# JOINT OPTIMIZATION OF SOURCE AND CHANNEL CODING BASED ON A NONLINEAR ESTIMATE RECEIVER

by

Fenghua Liu

B.A.Sc. National University of Defence Technology, 1983, China

M.E.Sc. Harbin Institute of Technology, 1986, China

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Engineering Science


SIMON FRASER UNIVERSITY
June. 17, 1995

# APPROVAL

**Name:** Fenghua Liu

**Degree:** Doctor of Philosophy

**Title of thesis :** Joint Optimization of Source and Channel Coding Based on a Nonlinear Estimate Receiver

**Examining Committee:** Dr. Andrew Rawicz, Associate Professor, Chairman

Dr. Paul Ho
Senior Supervisor, Associate Professor

Dr. Vladimir Cuperman
Senior Supervisor, Professor

Dr. Tho Le-Ngoc
External Examiner
Professor, Concordia University

Dr. Steve Hardy
Supervisor, Professor

Dr. Jacques Vaisey
Supervisor, Assistant Professor

Dr. Mehrdad Saif
Internal Examiner/Associate Professor

**Date Approved:** August 31 1995

## PARTIAL COPYRIGHT LICENSE



**Title of Thesis/Project/Extended Essay**

**"Joint Optimization of Source and Channel Coding Based on the Nonlinear Estimate Receiver"**

**Author:**

(signature)

Feng-Hua LIU
(name)

August 31, 1995
(date)

# Abstract

This thesis considers the problem of joint optimization of source and channel coding algorithms for communications over the additive white Gaussian noise (AWGN) and the Rayleigh fading channels. In the proposed system, the analog source signal is first compressed by a vector quantizer (VQ). The output of the VQ (the VQ index) is mapped directly into a signal vector in the modulation signal space, and the signal vector is transmitted over a noisy channel. A receiver based on a nonlinear conditional estimate is used to reconstruct a replica of the source signal directly from the received signal. The main blocks to be optimized in the joint source and channel coder are the VQ encoder, the mapping from the VQ index to the modulation constellation, the modulation constellation, and the decoder structure. Subject to constraints on the average energy and bandwidth, the objective is to minimize the mean-square error (MSE) between the original and the reconstructed signal.

Based on Bayes estimation theory, a soft decision vector quantizer (SDVQ) was developed. The optimal decoder for this system computes the conditional mean of the source signal given the received channel signal. The output of such an optimal decoder is a linear combination of the VQ centroids for the SDVQ partition, in which the weighting coefficients are nonlinear functions of the received signal. Several approximate implementations at various channel SNR were also studied. An iterative algorithm is presented for the joint design of the VQ and the modulation signal set. The algorithm first optimizes the VQ codebook for a fixed signal set, and then optimizes the signal set for a fixed VQ codebook. Iterating these two steps until convergence occurs will provide at least a locally optimal solution to the problem. The algorithm was used to design the VQ and signal constellation for a first order Gauss-Markov source operating in the AWGN and Rayleigh fading channels. The simulation results indicate that the system performance is significantly enhanced by

the joint design, especially when the channel signal-to-noise ratio is low. The improvement in the signal-to-noise ratio (SNR) for the reconstructed signal can be up to 5 dB.

Due to the constraints on the VQ encoder delay and VQ complexity, the source coder can not remove all the redundancy in the source. The residual redundancy is modeled as a first order Markov process. We further developed a sequential decoding algorithm to exploit the residual redundancy in order to improve the performance in noisy channels without any bandwidth expansion. The simulation results show that significant improvement can be obtained by using the sequential decoding strategy, especially in a Rayleigh fading channel.

# Dedication

*I dedicate this thesis to my wife, parents and son for their great love.*

# Acknowledgements

I would like to thank my senior supervisors, Dr. Paul Ho and Dr. Vladimir Cuperman, for their support and guidance throughout this thesis project.

I am grateful to the Natural Science and Engineering Research Council (NSERC) of Canada for research grants and the SFU Graduate Fellowship Committee for their financial support through scholarships and fellowships.

# List of Abbreviations

| | |
|---|---|
| AMPS | Advanced Mobile Phone System |
| AWGN | Additive white Gaussian noise |
| COVQ | Channel optimized vector quantizer |
| DPSK | Differential phase shift keying |
| FCC | Federal communication committee |
| FEC | Forward error correction |
| FM | Frequence modulation |
| GLA | Generalized Lloyd algorithm |
| GMSK | Gaussian-Minimum Shift-Keying |
| GSM | Global system for mobile |
| LBG | Linde-Buzo-Gray |
| LD-CELP | Low delay-code excited linear prediction |
| LPC | Linear prediction coefficient |
| LSP | Line spectral pair |
| ML | Maximum likelihood |
| MOR-VQ | Modulation-organized vector quantizer |
| MSE | Mean squared error |
| MSVQ | Multi stage vector quantizer |
| PCM | Pulse coding modulation |
| PSK | Phase shifting keying |
| QAM | Quadrature amplitute modulation |
| QPSK | Quadrature phase shifting keying |
| SD-SDVQ | Sequential decoding soft decision vector quantizer |
| SDVQ | Soft decision vector quantizer |

| | |
|---|---|
| SNR | Signal-to-noise ratio |
| SOVQ | Source optimized vector quantizer |
| TCM | Trellis coded modulation |
| VQ | Vector quantizer |
| VSELP | Vector sum excited linear prediction |
| WMSE | Weighted mean squared error |

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Since the introduction of digital cellular communication about a decade ago, wireless communications have experienced rapid growth. To improve the bandwidth efficiency in the mobile channel, new applications and approaches are being developed at an unprecedented rate, especially in the area of digital speech transmission, modulation, and multiple access technologies. These progresses have been accelerated by the advances of VLSI technologies and the development and application of advanced digital communication and signal processing techniques. Future generations of cellular and other forms of wireless communications will all use digital technologies.

In the United States, the FCC has opened up the existing cellular bands to essentially any technology that mobile communication providers want to use, as long as they continue to meet the needs of mobile users. This means that the new digital systems will likely co-exist with the current analog FM AMPS system [1]-[2] which was developed in the 1970s by Bell Lab. Driven by the cost of new cell sites and the limited radio spectrum, any new digital cellular standard should provide improved capacity over the existing analog counterpart. This can be achieved by adopting low bit rate speech coding, efficient digital modulation, and an effective error protection scheme.

If we look at the development of speech coding, we can see that the objective of speech coding has been to reduce the number of bits used to represent the speech signal. Several speech coding standards have been developed over the decades. The Global System for Mobile (GSM) for European vehicular digital cellular was driven by the need for a common mobile standard throughout Europe and the desire for digital transmission compatible with data privacy [3]. The 13 kbit/s full rate speech coding algorithm used in GSM is based on a regular pulse-excited linear prediction coding algorithm which includes long term prediction. In addition to the 13 kbit/s speech coding rate, channel coding (half-rate convolutional coding plus cyclic redundancy check) is also used, bringing the overall transmission bit rate of the full rate European digital cellular mobile telephone system to 22.8 kbit/s per user [4]. Gaussian-Minimum Shift-Keying (GMSK) is used as the modulation format.

The Northern American digital cellular standard, the IS-54 uses 8 kbit/s vector-sum excited linear prediction (VSELP) [5] as the speech coder and a rate 1/2 convolutional coder as the channel coding scheme. The modulation format is $\pi/4$-shifted QPSK. The immunity to co-channel interference for the chosen modulation is similar to that in the 30 kHz analog FM.

Other well-known speech coding standards include the CCITT 16 kbits/s low-delay code excited linear prediction (LD-CELP) speech coding standard [6], which provides toll-quality over the public telephone lines, and the U.S. Department of Defense 4.8 kbit/s speech coding standard FS-1016 [7], which is used for secure telephone communication.

Advanced digital signal processing techniques are used to achieve data compression in the speech coding standards mentioned above [8]-[10]. However, speech coding is usually designed for the noiseless channel. For the mobile radio channel, error-free transmission is not a practical design goal. In fact, the mobile channel often operates

at error rates as high as 0.001 or 0.01 which overwhelms ordinary speech coders. Typically, reasonable speech quality can be maintained at the 0.001 error rate, and tolerable quality at the 0.01 rate. While the search for lower bit rates to reduce the signal bandwidth drives new developments in speech coding, these new methods actually increase the relative importance of certain output bits of the speech coder and thus can intensify vulnerability to channel errors.

Traditional channel coding techniques [11]-[12], such as error correction or detection, can be employed to protect the transmitted signal from the effect of channel noise. Channel coding normally requires an increase in the transmission bite rate, as well as introducing an additional processing delay incurred in the process of calculating the redundant bits and decoding the information bits. For mobile communications, every bit allocated for channel coding implies sacrificing a bit that would otherwise help to improve the speech quality itself. Furthermore, error correction does not take into account the characteristics of the speech coder. This oversight can lead to an inefficient use of the system power and bandwidth. We can obtain some inspiration from the communication theory in order to design a jointly optimized source and channel coding system.

Generally speaking, digital communication theory deals with two problems: (1) *source coding*, whose purpose is to minimize the number of bits used to represent the signal, subject to a pre-determined fidelity requirement; and (2) *channel coding*, whose purpose is to ensure that the bits used to represent the signal are received correctly despite the existence of various types of interference over the channels. The basis of communication theory, which provides solutions to these two problems, was presenteded by Shannon in two remarkable papers published in 1948 and 1953 [13]-[14]. Shannon proved that source and channel coding could, in principle, be separated. As

a result, the traditional design philosophy for digital communication systems involving analog sources consists of three independent steps: design a source coder which satisfies the required compression ratio subject to the quality requirement; choose a suitable modulation scheme for the channel; and if necessary, design and implement a forward error correction (FEC) scheme as the channel coder [15]. The basic assumption for separating source and channel coding is that they both operate on sequences that are infinitely long. In many practical communication systems, it is impossible to process infinitely long sequences due to the limitations caused by the encoder delay and computational complexity. In this dissertation we consider the problem of joint source and channel coding operating in the noise and fading channels.

## 1.1 Literature Review

The issue of joint optimization of source and channel coding can be traced back to [16]-[18]. Fine [16] formulated the source-channel coding design problem as a joint optimization problem and gave the necessary conditions for an optimum digital encoder-decoder pair for transmission of continuous amplitude data over a digital channel. The model he considered was general enough to include a digital system operating over a noisy discrete-alphabet channel. This work was extended by Gibson and Fisher to include delayed encoding in the same system [20].

Kurtenbach and Wintz [17] considered the problem of designing a scalar quantizer for a binary symmetric channel. The optimum quantization levels and reconstruction values were determined for fixed codewords assigned to the quantization levels. It was found that the optimum quantizer depends on the probability density function (pdf) of the input signal and the channel transition matrix. The mean squared error (MSE) metric was used as the optimization criterion. Farvardin and Vaishampayan [19]

extended Kurtenbach's work by introducing a constraint condition on the quantizer threshold in order to guarantee convergence. Furthermore, they solved the problem of the optimization of codeword assignment using a simulated-annealing algorithm. They observed that not all the available codewords were used for transmission in the optimal system.

As opposed to the rather restrictive systems considered in [16] and [17], Wolf and Ziv [18] considered the optimization of a very general communication system based on the MSE criterion and derived the optimal encoder-decoder structure for this system. In addition, they showed that for any given encoder, the optimal decoder is one that calculates the conditional mean of the source signal given the received signal. Though very useful and relevant to the current study, Wolf and Ziv do not tell us how to select the most appropriate encoder.

Some of the more recent work in the area of combined source and channel coding are presented in [21]-[28]. Modestino and Daut [21] considered the transmission of still images through noisy channels. They demonstrated that in comparison to the uncoded transmission, convolutional coding, when applied according to the characteristic of the source signal, can dramatically improve the quality of the reconstructed images. These improvements were obtained by reducing the rate of the source coder (thus increasing the distortion introduced by the source coder), and using the available bandwidth for channel coding to combat the channel noise. Kumazawa et. al. [22] extended Kurtenbach's work [17] to include vector quantization. They introduced channel transition probability matrices to represent the effect of channel noise and discussed the problem of VQ design in a noisy channel. A similar study was conducted by Farvardin [23]. The key difference between [22] and [23] is that the latter also considered the optimization of the index assignment.

An important step toward robust vector quantization in noisy channels is index assignment optimization. For a vector quantizer with a fixed set of reproduction vectors, the performance in a noisy channel can be improved by optimizing the assignment of channel symbols (binary indices) to reproduction vectors. The codervectors (reproduction vectors) can be considered as points in an Euclidean space. Intuitively, indices which differ in as few bit positions as possible should be assigned to points which are close in the Euclidean space.

One of the first papers to demonstrate the importance of index assignment for scalar quantization was written by Rydbeck and Sundberg [24]. The problem of optimal index assignment for scalar quantization was considered by Farvardin and Vaishampayan [19]. Algorithms for improving the index assignment for a vector quantizer with a fixed codebook were introduced by De Marca et. al. [25]-[26]. A locally optimal solution based on a binary switching algorithm was introduced by Zeger and Gersho [27]-[28] and was called Pseudo-Gray coding because of its similarity to the well known Gray code. An efficient procedure for testing the performance of a bit assignment strategy was recently proposed by Knagenhjelm [29]. An alternative approach for solving the combinational problem generated by the bit assignment optimization is based on simulated annealing [30]. Kleijn and Sukkar [31] proposed a source-dependent channel coding technique based on the simulated annealing in which redundancy was introduced by having channel codewords which are not transmitted, but may be received.

Hagenauer, Seahardri and Sundberg [32] analyzed the effect of digital transmission errors on a family of variable-rate sub-band speech coders. Since different error sensitivities correspond to various source coders, a family of rate-compatible punctured convolutional coders with flexible unequal error protection capacities are used to match the speech coders. In a Rayleigh fading channel with differential phase shift

keying (DPSK) modulation, a 5 dB improvement in channel SNR can be obtained by using 4 levels of error protection. This gain is obtained without requiring extra bandwidth.

The generalized Lloyd algorithm was used by Dunham and Gray [33] and Ayanoglu and Gray [34] to design joint source and channel trellis and predictive trellis waveform coders for a variety of distortion criteria. They demonstrated that system performance could be significantly improved by jointly designing a trellis-based source and channel coding scheme, as opposed to tandeming a trellis vector quantizer (VQ) optimized for a clear channel and a trellis coded modulation scheme.

Recently, Skinnemoen [35] described a novel approach for designing a joint source-channel coder, called modulation-organized vector quantization (MOR-VQ). It was assumed that the modulation set is known when designing the VQ. The most important factor to increase the robustness is the choice of a good mapping from the source space to the modulation space. A solution to this problem is developed based on neural network theory. It was found that the MOR-VQ provides significant robustness against noise while not sacrificing the performance in a noise free channel. MOR-VQ is very simple to design and does not require knowledge of the channel.

All the work that we have made reference to so far is based on a fixed modulation system. One of the contributions of the present dissertation is in re-designing the modulation signal set according to the mean squared error (MSE) criterion which is associated with the reconstructed source vector.

The majority of the work on modulation signal design is mainly concerned with minimizing the probability of error [36]-[40]. The cost of error due to the channel noise is usually assumed to be the same no matter what types of error occur. Such an assumption is not true since the different bits in a digitized source signal may represent different signal parameters and hence have different levels of sensitivity

to errors. Another assumption is that signals in the signal set are used with an equal probability. This assumption is also questionable since most codewords in a practical source coding system do not occur with the same probability. As we know, the MSE criterion is usually used to measure the signal distortion, but standard modulation constellations such as QPSK and QAM are optimized to minimize the error probability. Generally, minimizing the error probability is not equivalent to minimizing the overall mean squared-error. Therefore, the overall distortion could be reduced by re-designing the modulation constellations according to the MSE criterion. Furthermore, as we will see later, the structure of the optimum source decoder is one that makes soft decision, not a maximum likelihood receiver plus an inverse vector quantizer (VQ).

Several studies have been carried out in designing signal sets in order to improve the MSE performance of zero-memory quantizers over a noisy channel. Wong, Steel and Sundberg [41] considered the problem of transmitting PCM signals by QAM modulation format. They defined an error sensitivity factor to measure the effect of the bit error patterns on the reconstructed source signal. By allocating different energy to the transmission of specific bits, it is possible to reduce the error probability for highly sensitive bits at the expense of less sensitive bits, thus minimizing the overall distortion. They reported that 1.85 dB improvement in the channel SNR were obtained for QPSK modulation, as well as 3 to 5 dB improvement for a weighted 16-level and 256-level QAM constellation.

## 1.2 Contributions of the Thesis

The major contribution of this dissertation is the development of an iterative algorithm for the joint design of a vector quantizer and modulation signal set in terms of

the MSE optimization criterion [45]-[47]. A non-linear conditional estimator (the optimum receiver under the MSE criterion) is used as a receiver. The results indicate that system performance on an additive white Gauss noise and Rayleigh fading channel improves significantly with the joint design, especially when the channel signal-to-noise ratio is low [48]. The joint design algorithm mentioned above was also used for the Rayleigh fading channel [49]. It was found that only a moderate improvement can be obtained by symbol-by-symbol decoding algorithm. However, the overall system performance can be further improved by using a sequential decoding strategy to exploit the residual correlation in the codervectors of the VQ [50].

Our approach to the modulation design is related to Farvardin and Vaishampayan's work on joint source and channel coding [44]. In their system, the source encoder was assumed to be a VQ, the channel encoder was a linear transformation from the source space to the modulation space and the decoder was also a linear transformation from the channel space to the source space. Constraints on the energy and bandwidth are imposed on the transmitted signal, and the encoder and decoder mappings are jointly optimized under the MSE criterion. This decoder structure, which is an approximation of the optimal decoder for the case of low channel SNR, is inherently sub-optimal, especially in the high channel SNR region. However, because of the linearity, Farvardin and Vaishampayan were able to adopt a semi-analytical approach in the joint source and channel coding design. This is in contrast to an algorithmic approach presented in this thesis. Note, however, that the decoder considered in this study is based on the optimal conditional estimate receiver rather than its approximate implementation in low SNR channel-linear receiver in [44].

The idea to jointly optimize the source and channel coding algorithm developed in this dissertation is similar to the non-linear optimization technique proposed by Lloyd [42]. However, the original Lloyd's algorithm cannot be directly applied to design a

vector quantizer, since its implementation requires integrating the source pdf over an irregular multi-dimensional Euclidean space. In order to overcome this problem, a training sequence-based approach was introduced by Linde, Buzo and Gray, which did not require the pdf of the source [43]. A generalized Lloyd iteration procedure was used in this thesis for joint source and channel coding design.

