathbf{V}\mathbf{R}_{\beta\beta\infty}\mathbf{V}^{T*}\right)\sum_{k=-\infty}^{\infty}\delta\left(f-\frac{k}{T_N}\right)\right]\\ &+ \frac{2}{T_N}\Re\left\{P_1(f)P_2^*(f)\left[\mathbf{V}\left(\mathbf{R}_{\beta\alpha0}-\mathbf{R}_{\beta\alpha\infty}\right)\mathbf{V}^{T*}\right.\right.\\ &+ \left.\left.\frac{1}{T_N}\left(\mathbf{V}\mathbf{R}_{\beta\alpha\infty}\mathbf{V}^{T*}\right)\sum_{k=-\infty}^{\infty}\delta\left(f-\frac{k}{T_N}\right)\right]\right\}.\end{aligned} \quad (5.60)$$

Let $\vec{b}^{(l)}=[b_1^{(l)},b_2^{(l)},\cdots,b_N^{(l)}]$ denote the codeword selected at time l, $b_i^{(l)}$, representing the ith digit of the selected codeword. is a complex number and can be written as $b_i^{(l)}=\alpha_i^{(l)}+j\beta_i^{(l)}$ ($\alpha_i,\beta_i\in\mathbb{R}$). The matrixes used in (5.60) can be also written in terms of $\alpha_i^{(l)}$ and $\beta_i^{(l)}$ as follows:

$$\begin{aligned}\mathbf{R}_{\alpha\alpha0}^{i,j} &= E\left\{\alpha_i^{(l)}\alpha_j^{(l)}\right\}\\ \mathbf{R}_{\beta\beta0}^{i,j} &= E\left\{\beta_i^{(l)}\beta_j^{(l)}\right\}\end{aligned}$$

$$
\begin{aligned}
\mathbf{R}^{i,j}_{\alpha\beta 0} &= E\left\{\alpha_i^{(l)}\beta_j^{(l)}\right\} \\
\mathbf{R}^{i,j}_{\alpha\alpha\infty} &= E\left\{\alpha_i^{(l)}\alpha_j^{(l+\infty)}\right\} = E\left\{\alpha_i^{(l)}\right\}E\left\{\alpha_j^{(l)}\right\} \\
\mathbf{R}^{i,j}_{\beta\beta\infty} &= E\left\{\beta_i^{(l)}\beta_j^{(l+\infty)}\right\} = E\left\{\beta_i^{(l)}\right\}E\left\{\beta_j^{(l)}\right\} \\
\mathbf{R}^{i,j}_{\alpha\beta\infty} &= E\left\{\alpha_i^{(l)}\beta_j^{(l+\infty)}\right\} = E\left\{\alpha_i^{(l)}\right\}E\left\{\beta_j^{(l)}\right\}
\end{aligned}
$$

for $0 \leq l < \infty$, since there is no memory involved between codewords. From (5.60), it is concluded that a sufficient condition for the discrete part of the power spectrum to vanish is

$$\mathbf{R}^{i,j}_{\alpha\alpha\infty} = \mathbf{R}^{i,j}_{\beta\beta\infty} = \mathbf{R}^{i,j}_{\alpha\beta\infty} = 0 \tag{5.61}$$

for $1 \leq i, j \leq N$. This can only be realized if

$$E\left\{\alpha_i^{(l)}\right\} = E\left\{\beta_j^{(l)}\right\} = 0 \tag{5.62}$$

for $1 \leq i \leq N$. In other words if the expected value of each complex digit of the output words of the encoder is equal to zero then the spectrum of such system will not contain any discrete components. In this case the power spectrum density function of the output signal sequence of a (M, N) block encoder can be further simplified:

$$
\begin{aligned}
\Phi(f) &= \frac{1}{T_N}|P_1(f)|^2\left[\Gamma^0_{\alpha\alpha 0} + 2\sum_{i=1}^{N-1}\Gamma^i_{\alpha\alpha 0}\cos 2\pi f iT\right] \\
&+ \frac{1}{T_N}|P_2(f)|^2\left[\Gamma^0_{\beta\beta 0} + 2\sum_{i=1}^{N-1}\Gamma^i_{\beta\beta 0}\cos 2\pi f iT\right] \\
&+ \frac{2}{T_N}\{P_1(f)P_2^*(f)\}\left[\Gamma^0_{\alpha\beta 0} + 2\sum_{i=1}^{N-1}\Gamma^i_{\alpha\beta 0}\cos 2\pi f iT\right]
\end{aligned} \tag{5.63}
$$

where Γ^k_x denotes the kth trace of the matrix $\mathbf{R}_x$.

(5.63) gives the power spectrum of any modulation system whose output symbol sequence consists of length N codewords, and the expected value of each digit in one block length output signal sequence is equal to zero.

5.2.1 Simulations for the CO Method

Example 1

The system shown in Fig. 3.4 is combined with a 128-point signal constellation $\bar{C}$ which is divided into two subconstellations, $\bar{C}^0$ and $\bar{C}^1$. A $\mathcal{G}_{8,2}$ greedy code is used for subconstellation selection. The 8 bit codewords of $\mathcal{G}_{8,2}$ are fed serially onto the subconstellation selector, where the "0" valued digits are mapped into "-1", representing $\bar{C}^0$, and the "1" valued digits represent $\bar{C}^1$. The "-1" , "1" sequence is called subconstellation selector sequence. An element of this sequence is fed into the signal point selector every symbol interval T. At the same time length 6 information-word enters the signal point selector to select one candidate signal point within $\bar{C}^0$. If the current input from the subconstellation selector is a "-1", then this candidate signal point becomes the selected signal point, and is transmitted over the channel. If the current input from the subconstellation selector is a "1", then the equivalent signal point in $\bar{C}^1$ to the candidate signal point becomes the selected signal point. The equivalence relation will be described shortly.

128-point signal constellation: Fig. 5.9 illustrates $\bar{C}$ with its subconstellations $\bar{C}^0$ and $\bar{C}^1$. Each subconstellation contains 64 signal points. Any point $\bar{r}_0$ in $\bar{C}^0$ is labeled with a 6 digit word such that

$$\begin{aligned} \bar{r}_0 &\in \bar{C}^0 \\ \bar{r}_0 &= (a_0, a_1, a_2, a_3, a_4, a_5) \quad a_i \in \{-1, 1\}. \end{aligned} \tag{5.64}$$

If $r_{0,x}$ and $r_{0,y}$ represent the coordinates of a point $\bar{r}_0$, then the coordinates of each point in $\bar{C}^0$ can be calculated from its label as follows:

$$\begin{aligned} r_{0,x} &= \frac{1}{2}(4a_0 + 2a_1 + a_2) \\ r_{0,y} &= \frac{1}{2}(4a_3 + 2a_4 + a_5). \end{aligned} \tag{5.65}$$

If $\bar{r}_1$ in $\bar{C}^1$ represents the equivalent point to $\bar{r}_0$, then the coordinates $r_{1,x}$ and $r_{1,y}$ of

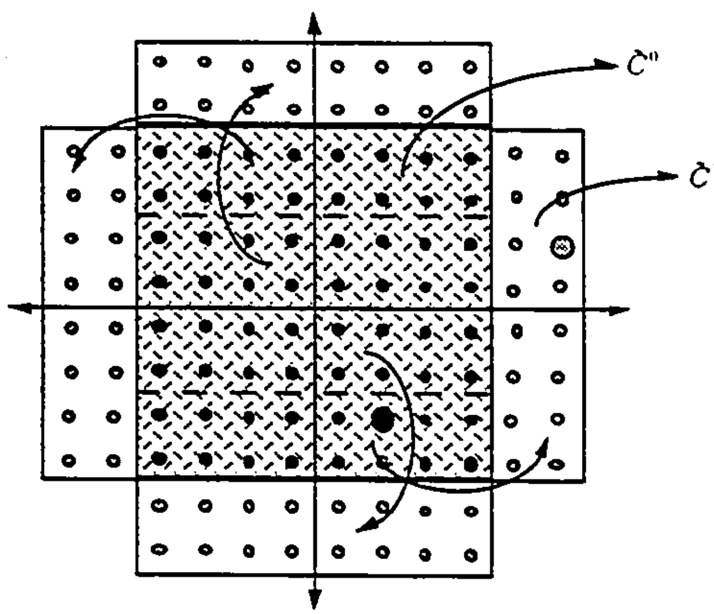


Figure 5.9: The division of a 64-point nonequiprobable constellation into two subconstellations and the matching of the subconstellations.