## 1.3 Organization of the Thesis

This thesis is partitioned into eight chapters. The optimum performance bound is described in Chapter 2. The vector representation of signal and noise is introduced in Chapter 3. Two noisy channel models which are extensively used in the thesis are also described. A joint VQ-modulation design based on a traditional hard decision receiver is given in Chapter 4. This includes the design of a robust channel optimized vector quantizer, the optimization of the modulation constellation, and the optimization of demodulation, all under the MSE criterion. The optimum soft decision receiver, in terms of the MSE criterion, is derived in Chapter 5. This receiver makes a conditional expectation of the source signal given the received signal. The constrainted gradient search algorithm is used to design the modulation constellation. In Chapter 6, a sequential reconstruction technique is developed to exploit the redundancy of the VQ output. Two applications of our joint source and channel coding technique are provided in Chapter 7: the design of the joint source and channel coding over a Rayleigh fading channel, and the transmission of a speech signal's line spectrum pair (LSP) over such channels. Finally, conclusions about this research are drawn and some suggestions for further work are presented in Chapter 8.

# Chapter 2

# A Performance Bound for Joint Source and Channel Coding

The purpose of this chapter is to present an overview of topics in information theory related to the subjects of data compression and transmission. The performance bounds for joint source and channel coding systems will be presented. The chapter starts with an overview of rate-distortion theory, which establishes a bound on the minimum rate required to obtain a particular fidelity, followed by a brief review of the channel capacity, which determines the maximum rate at which the information can be transmitted over a channel with an arbitrarily small probability of error. Based on the source rate distortion function and the channel capacity, the best performance achieved by the joint source and channel coding system can be bounded by setting the channel capacity equal to the rate given by the rate distortion theory for a particular fidelity. Such bounds will be derived in Section 2.3 and used as a reference for the performance of the proposed system.

Rate distortion theory just tells us the minimum rate required to encode the source satisfing the requirement on the distortion, but it does not show how to design a practical system to encode the source signal. We could use quantization theory to find the

best possible representation of a signal at a given data rate. The objective here is to find an optimal set of variables (generally called the reproduction alphabets or codewords) to represent the source signal and the corresponding regions associated with these codewords in the source signal space, according to a given distortion measure.

## 2.1 Rate Distortion Theory

This section presents some basic concepts on the source compression. For a complete discussion on the subject of the rate distortion theory, please refer to [53].

Let $U$ be a memoryless discrete source which generates symbols from a countable set $U \in [a_1, a_2, \cdots, a_K]$ with probabilities $P(a_1), P(a_2), \cdots, P(a_K)$. The entropy of this source is defined as

$$H(U) = \sum_{i=1}^{K} P(a_i) \log \frac{1}{P(a_i)} \tag{2.1}$$

It can be shown that if the code rate $R$ for the source is no less than the entropy $H(U)$, perfect reconstruction is possible. The objective of an efficient source coder is to operate at a rate as close as possible to the entropy of the source, within the complexity constraint.

For a continuous-amplitude source, the entropy of the source is defined as

$$H(u) = -\int p(u) \log p(u) du \tag{2.2}$$

where $p(u)$ is the probability density function of the source. Generally, it is impossible to reconstruct perfectly a continuous source with a finite rate code. In order to reproduce a source replica at the output, a certain amount of distortion must be accepted. Rate distortion theory gives the minimum rate $R$ for a given distortion $D$.

Rate distortion theory is based on minimizing the mutual information between the input and output for a given distortion requirement. The object to be optimized

is the conditional probability of the output for the given input variables. The mutual information measures the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one random variable due to the knowledge of the other. Mathematically, the mutual information $I(U,V)$ is the relative entropy between the joint distribution and the product distribution, i.e.,

$$I(U,V) = \sum_{U}\sum_{V} P(U,V)\log\frac{P(U,V)}{P(U)P(V)} \tag{2.3}$$

Similarly, the mutual information of a continuous source is defined as

$$I(u,v) = \int\int \log\frac{p(u,v)}{p(u)p(v)}dudv \tag{2.4}$$

We introduce a distortion function $d(u,v)$, which measures the cost of representing the symbol $u$ by its reproduction $v$. The squared error distortion is defined as $d(u,v) = (u-v)^2$ and is the most popular distortion measure due to its simplicity.

For an independent, identically distributed (i.i.d) source with a probability distribution $P(U)$ and a distortion measure $D$, a rate distortion function $R(D)$ is defined as

$$R(D) = \min_{P(V|U):\sum_{U,V} P(U)P(V|U)d(U,V)\leq D} I(U,V)$$

Note that the minimization considers all possible conditional distributions $P(V|U)$ for which the joint distribution $P(U,V) = P(U)P(V|U)$ satisfies the expected distortion constraint. From a practical viewpoint, the rate-distortion function for a continuous source is defined as the minimum bit rate required to code a signal at a given distortion.

The derivation of the rate distortion function is, unfortunately, an unsolved problem in many practical situations. Explicit expressions can be obtained for only a small number of source probability density functions and distortion measures. For

a continuous Gaussian source using the squared error distortion measure, the rate distortion function can be expressed explicitly [52] as:

$$R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & 0 \leq D \leq \sigma^2 \\ 0 & D > \sigma^2 \end{cases} ; \tag{2.5}$$

See Fig. 2.1.

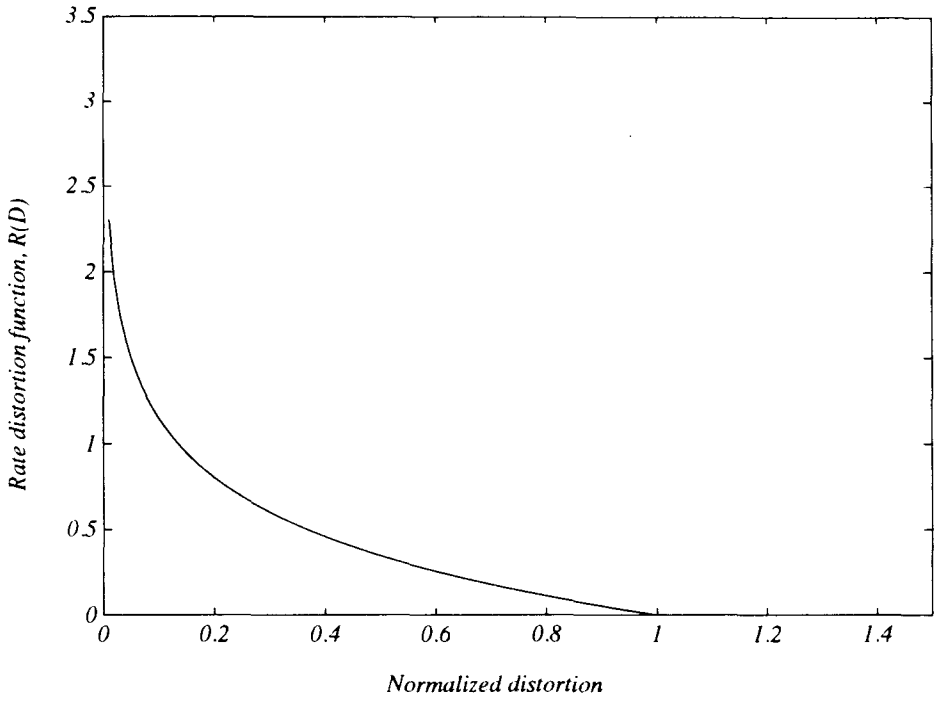


Figure 2.1: Rate distortion function for a Gaussian source

Since a mean squared-error larger than the signal variance can be avoided by simply decoding zero, the distortion is always less than the signal variance. We can re-write equation (2.5) to express the distortion in terms of the rate

$$D(R) = \sigma^2 2^{-2R} \tag{2.6}$$

Therefore, increasing the rate by one bit reduces the expected distortion by 6 dB.

Now we consider the rate distortion function for a first order Gauss-Markov source. This source model is used in the thesis for designing a soft decision vector quantizer.

A first order Gauss-Markov source is characterized by

$$u_n = \rho u_{n-1} + e_n \tag{2.7}$$

where $\rho$ is the normalized correlation coefficient, and $e_n$'s are independent, and identically distributed white Gaussian random variables. The autocorrelation matrix for a source vector $\mathbf{u} = [u_1, u_2, \cdots, u_L]$ formed from the Gauss-Markov source is given by

$$E[\mathbf{u}\mathbf{u}^T] = \begin{bmatrix} 1 & \rho & \cdots & \rho^{L-1} \\ \rho & 1 & \cdots & \rho^{L-2} \\ \rho^2 & \rho & \cdots & \rho^{L-3} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \rho^{L-1} & \rho^{L-2} & \cdots & 1 \end{bmatrix} \tag{2.8}$$

This matrix is symmetric and positive definite, therefore, there exits a set of $L$ orthogonal eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_L$ and a correspondent set of $L$ non-negative, eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_L$. A new source vector $\tilde{\mathbf{u}}$ can be formed by an orthogonal transform as follows:

$$\tilde{\mathbf{u}} = \mathbf{Q}^T\mathbf{u},$$

where $\mathbf{Q}$ is a matrix consisting of the $L$ orthogonal eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_L$. The covariance matrix of the transform vector $\tilde{\mathbf{u}}$ is $\mathbf{\Lambda} = E[\mathbf{Q}^T\mathbf{u}\mathbf{u}^T\mathbf{Q}] = diag\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$. The components of $\tilde{\mathbf{u}}$ are uncorrelated with variances equal to the eigenvalues. Since the average mutual information and average squared distortion are preserved in the orthogonal transform, the problem of encoding $\mathbf{u}$ with the squared distortion measure is equivalent to encoding $\tilde{\mathbf{u}}$ with the same distortion measure. Finally, the rate distortion function can be described parametrically in terms of $\theta$ by [53]

$$D_{opt}(\theta, L) = \frac{1}{L}\sum_{i=1}^{L} \min[\theta, \lambda_i] \tag{2.9}$$

and

$$R_{opt}(\theta, L) = \frac{1}{L}\sum_{i=1}^{L}\max[0, \frac{1}{2}\log_2\frac{\lambda_i}{\theta}], \text{ bits/per source sample} \qquad (2.10)$$

where the parameter $\theta$ is limited to $0 < \theta \leq \lambda_{\max}$. Note that the rate distortion bound can be approached arbitrarily closely by coding a long sequences of data samples, i.e., with vector quantization.

## 2.2 Channel Capacity

The channel capacity determines the maximum number of distinguishable signals that can be reliably transmitted over a channel. Mathematically, the channel capacity is defined as

$$C = \max_{P(U)} I(U, V) , \qquad (2.11)$$

where $I(U, V)$ denotes the mutual information between two random variables $U$ and $V$ and the maximum is taken over all possible input distributions $P(U)$. Shannon's channel coding theorem tells us that all rates below the channel capacity $C$ are obtainable, i.e., information can be transmitted reliably over a channel at all rates up to the channel capacity.

The channel capacity for a continuous channel is defined as

$$C = \max_{p(u)} I(u, v) \qquad (2.12)$$

where $p(u)$ is the source pdf. Generally, a continuous channel has some constraint conditions on the input probability distribution $p(u)$, the most common constraint is the average power constraint

$$\int_{-\infty}^{+\infty} u^2 p(u) du \leq P$$

The capacity of a continuous channel is the maximum value of the average mutual information over all probability distributions on the input subjected to the specific constraint.

For a Gaussian noise channel with an average power $P$ and noise variance $N_o$, the channel capacity is given by

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{N_o}\right) \tag{2.13}$$

Now let us consider the problem of communication over a band-limited Gaussian noise channel. Assuming that the bandwidth of the channel is $M/2$ and the spectral density function is $N_0/2$. According to the Nyquist sampling theorem, we can represent the channel by its sample values if the sampling rate is larger or equal to $M$. If the sampling rate is equal to $M$, we get $M$ samples per second, and these $M$ samples can be considered as an $M$ dimensional vector. The channel capacity of such an $M$ dimensional white Gaussian noise vector channel with a covariance matrix $\frac{N_0}{2}\mathbf{I}$ is given by [52]

$$C = \frac{M}{2}\log_2(1 + \frac{2P}{MN_0}) \tag{2.14}$$

If the source rate is $R$, the channel capacity must be larger than the source rate, i.e.,

$$\frac{M}{2}\log_2(1 + \frac{2P}{MN_0}) > R \tag{2.15}$$

For the source with a rate, $R$, the bit energy, $E_b$, is defined as

$$E_b = \frac{P}{R} \tag{2.16}$$

Submitting (2.16) into (2.15), and defining the spectral bit rate $r = 2R/M$ as the ratio of the source rate $R$ over the channel bandwidth $M/2$, we have

$$r < \log_2(1 + r\frac{E_b}{N_0}) \tag{2.17}$$

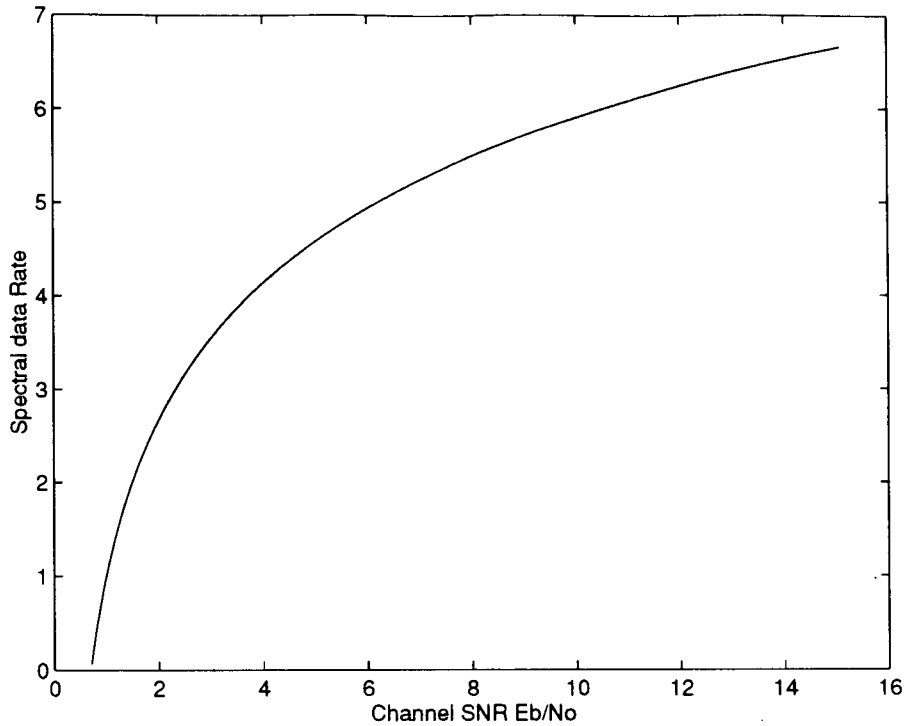


Figure 2.2: Capacity of a baseband Gauss Noisy Channel

Fig. 2.2 shows the boundary described by inequality (2.17). In principle, we can design a digital communication system for any point under the curve. The graph tells us that increasing the bit-rate per Hz increases the required energy per bit. This is the basis of the energy/bandwidth trade-off in digital communication theory, where increasing bandwidth at a fixed rate can reduce the power requirement.

## 2.3 Bounds for Joint Source and Channel Coding Systems

We have discussed both the rate distortion function for a first order Gauss-Markov source, which determines the minimum distortion for a given rate, and the channel capacity function, which determines the maximum rate that can be transmitted over

the channel. Therefore, the theoretically optimum performance for a first order Gauss-Markov process operating in a white Gaussian noisy channel can be determined by evaluating the rate-distortion function of the source at a rate equal to the channel capacity. Suppose that an $L$ dimensional signal vector is transmitted over an $M$ dimensional Gaussian noise channel. Let $e_{avg}$ denote the average channel energy per source sample, where $e_{avg} = P/L$. Using (2.14), the channel capacity per source sample is given by

$$C = \frac{M}{2L}\log_2(1 + \frac{2Le_{avg}}{MN_0}) \quad \text{bits per source sample} \tag{2.18}$$

Let the channel capacity given by (2.18) be equal to the rate-distortion function of (2.10). By solving the resulting equation for $e_{avg}$, the optimum performance is given parametrically in terms of $\theta$ by,

$$D_{opt}(\theta, L) = \frac{1}{L}\sum_{i=1}^{L}\min[\theta, \lambda_i] \tag{2.19}$$

and

$$e_{avg}(\theta, L, M) = \frac{N_0 M}{2L}((\prod_{i=1}^{L}\max[\lambda_i/\theta, 1])^{1/M} - 1) \tag{2.20}$$

The channel SNR per bit is given by

$$SNR_c = \frac{e_{avg}}{N_0/2}$$

Assuming the energy of the source signal is normalized, the synthesized source $SNR_s$ is given by

$$SNR_s = \frac{1}{D_{opt}}$$

Also let $B = M/L$ represent a bandwidth expansion factor, which is the ratio of the channel dimension to the source dimension. We then obtain the relationship between the channel SNR and the synthesized source SNR, in terms of parameter $\theta$, as

$$SNR_c = B\left(\left(\prod_{i=1}^{L}\max[\lambda_i/\theta, 1]\right)^{1/M} - 1\right) \tag{2.21}$$

$$SNR_s = \frac{1}{\frac{1}{L}\sum_{i=1}^{L}\min[\theta, \lambda_i]} \tag{2.22}$$

Now, let us look at how to determine a bound for a joint source and channel coding system based on (2.21) and (2.22). For a first order Gauss-Markov source, the covariance matrix is given by (2.8). We can compute the eigenvalues of the covariance matrix, also assuming that the eigenvalues are arranged in a descending order, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. First, if we let $\theta > \max\{\lambda_i, i = 1, 2, \cdots, L\}$, we get

$$SNR_s = \left(\frac{1}{L}\sum_{i=1}^{L}\lambda_i\right)^{-1} = 1$$

which is independent of the channel SNR; Second, if we let $\theta < \lambda_L$, we have

$$SNR_c = B\left(\left(\prod_{i=1}^{L}[\lambda_i/\theta]\right)^{1/M} - 1\right)$$

and

$$SNR_s = 1/\theta.$$

In the general case where $\lambda_m < \theta < \lambda_{m+1}$. We can express the channel SNR, $SNR_c$, and the synthesized source SNR, $SNR_s$, in terms of the parameter $\theta$ as

$$SNR_c = B\left(\left(\prod_{i=m+1}^{L}[\lambda_i/\theta]\right)^{1/M} - 1\right)$$

and

$$SNR_s = \frac{1}{\frac{1}{L}((L-m)\theta + \sum_{i=1}^{m}\lambda_i)}.$$

Now we consider the case of a continuous source. Let $\Phi(\omega)$ denote the power spectral density of a first order Gauss-Markov source, the bound for the joint source and channel coding system can be obtained by applying a theorem on the asymptotic

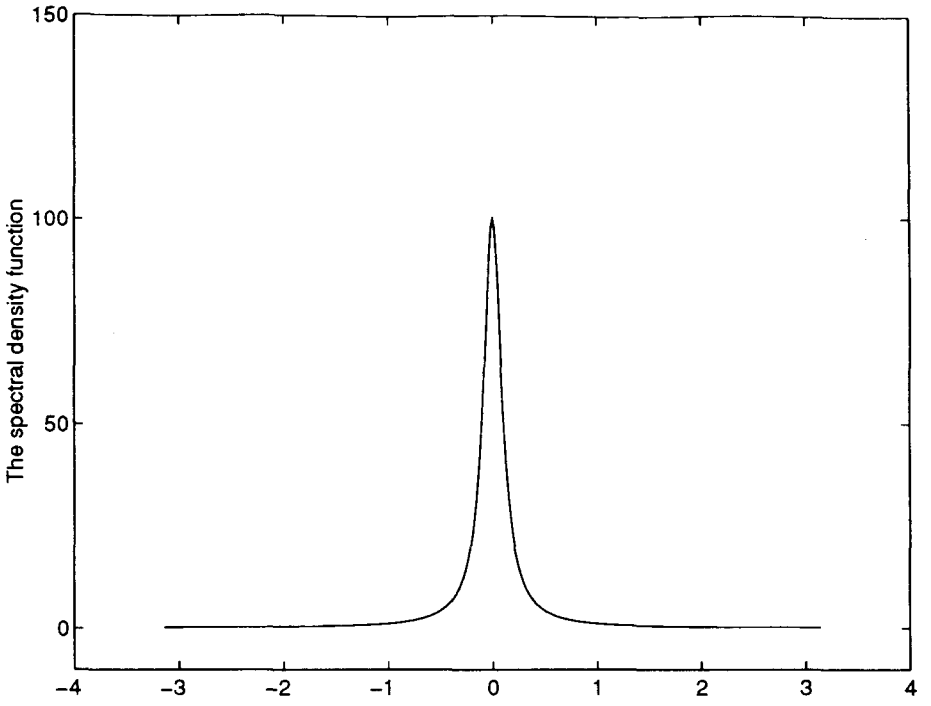


Figure 2.3: The spectral density function of a first order Gauss-Markov source with $\rho = 0.9$,

distribution of the eigenvalues of a Toeplitz form [54]. The bounds are given in terms of the spectral density function $\Phi(\omega)$ and $\theta$ as:

$$SNR_c(\theta, B) = B\left(2^{\frac{1}{2\pi B}\int_{-\pi}^{\pi} \max[0, \log_2 \frac{\Phi(\omega)}{\theta}]d\omega} - 1\right) \tag{2.23}$$

and

$$SNR_s = \frac{1}{\frac{1}{2\pi}\int_{-\pi}^{\pi} \min[\theta, \Phi(\omega)]d\omega} \tag{2.24}$$

For a first order Gauss-Markov source, the spectral density function is given by

$$\Phi(\omega) = \frac{1}{1 - 2\rho\cos\omega + \rho^2}$$

which is shown in Fig. 2.3.