$\bar{r}_1$ are given by:

$$\begin{aligned}
&\text{if} \quad 2 < r_{y,0} < 4 \quad \text{then} \quad \begin{cases} r_{x,1} = r_{y,0} - 8 \\ r_{y,1} = r_{x,0} \end{cases} \\
&\text{if} \quad 0 < r_{y,0} \le 2 \quad \text{then} \quad \begin{cases} r_{x,1} = r_{x,0} \\ r_{y,1} = r_{y,0} + 4 \end{cases} \\
&\text{if} \quad -2 < r_{y,0} \le 0 \quad \text{then} \quad \begin{cases} r_{x,1} = r_{x,0} \\ r_{y,1} = r_{y,0} - 4 \end{cases} \\
&\text{if} \quad -4 < r_{y,0} \le -2 \quad \text{then} \quad \begin{cases} r_{x,1} = r_{y,0} + 8 \\ r_{y,1} = r_{x,0} \end{cases}
\end{aligned} \tag{5.66}$$

The arrows in Fig. 5.9 represents the equivalence relation between $\bar{r}_0$'s and $\bar{r}_1$'s. For example, if $\bar{r}_0 = (1, -1, 1, -1, -1, 1)$ then the coordinates of $\bar{r}_0$ are $r_{x,0} = 1.5$ and $r_{y,0} = -2.5$. According to the equivalence relation, the coordinates of $\bar{r}_1$ are $r_{x,1} = 5.5$ and $r_{y,1} = 1.5$. Both of these points are marked boldly in Fig. 5.9.

One important variable in calculating the shape gain is the average signal energy of each subconstellation. For small constellations such as in this example, continuous approximation method is not required. The point by point calculations revealed that

$$\begin{aligned}
P_0 &= \frac{1}{4}\left[(\frac{1}{2})^2 + (\frac{3}{2})^2 + (\frac{5}{2})^2 + (\frac{7}{2})^2\right] \times 2 = \frac{21}{2}, \\
P_1 &= \frac{1}{4}\left[(\frac{1}{2})^2 + (\frac{3}{2})^2 + (\frac{5}{2})^2 + (\frac{7}{2})^2\right] + \frac{1}{2}\left[(\frac{9}{2})^2 + (\frac{11}{2})^2\right] = \frac{61}{2}.
\end{aligned} \tag{5.67}$$

The system specifications: The $\mathcal{G}_{8,2}$ code was already investigated in Chapter 3. It was found that the frequencies f_0 and f_1 of the elements "0" and "1" within the codewords of $\mathcal{G}_{8,2}$ are

$$\begin{aligned}
f_0 &= 0.784, \\
f_1 &= 0.216,
\end{aligned}$$

and the related code rate is $r(\mathcal{G}_{8,2}) = 0.651$. Consequently the nonequiprobable signal constellation is able to transmit $6+0.651 = 6.651$ bits per symbol, and a corresponding equiprobable signal constellation would have $2^{6.651} \approx 100$ signal points. A square shaped 100-point equiprobable signal constellation has an average signal energy of

16.5. The average signal energy of the nonequiprobable signal constellation is found as 14.82; thus the total shape gain is calculated as 0.47 dB. The peak to average power ratio is

$$\overline{PAR}_2(\tilde{C}) = \frac{\left[(\frac{11}{2})^2 + (\frac{7}{2})^2\right]}{\overline{P}_{av}(\tilde{C})} = \frac{170/4}{14.82} = 2.87$$

and the constellation expansion ratio

$$\overline{CER}_2 = \frac{128}{100} = 1.28$$

Example 2

This time we designed a modulation system with a $\mathcal{G}_{7,2}$ code for shaping purposes and $(7,3)$ Reed-Solomon code for conventional coding purposes. A 16 point one dimensional signal constellation is divided into two subconstellations $\tilde{C}^0$ and $\tilde{C}^1$ as shown in Fig. 5.10. The average signal energy of these subconstellations were found to be $P_0 = 21/4$ and $P_1 = 149/4$.

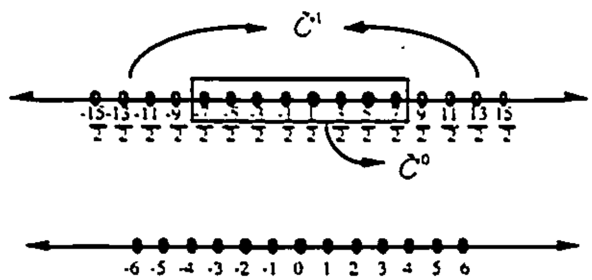


Figure 5.10: A 16-point nonequiprobable 1-dimensional constellation (above), and 13-point equiprobable constellation.

System specifications: From (3.64) and (3.66) the frequencies f_0 and f_1 and the code rate are calculated as follows:

$$r(\mathcal{G}_{7,2}) = \frac{1}{7}\log_2 \sum_{j=0}^{2} \binom{7}{j} = 0.694$$

$$
\begin{aligned}
f_0 &= \frac{\sum_{j=0}^{2}\binom{7}{j}(7-j)}{\left(\sum_{j=0}^{2}\binom{7}{j}\right)7} = 0.759 \\
f_1 &= 1 - f_1 = 0.241
\end{aligned}
$$

The average signal energy of the system is found as $\overline{P}_{av}(\tilde{C}) = 12.97$. An equiprobable constellation would require $2^{3+0.694} \approx 13$ signal points in order to have same information rate per symbol as the nonequiprobable constellation has. A 13-point equiprobable one dimensional constellation has an average signal energy $P_{av}(\tilde{C}) = 14.00$; therefore the total shape gain obtained by this system is found to be 0.33 dB.

The equivalence relation between points $\bar{r}_0$ in $\tilde{C}^0$ and the points $\bar{r}_1$ in $\tilde{C}^1$ is defined as

$$
r_{1,x} = \begin{cases} r_{0,x} + 8, & \text{if } \; -4 < r_{0,x} < 0 \\ r_{0,x} - 8, & \text{if } \; 0 < r_{0,x} < 4 \end{cases} \tag{5.68}
$$

The peak to average power ratio is

$$
\overline{PAR}_2(\tilde{C}) = \frac{(15/2)^2}{12.97} = 4.34,
$$

and the constellation expansion ratio is

$$
\overline{CER}_s(\tilde{C}) = \frac{16}{13} = 1.23
$$

Simulation Results

The correlation matrices used in equation (5.63) are first calculated for each system. It has been found that all elements of $\mathbf{R}_{\alpha\alpha\infty}$, $\mathbf{R}_{\beta\beta\infty}$, $\mathbf{R}_{\alpha\beta\infty}$, $\mathbf{R}_{\alpha\beta 0}$, are equal to zero. Furthermore, $\mathbf{R}_{\alpha\alpha 0}$ and $\mathbf{R}_{\beta\beta 0}$ contained nonzero elements only on their diagonals $(\mathbf{\Gamma}^0_{\alpha\alpha 0})$ such that

$$
\begin{aligned}
\mathbf{\Gamma}^i_{\alpha\alpha 0} &= 0 \qquad \text{for } \; i \neq 0 \\
\mathbf{\Gamma}^i_{\alpha\alpha 0} &= 0 \qquad \text{for } \; i \neq 0
\end{aligned}
$$

so that frequency components and the statistics components of the signal points are separable in (5.63). Therefore nonequiprobable signalling does not shape the power spectrum of the output symbol sequence. In the next section we will give a theoretical reason for this fact.

Discussion

As explained in Chapter 3, the CO method consists of any kind of encoder used for shaping purposes, and a conventional encoder, which can either be a block encoder to generate lattice codes or a convolutional encoder to generate trellis codes. The output sequence of a trellis code can be assumed to be an infinite length block code. This assumption allows the use of formula given in (5.63) to investigate the power spectrum of the output signal of the transmitter, which implements the Calderbank & Ozarow method. The sufficient condition for the necessary bandwidth of a nonequiprobable signalling scheme to be the same as that for equiprobable signalling is that the components depending on the statistics of the signal points and the frequency components in the power spectrum functions for both cases should be separable.

Let $\bar{b}^{(l)} = [b_1^{(l)}, b_2^{(l)}, \cdots, b_N^{(l)}]$ denotes the codeword selected at time l, where $b_i^{(l)}$, representing the ith digit of the selected codeword, is a complex number and can be written as $b_i^{(l)} = \alpha_i^{(l)} + \jmath\beta_i^{(l)}$. A sufficient condition that the statistical components and the frequency components in the (5.63) are separable is as follows:

for all $l \geq 0$

$$E\left\{\alpha_i^{(l)}\right\} = E\left\{\beta_i^{(l)}\right\} = 0 \qquad \text{for} \quad 1 \leq i \leq N; \tag{5.69}$$

$$E\left\{\alpha_i^{(l)}\beta_j^{(l)}\right\} = 0 \qquad \text{for} \quad 1 \leq i, j \leq N; \tag{5.70}$$

$$E\left\{\alpha_i^{(l)}\alpha_j^{(l)}\right\} = E\left\{\beta_i^{(l)}\beta_j^{(l)}\right\} = 0 \qquad \text{for} \quad 1 \leq i, j \leq N, \quad i \neq j; \tag{5.71}$$

$$E\left\{\left(\alpha_i^{(l)}\right)^2\right\} = E\left\{\left(\beta_i^{(l)}\right)^2\right\} \qquad \text{for} \quad 1 \leq i \leq N \tag{5.72}$$

Next the meaning of these conditions are investigated. According to the CO method the conventional encoder selects one signal point $\bar{r}_0 = r_{0,x} + \jmath r_{0,y}$ within $\bar{C}^0$. Afterwards, if the subconstellation selector activates say $\bar{C}^s$, then the equivalent point of $\bar{r}_0$ in $\bar{C}^s$, which is denoted as $\bar{r}_s = r_{s,x} + \jmath r_{s,y}$, is selected and transmitted over the channel. The equivalence relation between $\bar{r}_0$ and $\bar{r}_s$ can be formulated as

$$\bar{r}_s = g_s(\bar{r}_0) \quad , \tag{5.73}$$

or equivalently

$$\begin{aligned} r_{s,x} &= g_{\Re,s}(\bar{r}_0) \\ r_{s,y} &= g_{\Im,s}(\bar{r}_0) \end{aligned} \tag{5.74}$$