When $\theta > (1-\rho)^{-2}$, we get

$$SNR_s = 1$$

which is independent of the channel SNR $SNR_c$; When $\theta < (1+\rho)^{-2}$, we have

$$SNR_c = B(\theta^{1/B} - 1)$$

and

$$SNR_s = 1/\theta.$$

In the general case where $(1+\rho)^{-2} < \theta < (1-\rho)^{-2}$, let $\omega_\theta$ be a particular value that makes $\Phi(\omega)$ equal to $\theta$. Then it can be shown that

$$\theta = \frac{1}{1 - 2\rho\cos\omega_\theta + \rho^2}$$

where $\omega_\theta$ can be expressed in terms of $\theta$ as:

$$\omega_\theta = \cos^{-1}\left(\frac{1+\rho^2 - 1/\theta}{2\rho}\right).$$

The corresponding synthesized source SNR, $SNR_s$, is given by

$$SNR_s = \frac{\pi}{\left(\theta\omega_\theta - \int_{\omega_\theta}^{\pi} \frac{1}{1-2\rho cos\omega+\rho^2} d\omega\right)} \tag{2.25}$$

and the channel SNR $SNR_c$ is given by

$$SNR_c(\theta, B) = B\left(2^{\frac{1}{\pi B}\int_0^{\omega_\theta} log_2 \frac{1}{\theta(1-2\rho cos\omega+\rho^2)} d\omega} - 1\right). \tag{2.26}$$

A numerical method can be used to calculate the integral of equation (2.25) and (2.26). Fig. 2.4 and Fig. 2.5 show the bounds for a source block length of $L$=2 and a block length of infinity at different bandwidth expansion factors. We can see that in the low channel SNR region, the difference in the synthesized signal SNR for different bandwidth expansion factors is quite small, however, in the high channel SNR region, the difference becomes significant. For example, in Fig. 2.4, at a channel SNR of 0 dB, the synthesized source SNR is 5.3 dB for $B = 0.5$ and 6.7 dB for $B = 1$, and the difference in synthesized source SNR is just 1.4 dB. But when the channel SNR is 15

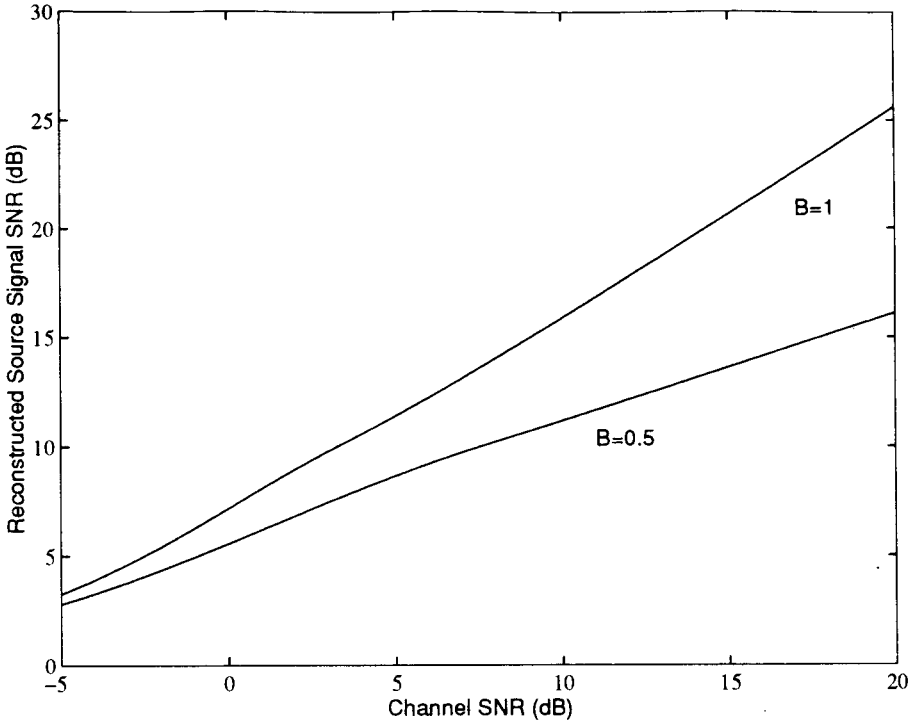


Figure 2.4: The bound for the joint source and channel coding system. Block length $L$ is equal to 2. The source is a first order Gauss-Markov source with $\rho = 0.9$, the channel is a Gaussian noisy channel with the bandwidth expansion factor $B$ equal to 0.5 and 1 respectively.

dB, the synthesized source SNR is 14 dB for $B = 0.5$ and 21 dB for $B = 1$ and the difference is 7 dB. The improvement in the synthesized source SNR is obtained at the cost of the channel bandwidth efficiency.

We can observe a similar phenomenon in Fig. 2.4. Comparing Fig. 2.3 with Fig. 2.4, we can see that with the increase of the code length, the synthesized source SNR can be improved. For L=2, B=0.5 and the channel SNR equal to 15 dB, the synthesized SNR is 14 dB; but for the same channel SNR and bandwidth factor, if the block length $L$ tends to infinity, the synthesized SNR is 17 dB. By increasing the block length, we can get a 3 dB improvement in the synthesized source SNR. These figures suggest that by coding a block of signal samples, i.e., vector quantization,

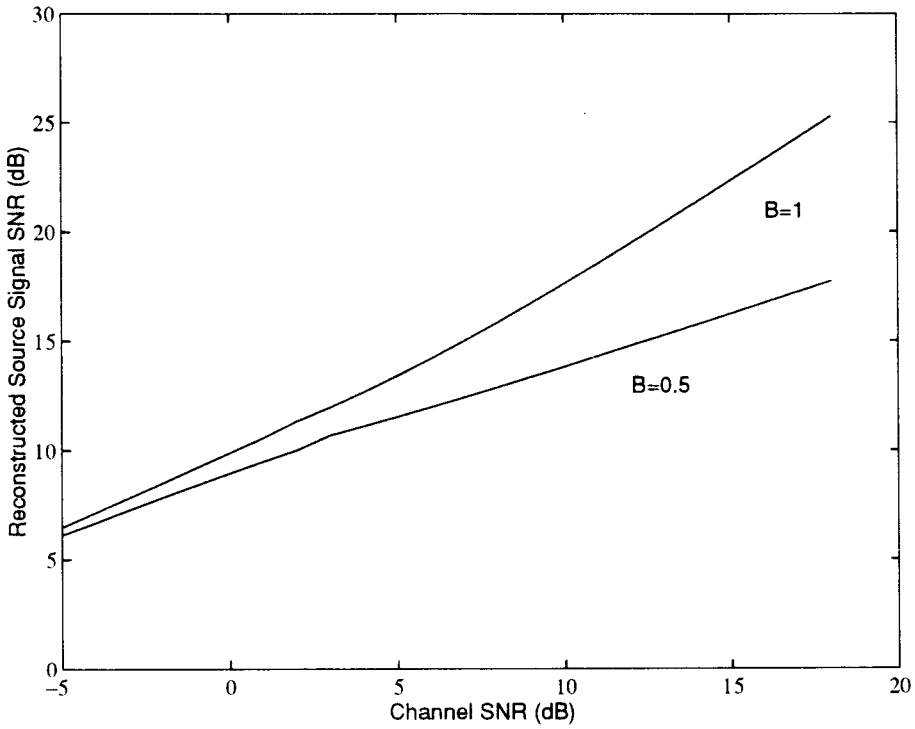


Figure 2.5: The bound for the joint source and channel coding system. Block length $L$ is infinite. The source is a first order Gauss-Markov source with $\rho = 0.9$, the channel is a Gaussian noisy channel with the bandwidth expansion factor $B$ equal to 0.5 and 1 respectively.

we can obtain a significant performance improvement. We will discuss the vector quantization in detail in the next section.

The bounds derived in this section will be used in the following chapter as a reference for the proposed system performance.

## 2.4 Vector Quantization

Rate Distortion theory does not tell us how to design a practical system to encode the signal. In order to approach the rate-distortion function for the given source, we consider the joint quantization of a block of signal samples in this section. This type of quantization is called *vector quantization* [56, 57].

A fundamental result of rate-distortion theory is that better performance can be achieved by quantizing vectors instead of scalars, even if the source is memoryless. A vector quantizer(VQ) will work even better if the signal samples are statistically dependent.

A vector quantizer can be viewed as a mapping of the input signal vector into a discrete number of output vectors in the $k$-dimensional Euclidean space $\Re^k$ in a way that optimizes a given fidelity criterion, such as the mean squared error. The input signal is a $k$-dimensional vector $\mathbf{x}$, and the output space is defined to be $N$ distinct points in $\Re^k$. Denote the mapping by $\mathrm{Q}(\mathbf{x})$. The mean-squared error (MSE) distortion is defined as

$$e = E\{\|\mathbf{x} - Q(\mathbf{x})\|^2\} \tag{2.27}$$

A $k$-dimensional, $N$-level VQ is defined by the partition set $\mathbf{\Omega} = \{\Omega_1, \Omega_2, \cdots, \Omega_N\}$ and the codebook $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N\}$, where $\Omega_i$ is a partition region and $\mathbf{c}_i$ is the corresponding centroid or codevector. The operation of the quantizer is to map every point in $\Omega_i$ into $\mathbf{c}_i$. Clearly, all of the $\Omega_i$ should be disjoint regions which cover the

total input space

$$\Omega_i \bigcap \Omega_j = \mathbf{\Phi} \quad \text{for } i \neq j;$$

and

$$\bigcup_{i=1}^{N} \Omega_i = \Re^{\mathbf{k}}$$

where $\mathbf{\Phi}$ denotes an empty set. Using these facts, the mean-squared error associated with the vector quantizer can be written as

$$e = \sum_{i=1}^{N} \int_{\Omega_i} \|\mathbf{x} - \mathbf{c_i}\|^2 p(\mathbf{x}) d\mathbf{x} \tag{2.28}$$

In order to find out the $\Omega_i's$ and $\mathbf{c}_i$'s, we first assume that all the partition regions $\Omega_i$ are fixed in the source space. The corresponding codevector $\mathbf{c}_i$ must then satisfy

$$\mathbf{c}_i = \frac{\int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_i} p(\mathbf{x}) d\mathbf{x}} \tag{2.29}$$

This expression states that each codevector $\mathbf{c}_i$ should be the centroid of its associated region. Equation (2.29) is called *the centroid condition.*

Next, we assume that the codevectors $\mathbf{c}_i, i = 1, 2, \cdots, N$, are fixed. To minimize the total MSE, an input signal vector should be assigned to the $i$-th partition $\Omega_i$ according to *the nearest neighbor* criterion:

$$Q(\mathbf{x}) = \mathbf{c_i} \quad \text{if} \quad |\mathbf{x} - \mathbf{c}_i| < |\mathbf{x} - \mathbf{c}_j| \quad \text{for any } i \neq j \tag{2.30}$$

Since the pdf of the signal is unknown in many practical sources, direct analytical calculation of the codevectors is usually impossible. However, if a set of input vectors $\{\mathbf{x}_j, j = 1, 2, \cdots, \}$, called the training set, is given, an iterative training procedure known as the generalized Lloyd algorithm (GLA) [43] can be used to design a VQ codebook. Starting from an initial set of codevectors, this algorithm iteratively uses the centroid condition (2.29) and the nearest neighbor condition (2.30) to generate a local VQ codebook. The basic design steps are as follows:

1. Set iteration counter $m = 0$ and choose an initial VQ codebook $\mathcal{C}^0$.

2. m=m+1. Cluster the set of training vectors into partition $\mathbf{\Omega}^m$ for the given VQ codevectors by applying the nearest neighbor condition.

3. Compute the centroids or codevectors $\mathcal{C}^m$ according to the new partitions. For the mean squared error distortion measure, the codevector $c_i$ is given by

$$\mathbf{c}_i = \frac{1}{M_i} \sum_{\mathbf{x}_j \in \Omega_i} \mathbf{x}_j$$

where $M_i$ denotes the number of training vectors in $\Omega_i$.

4. Compute the average distortion. Check whether the convergence condition is satisfied; If yes then stop; Otherwise go to step 2.

The algorithm will converge because the distortion will not increase in successive iteration.

There are several ways to choose the initial codevectors. A well-known method is to generate the codevectors with a VQ codebook size of $K$ by perturbating the codevectors of a codebook size of $K/2$, which was proposed by Linde, Buzo and Gray [43], and generally is called the LBG algorithm. In this way, a VQ codebook size of $K$ can be designed starting from a one-vector codebook containing only the centroid of the training set [43].

The complexity of the VQ design is much greater than that of scalar quantization. A considerable amount of memory is also required for storing the VQ codebook. Several methods can be used to reduce the complexity, such as multi-stage VQ, gain-shape VQ, and tree-searched VQ [57].

# Chapter 3

# Communication System Modeling

In this chapter we will consider the vector space description of communication systems. Although all communication channels are real waveform channels, it turns out that the most efficient way to describe and interpret such channels is through vector representation. Two types of the channel model used in this thesis are also introduced, based on the vector representation.

## 3.1 Vector Representation of Signal and Noise

We first discuss the vector representation of the signal and noise [65]-[66]. For any set of signals $s_i(t)$, $i = 1,\ 2,\ \cdots,\ L$, in the interval $0 < t < T$, with finite energy

$$E_i = \int_o^T |s_i(t)|^2 dt \quad i = 1, 2, \cdots, L, \tag{3.1}$$

a set of $M \leq L$ orthonormal functions $\phi_1(t),\ \phi_2(t),\ \cdots,\ \phi_M(t)$ can always be found such that

$$s_i(t) = \sum_{m=1}^{M} s_{im}\phi_m(t), \quad i = 1, 2, \cdots, L \tag{3.2}$$

where

$$s_{im} = \int_0^T s_i(t)\phi_m(t)dt \tag{3.3}$$

The functions $\phi_m(t)$, $m = 1, 2, \cdots, M$, are orthonormal and can be constructed from the $s_i(t)$'s through the Gram-Schmidt procedure [66].

It is observed from equation (3.2) that once the functions $\phi_m(t)$ have been determined, each waveform $s_i(t)$ is completely defined by its coefficients $s_{im}$. We could visualize each set of coefficients as an $M$-dimension vector and the collection of all such vectors forms the signal constellation, defined in an $M$-dimensional signal space. Each axis or coordinate in the signal space is associated with one of the orthogonal functions. Therefore, the $i$-th signal $s_i(t)$ is defined by the vector $\mathbf{s}_i=[s_{i1}, s_{i2}, \cdots, s_{iM}]^T$ in the $M$-dimension signal space. Note that the energy in $s_i(t)$ is related to its vector coordinates as follows

$$E_i = \sum_{i=1}^{M} s_{in}^2 = \|\mathbf{s}_i\|^2. \tag{3.4}$$

A white Gaussian noise random process can also be represented by the orthonormal representation as [66]

$$n_w(t) = \sum_{j=1}^{\infty} n_j \phi_j(t) \tag{3.5}$$

where $n_j$ is a zero-mean stationary independent Gaussian variable with properties

$$n_j = \int_0^T n(t)\phi_j(t)dt \tag{3.6}$$

and

$$E(n_j n_m) = \begin{cases} \frac{N_0}{2} & \text{j=m} \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

where $N_0/2$ is the power spectrum density function of the noise.

Noise terms in (3.5) for $j > M$ exist in coordinates orthogonal to all coordinates for $j \leq M$. These terms therefore have no affect on the coordinates in which the signals are defined, and it is known that the terms are irrelevant to the optimum receiver [66]. On dropping the irrelevant noise, we denote the relavent noise by

$$n(t) = \sum_{j=1}^{M} n_j \phi_j(t) \tag{3.8}$$

## 3.2 Two Channel Models for Digital Communications

Traditionally, source coding is optimized for a given source signal, it is implicitly assumed that this coder is transimitted over the noiseless channel. However, this is not the case in physical channels. Typical communication channels such as telephone lines, mobile radio links, microwave links, etc introduce various types of the noise and interference. We focus in this study on the additive white Gaussian Noise (AWGN) channel and the Rayleigh flat fading channel.

### 3.2.1 Additive White Gaussian Noise Channel

An additive white Gaussian noisy (AWGN) channel is the most commonly used channel model in the analysis of communication systems. Such a channel model is appropriate when thermal noise is a dominant noise source in the system. As its name suggested, the power spectral density function of the white noise is a constant, independent of the frequency, and equal to $N_0/2$. The noise amplitude has a Gaussian probability density function (pdf):

$$p(\mathbf{n}) = \frac{1}{(\pi N_0)^{\frac{M}{2}}} \exp(-\frac{\mathbf{n}^T \mathbf{n}}{N_0}) \tag{3.9}$$

where $\mathbf{n}$ is the vector representation of the channel noise, and $\mathbf{n}^T$ is a transpose vector of $\mathbf{n}$. In such a channel, the noise component adds linearly along with the desired signal. The transmitted and the received signals can be described in a vector format

$$\mathbf{r} = \mathbf{s} + \mathbf{n}$$

where **s** denotes the transmitted signal, **n** denotes the channel noise contribution and **r** denotes the received signal. The received signal **r** also has a Gaussian distribution.

### 3.2.2 Rayleigh Fading Channel

In the mobile channel, the signal can enter the receiver via more than one path. The effect of the multipath is caused by the reflection and scattering from buildings, trees and other obstacles. The received signal is the sum of the signals with different amplitudes and phases. It is plausible to assume that the phases of the scattered signals are uniformly distributed between 0 and $2\pi$ and the amplitudes and phases are statistically independent of each other. Since the number of the scattered signals is quite large, the quadrature components have Gaussian distributions according to the central limit theorem. As a result, the envelope, $a$, defined as the square root of the norm of the quadrature components, has a Rayleigh probability density function [68], [69]

$$p(a) = \frac{a}{\sigma^2} \exp(-\frac{a^2}{2\sigma^2}) \tag{3.10}$$

where $\sigma^2$ is the noise power of the quadrature components.

If the receiver or transmitter is in relative motion, the received signal also experiences a Doppler shift. The Doppler shift $f_D$ is equal to [69]

$$f_D = \frac{v}{\lambda} \cos \alpha$$

where $v$ is the velocity at which a vehicle is moving, $\lambda$ is the wavelength of the electromagnetic wave, and $\alpha$ is the spatial angle between the direction of arrival of the wave and the direction of the vehicle motion.

Assume that the power of the received signals is uniformly distributed in the range $-\pi \leq \alpha \leq \pi$, i.e., the pdf of $\alpha$ is given by

$$p(\alpha) = \frac{1}{2\pi}$$

for a mobile station with an omnidirectional mobile antenna, the power spectrum of the multipath fading channel is given by [67]

$$P(f) = \frac{3W_0}{2\pi f_m}\left[1 - \frac{(f - f_c)^2}{f_m^2}\right]^{1/2} \tag{3.11}$$

where $W_0$ is a constant, $f_m = v/\lambda$ is the maximum Doppler frequency shift. By taking the inverse Fourier transform of the spectrum function (3.11), we obtain the autocorrelation function of the fading channel as

$$\rho(\tau) = BJ_0(2\pi f_D \tau) \tag{3.12}$$

where $B$ is a constant, $J_0(.)$ is the Bessel function of the first kind and the zero order. The autocorrelation is shown in Fig. 3.1. There is a rapid decorrelation, showing that space diversity can be implemented at the mobile end to reduce the fading effect.

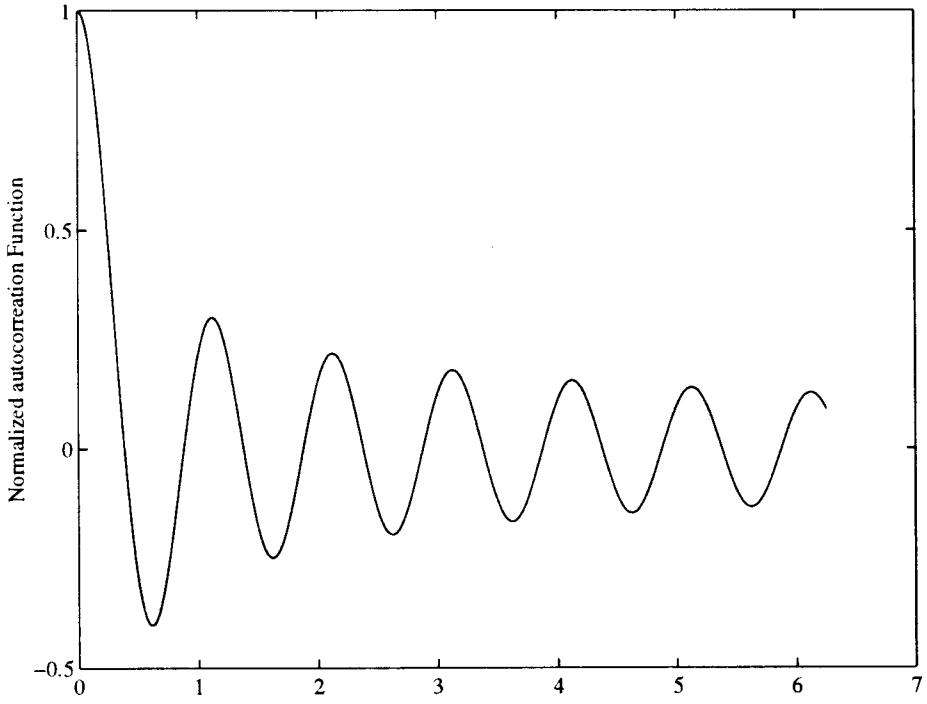


Figure 3.1: The normalized autocorrelation function of the signal for Rayleigh fading channel

Since every two dimensions of an $M$ dimensional vector can be viewed as a quadrature amplitude modulation (QAM) symbol, it is convenient to treat the $M$ dimensional signal as a complex vector with $M/2$ components. In this way, the received signal over the Rayleigh fading channel can be expressed as

$$\mathbf{r}^* = \mathbf{G}^*\mathbf{s}^* + \mathbf{n}^* \tag{3.13}$$

where $\mathbf{G}^* = dial[g_1^*, g_2^*, ..g_{M/2}^*]$ is a diagonal matrix, its diagonal element $g_i^*$ is a zero mean, complex, Gaussian fading process, $\mathbf{n}^*$ is a zero-mean, $M/2$ dimensional complex Gaussian noise vector with a covariance matrix of $\frac{N_0}{2}\mathbf{I}$, and $\mathbf{s}^*$ is the transmitted complex signal.

## 3.3 Formulation of the Joint Source and Channel Coding Design Problem

Applying the concept of finite dimensional representation of signal and noise, the general communication models we want to explore are shown in Fig. 3.2 and Fig.3.3 respectively. The source is a zero-mean stationary random process. Let $\mathbf{x}$ be a $k$-dimensional vector derived from the random source with a pdf $p(\mathbf{x})$. The vector $\mathbf{x}$ is represented by a VQ codeword chosen from the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$, where $N$ is the size of the codebook. Assuming $\mathbf{c}_i$ was chosen, then the index $i$ is mapped directly into the modulation signal points $\mathbf{s}_i$. The set of $\mathbf{s}_i$, $i = 1, \ldots, N$ forms the signal constellation. The dimension of this constellation, $M$, is an important design parameter as it is proportional to the bandwidth of the communication channel [15]. The modulated signal $\mathbf{s}_i$ is transmitted over an additive white Gaussian noise channel (AWGN), as shown in Fig. 3.2, or a Rayleigh fading channel, as shown in Fig. 3.3. In a vector notation, the noise waveform is represented by an $M$-dimensional Gaussian random vector $\mathbf{n}$. The different components of $\mathbf{n}$ are independent and identically distributed

(i.i.d), each having a zero mean and a variance of $N_0/2$. The received signal $\mathbf{y}$ is

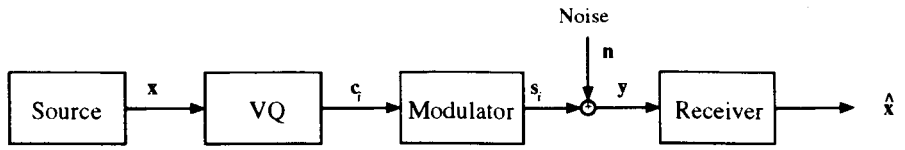


Figure 3.2: General communication system model in the AWGN channel.

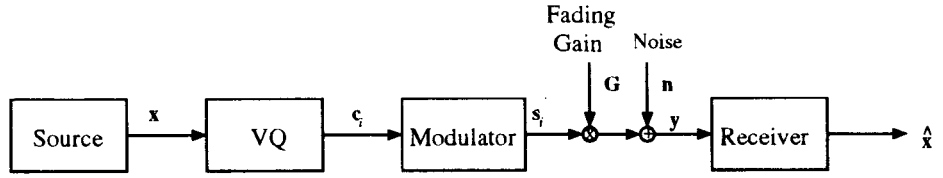


Figure 3.3: General communication system model in the Rayleigh fading channel.

simply the sum of the transmitted signal $\mathbf{s}_i$ and the noise vector $\mathbf{n}$ in the AWGN channel or the sum of the faded signal by the fading channel $\mathbf{G}\mathbf{s}_i$ and the noise vector $\mathbf{n}$. Given the signal $\mathbf{y}$, the decoder will provide the "best" estimate of the original signal $\mathbf{x}$. This estimate, denoted by $\hat{\mathbf{x}}$, is then delivered to the final destination.

The problem we want to solve is to minimize the per sample MSE, $\frac{1}{k}E(\|\mathbf{x}-\hat{\mathbf{x}}\|^2)$, by appropriately selecting the source VQ encoder, the modulation signal constellation and demodulator, subject to constraints in the average transmitting signal energy and the channel bandwidth. It should be pointed out that the choice of the receiver structure is included in the optimization process, which gives us a maximum degree of freedom to design an optimal communication system.

# Chapter 4

# Joint Optimization Based on the Signal Detection Receiver

Traditionally, a source coder such as vector quantizer (VQ) is designed for a noiseless channel. The resulting approach is called a *source optimized vector quantization* (SOVQ). The performance of SOVQ degrades significantly in the presence of transmission errors. In noisy channels, the received index may differ from the transmitted one and, as a result, the reproduction vector may be a poor representation of the original source vector. In this chapter, we consider the problem of joint optimization of source and channel coding with a conventional maximum likelihood (ML) receiver. The system is composed of a VQ, a modulator, and an ML receiver. The objective is to design an optimal VQ, modulation constellation, and receiver which will minimize the average distortion between the source vector and its replica at the receiver. An iterative optimization strategy is used to design the VQ, the modulator, and the receiver, while maintaining constraints on the signal energy and transmission bandwidth.

One method toward the robust vector quantization in the noise channel is to optimize the VQ codebook for the given channel transition probabilities. This approach leads to a *channel optimized vector quantizer* (COVQ) [22], [23]. In a traditional

communication system, an ML receiver is used to obtain an index from the demodulated data. However, the ML receiver is not optimal under the criterion of the mean squared-error (MSE). We use a generalized Bayes receiver as the demodulator that minimizes the overall MSE.

This chapter is organized into eight sections as follows. In Section 4.1, a general communication system model is described. The design steps for the channel optimized VQ are reviewed in Section 4.2, followed by a discussion on the VQ index assignment using simulated annealing. Optimizationof the modulation constellation is discussed in Section 4.4. The optimal Bayes receiver in terms of the MSE criterion is developed in Section 4.5. An iterative optimization algorithm for the joint optimization of the source and channel coding is then introduced based on a straightforward extension of the Lloyd algorithm. In Section 4.7 numerical results are presented for the performance of the jointly optimized system for a first order Gauss-Markov source. These results are compared with the performance of standard communication systems with a SOVQ and using different modulation schemes such as quadrature phase shift keying (QPSK) and the trellis-coded modulation (TCM) for 8-phase shift keying (PSK). Finally a summary is provided, and several possible improvements are discussed in Section 4.8.

## 4.1 System Structure

The general diagram of the communication system is shown in Figure 4.1. Let $\mathbf{x}$ be a

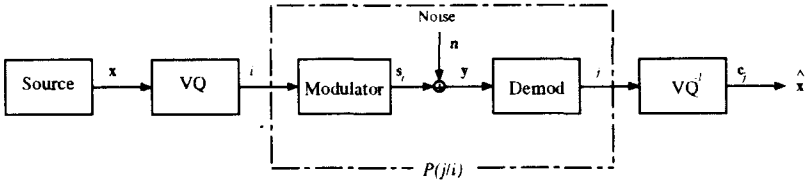


Figure 4.1: Communication system model with a transmission probability $P(j/i)$

$k$-dimensional vector derived from a random source with a probability density function (pdf) $p(\mathbf{x})$. A $k$-dimensional $N$-level vector quantizer (VQ) is defined by a codebook $\mathcal{C} = \{\mathbf{c}_i, i = 1, 2, \cdots, N\}$ consisting of $N$ $k$-dimensional reconstruction vectors $\mathbf{c}_i$ and the partition set $\mathbf{\Omega} = \{\Omega_i, i = 1, 2, \cdots, N\}$ of the $k$-dimensional Euclidean space $\Re^{\mathbf{k}}$. The vector quantizer is actually a mapping, $Q(.)$, from the k-dimensional Euclidean space $\Re^{\mathbf{k}}$ into a finite reconstruction vector space $\{\mathcal{C}\}$. A VQ is said to be optimal with respect to a distortion measure $d$ if the average distortion

$$D = \frac{1}{k} E[d(\mathbf{x}, Q(\mathbf{x}))] \tag{4.1}$$

is minimized over all possible partitions $\mathbf{\Omega}$ of $\Re^{\mathbf{k}}$ and reproduction codebooks $\mathcal{C}$. The necessary conditions for optimality for a noise-free channel are

$$\Omega_i = \{\mathbf{x} : d(\mathbf{x}, \mathbf{c}_i) \leq d(\mathbf{x}, \mathbf{c}_j)\} \quad \text{for any } j \tag{4.2}$$

$$\mathbf{c}_i = \arg\min_{\mathbf{z}} \{E[d(\mathbf{x}, \mathbf{z}) \mid \mathbf{x} \in \Omega_i]\} \qquad \mathbf{z} \in \Re^{\mathbf{k}} \tag{4.3}$$

Equation (4.2) gives the so-called nearest-neighbor condition, while (4.3) is the centroid condition. The necessary conditions specified by Equations (4.2) and (4.3) do not lead to an analytical solution for the optimum codebook and partition. However, an iterative algorithm can be used to design a locally optimal vector quantizer as in the work of Linde, Buzo, and Gray (LBG) [43].

In many applications, a source vector is quantized and then transmitted over a noisy communication channel. The source compression system uses a vector quantizer (VQ) to map an input signal vector $\mathbf{x}$ into its corresponding codevector $\mathbf{c}_i$. The index of the corresponding codevector, $i$ is directly mapped into a modulation signal vector $\mathbf{s}_i$. Then $\mathbf{s}_i$ is transmitted over a noisy channel. The received signal is the sum of the

modulation signal $\mathbf{s}_i$ and the channel noise $\mathbf{n}$. In a conventional system, as shown by Fig. 4.1, the received channel symbol $\mathbf{y}$ is quantized into a received index, $j$, according to a decision rule. The information transmitted to the user is then $\hat{\mathbf{x}} = \mathbf{c}_j$. Due to the channel noise, the received index $j$ may differ from the transmitted index $i$, creating a relatively large distortion due to a poor representation of the original signal vector by $\mathbf{c}_j$. The channel state transition probabilities $P(j/i)$ can be used to represent the effect of the channel noise.

## 4.2 Channel Optimized VQ

As shown in [23] and [70], channel noise may significantly degrade the performance of a source optimized VQ. Assuming some knowledge of the noise statistics, a better alternative is to optimize the VQ for a given noisy channel. This approach leads to the concept of a *channel optimized vector quantizer* (COVQ) introduced by Kumazawa et al and Farvardin et al [22], [23]. Farvardin and Vaishampayan researched the structure of the COVQ in detail in [70]. Let $P(j|i)$ be the channel transition probability from index $i$ to index $j$, which depends on the modulation signal set. The average distortion is then given by

$$D = \frac{1}{k} \sum_{i=1}^{N} \sum_{j=1}^{N} P(j|i) \int_{\Omega_i} \|\mathbf{x} - \mathbf{c}_j\|^2 p(\mathbf{x}) d\mathbf{x} \tag{4.4}$$

The optimum VQ should minimize the average distortion for a given VQ codebook size $N$ and channel transition probabilities $P(j|i)$.

First, assuming that the VQ partition, $\mathbf{\Omega}$, is given, the necessary condition for the optimal codebook $\mathcal{C}$ can be found by setting the derivative of the average distortion (4.4) with respect to $\mathbf{c}_j$ to zero, i.e.,

$$\frac{\partial D}{\partial \mathbf{c}_j} = 0 \tag{4.5}$$

Correspondingly, the optimal codevector $\mathbf{c}_j$ is

$$\mathbf{c}_j = \frac{\sum_{i=1}^{N} P(j|i) \int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\sum_{i=1}^{N} P(j|i) \int_{\Omega_i} p(\mathbf{x}) d\mathbf{x}} \tag{4.6}$$

When a training sequence $\mathbf{x}_m, (m = 1, 2, \cdots, )$ is given, a discrete version of equation (4.6) is given by

$$\mathbf{c}_j = \frac{\sum_{i=1}^{N} P(j|i) \sum_{m:\mathbf{x}_m \in \Omega_i} \mathbf{x}_m}{\sum_{i=1}^{N} P(j|i) m_i} \tag{4.7}$$

where $m_i$ denotes the number of training sequences belonging to the $i$-th partition $\Omega_i$

Next, we must determine the optimum partition $\Omega_i$ for the given VQ codebook $\mathcal{C}$. The average distortion (4.4) can be re-written as

$$D = \frac{1}{k} \sum_{i=1}^{N} \int_{\Omega_i} \left( \sum_{j=1}^{N} P(j|i) \|\mathbf{x} - \mathbf{c}_j\|^2 \right) p(\mathbf{x}) d\mathbf{x} \tag{4.8}$$

Note that all the terms inside the bracket of equation (4.8) are positive and so is the probability density function $p(\mathbf{x})$. The only variable quantity is the integration region $\Omega_i$. Consequently, the average distortion is minimized by selecting the decision region $\Omega_i$ to include only those points of $\mathbf{x}$ for which the term inside the bracket is minimum. This process leads to the following optimum partition:

$$\Omega_i = \left\{ \mathbf{x} : \sum_{j=1}^{N} P(j|i) \|\mathbf{x} - \mathbf{c}_j\|^2 < \sum_{j=1}^{N} P(j/m) \|\mathbf{x} - \mathbf{c}_j\|^2 \quad \text{for } \forall m \neq i \right\} \tag{4.9}$$

A modified GLA procedure can be used to design a COVQ based on (4.7) and (4.9). The design algorithm for the COVQ is as follows:

1. Set iteration counter $m = 0$. For the given channel transition probabilities $P(j|i)$, VQ codebook size $N$, and training sequences, choose an initial VQ codebook $\mathcal{C}^0$ and a threshold $\epsilon$.

2. $m = m + 1$. Cluster the set of training vectors into a partition $\mathbf{\Omega}^m$ for the given VQ codevectors by applying the modified nearest neighbor condition (4.9).