Similarly, we can write

$$b_i^{(l)} = \alpha_i^{(l)} + \jmath\beta_i^{(l)} = g_s(b_{i,0}^{(l)}) = g_{\Re,s}(b_{i,0}^{(l)}) + \jmath g_{\Im,s}(b_{i,0}^{(l)}) \tag{5.75}$$

where $b_{i,0}^{(l)} = \alpha_{i,0}^{(l)} + \jmath\beta_{i,0}^{(l)}$, and it represents the (from the conventional encoder) selected signal point within $\bar{C}^0$, while $b_i^{(l)}$ is the transmitted signal point at time l at digit i. Rewriting the crosscorrelation equation

$$E\left\{\alpha_i^{(l)}\alpha_j^{(l)}\right\} = \sum_{e=0}^{G-1}\sum_{f=0}^{G-1} E\left\{g_{\Re,e}(b_{i,0}^{(l)})g_{\Re,f}(b_{j,0}^{(l)})\right\} P(k_i^l = e, k_j^l = f) \quad \text{for } i \neq j, \tag{5.76}$$

where G denotes the number of subconstellations in $\bar{C}$, k_i denotes the selected subconstellation at the ith output digit at time l, and $P(k_i^l = e, k_j^l = f)$ the probability of having selected eth subconstellation at digit i and fth subconstellation at digit j. To make (5.76) satisfy the condition stated in (5.71),

$$E\left\{g_{\Re,e}(b_{i,0}^{(l)})g_{\Re,f}(b_{j,0}^{(l)})\right\} = 0, \tag{5.77}$$

is sufficient for $0 \leq e, f \leq G-1$. Let $\bar{r}_0$ and q_0 be two points in $\bar{C}^0$; then (5.77) is rewritten as

$$\sum_{\substack{\bar{r}_0\in\bar{C}^0 \\ q_0\in\bar{C}^0}} g_{\Re,e}(\bar{r}_0)g_{\Re,f}(q_0)P(b_{i,0}^{(l)} = \bar{r}_0, b_{j,0}^{(l)} = q_0) = 0, \tag{5.78}$$

equivalently

$$\sum_{q_0\in\bar{C}^0} g_{\Re,e}(\bar{r}_0)\left(\sum_{\bar{r}_0\in\bar{C}^0} g_{\Re,f}(q_0)P(b_{i,0}^{(l)} = \bar{r}_0 \mid b_{j,0}^{(l)} = q_0)\right) P(b_{j,0}^{(l)} = q_0) = 0 \tag{5.79}$$

Again, the sufficient condition for the equality in (5.79) is

$$\sum_{\bar{r}_0 \in \bar{C}^0} g_{\Re,f}(q_0) P(b_{i,0}^{(l)} = \bar{r}_0 \mid b_{j,0}^{(l)} = q_0) = 0 \qquad \text{for } 0 \le f \le G-1 \text{ and } 1 \le i,j \le N \quad (5.80)$$

Equation (5.80) requires knowledge of the conditional probabilities of the elements of the output signal sequence generated by a conventional encoder. Let's assume that an (M, N) block encoder is utilized as conventional encoder. The relation between the input words and the codewords is then given as

$$\mathbf{U}\mathbf{G} = \mathbf{V}, \qquad (5.81)$$

where $\mathbf{U}$ and $\mathbf{V}$ represents the length M input, and length N output words, respectively. $\mathbf{G}$ denotes the generator matrix of the block code. More specifically, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{G}$ can be written as

$$\begin{aligned} \mathbf{U} &= [u_1, u_2, \cdots, u_M] \\ \mathbf{V} &= [v_1, v_2, \cdots, v_N] \\ \mathbf{G} &= \begin{vmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,N} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ g_{M,1} & g_{M,2} & \cdots & g_{M,N} \end{vmatrix} \end{aligned} \qquad (5.82)$$

Consequently, for a given input word $\mathbf{U}$, the ith and jth digits of the corresponding codeword are

$$\begin{aligned} v_i &= \sum_{l=1}^{M} g_{l,i} u_l \\ v_j &= \sum_{l=1}^{M} g_{l,j} u_l \end{aligned} \qquad (5.83)$$

for $i, j = 1, 2, \cdots, N$. Although the symbols of the input words and the output words can be selected from two different alphabets, it is assumed that all u_i's, and v_i's are elements of $GF(L)$, where L is a power of 2. Furthermore, define two new variable sets $\hat{v}_i$ and $\hat{v}_j$

$$\hat{v}_i = v_i - \sum_{\substack{l=1 \\ l \ne p,r}}^{M} g_{l,i} v_l = g_{p,i} v_p + g_{r,i} v_r$$

$$\hat{v}_j = v_j - \sum_{\substack{l=1 \\ l \neq p,r}}^{M} g_{l,j} v_l = g_{p,j} v_p + g_{r,j} v_r \tag{5.84}$$

where p, and r can take any value between 1, and N. They should be selected in such a way that the equations

$$\hat{v}_i = g_{p,i} v_p + g_{r,i} v_r \tag{5.85}$$

$$\hat{v}_j = g_{p,j} v_p + g_{r,j} v_r \tag{5.86}$$

form a system of linear equations with a unique solution. The only case when this is not true is when two columns of G are linearly dependent. Otherwise, r and p satisfying the above conditions do exist. To calculate the number of codewords whose ith and jth digits are

$$v_i = a$$
$$v_j = b,$$

with $a, b \in L$-ary alphabet. assume that $\hat{v}_i$ is also fixed, such that

$$\hat{v}_i = x$$

with $x \in L$-ary alphabet, then from (5.84)

$$\sum_{\substack{l=1 \\ l \neq p,r}}^{M} g_{l,i} u_l = v_i - \hat{v}_i = a - x. \tag{5.87}$$

Since the number of degrees of freedom of U is $M - 3$ in this equation, there are L^{M-3} possible combinations of u_l's, which satisfy (5.87). Any of these combinations of u_l's, will generate a $\hat{v}_j$, such that

$$\hat{v}_j = b - \sum_{\substack{l=1 \\ l \neq p,r}}^{M} g_{l,j} v_l = b - a + x = y \tag{5.88}$$

with y being an element of $GF(L)$. If (5.85) and (5.86) form a set of linear equations with a unique solution. then for given $\hat{v}_i = x$ and $\hat{v}_j = y$. unique values for u_p and u_r can be found; therefore concluded that if $v_i = a$. $v_j = b$. and $\hat{v}_i = x$. then there exist L^{M-3} different codewords satisfying equations (5.84). (5.85). and (5.86). For L possible values of $\hat{v}_i$, (5.87) will yield L^{M-3} different combinations of u_i's. As a result there are L^{M-2} codewords with $v_i = a$. and $v_j = b$. The total number of codewords is already known to be L^M. because every digit of a length M input word can take one of the L different values within $GF(L)$. Since every codeword is equally likely to be selected by a random input word sequence. the probability of transmitting a codeword V with $v_i = a$. and $v_j = b$ is

$$P(v_i = a, v_j = b) = \frac{L^M - 2}{L^M} = \frac{1}{L^2} \tag{5.89}$$

Equation (5.89) holds for all $a, b \in L$-ary alphabet. Consequently.

$$P(v_i = a) = \sum_{b \in GF(L)} P(v_i = a. v_j = b) = L\frac{1}{L^2} = \frac{1}{L}. \tag{5.90}$$

Assume that every element $a_0. a_1, \ldots, a_{L-1}$ of the L-ary alphabet is mapped into one of the signal points $c_0. c_1. \ldots. c_{L-1}$. with

$$c_i = \alpha_i + j\beta_i.$$

The expected value of the mth output digit $b_{m.0}^{(l)}$ of the encoder at time l is

$$E\left\{b_{m,0}^{(l)}\right\} = \sum_{i=0}^{L-1} c_i\, P(b_{m,0}^{(l)} = c_i) = \frac{1}{L}\sum_{i=0}^{L-1} c_i. \tag{5.91}$$

Also, the crosscorrelation of two output digits is

$$\begin{aligned} E\left\{b_{m,0}^{(l)} b_{m,0}^{(l)}\right\} &= \sum_{i,j=0}^{L-1} c_i c_j P(b_{m,0}^{(l)} = c_i, b_{m,0}^{(l)} = c_j) = \frac{1}{l^2}\sum_{i,j=0}^{L-1} c_i c_j \\ &= \left(\frac{1}{L}\sum_{i=0}^{L-1} c_i\right)\left(\frac{1}{L}\sum_{j=0}^{L-1} c_j\right). \end{aligned} \tag{5.92}$$

Interestingly, by inserting (5.91) into (5.92) we find

$$E\left\{b_{m,0}^{(l)}b_{n,0}^{(l)}\right\} = E\left\{b_{m,0}^{(l)}\right\}E\left\{b_{n,0}^{(l)}\right\} = E^2\left\{b_{m,0}^{(l)}\right\} = E^2\left\{b_{n,0}^{(l)}\right\}. \tag{5.93}$$

As a result, if the ith and the jth columns of the generator matrix G of a linear block code consist of two linearly independent, nonzero vectors, then the ith and the jth output digits of the encoder output are mutually independent, and therefore are not correlated.