3. Compute the VQ codebook of the $m$-th iteration, $\mathcal{C}^m$, according to the new partition $\mathbf{\Omega}^m$. For the mean squared error distortion measure, the VQ codevector $c_i$ is given by equation (4.7).

4. Compute the average distortion $D^m$ resulting from the $m$-th iteration. Check whether the convergence condition is satisfied, i.e. if $\frac{|D^m - D^{m-1}|}{D^{m-1}} < \epsilon$. If yes, then stop; otherwise go back to step 2. Since the average distortions in successive iterations are non-increasing, the training process is guaranteed to converge.

The design result of the source optmized VQ could be used as an initial codition for design of the channel optmized VQ. When used in a noisy channel, a channel optimized VQ usually achieves better performance than a source optimized VQ. It could provide up to 1.5 dB improvement in the synthesized source SNR [23].

## 4.3 Optimal Index Assignment

In this section we consider the problem of how to assign the VQ indices to the modulation signals for a given source coder VQ [23]. The objective is to minimize the average distortion $D$ of (4.4) over the set of all possible index assignments, i.e., optimizing the VQ-modulation mapping. Rydbeck and Sundberg [24] presented one of the first papers to demonstrate the importance of index assignment for scalar quantization. The problem of optimal index assignment for scalar quantization was also considered by Farvardin and Vaishampayan [19]. An algorithm for improving the index assignment for a vector quantizer with a fixed codebook was introduced by De Marca et al [25], [26]. A locally optimal solution based on a binary switching algorithm was introduced by Zeger and Gersho [27], [28] and called Pseudo-Gray coding to acknowledge its similarity to the well known Gray code. For small values of $N$, optimization can be accomplished simply by trying all the possible permutations of the $N$ indices.

The number of computations increases very quickly with $N$, making an exhaustive search impractical even for moderate $N$ values. An alternative approach for solving the combinatorial problem generated by the bit assignment optimization is based on the simulated annealing algorithm [30]. A detailed description of this algorithm can be found in [73]. In this section, the minimization technique is based on simulated annealing [74]. The basic idea of the simulated annealing algorithm is to randomly perturb the state of the system at each iteration of the algorithm. The algorithm unconditionally accepts perturbations that reduce the distortion and probabilistically accepts perturbations that increase the distortion.

In the context of our current study, the algorithm works as follows. Assume that we begin by arbitrarily labeling the modulation points and storing them in an array $\mathbf{E}$ of size $M \times N$; $\mathbf{E} = (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N)$. Also, let $\mathbf{F}$ be a permutation matrix of size $M \times M$. Then any mapping of VQ codevectors to the modulation points can be described in terms of the $\mathbf{F}$ and $\mathbf{E}$ matrices as follows:

$$\mathbf{S}_{mod} = \mathbf{FE}$$

The simulated annealing algorithm will attempt to find the permutation matrix $\mathbf{F}$ that minimizes the distortion $D$ of (4.4). Here, the average distortion is considered as a function of the permutation matrix $D(\mathbf{F})$. In the simulated annealing language, $\mathbf{F}$ represents the current state of the system. The evolution from the current state $\mathbf{F}$ to a new state $\mathbf{F}'$ is obtained by using a "perturbation" defined by

$$\mathbf{F}' = \mathbf{BF}$$

where $\mathbf{B}$ is a randomly generated permutation matrix that contains only two non-zero elements. Multiplying $\mathbf{F}$ by $\mathbf{B}$ is thus equivalent to swapping the mapping for two indices.

The steps in the simulated annealing algorithm are :

1) For the $m$-th iteration, define an abstract system "temperature" $T_m$. The initial temperature $T_o$ is usually relatively high. Also randomly arrange an initial state, $\mathbf{F}$, from the given modulation constellation.

2) Randomly choose a perturbation matrix $\mathbf{B}$ of the state $\mathbf{F}$ and compute the energy variation $\delta D = D(\mathbf{F}) - D(\mathbf{F}')$.

a) if $\delta D \leq 0$, replace $\mathbf{F}$ by $\mathbf{F}'$, go to step 3.

b)if $\delta D > 0$, replace $\mathbf{F}$ by $\mathbf{F}'$ with probability $\exp(-\delta D/T_m)$ and go to step 3.

3) If the number of energy drops exceeds a prescribed number or if the number of unsuccessful perturbations (perturbations not resulting in drops of energy) exceed a prescribed threshold, go to step 4. Otherwise, go back to step 2.

4) Lower temperature $T_m$. If a stable state has been reached, stop and $\mathbf{F}$ is the optimal assignment we need; otherwise go back to step 2.

A particular perturbation algorithm is used to change the state of the system in order to reduce the number of computations required to evaluate the average distortion [19]. In step 2, the perturbation of the state is implemented by selecting two components of an index vector $\mathbf{b}$ randomly and by permuting them. For instance $\mathbf{b}' = (9,4,6,8,2)$ is a perturbation of $\mathbf{b} = (9,8,6,4,2)$ in which the second and the fourth elements in the state vector are altered. This choice of perturbation allows us to move from any state to any other state in a finite number of perturbations. In step 2 (a), $\delta D \leq 0$ corresponds to a decrease of the distortion and therefore it is accepted because the goal is to minimize the distortion. On the other hand $\delta D > 0$ corresponds to an increase of the distortion, a probabilistic decision is made whether or not to accept the trial perturbation, and the probability of acceptance decreases

exponentially with the temperature, i.e.,

$$\text{Prob(accept)} = \begin{cases} \exp(-\delta D/T_m) & \text{if } \delta D > 0 \\ 1 & \delta D < 0 \end{cases}$$

as suggested by Farvardin [23]. In the beginning when the system is "hot", almost all perturbations (whether they reduce or increase the distortion) are accepted. However, as the temperature is reduced, those perturbations that cause an increase of the distortion will be accepted with diminishingly small probabilities. This process allows the algorithm to climb out of local minima when the temperature is high in the hope that as the system is cooled the state falls to the global minimum. In the limited case , $\lim_{m\to\infty} T_m = 0$, the system is no longer able to escape from the global minimum. The ability of the system to move from the state of higher distortion guarantees that the algorithm avoids non-optimal global minima. It can be shown that if the initial melting temperature $T_0$ is large enough, the simulated algorithm guarantees convergence to the global minimum in probability with a cooling schedule described by $T_m = c/\log(m+1)$ [74]. Since this cooling schedule is very slow, a faster cooling schedule such as $T_m = \alpha T_{m-1}$ , where $0 < \alpha < 1$, has been used in practical applications of simulated annealing algorithm. If the cooling speed is too fast(corresponding to a very small $\alpha$) we may have not the sufficient perturbations. On the other hand, if $\alpha$ is near to 1, the cooling schedule is too slow. We find empirically that $\alpha = 0.9$ is a good compromise between the sufficient perturbations and the cooling speed.

## 4.4 Optimization of the Modulation Constellation

From equation (4.4) we can see that the overall system distortion depends on both the source coder, i.e., the VQ codebook $\mathcal{C}$ and partition $\mathbf{\Omega}$, and the channel transition

probability $P(j|i)$. The transition probability is also related to the modulation signal constellation as shown by

$$P(j|i) = \int_{\Theta_j} p(\mathbf{y}/\mathbf{c}_i)d\mathbf{y} \tag{4.10}$$

where $\Theta_j$ is the $j$-th detection partition region associated with the signal $\mathbf{s}_j$ in the modulation signal space. Therefore, the overall system distortion can be further reduced by appropriately designing the modulation signal sets. To realize this objective, let us express the overall system distortion as

$$D = \frac{1}{k}\sum_{i=1}^{N}\sum_{j=1}^{N} P_i P(j|i)\|\mathbf{c}_i - \mathbf{c}_j\|^2 \tag{4.11}$$

where $\mathbf{c}_i$ denotes the transmitted codeword, $\mathbf{c}_j$ denotes the reconstruction codeword, and $P(j|i)$ denotes the transition probability that the demodulator chooses $\mathbf{c}_j$ when $\mathbf{c}_i$ is sent. The demodulation can be thought of as a partition of the $M$-dimension signal space into $N$ regions $\mathbf{\Theta} = \{\Theta_1, \Theta_2, \ldots, \Theta_N\}$. For the AWGN channel, the transition probability $P(j|i)$ can be expressed by

$$P(j|i) = \int_{\Theta_j} \frac{1}{(\pi N_0)^{\frac{M}{2}}} exp(-\frac{1}{N_0}\|\mathbf{y} - \mathbf{s}_i\|^2)d\mathbf{y} \tag{4.12}$$

where $M$ is the dimensional number of the modulation signal. It is necessary to impose a constraint condition on the average energy:

$$P_{avg} = \sum_{i=1}^{N} P_i\|\mathbf{s}_i\|^2. \tag{4.13}$$

We want to find the optimal modulation signal constellation to minimize the average distortion under the constraint condition (4.13). In formulating the optimization problem with a constraint, this constraint can be incorporated using a Lagrange multiplier.

Let us first construct another unconstrained problem with the target function

$$D' = kD + \gamma(\sum_{i=1}^{N} P_i\|\mathbf{s}_i\|^2 - P_{avg})$$

where $\gamma$ is a Lagrange multiplier. It can be shown easily that

$$D' = \sum_{i=1}^{N}\sum_{j=1}^{N}\int_{\Theta_j}\frac{1}{(\pi N_0)^{\frac{M}{2}}}exp(-\frac{1}{N_0}\|\mathbf{y}-\mathbf{s}_i\|^2)d\mathbf{y} + \gamma(\sum_{i=1}^{N}P_i\|\mathbf{s}_i\|^2 - P_{avg})$$

The optimal constellation $\mathbf{s}_i$ can be obtained by setting the derivatives of $D'$ with respect to $\mathbf{s}_i$ to zero

$$\frac{\partial D'}{\partial \mathbf{s}_i} = 0$$

Actually, the decision space partition $\mathbf{\Theta}= \{\Theta_1, \Theta_2, \ldots, \Theta_N\}$ is related to the modulation signal constellation $\mathbf{s}_1, \mathbf{s}_2 ..., \mathbf{s}_N$. Because no analytical expression for the decision partition is available, we resort the GLA strategy, i.e., assuming the decision partition region $\mathbf{\Theta}$ is fixed. As a result, we find the optimal necessary condition for $\mathbf{s}_i$, i.e.,

$$\sum_{j=1}^{N}\left[\int_{\Omega_j}\frac{1}{(\pi N_0)^{\frac{M}{2}}}exp(-\frac{1}{N_0}\|\mathbf{y}-\mathbf{s}_i\|^2)\mathbf{y}d\mathbf{y}\right]\|\mathbf{c}_j-\mathbf{c}_i\|^2$$

$$-\sum_{j=1}^{N}\left[\int_{\Omega_j}\frac{1}{(\pi N_0)^{\frac{M}{2}}}exp(-\frac{1}{N_0}\|\mathbf{y}-\mathbf{s}_i\|^2)d\mathbf{y}\right]\|\mathbf{c}_j-\mathbf{c}_i\|^2\mathbf{s}_i + \gamma P_i\mathbf{s}_i = 0 \qquad (4.14)$$

Now if we define

$$\xi_i \triangleq \left[\sum_{j=1}^{N}\int_{\Omega_j}\frac{1}{(\pi N_0)^{\frac{M}{2}}}exp(-\frac{1}{N_0}\|\mathbf{y}-\mathbf{s}_i\|^2)\mathbf{y}d\mathbf{y}\right]\|\mathbf{c}_j-\mathbf{c}_i\|^2, \qquad (4.15)$$

$$\eta_i \triangleq \sum_{j=1}^{N}\int_{\Omega_j}\frac{1}{(\pi N_0)^{\frac{M}{2}}}exp(-\frac{1}{N_0}\|\mathbf{y}-\mathbf{s}_i\|^2)d\mathbf{y}\|\mathbf{c}_j-\mathbf{c}_i\|^2, \qquad (4.16)$$

where $\xi_i$ is an $M$-dimension vector, and $\eta_i$ is a scalar, we obtain

$$\xi_i = (\eta_i - P_i\gamma)\mathbf{s}_i$$

and the optimal modulation signal vector is given by

$$\mathbf{s}_i = \frac{\xi_i}{\eta_i - P_i\gamma} \qquad (4.17)$$

Substituting equation (4.17) into the constrained condition (4.13) we can determine the corresponding Lagrange multiplier $\gamma$

$$\sum_{i=1}^{N} \frac{\|\xi_i\|^2}{(\eta_i - P_i\gamma)^2} = P_{avg} \tag{4.18}$$

If we use equation (4.18) to determine the Lagrange multiplier $\gamma$, we get an nonlinear equation of $2N$-order. Correspondingly we have $2N$ Lagrange multipliers $\gamma_i$, $i = 1, 2, \ldots, N$, which means that we need to test a total of $2N$ modified signal vectors to determine the best candidate. This method is not very efficient.

Alteratively, we can use a constrained gradient-search algorithm to determine the optimal constellation as Foschini did in the optimization of signal constellation under the criterion of minimum error detection probability [39]. To describe this approach, it is convenient to introduce a vector presentation

$$\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{s}_N \end{bmatrix}$$

where $\mathbf{S}$ is an $NM \times 1$ vector. The channel signal-to-noise ratio per bit $SNR_c$ is defined as

$$SNR_c = 10\frac{P_{avg}}{\frac{N_0}{2}\log_2 N}$$

where $\frac{N_0}{2}$ denotes the spectral density of the channel's additive white noise. Let

$$\mathbf{S}_m = \begin{bmatrix} \mathbf{s}_1^{(m)} \\ \mathbf{s}_2^{(m)} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{s}_N^{(m)} \end{bmatrix}$$

denote an $NM$ dimension modulation signal vector obtained at the $m$th step of the iteration process. A conventional unconstrained gradient search algorithm can be described by

$$\mathbf{S}_{m+1} = \mathbf{S}_m - \mu \frac{\partial D}{\partial \mathbf{S}_m}$$

where $\mu$ is a positive step size, and $\frac{\partial D}{\partial \mathbf{S}_m}$ is the gradient of $D$ with respect to $\mathbf{S}_m$. The energy of $\mathbf{S}_m$ will increase or decrease with the iteration number $m$ unless an energy constraint condition is incorporated. To satisfy the constraint on the energy condition, the signal vectors are re-normalized at each step. The signal vector $\mathbf{S}_m$ is updated according to

$$\mathbf{S}'_{m+1} = \mathbf{S}_m - \mu \frac{\partial D}{\partial \mathbf{S}_m}, \tag{4.19}$$

and

$$\mathbf{S}_{m+1} = \sqrt{\frac{P_{avg}}{\sum_{i=1}^{N} P_i \|\mathbf{S}'_{m+1}\|^2}} \mathbf{S}'_{m+1}. \tag{4.20}$$

A heuristic discussion on the convergence of the constrained gradient-search algorithm can be found in Foschini's paper [39].

When we use the constrained gradient-search optimization algorithm to solve for the optimal signal constellation, we need to determine the gradient of $D$ with respect to $\mathbf{S}_m$. The gradient can be written as

$$\frac{\partial D}{\partial \mathbf{S}_m} = \xi_i - \eta_i \mathbf{s}_i \tag{4.21}$$

Two factors prevent us from using (4.21) to evaluate the gradient directly. First, obtaining an analytic expression for the source VQ partition $\mathbf{\Omega}$ is difficult. Although we can determine, in principle, the $N$ disjoint source partitions from the nearest neighbor condition, in general, analytical expressions of the partition boundaries can not be found for arbitrary source distributions. The second factor is related to the

calculation of a numerical integration over an $N$-dimension volume. It is generally believed that only two- or three-dimensional integral can be efficiently solved by a numerical method. If the dimension of the integral is over four, it is advantageous to use the Monte-Carlo simulation method to evaluate the integration. Therefore, it is useful to find a numerical method to determine the optimal constellation without resorting to an exact analytic description on the partitions $\Omega_1, \Omega_2, \ldots, \Omega_N$.

What we need to do is to find a way to calculate the integrals in equation (4.15) and (4.16). It is easy to see that

$$P(j|i) = \int_{\Theta_j} \frac{1}{(\pi N_0)^{\frac{M}{2}}} exp(-\frac{1}{N_0}\|\mathbf{y} - \mathbf{s}_i\|^2) d\mathbf{y}, \tag{4.22}$$

$$E\{\mathbf{y}(j/i)\} = \int_{\Theta_j} \frac{1}{(\pi N_0)^{\frac{M}{2}}} exp(-\frac{1}{N_0}\|\mathbf{y} - \mathbf{s}_i\|^2) \mathbf{y} d\mathbf{y}, \tag{4.23}$$

where $P(j|i)$ denotes the transition probability and $E\{\mathbf{y}(j/i)\}$ denotes the conditional mean of the received signal under the condition that the output of the demodulator is $j$ while the index $i$ is transmitted. When the Monte-Carlo simulation is used to estimate the transition probability, an empirical version of (4.22) is adopted

$$\hat{P}(j/i) = \frac{1}{N_t} \sum_{i \in \Theta_j} i \tag{4.24}$$

Similarly, an empirical counterpart of (4.23) is

$$E\{\hat{\mathbf{y}}(j/i)\} = \frac{1}{N_t} \sum_{i \in \Theta_j} \mathbf{y}_i \tag{4.25}$$

where $N_t$ denotes all the experiment data, and $\Theta_j$ denotes the decision region associated with the $j$-th modulation signal point $\mathbf{s}_j$. It is easy to show that the estimates of $P(j|i)$ and $E\{\mathbf{y}(j/i)\}$ are unbiased. For the unbiased estimate, the variance is a good measure of the estimated performance [75]. Generally speaking, the variance is reduced in proportion to $1/N_t$.

Table 4.1: Transition probability by simulation and numerical method

| *Probability* | *Theory Value* | *Simulation Result* |
|---|---|---|
| $P(1/1)$ | 0.848878 | 0.847942 |
| $P(2/1)$ | 0.072464 | 0.0721237 |
| $P(3/1)$ | 0.006194 | 0.005338 |
| $P(4/1)$ | 0.072464 | 0.071915 |

Table 4.2: Condition mean by simulation and numerical method

| *Condition Mean* | *Theory Value* | *Simulation Result* |
|---|---|---|
| $x(1/1)$ | 0.981241 | 0.983191 |
| $y(1/1)$ | 0 | 0.002563 |
| $x(2/1)$ | 0.028737 | 0.002746 |
| $y(2/1)$ | 0.051888 | 0.051912 |
| $x(3/1)$ | 0 | 0.00748 |
| $y(3/1)$ | 0.001976 | 0.002713 |
| $x(4/1)$ | -0.028737 | -0.028297 |
| $y(4/1)$ | 0.051885 | 0.052515 |

To investigate the numerical accuracy of the Monte-Carlo simulation, let us look at a simple system in which four codewords are transmitted using a QPSK modulation constellation. The experimental parameters are as follows. The channel SNR is 7 dB, and 150,000 random numbers were used to get the estimated value of the transition probability (4.24) and the conditional mean (4.25). The simulated annealing algorithm was used to optimize the VQ index assigned to the modulation constellation. The results are shown in Tables 4.1 and 4.2. Note the theoretical values of the transition probability and conditional mean are obtained by using a two-dimensional numerical integration method [76].

Using the Monte-Carlo simulation method we need not depend on the analytical expressions of the VQ partitions $\mathbf{\Omega}$ and the decision regions $\mathbf{\Theta}$. Therefore the technique is very suitable for designing an arbitrary $N$ point modulation constellation in an M-dimension signal space.

## 4.5 Demodulation Optimization

As we see from equation (4.4), the overall system distortion is also a function of the channel transition probability $P(j|i)$. The channel transition probability $P(j|i)$ itself is determined by the rule of the demodulator. Therefore, the average distortion $D$ can also be reduced by appropriately optimizing the decision rule in the demodulator. In this section, we develop an optimal demodulator to minimize the MSE distortion.

Optimizing demodulation is well-known in signal detection theory [65]-[66] and is accomplished if the minimization of error probability is used as an optimization criterion. This criterion leads to an maximum likelihood (ML) receiver if the priori probabilities of the source are equal. In the joint source and channel coding system, the MSE, instead of the bit error probability, is used as the optimization criterion due

to different error bits corresponding to various distortions [65]. The average distortion can be expressed as

$$D = \sum_{i=1}^{N} \sum_{j=1}^{N} P_i P(j|i) d_{ij} \tag{4.26}$$

where $d_{ij}$ is the cost of choosing index $j$ when the codeword $i$ is transmitted. In our case

$$d_{ij} = \|\mathbf{c}_i - \mathbf{c}_j\|^2$$

Since no redundancy is introduced, the output of the source coder is directly mapped into the modulation signal point. Replacing the transition probability by an integral of the conditional pdf, we have

$$D = \sum_{j=1}^{N} \int_{\Theta_j} \left( \sum_{i=1}^{N} P_i \frac{1}{(\pi N_0)^{\frac{M}{2}}} exp(-\frac{1}{N_0}\|\mathbf{y} - \mathbf{s}_i\|^2) \|\mathbf{c}_i - \mathbf{c}_j\|^2 \right) d\mathbf{y} \tag{4.27}$$

Note that every item in the parentheses is positive. If we want to minimize the distortion, we should assign $\mathbf{y}$ to the region $\Omega_j$ in which each item in the bracket is the minimum. Let us define a decision function $I_j(\mathbf{y})$ as

$$I_j(\mathbf{y}) = \sum_{i=1}^{N} P_i \frac{1}{(\pi N_0)^{\frac{M}{2}}} exp(-\frac{1}{N_0}\|\mathbf{y} - \mathbf{s}_i\|^2) \|\mathbf{c}_i - \mathbf{c}_j\|^2. \tag{4.28}$$

It is simple to show that the optimum decision rule is given by:

$$\text{if} \quad I_j(\mathbf{y}) < I_m(\mathbf{y}) \quad \text{for } m = 1, 2, \ldots N \quad \text{then} \quad \mathbf{y} \in \Omega_j \tag{4.29}$$

The receiver described by equation (4.28) and equation (4.29) is called the generalized Bayes receiver in the signal detection theory literature [65]. In fact, the ML receiver is just a special case of the general Bayes receiver. That is, if we define the distortion cost $d_{ij}$ as follows

$$d_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{otherwise} \end{cases} \tag{4.30}$$

for every correct decision the cost is zero and for every error decision the cost is equal to one. Also assume that the apriori probability for each signal is the same, i.e.

$$P_i = \frac{1}{N} \quad \text{for } i = 1, \ldots, N$$

Then, the decision rule can be expressed as [65]

$$\text{if} \quad \|\mathbf{y} - \mathbf{s}_i\|^2 < \|\mathbf{y} - \mathbf{s}_j\|^2 \quad \text{for } \mathrm{j} = 1, 2, \ldots, \mathrm{N}, \text{ then } \mathbf{y} \in \Omega_i$$

## 4.6 An Iterative Optimization Procedure for the Source and Modulation Constellations

We have shown in the previous two sections how to optimize the VQ and the modulation constellation separately. In this section, a strategy for jointly optimizing these two subsystems is discussed. Once again, an algorithmic approach is adopted. The algorithm is essentially a general form of the Lloyd algorithm. The key steps are summarized below.

1. For the given continuous source, design an $N$-level VQ for a noiseless channel using the Linde-Buzo-Gray(LBG) algorithm, or source optimized VQ [43]. Choose also a standard modulation constellation. These initial VQ and modulation systems will be updated in each iteration of the design process.

2. For the given modulation scheme, optimize the VQ partitions and the index mapping according to the procedure outlined in sections 4.3 and 4.4.

3. For the given VQ partitions and mapping, optimize the modulation constellation according to section 4.5.

4. Test whether or not the convergence has occurred. If so, declare the current VQ encoder and the modulation constellation as being jointly optimal. Otherwise go back to step 2.

It should be pointed out that the overall system distortion is non-increasing at every iteration. Consequently, the iterative procedure will always converge, at least to a local minimum. Since only a local optimal solution is guaranteed by this design procedure, it is therefore important to choose an appropriate initial VQ and modulation signal set. Both the convergence rate and the performance of the final system can be improved substantially if the initial state of the design process is set properly. We found empirically that a standard QAM signal set and a VQ codebook optimized for a noiseless channel are an appropriate choice for the algorithm initialization.

## 4.7 Performance Results

We present in this section the simulation results for our jointly optimized source-channel coding system. These results are obtained for a first order Gauss-Markov source for a variety of channel signal-to-noise ratios and bandwidths. Comparisons are made against the standard systems consisting of a source optimized vector quantizer (SOVQ) and using a QAM or trellis-coded modulation (TCM).

The source is assumed to be modeled by a zero-mean, first-order, Gauss-Markov random process. This source is represented by the equation

$$u_i = \rho u_{i-1} + e_i$$

where $e_i$ is an independent, and identically distributed Gaussian process with variance $E\{e_i^2\} = (1 - \rho^2)$, and $\rho$ is the normalized correlation coefficient between successive samples. If we want $u_n$ to have some sort of characteristic of the speech signal, then a typical value of $\rho$ is 0.9. The channel is modeled as a stationary, independent, vector Gaussian random process of dimension $M$ with a covariance matrix $\frac{N_0}{2}\mathbf{I}$, where $\mathbf{I}$ is

an $M \times M$ identity matrix. The channel $SNR_c$ per bit is defined as following

$$SNR_c = 10 \log_{10} \frac{P_{av}}{\frac{N_0}{2} \log_2 N}$$

For the purposes of comparison, we consider two simple communication systems in which a source optimized VQ is used as a source coder. It is assumed that the source encoder rate and the source vector dimension is the same as that use in the jointly optimized system. In these two systems, the index of the VQ is directly mapped into a modulation signal point and then transmitted over an AWGN channel. For the first reference system, a standard QAM signal constellation is used as the modulation. In order to maintain the same bandwidth with the optimized signal constellation, a $J$-fold Cartesian product of the QPSK signal constellation is used. The receiver consists of a conventional maximum-likelihood receiver followed by a source decoder that maps the decoded index back to its corresponding centroid. The simulated annealing optimization algorithm is employed to find the best mapping from the VQ index to the signals in the modulation signal sets. In our second reference system, we use the trellis-coded-modulation (TCM) with an 8-PSK constellation. Soft decision Viterbi decoding algorithm is used for the TCM decoding.

The design of the optimal source and channel coding system according to the presented procedure requires the knowledge of channel SNR. Strictly speaking, we should design the optimal modulation and VQ codebook for every possible channel SNR condition, which requires channel SNR estimation and coordination between the transmitter and the receiver. An alternative method is adopted to avoid channel SNR estimation. We choose a fixed channel SNR of 5.5 dB as a design parameter, since it gives a good compromise between the system performance at the high channel SNR and the low channel SNR. Otherwise, if the designed SNR is too high or too low, the loss due to the channel mismatch will become larger. The resultant system is then

applied in a variety of SNR values to evaluate the system performance.

Figures 4.2 and 4.3 present the VQ codebook constellation and the corresponding modulation signal constellation for the codebook size $N = 16$, the block length $k = 2$, and the signal dimension $M = 2$. Figures 4.4-4.9 present the optimal system performance compared with the QAM and TCM systems. These figures show that the jointly designed system provides significant improvement when the channel SNR is less than 8 dB. For instance, in Figure 4.6, when the channel SNR is 6 dB, the source codebook size $N$ is 256, the block length $k$ is 2, the 8-dimension optimal modulation system has 3.5 dB gain in source SNR compared to a traditional system that uses 4-fold QPSK modulation. The source SNR improvement decreases with the VQ codebook size decreasing. For the same channel SNR, when the codebook size is $N = 16$, we obtain just about 1 dB improvement. If we examine the figures in the higher channel SNR region, we will find an interesting phenomenon. The conventional QAM modulation system is better than the optimized system in this region. The phenomenon can be explained as follows. When we optimize the modulation constellation in terms of the MSE criterion, the Euclidean distance between the modulation signal points corresponding to the smaller source distortion is decreased. In this way, the optimized system is quite robust to the decision errors caused by the channel noise. However, at high channel SNR, this strategy has the disadvantage that more signal energy is needed to distinguish the modulation signal points with smaller Euclidean distances. It should be pointed out that this phenomenon is due to the mismatch in the channel SNR with respect to the designed SNR.

We can also see that the TCM system is better than the proposed system in the high channel SNR region since the TCM system combines the channel coding and modulation to increase the Euclidean distance of the modulation signal. Therefore, it can combat the channel noise without extra bandwidth. At receiver, it uses a soft

decision Viterbi decoding algorithm [77] as the demodulation scheme. This scheme chooses the most likely transmitted sequence based on the received sequence.

We also studied the effect of the dimension of modulation signal on the optimized gain for a fixed VQ source coder. Figs. 4.7-4.9 present the results. It is found that the optimized gain increases with the increase of the modulation signal dimension, but the rate of increasing decreases. For instance, when the codebook size $N$ is equal to 256, the VQ dimension $k$ is equal to 2, the channel SNR is 6 dB, and the dimension of the modulation signal increases from 2 to 4, the optimized gain is 2.5 dB. However, the optimized gain is only 1 dB for the same source encoder if the dimension $M$ changes from 6 to 8. Using high dimension signal constellation will result in better performance but it increases the complexity of the modulation system. We should trade-off between the system complexity and the performance improvement.

## 4.8 Conclusions

In this chapter, we considered the problem of joint optimization of source and channel coding with a hard decision receiver. The source signal is compressed by a VQ and the VQ index is directly mapped into a signal point in the modulation constellation. The objective is to design an appropriate VQ and modulation constellation to minimize the average distortion between the source vectors and its corresponding reproduction at the receiver. The strategy we used is to optimize the encoder, the modulator, and the demodulator iteratively while maintaining constraint conditions on the signal energy and transmission bandwidth. A hard decision rule that minimizes the overall system MSE is derived from the generalized Bayes receiver. It turns out the receiver depends not only on the channel statistics but also on the source character. A constrained gradient search algorithm is introduced to optimize the modulation constellations in

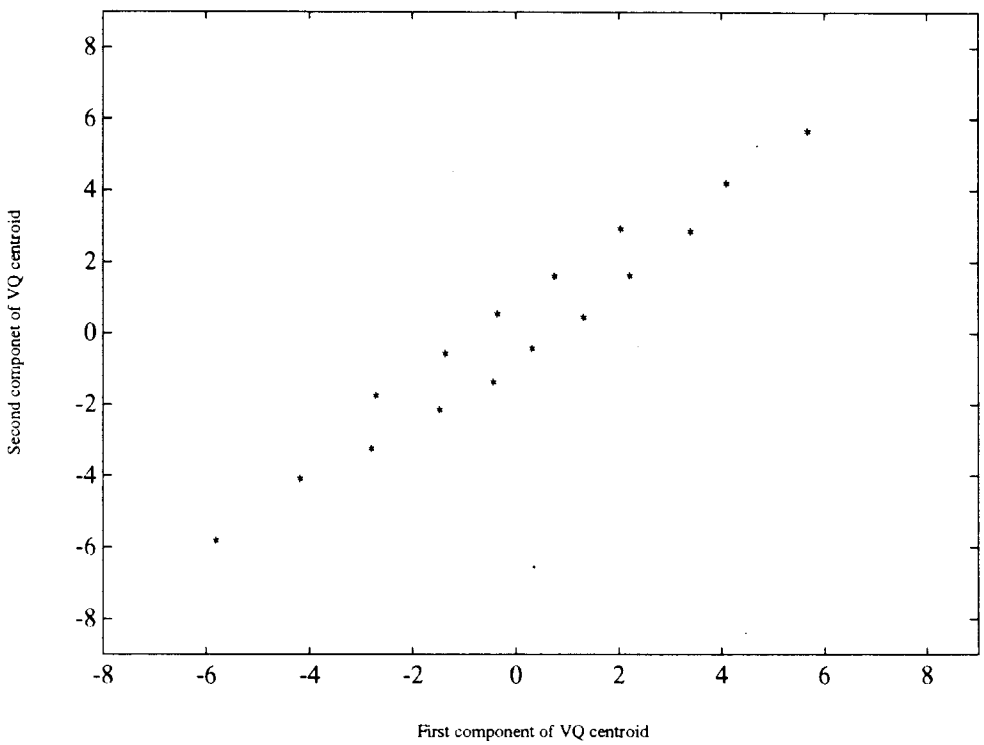


Figure 4.2: A 2-dimension VQ codebook constellation. VQ codebook size $N = 16$, block length $k = 2$, Source AR with coefficient $\rho = 0.9$

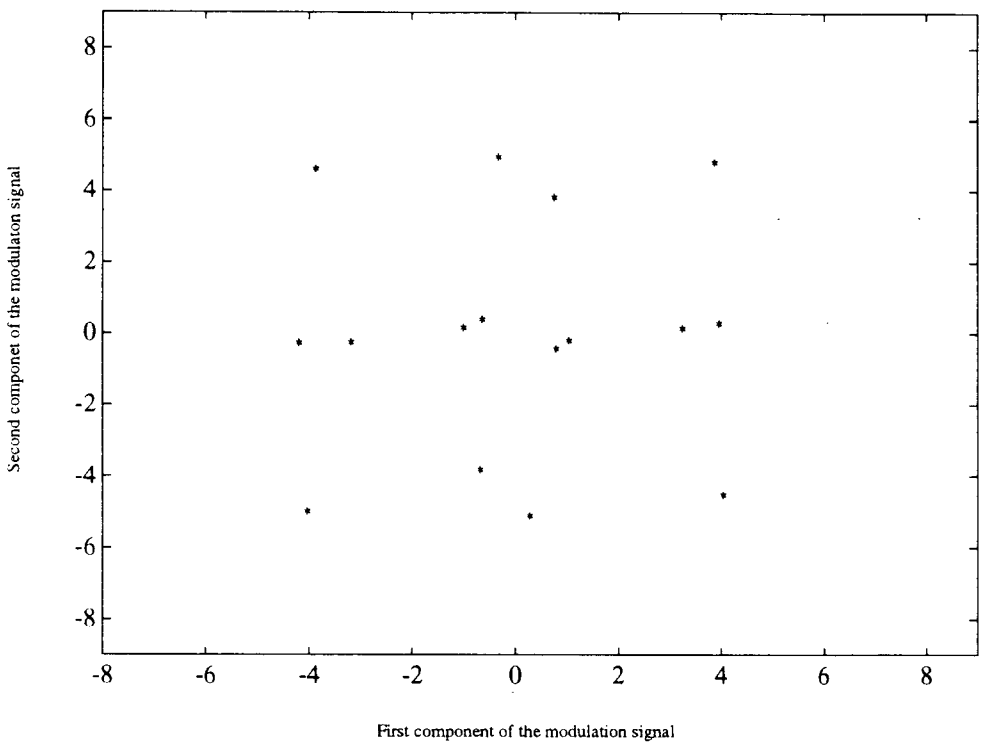


Figure 4.3: A 2-dimension modulation Constellation. VQ codebook size $N = 16$, block length $k = 2$, Source **AR**-1 with $\rho = 0.9$

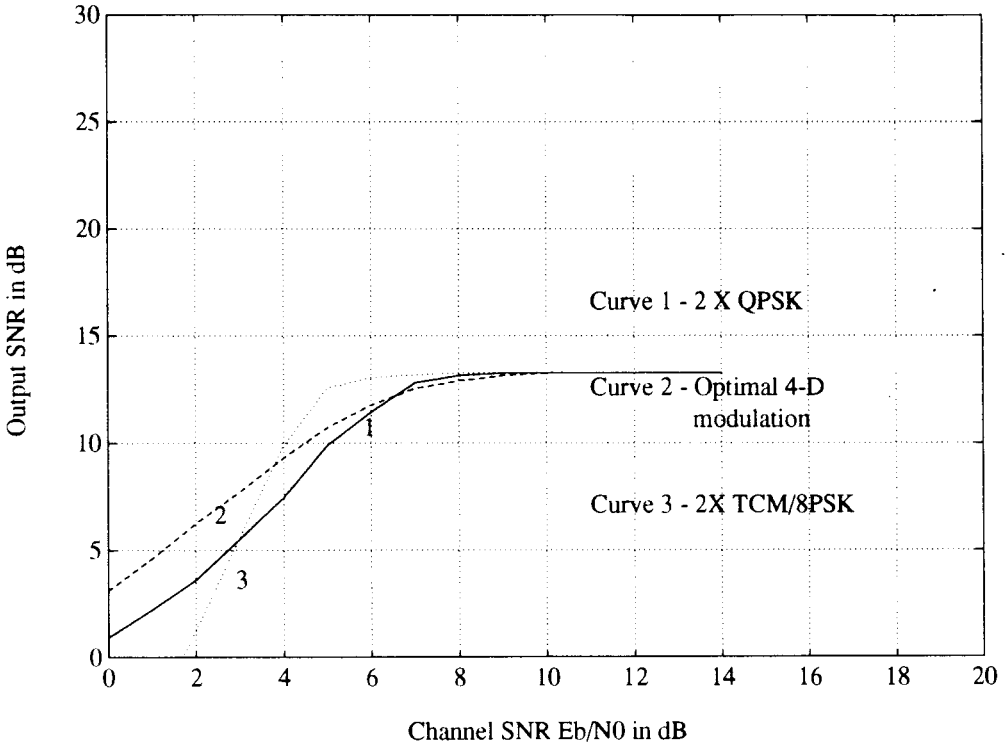


Figure 4.4: Comparison amongst various combined source and channel coding systems. The modulation signal dimension $M$ is 4. Curve 2 is the performance curve of the jointly optimized system. Curve 1 corresponds to the case where 2 QPSK symbols are used to represent an index in the VQ codebook. Curve 3 is similar to curve 1 except that 2 trellis-code 8PSK symbols are used instead. VQ codebook size $N = 16$, block length $k$=2, Source **AR**-1 with $\rho$=0.9