Returning to (5.80) the condition on the probability function can be disregarded, and so can be rewritten

$$\sum_{\bar{r}_0 \in \bar{C}^0} g_{\Re,f}(q_0)P(b_{i,0}^{(l)} = \bar{r}_0) = 0, \tag{5.94}$$

for $0 \leq f \leq G-1$ and $1 \leq i \leq N$. From (5.90) since $P(b_{i,0}^{(l)} = \bar{r}_0)$ is constant for all $\bar{r}_0 \in \bar{C}^0$, (5.94) is finally simplified to

$$\sum_{q_0 \in \bar{C}^0} g_{\Re,f}(q_0) = 0 \tag{5.95}$$

for $0 \leq f \leq G-1$.

If the conventional encoder is a convolutional encoder, than it is slightly more complicated to simplify (5.80). Given the state of the encoder entering at time interval l, all the branches diverging from that state are equally likely to be selected. Therefore (5.80) can be simplified to

$$\sum_{q_0 \in s_h} g_{\Re,f}(q_0) = 0 \qquad \text{for } 0 \leq f \leq G-1, \text{ and all encoder states } h, \tag{5.96}$$

where s_h denotes the set of signals associated with the branches diverging from state h.

Equations (5.95) and (5.96) can be combined to construct a common sufficient condition for $E\left\{\alpha_i^{(l)}\alpha_j^{(l)}\right\} = 0$ for $i \neq j$: For every possible selection of the subconstellation and for every possible state of the conventional encoder, the summation of the real coordinates of all possible signal points that can be selected by the current input bits of the conventional encoder must be equal to zero.

A similar sufficient condition can be stated for $E\left\{\beta_i^{(l)}\beta_j^{(l)}\right\} = 0$ for $i \neq j$, only considering the imaginary coordinates of the above mentioned signals, and it can also easily be shown that these conditions are also sufficient for (5.69), and for (5.70) when $i \neq j$. Now we want to find the condition for

$$E\left\{\alpha_i^{(l)}\beta_i^{(l)}\right\} = 0 \qquad \text{for } \ 1 \leq i \leq N. \tag{5.97}$$

By modifying (5.76)

$$E\left\{\alpha_i^{(l)}\beta_i^{(l)}\right\} = \sum_{e=0}^{C-1} E\left\{g_{\Re,e}(b_{i,0}^{(l)})g_{\Im,e}(b_{i,0}^{(l)})\right\} P(k_i = e). \tag{5.98}$$

Furthermore

$$E\left\{g_{\Re,e}(b_{i,0}^{(l)})g_{\Im,e}(b_{i,0}^{(l)})\right\} = \sum_{\bar{r}_0 \in s_h} g_{\Re,e}(\bar{r}_0)g_{\Im,e}(q_0)P(b_{i,0}^{(l)} = \bar{r}_0) = 0. \tag{5.99}$$

Since every point in $\tilde{C}^0$ is equiprobable

$$\sum_{\bar{r}_0 \in s_h} g_{\Re,e}(\bar{r}_0)g_{\Im,e}(q_0) = 0. \tag{5.100}$$

By following the same procedure the condition stated in (5.72) is modified to

$$\sum_{\bar{r}_0 \in s_h} \left(g_{\Re,e}(\bar{r}_0)\right)^2 = \sum_{\bar{r}_0 \in s_h} \left(g_{\Im,e}(\bar{r}_0)\right)^2. \tag{5.101}$$

The sufficient condition that (5.100) and (5.101) hold is that every subconstellation should have 90 ° rotational symmetry. If this condition is combined with the previous condition the final conclusion is:

A transmission system which utilizes a linear conventional code combined with the CO method to achieve some shape gain will not suffer from any bandwidth expansion or from any undesirable behavior regarding its spectral properties because of nonequiprobable signalling if *for any given selection of the subconstellations and for any given state of the conventional encoder the set of signal points that can be selected by the current input bits of the conventional encoder, forms a signal constellation which has* 90° *rotational symmetry.*

5.2.2 Simulations of Trellis Shaping

Trellis shaping , a very powerful method to select signal points within a constellation according to an approximated Gaussian distribution, was first suggested by Forney in [3] in order to minimize average signal power, and was discussed in Chapter 3. Forney also noted that with a simple 4-state code it is possible to achieve a 1 dB shape gain in expense of a small increase in complexity. In this thesis we question whether trellis shaping also causes an increase in bandwidth or equivalently whether trellis shaping has any effect on the power spectrum of the output signal sequence.

The most practical way of investigation is simulation of the output signal sequence generated by different modulators equipped with trellis shaping. 16 different simulations have been accomplished by utilizing the combinations of two different convolutional codes for shaping purposes, four different signal constellations and two different initial regions, i.e., the region which contains only the signal points selected by (y_j, w_j, t_j) bit sequence as shown in Fig. 3.9. All these systems are combined with QAM. In the following section each of these alternatives will be explained and the simulation results summarized in a table in order to reach some conclusions.

Components of Modulation System with Trellis Shaping

The block diagram of the system which is used for simulation of the output signal sequence is given in Fig. 3.9. Now, let's see the components of the system.

$\mathbf{G}_c$ encoder for code $\mathbf{C}_c$: $\mathbf{G}_c$ is a rate 1/2 ordinary convolutional encoder, that is placed to attain conventional coding gain. The generator matrix is given in D-notations as

$$\mathbf{G}_c = [1 + D^2, D] \tag{5.102}$$

$(\mathbf{H}_s^{-1})^T$ Coset representative generator: Throughout the simulation chain we used two different convolutional codes for shaping purposes which are labeled as $\mathbf{C}_{s1}$ and $\mathbf{C}_{s2}$. Those two codes are characterized by their generator matrixes as

$$\mathbf{G}_{s1} = [1 + D^2, 1 + D + D^2]$$

$$\mathbf{G}_{s2} = [1 + D^2, D] \tag{5.103}$$

The minimal feedbackfree syndrome-former generators $(\mathbf{H}_s)^T$ for both codes are

$$\begin{aligned}(\mathbf{H}_{s1})^T &= [1 + D + D^2, 1 + D^2]^T \\ (\mathbf{H}_{s2})^T &= [D, 1 + D^2]^T\end{aligned} \tag{5.104}$$

Possible left inverses $(\mathbf{H}_s^{-1})^T$ for these syndrome former matrices that may be used as a coset representative generator matrix in our simulations include

$$\begin{aligned}(\mathbf{H}_{s1}^{-1})_a^T &= [1 + D + D^2, D^2] \\ (\mathbf{H}_{s1}^{-1})_b^T &= [0, \frac{1}{1 + D^2}] \\ (\mathbf{H}_{s2}^{-1})_a^T &= [D, 1] \\ (\mathbf{H}_{s2}^{-1})_b^T &= [0, \frac{1}{1 + D^2}] = (\mathbf{H}_{s1}^{-1})_b^T\end{aligned} \tag{5.105}$$

If the modulation system did not include decoder for C_s, i.e., no trellis shaping would be employed, then the n_c coded bits, n_u uncoded bits and n_s bits of the output of the coset representative generator circuit $(\mathbf{H}_s^{-1})^T$ would determine the signal point to be transmitted. Therefore decoder for C_s can also be interpreted as a circuit which replaces the by-the-rest-of-the-system-selected-candidate sequence by one of its equivalent, in order to decrease average transmitted power. The set of possible such candidate signals is called as initial region. In the cases that $(\mathbf{H}_{s1}^{-1})_a^T$ and $(\mathbf{H}_{s2}^{-1})_a^T$ are used as coset representative generator, the initial region R_i is the whole signal constellation, but if $(\mathbf{H}_{s1}^{-1})_b^T$ or $(\mathbf{H}_{s2}^{-1})_b^T$ is applied then the first bit of the output word of the coset representative generator will be always zero; consequently, the initial region R_i will be only the half of the total signal constellation.

Signal Constellations: Four different signal constellations with different subregions are utilized. Fig 5.11 shows the shaping subregions of the constellations. Here are the brief descriptions of these constellations:

Sig32a: This 32-point constellation consists of points within two dimensional half integer lattice translate $\mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$ and is the same constellation as in Fig. 3.10. This constellation has 90 ° rotational symmetry property.

Sig32b: This is 0.2 dB more power efficient version of Sig32a, but it has only 180 ° rotational symmetry.

Sig256a: It is a 256-point two-dimensional square shaped constellation, with 180 ° rotational symmetry.

Sig256b: This constellation is similar to Sig256a, but the subregions do not have any kind of rotational symmetry.