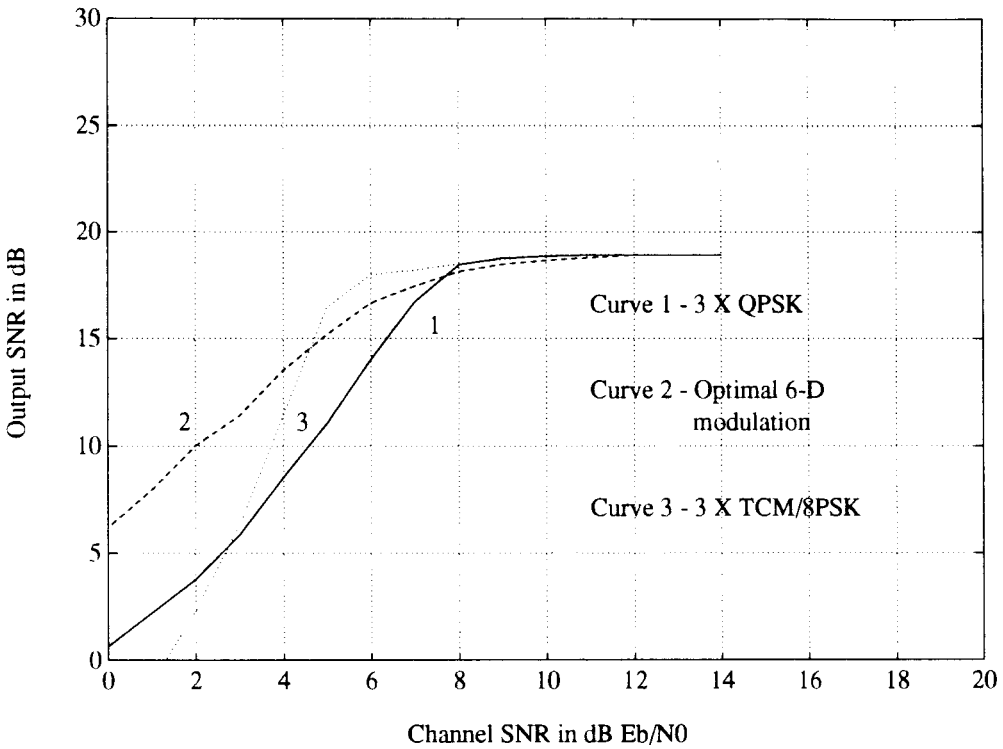


Figure 4.5: Comparison amongst various combined source and channel coding system. The total number of dimension $M$ in the signal space is 6. Curve 2 is the performance curve of the jointly optimized system. Curve 1 corresponds to the case where 3 QPSK symbols are used to represent an index in the VQ codebook. Curve 3 is similar to curve 1 except that 3 trellis-code 8PSK symbols are used instead. VQ codebook size $N = 64$, block length $k$=2, Source **AR**-1 with $\rho$=0.9

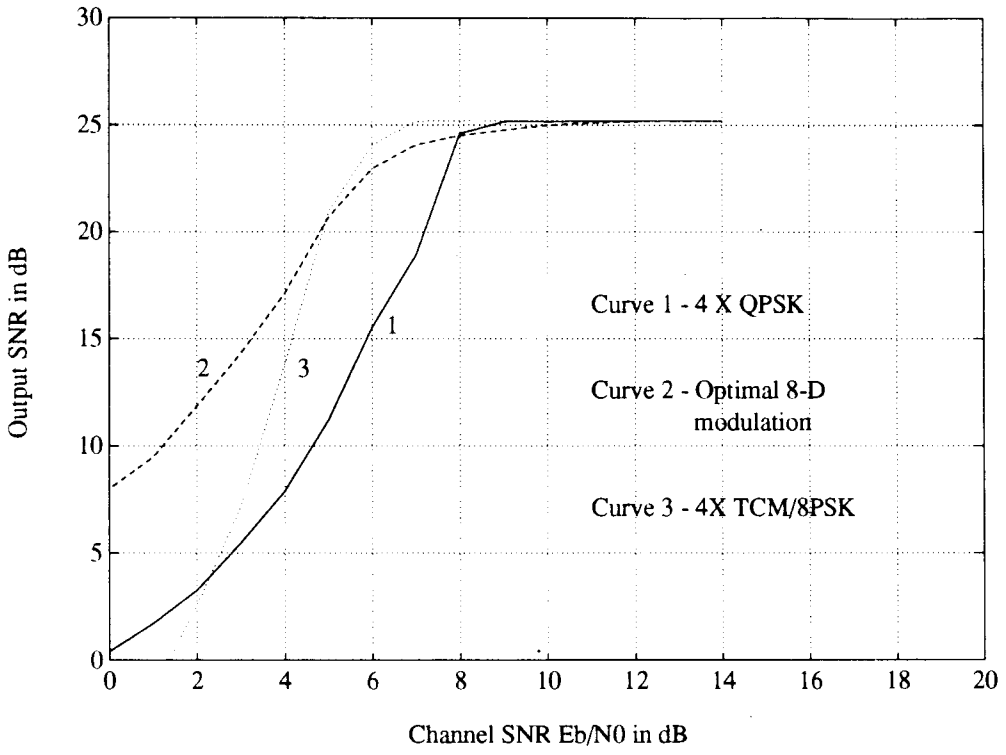


Figure 4.6: Comparison amongst various combined source and channel coding system. The total number of dimension $M$ in the signal space is 8. Curve 2 is the performance curve of the jointly optimized system. Curve 1 corresponds to the case where 4 QPSK symbols are used to represent an index in the VQ codebook. Curve 3 is similar to curve 1 except that 4 trellis-code 8PSK symbols are used instead. VQ codebook size $N = 256$, block length $k$=2, Source **AR**-1 with $\rho$=0.9

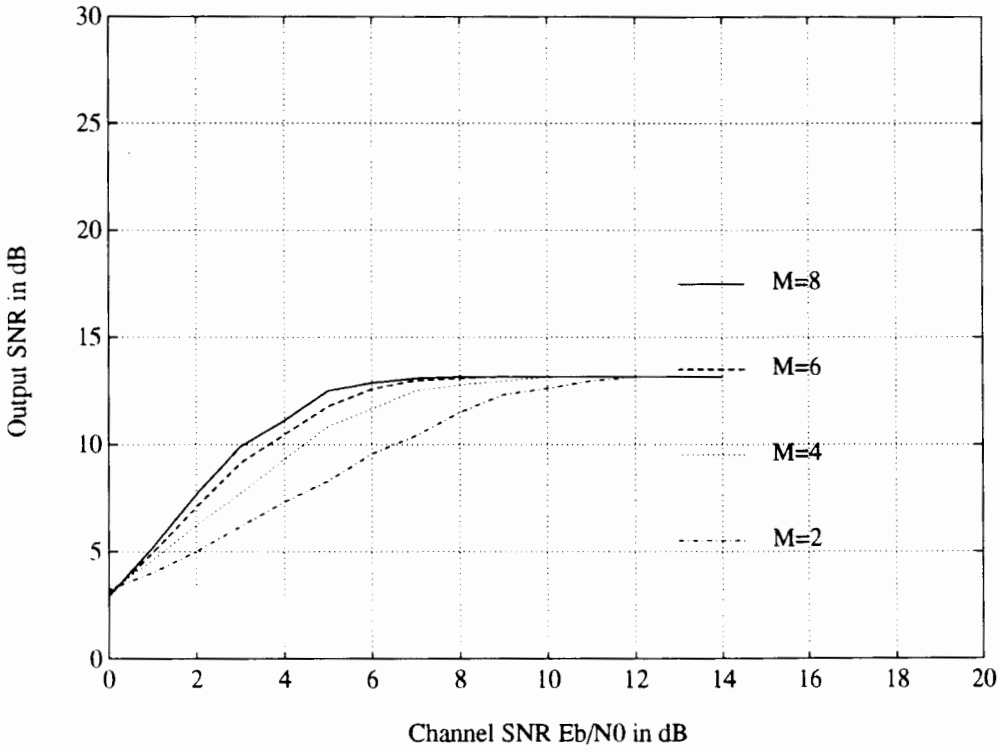


Figure 4.7: Output SNR changes with respect to various channel SNR for different signal space dimension $M$. VQ codebook size $N = 16$, block length $k$=2, Source **AR**-1 with $\rho$=0.9

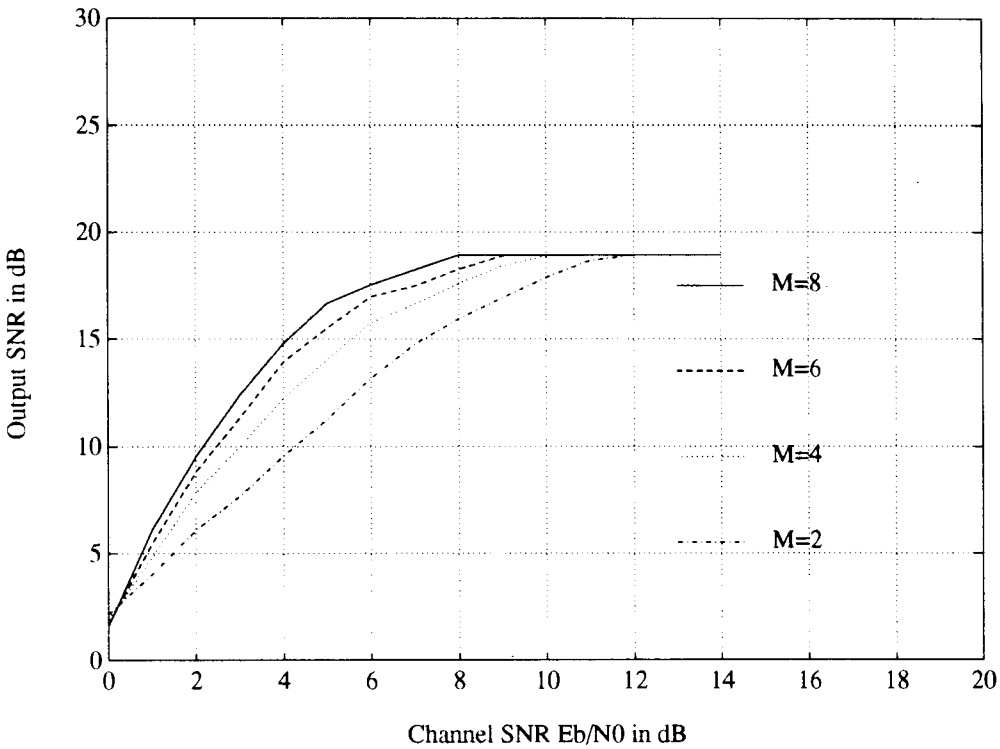


Figure 4.8: Output SNR changes with respect to various channel SNR for different signal space dimension $M$. VQ codebook size $N = 64$, block length $k$=2, Source **AR**-1 with $\rho$=0.9

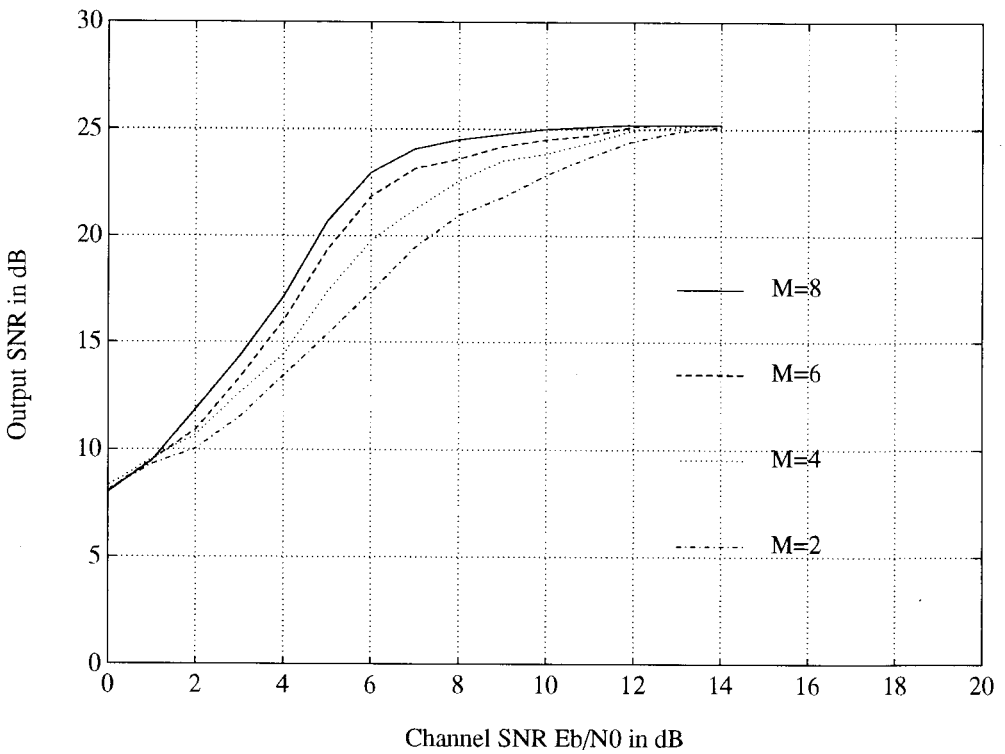


Figure 4.9: Output SNR changes with respect to various channel SNR for different signal space dimension $M$. VQ codebook size $N = 256$, block length $k$=2, Source **AR**-1 with $\rho$=0.9

terms of MSE criterion. We also considered the problem of VQ index assignment and used the simulated annealing algorithm for the optimization of the VQ index assignment. An iterative optimization algorithm for jointly optimizing source and channel coding was developed based on a straightforward extension of Lloyd algorithm. Simulation results show that the joint design of source and channel coding improves the system performance significantly, especially at the low channel SNR regions.

From estimation theory [65], we know that the optimal receiver in terms of the MSE criterion should be a conditional mean of the source given the received signal. We will consider the problem of joint design of VQ and modulation constellation based on this result.

# Chapter 5

# Joint VQ and Modulation Signal Design Based on the Conditional Estimate

The problem of the joint optimization of source and channel coding with a maximum likelihood receiver was considered in Chapter 4. In this chapter we go further to consider the problem of the joint design of the VQ and modulation constellations in a system that uses a "soft decision" decoder. Note that in a conventional system, it is implicitly assumed that the output of the demodulator is one of the VQ codebook entries. In this case, the design procedure aims to optimize the VQ codebook and the corresponding partition in order to achieve the best representation of the source vector at the receiver. However, there is no reason to constrain the output of demodulator to be one of the codevectors. This chapter introduces a "soft decision" decoder which uses a linear combination of the VQ codevectors to reconstruct the source signal. The weighting coefficients used in the reconstruction process depend on the Euclidean distance between the received signal vector and the modulation signal constellation (in the modulation signal space). They also depend on the channel statistics.

The objective is to suitably design the VQ, the modulation scheme, and the decoder structure in order to minimize the over all system distortion subject to the constraints on the signal energy and bandwidth. This chapter is organized as follows. We derive in Section 5.1 the optimum MSE decoder for a given source encoder (VQ) and a channel encoder (signal constellation) pair. This is followed by a derivation of the optimum VQ encoder for a fixed modulation constellation. In Section 5.3, the necessary conditions for optimality about the modulation constellation are derived for a fixed VQ. A constrained gradient-search optimization algorithm is introduced to find the optimal constellation. In Section 5.4, an iterative algorithm is presented to jointly optimize the source and channel coding. Numerical results for a first order Gauss-Markov source are also shown. Finally, a summary of the chapter is given in Section 5.5.

## 5.1 The Decoder Design

The block diagram of the communications system to be investigated is shown in Fig. 5.1. Remember that a vector representation of signals is used in this thesis. The source

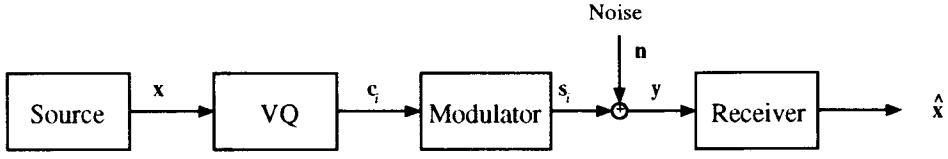


Figure 5.1: A general communication system model.

in Fig. 5.1 is a zero-mean stationary random process. Let $\mathbf{x}$ be a $k$-dimensional vector derived from the random source. The pdf of $\mathbf{x}$ is $p(\mathbf{x})$. The vector $\mathbf{x}$ is represented by a VQ codevector chosen from the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$, where $N$ is the codebook size. Assuming that the codevector $\mathbf{c}_i$ is chosen, then its index $i$ is mapped directly into the modulation signal point $\mathbf{s}_i$. The set of $\mathbf{s}_i$, $i = 1, \ldots, N$, forms the signal constellation.

The dimension of this constellation, $M$, is an important design parameter as it is proportional to the bandwidth of the communication channel [15]. The modulated signal $\mathbf{s}_i$ is transmitted over an additive white Gaussian noisy channel (AWGN) with a power spectral density of $N_0/2$. In vector notation, the noise waveform is represented by an $M$-dimensional Gaussian random vector $\mathbf{n}$. The different components of $\mathbf{n}$ are independent and identically distributed (i.i.d), each having a zero mean and a variance of $N_0/2$. The received signal $\mathbf{y}$ is simply the sum of the transmitted signal $\mathbf{s}_i$ and the noise vector $\mathbf{n}$. Given the signal $\mathbf{y}$, the decoder should provide the "best" estimate of the original signal $\mathbf{x}$. This estimate, denoted by $\hat{\mathbf{x}}(\mathbf{y})$, is then delivered to the final destination.

The performance of the above system is measured in terms of the (per sample) mean squared error (MSE) between the original and the reconstructed signal. Let $p(\mathbf{x}, \mathbf{y})$ denote the joint probability density function of $\mathbf{x}$ and $\mathbf{y}$. Then the average distortion can then be written as

$$D = \frac{1}{k} E(\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2) = \frac{1}{k} \int \int \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2 p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \tag{5.1}$$

where $k$ is the dimension of the source vector $\mathbf{x}$. The problem we want to address is how to design the VQ codebook, the modulation signal constellation, and the decoder structure in order to minimize the average distortion while satisfying the constraints on the average signal energy and bandwidth. In this section, we first look into the optimization of the decoder for a fixed VQ and modulation scheme.