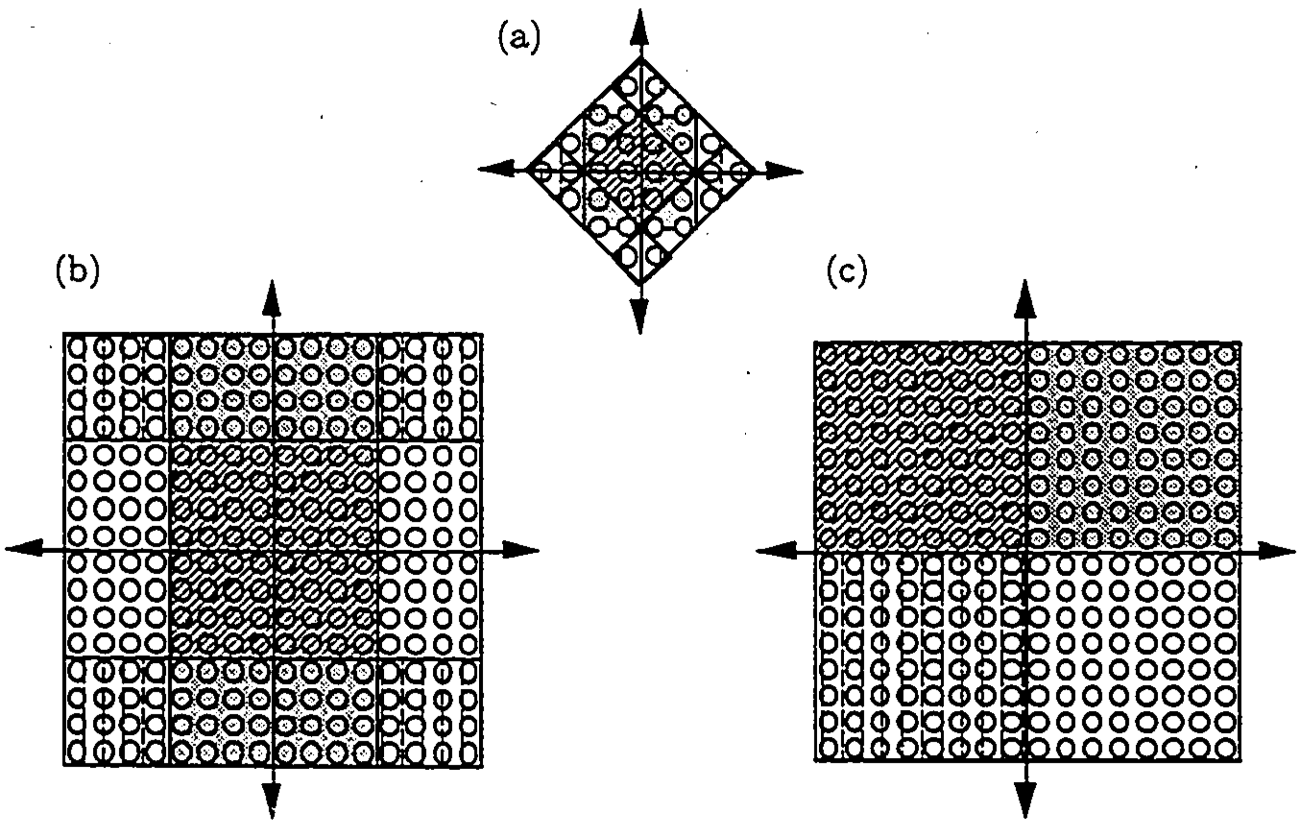


Figure 5.11: Illustration of division of constellations into subregions. (a) Sig32b; (b) Sig256a; (c) Sig256b.

Simulation Results

After designing each system, a very long ($\approx$ 400000 symbols long) output symbol sequence was simulated. Thereafter, using the numerical methods explained in Chapter 4, the power spectrum of each output sequence, and 90 % fractional power bandwidth was calculated. These results were compared with the results obtained from systems which do not employ trellis shaping. The consequences of this simulation are summarized in Table 5.6.

Table 5.6: Simulation results. The corresponding 16-point equiprobable signalling has the following parameters: $H(\bar{C}) = 4.0$, $P_{av}(\bar{C}) = 2.5$, $BW = 0.51$ rad, and 128 point equiprobable signalling: $H(\bar{C}) = 7.0$, $P_{av}(\bar{C}) = 20.5$, $BW = 0.51$ rad.

constellation $(\mathbf{H}_s^{-1})^T$	Sig32a	Sig32b	Sig256a	Sig256b
$(\mathbf{H}_{s1}^{-1})_a^T$	4.18 2.21 0.51	4.15 2.15 0.51	7.15 17.25 0.51	7.23 18.15 0.50
$(\mathbf{H}_{s1}^{-1})_b^T$	4.16 2.23 0.51	4.27 2.34 0.51	7.28 18.90 0.50	7.23 18.13 0.51
$(\mathbf{H}_{s2}^{-1})_a^T$	4.17 2.21 0.51	4.15 2.15 0.51	7.15 17.29 0.51	7.23 18.15 0.51
$(\mathbf{H}_{s2}^{-1})_b^T$	4.17 2.23 0.51	4.28 2.35 0.51	7.28 18.87 0.51	7.23 18.13 0.51
How to read the table	$\overline{H}(\bar{C})$ $\overline{P}_{av}(\bar{C})$ BW in (rad)			

Discussion

Every system simulated, regardless of the constellation, initial region R_1 and the code used for shaping purposes, achieves some power efficiency by trellis shaping over equiprobable signalling. This gain can reach up to 0.66 dB if the signal points are selected within a 32-point signal constellation, or 0.75 dB if a 256-point signal constellation is used.(The latter gain represents the power efficiency of 256-point nonequiprobable signalling via trellis shaping over equiprobable signalling with 128-point cross shaped constellation. If it is compared with signalling by a square shaped 128-point equiprobable constellation, then the gain obtained by trellis shaping becomes 0.92 dB.)

As the size of the constellation increases the shape gain also increases. This is due to a better approximation of the probability distribution of points to a Gaussian distribution.

There has been no change observed in the shape gain due to the selection of the initial region R_1, but the selection of the shaping code has some small effect on the efficiency of the system.

As a final remark, all the simulated systems had the same 90% fractional power bandwidth, with the result that trellis shaping most probably does not effect the bandwidth of the system. In other words, it is possible to obtain shaping gain via trellis shaping without any expense in terms of required transmission bandwidth for reliable communications.

Intuitively, it can be concluded that the conditions derived for the CO method are also valid for systems employing trellis shaping, since all systems obey these conditions. Note that although the sig256b constellation does not have any kind of symmetry, for every symbol interval the set of signal points that can be selected with current input words, regardless of the state of the overall transmission system has 180° rotational symmetry which is enough to assure that no bandwidth expansion will occur if the constellation is combined with QAM.

In the following two figures we draw the power spectrum found using the periodogram averaging method. The graphs correspond only to the first part of the (4.33) since the second factor C_n corresponds to the power spectrum of the windowing function. Clearly, these graphs reveal that the spectrum of output signal sequence in both cases, equiprobable or nonequiprobable signalling, is simply flat or white, therefore the elements of the output signal sequence, which are uncorrelated with each other in the case of equiprobable, remain uncorrelated after trellis shaping is applied. This fact explains why trellis shaping does not affect the power spectrum function.

The last two figures show the complete power spectral density function for equiprobable and nonequiprobable signalling. The only noticable difference we notice is that there is less than 1 dB shift between both cases which corresponds to the shape gain obtained by trellis shaping.

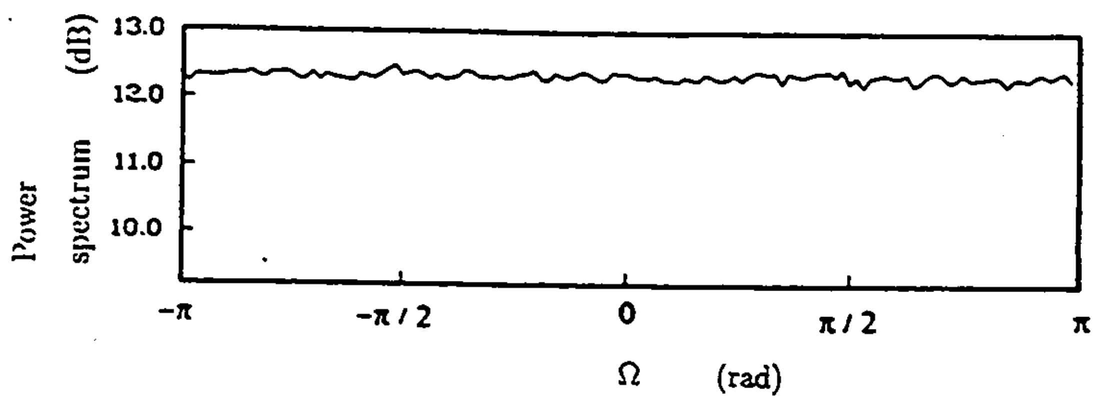


Figure 5.12: Power spectrum of nonequiprobable signalling with trellis shaping without considering the effect of the window.

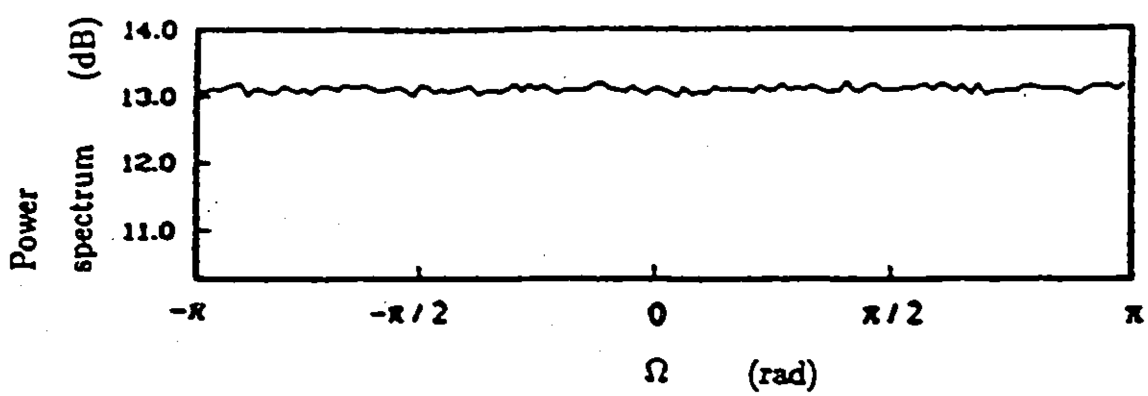


Figure 5.13: Power spectrum of equiprobable signalling without considering the effect of the window.

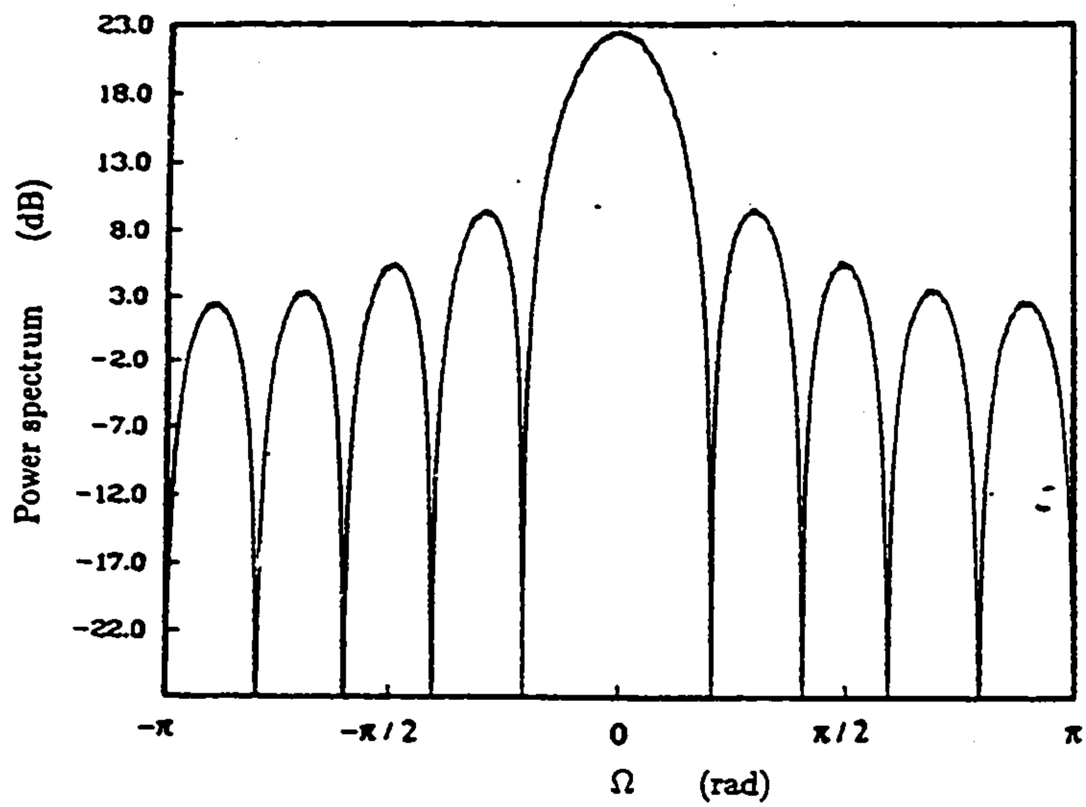


Figure 5.14: Power spectrum of nonequiprobable signalling with trellis shaping .

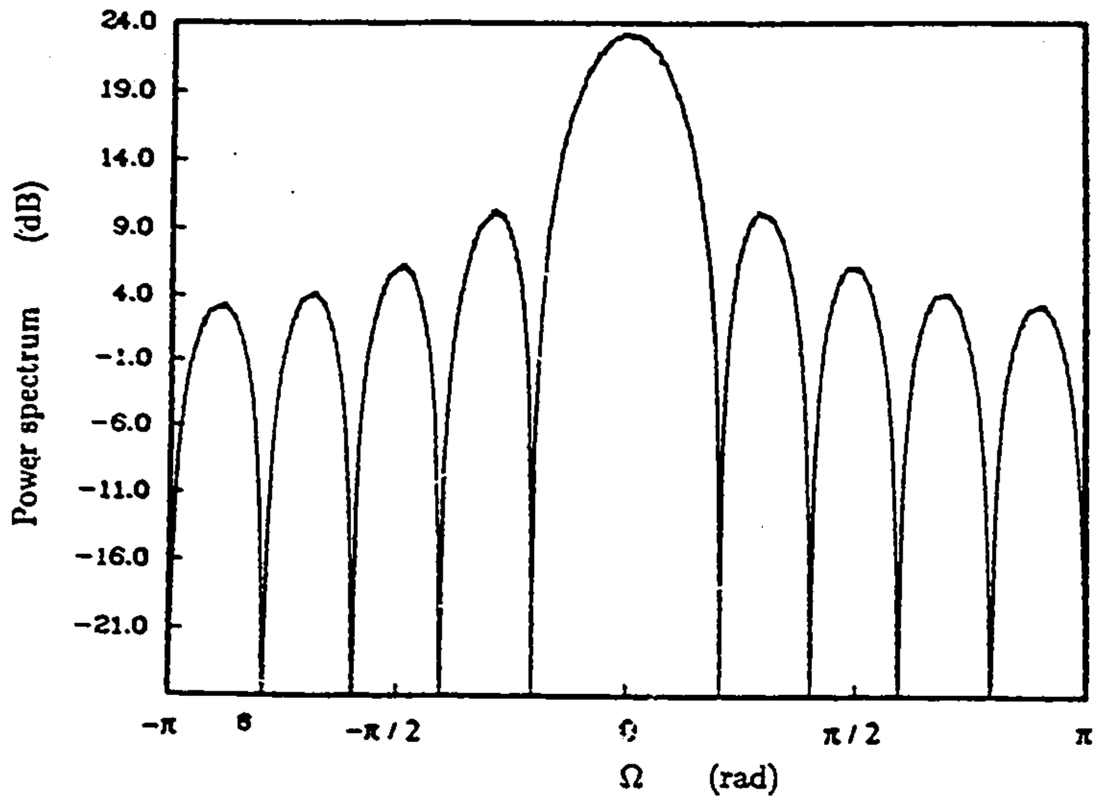


Figure 5.15: Power spectrum of equiprobable signalling .

Chapter 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

In this thesis the power spectrum of channel signals generated by nonequiprobable signallinghas been investigated.

The first half of this work examined the effect of nonequiprobable distribution of points in a signal constellation on the power spectrum. Transmission systems combined with various equiprobable and nonequiprobable constellations were designed. Using the Markov chain model, the power spectrum of the output signal sequence of each system was calculated and used to determine the fractional 90% power bandwidth. The results of these simulations and theoretical calculations revealed the following.

The probability distribution function of points in a constellation has a direct effect on the power spectrum of the output signal sequence. This effect is usually an expansion of the transmission bandwidth. However if certain design rules are obeyed, this bandwidth expansion does not occur. The sufficient conditions that were established are:

(i) If any orthogonal signal pair is used to represent a two dimensional constellation, any 90 ° rotation of the signal constellation should not effect the system.

(ii) If QAM is used to generate the output signal sequence, the system should be invariant to a 180 ° rotation of the constellation.

The second half of the research was to investigate the effects on the power spectrum of the correlation between the elements of the output signal sequence, due to implementation of shaping codes.

The calculation of the power density function by use of a Markov chain model was modified to calculate the power spectrum of the sequence generated by a block encoder that is combined with a two-dimensional constellation.

Various systems employing the Calderbank & Ozarow method and trellis shaping were simulated. In these simulations, only linear codes were considered for both conventional coding and shaping purposes.

Simulation results and theoretical calculations reveal that systems employing the CO method are guaranteed to generate the same power spectrum function as those which do not employ CO method, if a coset code is used for conventional coding purposes. In addition,

(i) for any given selection of the subconstellation and for any given state of the conventional encoder, the set of signal points that can be selected within any two dimensional constellation by the current input bits of the conventional encoder, must form a signal constellation that has 90 ° rotational symmetry, or

(ii) for a QAM, the set of signal points those are possible to be selected by the current input bits must form a 180 ° rotational symmetric signal constellation.

The simulation results of the systems with trellis shaping revealed the following:

As the size of the signal constellation increases the shape gain obtained by the system also increases. This is due to imposing a better approximation of a truncated Gaussian function on the distribution of points in the constellation.

The selection of the initial region does not effect the shape gain but the code applied as a shaping code has a small effect on the magnitude of the gain.

It is believed that if the conditions stated for systems designed using the CO method are satisfied for systems with trellis shaping, the power spectrum of the output signal sequence will be indifferent to the changes in the distribution of the signal points.

Consequently, it is concluded that geometrical symmetry of a signal constellation is an important factor in the band efficiency of a memoryless transmission system. Since blockcoding combined with two-dimensional signalling is equivalent to memoryless multidimensional signalling and trellis coding with two-dimensional signalling is equivalent to memoryless infinite dimensional signalling, the code itself should also have a contribution to the symmetry of the overall equivalent constellation, and hence it should have an impact on the band efficiency of the system. From the results of this thesis it is to conjectured that linearity of a code is sufficient to obtain an overall symmetric equivalent constellation if the two-dimensional constellation combined with this code has some geometrical symmetry.

6.2 Suggestions for Further Research

The conditions stated in this thesis can be applied to any system which uses nonequiprobable signalling. Further research may reveal more flexible conditions for specific applications.

The conditions given for systems using trellis shaping are based on simulation results and on the senses developed in deriving such conditions for previously described systems. Therefore a possible area of research is to derive these conditions theoretically.

Nonlinear codes can also yield shape gain with trellis shaping. The effect on the power spectrum of such systems bears further investigation.

Appendix A

Continuous Approximation

The average signal power $P(\tilde{C})$ of a large constellation $\tilde{C}$ of equiprobable points can be calculated by approximating the discrete distribution of points in the constellation by a continuous distribution that is uniform within the boundary region R_B of $\tilde{C}$, and zero elsewhere, such that

$$P(\tilde{C}) \simeq P(R_B) \triangleq \frac{2}{NV(R_B)} \int_{R_B} \|r\|^2 \, dV \tag{A.1}$$

In this equation, $\|r\|^2$ denotes the energy of the point r within R_B, and $V(R_B) = \int_{R_B} dV$ is the volume of the region enclosed by R_B. Equation (A.1) can also be written in the form

$$P(\tilde{C}) = 2G(R_B)V(R_B)^{2/N}, \tag{A.2}$$

where

$$G(R_B) = \frac{\int_{R_B} \|r\|^2 \, dV}{NV(R_B)^{1+2/N}} \tag{A.3}$$

denotes the normalized or dimensionless second moment. $G(R_B)$ results from taking the average squared distance from a point in R_B to the centroid, and normalizing to obtain a dimensionless quantity [22]. The second moment $G(R_B)$ is invariant to scaling operation, since

$$\int_{R_B} \|\alpha r\|^2 \, d(\alpha V) \quad = \quad [\int_{R_B} \|r\|^2 \, dV] \alpha^{2+N}$$

$$
\begin{aligned}
V(\alpha R_B)^{1+2/N} &= V(R_B)^{1+2/N}\alpha^{N(1+2/N)} \\
&= V(R_B)^{1+2/N}\alpha^{N+2}
\end{aligned} \tag{A.4}
$$

yielding,

$$
G(\alpha R_B) = \frac{\left[\int_{R_B} \|r\|^2 \, dV\right] \alpha^{2+N}}{N V(R_B)^{1+2/N} \alpha^{2+N}} = G(R_B). \tag{A.5}
$$

$G(R_B)$ is also invariant to orthogonal transformations of R_B, and to the cartesian product operation [18].

Appendix B

Limits on Shape Gain

The volume of an N-dimensional (N is assumed to be even) spherical region $\otimes_N$ with a radius R is given in [18] to be

$$V(\otimes_N) = \frac{\pi^{N/2} R^N}{(N/2)!}, \tag{B.1}$$

and the average signal power of the constellation bounded by this region is found by the continuous approximation method (see Appendix A) to be

$$P_{av}(\otimes_N) = \frac{R^2}{(N/2+1)}. \tag{B.2}$$

According to the definition of shape gain given by (3.17), the shape gain of an N-dimensional spherical region is

$$\gamma_s(\otimes_N) = \frac{(V(\otimes_N))^{2/N}}{6P_{av}(\otimes_N)} = \frac{\pi(\frac{N}{2}+1)}{6[(\frac{N}{2})!]^{2/N}}. \tag{B.3}$$

For large N, $(N/2)!$ can be approximated by the Stirling formula

$$(\frac{N}{2})! \sim \sqrt{N\pi}(\frac{N}{2e})^{N/2}, \tag{B.4}$$

where $\sim$ denotes asymptotically equal and means that the expression in (B.4) approaches equality as N becomes arbitrarily large [32]. Using (B.4) in (B.3), yields for large N

$$\gamma_s(\otimes_N) \sim \frac{\pi(\frac{N}{2}+1)}{6(N\pi)^{1/N}(\frac{N}{2e})} \sim \frac{2\pi e(\frac{N}{2}+1)}{6N} = \frac{\pi e}{6} = 1.53 \text{ dB}. \tag{B.5}$$

Equation (B.5) illustrates that the shape gain obtained by a spherical boundary region of an N-dimensional constellation $\bar{C}$ approaches 1.53dB as N becomes arbitrarily large. The constellation expansion ratio of the 2-dimensional constituent constellation $\bar{C}_2$ induced by $\bar{C}$, and the peak to average power ratio in two dimensions are given in [18] for the spherically shaped constellations as

$$CER_{s,2}(\otimes_N) = \frac{\pi R^2}{(V(\otimes_N))^{2/N}} = (\frac{N!}{2})^{2/N} \tag{B.6}$$

$$\begin{aligned} PAR_2(\bar{C}) &= \frac{R^2}{P_{av}(\otimes_N)} = (\frac{N}{2}+1) \\ &= (\frac{6}{\pi})\, \gamma_{s,\otimes}\, CER_{s,2}(\otimes_N) \end{aligned} \tag{B.7}$$

From (B.7), the peak to average power ratio in two dimensions grows without bound as N becomes arbitrarily large. This is mainly because of the increase in $CER_{s,2}(\otimes_N)$ as N gets larger. High $PAR_2(\bar{C})$, and $CER_{s,2}(\otimes_N)$ are undesirable in communications systems. The optimum trade-off between shape gain and $PAR_2(\bar{C})$, or $CER_{s,2}(\otimes_N)$ is calculated by bounding the peak energy of the constituent 2-dimensional constellation $\bar{C}_2$ within a circle. As a result, the optimum trade-off is reached in finite dimensions when the distribution probability of points in $\bar{C}_2$ approaches a Gaussian distribution [18].

Appendix C

Calculation of the Average Signal Power of QAM

Quadrature Amplitude Modulation systems are widely used for digital communications purposes. The general structure of the signal constellation utilized with QAM is given in section 5.1.1. Every signal point in the constellation is characterized by its two coordinates, and the two orthogonal base signals $p_1(t)$ and $p_2(t)$ of the constellation. Generally, a sine and cosine signal are used as base signals:

$$p_1(t) = \begin{cases} \cos\omega_c t & t \in [0, T) \\ 0 & \text{elsewhere} \end{cases}$$

$$p_1(t) = \begin{cases} -\sin\omega_c t & t \in [0, T) \\ 0 & \text{elsewhere} \end{cases}$$

The channel waveform assigned to the mth signal point with signal space coordinates (A_{mc}, A_{ms}) is then formulated as

$$s_m(t) = \sqrt{\frac{2\varepsilon}{T}}[A_{mc}\cos\omega_c t - A_{ms}\sin\omega_c t]p(t) \tag{C.1}$$

where $p(t)$ is the time-limited square pulse with unit amplitude in the interval $t \in [0, T)$. In equation (C.1) ε denotes a constant representing the energy of base signals; and it directly determines the minimum separation between the signal points in the constellation according to the equation

$$d_{\min} = 2\sqrt{\varepsilon}. \tag{C.2}$$

The minimum separation, or minimum distance between the signal points d_{min}, is a very important variable determining the performance of a transmission system. Conventionally, ε is adjusted so that the minimum separation is equal to 1 for all constellations, before any comparisons are made. Finally, the carrier frequency of the modulator is labeled f_c.

The channel signal sequence is a superposition of the individual waveforms transmitted every signalling interval T.

$$\begin{aligned} s(t) &= \sum_{i=-\infty}^{\infty} s_m^{(i)}(t-iT) \\ &= \sum_{i=-\infty}^{\infty} \sqrt{\frac{2\varepsilon}{T}}[A_{mc}^{(i)} \cos\omega_c(t-iT) - A_{ms}^{(i)} \sin\omega_c(t-iT)]p(t-iT) \end{aligned} \quad \text{(C.3)}$$

The average power of $s(t)$ is then,

$$\begin{aligned} P_{av} &= \langle |s(t)|^2 \rangle \\ &= \lim_{N\to\infty} \frac{1}{2NT} \int_{-NT}^{NT} \frac{2\varepsilon}{T} \sum_{i=-\infty}^{\infty} \langle [A_{mc}^{(i)} \cos\omega_c(t-iT) - A_{ms}^{(i)} \sin\omega_c(t-iT)]^2 \rangle p^2(t-iT)\, dt, \end{aligned} \quad \text{(C.4)}$$

where the notation $\langle x \rangle$ is used to indicate the ensamble average. After certain simplifications,

$$\begin{aligned} P_{av} &= \lim_{N\to\infty} \frac{1}{2NT} \frac{2\varepsilon}{T} \sum_{i=-N}^{N-1} \int_0^T \langle [A_{mc}^{(i)} \cos\omega_c t - A_{ms}^{(i)} \sin\omega_c t]^2 \rangle\, dt \\ &= \lim_{N\to\infty} \frac{2\varepsilon}{2NT^2} \sum_{i=-N}^{N-1} \left(\int_0^T \langle A_{mc}^{(i)\,2} \rangle \cos^2\omega_c t\, dt + \int_0^T \langle A_{ms}^{(i)\,2} \rangle \sin^2\omega_c t\, dt \right. \\ &\quad \left. - \int_0^T \langle A_{mc}^{(i)} A_{ms}^{(i)} \rangle \sin^2 2\omega_c t\, dt \right) \\ &= \lim_{N\to\infty} \frac{2\varepsilon}{2NT^2} \sum_{i=-N}^{N-1} \frac{T}{2} \left(\langle A_{mc}^{(i)\,2} \rangle + \langle A_{ms}^{(i)\,2} \rangle \right). \end{aligned} \quad \text{(C.5)}$$

Since the expected values of $A_{mc}^{(i)\,2}$ and $A_{ms}^{(i)\,2}$ are time invariant, the average signal power of a QAM system is found to be

$$P_{av} = \frac{\varepsilon}{T} \left(\langle A_{mc}^2 \rangle + \langle A_{ms}^2 \rangle \right). \quad \text{(C.6)}$$

Appendix D

Shape Gain of a 2D Circular Boundary Region

The average power of a square shaped constellation with a uniform distribution of points can be calculated using the continuous approximation method (see Appendix A):

$$\begin{aligned} P\left(R_{B,sq}\right) &= \frac{2}{2V\left(R_{B,sq}\right)} \int_{R_{B,sq}} \|r\|^2 \, dV; \\ &= \frac{1}{a_{sq}^2} \int_{-\frac{a_{sq}}{2}}^{\frac{a_{sq}}{2}} \int_{-\frac{a_{sq}}{2}}^{\frac{a_{sq}}{2}} \left(x^2 + y^2\right) dx \, dy = \frac{a_{sq}^2}{6}. \end{aligned} \tag{D.1}$$

In the equation above, a_{sq} denotes the length of one side of the square boundary region $R_{B,sq}$. Similarly, the average power $P(R_{B,cir})$ of a circular shaped constellation is calculated as

$$P\left(R_{B,cir}\right) = \frac{2}{2v\left(R_{B,cir}\right)} \int_0^{2\pi} \int_0^{a_{cir}} r^2 r \, dr \, d\theta = \frac{2\pi a_{cir}^4}{4\pi a_{cir}^2}. \tag{D.2}$$

In order to make a fair comparison, the volume of each constellation should be equalized so that the entropies of each constellation are equal. The volume (in 2D vector space the volume of a constellation actually corresponds to the area of itself) of a circle with a radius a_{cir} is equal to the volume of a square with a side length a_{sq} if

$$a_{cir} = \frac{1}{\sqrt{\pi}} a_{sq}. \tag{D.3}$$

Inserting equation (D.3) into (D.2) yields

$$p(R_{B,cir}) = \frac{a_{sq}^2}{2\pi}, \tag{D.4}$$

and the resultant shape gain of the circular boundary region in 2 dimensions is

$$\gamma_{s,R_{B,cir}} = \frac{P(R_{B,sq})}{P(R_{B,cir})} = \frac{\frac{a_{sq}^2}{6}}{\frac{a_{sq}^2}{2\pi}} = \frac{\pi}{3} = 0.2\text{dB}. \tag{D.5}$$

Bibliography

[1] E. Biglieri, "Ungerboeck Codes Do Not Shape the Signal Power Spectrum", *IEEE Transactions on Information Theory*, Vol. IT-32, No. 4, pp. 595-596, July 1986.

[2] A. R. Calderbank, L. H. Ozarow, "Nonequiprobable signalling on the Gaussian Channel", in preparation.

[3] G. D. Forney, Jr., "Trellis Shaping", *IEEE Information Theory Workshop*, Ithaca, NY, June 1989.

[4] G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets Part I: Introduction", *IEEE Communications Magazine*, Vol. 25, No. 2, pp. 5-11, February 1987.

[5] G. Ungerboeck, "Trellis-Coded Modulation with Redundant Signal Sets Part II: State of the Art", *IEEE Communications Magazine*, Vol. 25, No. 2, pp. 12-21, February 1987.

[6] P. Galko, and S. Pasupathy, "The Mean Power Spectral Density of Markov Chain Driven Signals", *IEEE Transactions on Information Theory*, Vol. IT-27, No. 6, pp. 746-754, November 1981.

[7] W. R. Bennett, and S. O. Rice, "Spectral Density and Autocorrelation Functions Associated with Binary Frequency-Shift Keying", *The Bell System Technical Journal*, pp. 2355-2385, September 1963.

[8] B. S. Bosik, "The Spectral Density of a Coded Digital Signal", *The Bell System Technical Journal*, Vol. 51, No. 4, pp. 921-932, April 1972.

[9] H. Yasuda, and H. Inose, "Direct Calculation Method for Power Spectra of Pulse Sequences by Means of Transition Probability Matrices", *Electronics and Communications in Japan*, Vol. 53-A, No. 11, pp. 33-39, 1970.

[10] H. Yasuda, "Calculation of Power Spectra of Pulse Sequences by Means of Transition-Probability Matrices", *Electronics and Communications in Japan*, Vol. 54-A, No. 11, pp. 1-7, 1971.

[11] G. L. Cariolaro, and G. P. Tronca, "Spectra of Block Coded Digital Signals", *IEEE Transactions on Communications*, Vol. COM-22, No 10, pp. 1555-1563, October 1974.

[12] D. Divsalar, M. K. Simon, "Spectral Characteristics of Convolutionally Coded Digital Signals", *IEEE Transactions on Communications*, Vol. COM-28, No. 2, pp. 173-186, February 1980.

[13] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient Modulation for Bandlimited Channels", *IEEE Journal on Selected Areas in Communications*, Vol. SAC-2, pp. 632-647, 1984.

[14] G. D. Forney, "Coset Codes - Part I: Introduction and Geometrical Classification", *IEEE Transactions on Information Theory*, Vol. 34, No. 5, pp. 1123-1151, September 1988.

[15] G. D. Forney, "Coset Codes - Part II: Binary Lattices and Related Codes", *IEEE Transactions on Information Theory*, Vol. 34, No. 5, pp. 1152-1187, September 1988.

[16] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals", *IEEE Transactions on Information Theory*, Vol. IT-28, No. 1, pp. 55-67, January 1982.

[17] L.-F. Wei, "Trellis-Coded Modulation with Multidimensional Constellations", *IEEE Transactions on Information Theory*, Vol. IT-33, No. 4, pp. 483-501, July 1987.

[18] G. D. Forney, Jr.and L.-F. Wei, "Multidimensional Constellations-Part I: Introduction, Figures of Merit, and Generalized Cross Constellations", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 6, pp. 877 -892, August 1989.

[19] G. D. Forney, Jr.and L.-F. Wei, "Multidimensional Constellations-Part II: Voronoi Constellations", *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 6, pp. 941-958, August 1989.

[20] J. N. Livingston, "Shaping Gains Using Nonequiprobable Signalling", in preparation.

[21] A. R. Calderbank. "Multilevel Codes and Multistage Decoding". *IEEE Transactions on Communications*. Vol. 37. No. 3. pp. 222-229. March 1989.

[22] A. R. Calderbank. "The Mathematics of Modems". Article intended for publication in the *Mathematical Intelligencer.*

[23] A. R. Calderbank, chapter 4. from a book in preparation.

[24] A. R. Calderbank and N J.A. Sloane, "New Trellis Codes Based on Lattices and Cosets", *IEEE Transactions on Information Theory*. Vol. IT-33. No. 2, pp. 177-195, March 1987.

[25] A. R. Calderbank and N.J.A. Sloane. "Four-Dimensional Modulation with an Eight-State Trellis Code". *AT&T Technical Journal* Vol. 64, No. 5, pp. 1005-1018, May-June 1985.

[26] S. L. Portnoy and A. N. Kozlenko, "Band-Efficient Coding for a Gaussian Channel", *Radiotekhnika*, No. 5, pp. 33-38, 1987.

[27] S.I. Sayegh, "A Class of Optimum Block Codes in Signal Space", *IEEE Transactions on Communications*, Vol. COM-34, No. 10, pp. 1043-1045, October 1986.

[28] S. G. Wilson and M. Lakshman, "Autocorrelation and Power Spectrum of Hadamard Signalling". *IEE Proceedings*, Vol. 135. No. 3, pp. 258-261, June 1988.

[29] F. Amoroso. "The Bandwidth of Digital Data Signals". *IEEE Communications Magazine*, pp. 13-24. November 1980.

[30] P. Galko, "Stochastic Process". *Course Notes* for ELG 5119. Fall 1989.

[31] S. Lin and D.J. Costello,Jr., *Error Control Coding: Fundamentals and Applications*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632, 1983.

[32] E. Kreyszig, *Advanced Engineering Mathematics*, Fifth Edition, John Wiley & Sons, Inc., 1983.

[33] A. V. Oppenheim, R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey 07632, 1989.

[34] J. G. Proakis, *Digital Communications*, Mc Graw Hill, New York, 1989.