According to the Bayesian estimation theory [65], the optimum decoder under the minimum MSE criterion should compute the conditional mean of the source vector $\mathbf{x}$ given the received signal $\mathbf{y}$. In other words

$$\hat{\mathbf{x}}(\mathbf{y}) = E[\mathbf{x}|\mathbf{y}] = \frac{1}{p(\mathbf{y})} \int \mathbf{x} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}) d\mathbf{x} \tag{5.2}$$

where $p(\mathbf{x})$, $p(\mathbf{y})$ are the pdf of $\mathbf{x}$ and $\mathbf{y}$ respectively, and $p(\mathbf{y}|\mathbf{x})$ is the conditional pdf of $\mathbf{y}$ given $\mathbf{x}$. Let $\Omega_i$, $i = 1, 2, \cdots, N$, denote the $i$th partition in the VQ, and $p(\mathbf{n})$ be the pdf of the Gaussian noise vector $\mathbf{n}$. Then it can be shown that (5.2) can be rewritten as

$$\hat{\mathbf{x}}(\mathbf{y}) = \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p(\mathbf{y}|\mathbf{s}_i)}{\sum_{i=1}^{N} P_i p(\mathbf{y}|\mathbf{s}_i)}, \tag{5.3}$$

where

$$p(\mathbf{y}|\mathbf{s}_i) = p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_i) = \frac{1}{(N_0\pi)^{M/2}} \exp(-\frac{\|\mathbf{y} - \mathbf{s}_i\|^2}{N_0}) \tag{5.4}$$

is the conditional pdf of $\mathbf{y}$ given that $\mathbf{s}_i$ was transmitted,

$$\mathbf{c}_i = \frac{\int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_i} p(\mathbf{x}) d\mathbf{x}} \tag{5.5}$$

is the VQ's centroid, and

$$P_i = \int_{\Omega_i} p(\mathbf{x}) d\mathbf{x} \tag{5.6}$$

is the probability that the source signal lies in $\Omega_i$.

Equation (5.3) depicts the structure of the optimal decoder for a given $\mathbf{c}_i, \Omega_i$ and $\mathbf{s}_i$. The optimal decoder generates a signal that is a weighted sum of centroids for the VQ partitions. The weighting coefficients are exponential functions of the distance between the received signal and the modulation signal points. Because of these exponential functions, the optimal decoder in (5.3) is inherently nonlinear. In the limiting case when the channel is noiseless, i.e. when $N_0 = 0$, only the exponential function $p(\mathbf{y}|\mathbf{s}_i)$ is non-zero (assuming $s_i$ is sent). This means that $\hat{\mathbf{x}}(\mathbf{y}) = \mathbf{c}_i$ - the output one would expect for a noise-free channel.

The decoder given in Fig 5.1 can be called a soft decision decoder since it operates directly on the unquantized received signal $\mathbf{y}$. Therefore, we call the vector quantizer based on a soft decision as a *soft decision vector quantizer* (SDVQ). In contrast, a hard decision decoder consists of a decision device (or in that matter a quantizer) followed

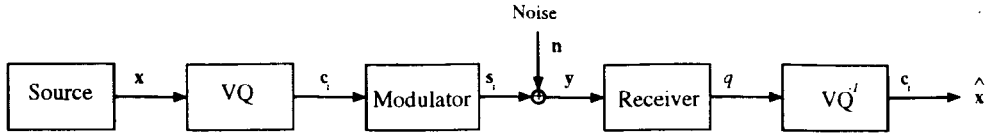


Figure 5.2: Communication system with a Hard Decision Device

by a source decoder structure in (5.3); see Fig. 5.2. However in this case, the decoder structure in (5.3) operates on the output $q$ of the decision device, rather than directly on the received signal $\mathbf{y}$. This means the pdf's $p(\mathbf{y}|\mathbf{s}_i)$, $i = 1, \ldots, N$ in (5.3) should be replaced by the conditional probabilities $P(q|\mathbf{s}_i)$'s. The hard decision decoder, though perhaps more practical, is in general inferior to the soft decision decoder. However, the degradation in performance is practically zero at the high channel signal-to-noise ratio region. This stems from the fact that at large SNR, the conditional probability $P(q = \mathbf{s}_i/\mathbf{s}_i)$ will tend to be 1 and other conditional probabilities will tend to be 0. Consequently, the output of the decoder structure in (5.3) is simply $\mathbf{c}_i$ - approximately the same output we would get from a soft decision decoder for large channel SNR. The channel optimized vector quantizer (COVQ) discussed in Chapter 4, and [22], [19] falls into the category of the hard decision decoder.

To compare SDVQ with COVQ, we note that the transition probabilities can be expressed as

$$P(j|i) = \int_{\Theta_j} p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_i)d\mathbf{y}$$

where $\Theta_j$, $j = 1, 2, ..., N$, are the decision regions in the modulation signal space. By comparing equation (5.3) with the following equation (4.6)

$$\mathbf{c}_j = \frac{\sum_{i=1}^{N} P(j|i) \int_{\Omega_i} \mathbf{x}p(\mathbf{x})d\mathbf{x}}{\sum_{i=1}^{N} P(j|i) \int_{\Omega_i} p(\mathbf{x})d\mathbf{x}},$$

it is apparent that the COVQ centroids can be obtained from the optimal decoder of our SDVQ by replacing the pdf $p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_i)$ by their average values over the decision region $\Theta_j$. The SDVQ receiver becomes equivalent to the COVQ receiver if $p_{\mathbf{n}}(\mathbf{y} - \mathbf{s}_i)$

is uniform over $\Theta_j$ and the decision regions have equal volumes.

Finally, before leaving this section, we would like to provide a suboptimal decoder for operation in the low channel SNR region. To proceed, we note that the conditional pdf $p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}_i)$ in (5.4) can be expanded as a Taylor series around the point $\mathbf{y}=\mathbf{s}_i$. Retaining only the first term in the expansion, we have

$$p_{\mathbf{n}}(\mathbf{y}-\mathbf{s}_i) \approx p_{\mathbf{n}}(\mathbf{y})(1+\frac{2}{N_0}\mathbf{s}_i^T\mathbf{y})$$

and

$$\hat{\mathbf{x}}(\mathbf{y}) \approx \frac{\sum_{i=1}^N P_i\mathbf{c}_i + (2/N_0)\sum_{i=1}^N P_i\mathbf{c}_i\mathbf{s}_i^T\mathbf{y}}{1+(2/N_0)\sum_{i=1}^N P_i\mathbf{s}_i^T\mathbf{y}}$$

where $\mathbf{s}_i^T$ is the transpose of the signal vector $\mathbf{s}_i$. Since the mean of the source signal is zero, i.e. $\sum_{i=1}^N P_i\mathbf{c}_i = 0$, and $\frac{2}{N_0}\sum_{i=1}^N P_i\mathbf{s}_i^T\mathbf{y} << 1$ for the low channel SNR, this means that $\hat{\mathbf{x}}(\mathbf{y})$ can be written as

$$\hat{\mathbf{x}}(\mathbf{y}) \approx G\cdot\mathbf{y} \tag{5.7}$$

where

$$G = \frac{2}{N_0}\sum_{i=1}^N P_i\mathbf{c}_i\mathbf{s}_i^T \tag{5.8}$$

In other words, the optimum MSE decoder is approximately linear for the low channel SNR condition. This result is attributed to Gardner [71].

## 5.2 Optimization of the VQ Partitions

We consider in this section the problem of how to design the optimal soft decision vector quantizer for a given modulation scheme and communication channel. The decoder is assumed to be the optimal decoder given in (5.3). There are two key steps in the design process

1. determine the optimal partitions $\Omega_i, i=1,2,\cdots,N$ in the VQ,

2. assigning the VQ codewords to the channel signal vectors $s_i, i = 1, 2, \cdots, N$.

Below is a detailed description of the design process.

First we show how to determine the optimal VQ partitions for a given modulation constellation and mapping from the VQ index to the modulation signal vector. The channel is assumed to be an AWGN channel and the decoder is the optimal decoder given in (5.3).

Like in the design of a source optimized VQ by the LBG algorithm [43], we assume a set of initial centroids for the VQ, $\{\mathbf{c}_i, i = 1, \cdots, N\}$, is given. Moreover, we assume that the centroid $\mathbf{c}_i$ is mapped to the modulation signal vector $\mathbf{s}_i$. This implies that after substituting (5.3) into (5.1), the average distortion $D$ can be expressed as

$$D = \frac{1}{k} \sum_{m=1}^{N} \int_{\Omega_m} \left\{ \int_n \|\mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}{\sum_{i=1}^{N} P_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})} \|^2 p_{\mathbf{n}}(\mathbf{n}) d\mathbf{n} \right\} p(\mathbf{x}) d\mathbf{x} \tag{5.9}$$

where $\Omega_m, m = 1, \cdots, N$, denotes the VQ partitions. For the VQ encoder design, we have to impose an energy constraint during the optimization. Specifically, we set the average energy as

$$E_{avg} = \sum_{m=1}^{M} P_m \|\mathbf{s}_m\|^2 \tag{5.10}$$

and the objective function is

$$D' = \frac{1}{k} \sum_{m=1}^{N} \int_{\Omega_m} \int_n \|\mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}{\sum_{i=1}^{N} P_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})} \|^2 p_{\mathbf{n}}(\mathbf{n}) p(\mathbf{x}) d\mathbf{n} d\mathbf{x}$$
$$+ \gamma \left( \sum_{m=1}^{N} P_m \|\mathbf{s}_m\|^2 - E_{avg} \right) \tag{5.11}$$

where $P_m$ is the probability that $\mathbf{s}_m$ is transmitted and $\gamma \geq 0$ is a positive Lagrange multiplier. Equation (5.11) can be rewrittern as following

$$D = \frac{1}{k} \sum_{m=1}^{N} \int_{\Omega_m} \left\{ \int_{\mathbf{n}} \|\mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}{\sum_{i=1}^{N} P_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})} \|^2 p_{\mathbf{n}}(\mathbf{n}) d\mathbf{n} + k\gamma \|\mathbf{s}_m\|^2 \right\} p(\mathbf{x}) d\mathbf{x} - \gamma E_{avg} \tag{5.12}$$

Since each inner integral term

$$I_m = \int_{\mathbf{n}} \|\mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}{\sum_{i=1}^{N} P_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}\|^2 p_{\mathbf{n}}(\mathbf{n}) d\mathbf{n} + k\gamma \|\mathbf{s}_m\|^2 \qquad (5.13)$$

in the above equation is non-negative, this implies the average distortion is a minimum if we adopt the following partitioning rule for the VQ :

$$\mathbf{x} \in \Omega_m \quad \text{if} \quad I_m \leq I_j \quad \text{for all} \quad j \neq m \qquad (5.14)$$

The partitions defined by (5.13) and (5.14) depend on the initial centroids $\mathbf{c}_1, \ldots, \mathbf{c}_N$, which in turn depend on the partitions according to the following

$$\mathbf{c}_i = \frac{\int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_i} p(\mathbf{x}) d\mathbf{x}} \qquad (5.15)$$

It should be evident at this point that the optimal centroids and partitions can be obtained, in principle, by re-iterating (5.13) and (5.15) until convergence occurs. This approach, generally known as the generalized Llyod algorithm [42], was first suggested in [43] for designing vector quantizers for the noiseless channels. Convergence is guaranteed since the distortion can either decrease or remain the same at each step of the iteration. However due to the nature of iterative optimization, the algorithm can not guarantee to generate a global optimum solution .

As we see from equation (5.13), the VQ partitions $\Omega_m, m = 1, 2, \cdots, N$, are related to the Lagrange multiplier $\gamma$. Because an analytical method to determine $\gamma$ is not available, a numerical method is adopted to find the multiplier $\gamma$. In the beginning, we assume that $\gamma$ is equal to zero, then get the VQ centroids and compute the corresponding modulation signal energy. This is followed by checking whether the energy constraint condition is satisfied. If yes, then stop; otherwise update the multiplier according to the equation: $\gamma_{new} = \gamma_{old} + \delta$, where $\delta$ is a fixed step-size, and repeat the partition of the VQ and the computation of the signal energy.

For a joint source and channel coding system, the VQ indices are mapped into the modulation signal vectors for transmission. As in Chapter 4, we use a simulated annealing algorithm to optimize the VQ indices mapping.

Once the optimal mapping is found, the design process returns to the determination of the optimal VQ partitions for the given mapping. If, as a result, there is no reduction in distortion, the design process stops. Otherwise it goes back to the optimization of the VQ index mapping.

## 5.3 Optimization of the Modulation Constellation

The distortion in (5.9) is a function of the VQ partitions $\{\Omega_m, m = 1, \cdots, N\}$ and the modulation constellation $\{\mathbf{s}_m, m = 1, \cdots, N\}$. We showed in Section 5.2 how to optimize the VQ partitions for a given modulation constellation. In this section, the problem of optimizing the modulation constellation for a given VQ is considered. We assume that the VQ partitions and the VQ-Modulation mapping are fixed when dealing with the optimization of the modulation constellation.

The average distortion for the system can be written as

$$D = \sum_{m=1}^{N} \int_{\Omega_m} \int_{\mathbf{n}} \left\| \mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}}}{\sum_{i=1}^{N} P_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}}} \right\|^2 p_{\mathbf{n}}(\mathbf{n}) p(\mathbf{x}) d\mathbf{n} d\mathbf{x} \qquad (5.16)$$

where a Gaussian pdf is substituted into equation (5.9). Once again we should consider the energy constraint, see equation (5.10):

$$E_{avg} = \sum_{m=1}^{M} P_m \|\mathbf{s}_m\|^2$$

Due to the difficulty to find the optimum modulation constellation by an analytic approach, a constrained gradient-search algorithm is again used to determine the

optimal constellation, see Chapter 4 for detailed discussion. In the current application, the algorithm will provide us with at least a local optimum solution. This stems from the fact that the distortion can never increase in each iteration and the average distortion is bounded by zero. Also, it should be noted that the partial deviative of the average distortion $D$ with respect to the modulation signal vector $\mathbf{s}_j$ is given by

$$\frac{\partial D}{\partial \mathbf{s}_j} = \sum_{m=1}^{N} \int_{\Omega_m} \int_n \left( x^T - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i^T e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}}}{\sum_{i=1}^{N} P_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}}} \right) \frac{1}{\left( \sum_{i=1}^{N} P_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}} \right)^2}$$

$$\left\{ \sum_{i=1}^{N} P_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}} (c_i - c_j) \right\} P_j e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}} (\mathbf{s}_m - \mathbf{s}_j + \mathbf{n})\, p(\mathbf{x}) p_{\mathbf{n}}(\mathbf{n}) d\mathbf{x} d\mathbf{n} +$$

$$\int_{\Omega_j} \int_n \left( \mathbf{x}^T - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i^T e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}}}{\sum_{i=1}^{N} P_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}}} \right) \frac{1}{\left( \sum_{i=1}^{N} P_i e^{-\frac{\|\mathbf{s}_m - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}} \right)^2} \tag{5.17}$$

$$\left\{ \sum_{i=1}^{N} \sum_{l=1}^{N} e^{-\frac{\|\mathbf{s}_j - \mathbf{s}_i + \mathbf{n}\|^2}{N_0}} e^{-\frac{\|\mathbf{s}_j - \mathbf{s}_l + \mathbf{n}\|^2}{N_0}} \mathbf{c}_i P_i P_l (\mathbf{s}_l - \mathbf{s}_i) \right\} p(\mathbf{x}) p_{\mathbf{n}}(\mathbf{n}) d\mathbf{x} d\mathbf{n}$$

The computation of the partitial derivative (5.17) is implemented by the Monte Carlo simulation.

## 5.4 Joint Source-Modulation Optimization and Numerical Results

We discussed in the last two sections how to optimize individually the VQ and the modulation constellation. In this section, a strategy of jointly optimizing these two subsystems will be discussed. Once again, an algorithmic approach is adopted. The algorithm used is essentially a general form of the Lloyd algorithm (GLA). For the

given modulation scheme, we first optimize the VQ partitions and the mapping according to the procedure outlined in Section 5.3, and then optimize the modulation constellation according to Section 5.4 for the given VQ partitions and mapping; Iterate these two steps until the convergence occurs.

The design procedure described above was applied to a first order Gauss Markov source. The correlation coefficient between successive samples is set to 0.9. The corresponding numerical results are shown in Figs 5.3 through Figs. 5.11 for different VQ size $N$, signal space dimension $M$, and VQ rate $R_s$, defined as

$$R_s = \frac{\log_2(N)}{k} \tag{5.18}$$

Table 5.1 summarizes the values of the various parameters used in these figures. For each set of system parameters, we also include in the same figure the upperbound on performance, as well as the performance of the corresponding reference system.

Each reference system is constructed as follows. For given $N$, $M$, and $R_s$, we first design the optimal VQ for the AR-1 source in a noiseless channel, using the GLA algorithm [43]. Then from the value of $N$ and $M$, determine the most appropriate basic QAM constellation. The Cartesian product of the basic constellation $M/2$ times with itself results in a $M$ dimensional constellation with $N$ points. For example, let $N = 256$ and $M = 4$. Then the most appropriate basis constellation is a 16QAM constellation. Alternatively, one can use 16PSK. The decoder in the reference system is assumed to perform hard decision decoding, i.e., it will determine the most likely transmitted VQ codeword from the received signal $\mathbf{y}$. It should be pointed out the simulated annealing algorithm is used to find the optimal VQ-modulation mapping for the reference system.

For comparison, we also show the bounds for the joint source and channel coding system. These bounds are derived according to the information theory, see Chapter 2

for a detail discussion. We show in Figures 5.3-5.5 that the performance of our systems as a function of the channel signal to noise ratio, $E_{avg}/N_0$, in dB. The difference in these three figures are the number of codewords, $N$, in the VQ. The performance measure is the reconstructed source signal-to-noise ratio, defined as

$$SNR_s = \frac{E(|\mathbf{x}|^2)}{E(|\mathbf{x} - \hat{\mathbf{x}}|^2)} \tag{5.19}$$

The modulation schemes in these figures all have a bandwidth efficiency of $\eta = 2$ bits/sec/Hz. The system optimization was performed at a channel SNR of 5 dB. The resultant systems were then used over the entire SNR range. The reason why we chose 5 dB SNR value as an optimization parameter is that it gives a good compromise between the performance in the high SNR and the low SNR regions. Otherwise, if the designed SNR is too high or too low, the overall distortion due to the mismatch of the channel SNR will increase. It should be pointed out the chioce of the designed channel SNR value is related to the source and channel model. For example, we select 5 dB as a design channel SNR value for a first order Gaussian-Markov source in AWGN channel. However, for the same source model but a Rayleigh fading channel, we select 17 dB as a designed channel SNR; see Chapter 7. Similarly, for the same channel but different sources, the design SNR would be different. Note that in principle we can optimize the system at every channel SNR. However this would require an adaptive system with SNR estimation at the receiver. Three observations are made :

1. The larger $N$ is, the better the synthesized signal-to-noise ratio (SSNR) achievable. The value of $N$ apparently does not have a strong influence at the lower channel SNR region.

2. The proposed systems significantly outperform the reference (or standard) systems at the lower SNR region. For example, at the channel SNR equal to 2 dB,

the gain in SSNR is roughly 2 dB in each of the 3 figures. This clearly illustrates the benefit of joint source and channel coding.

3. At low channel SNR (below 5 dB), the performance curves of our systems run roughly parallel to the bound of the joint source and channel coding system. The difference in SSNR ranges from 5-7 dB. At higher channel SNR, our performance curves deviate from the bound. This is due to the fact that only a relatively small number of codewords are used in the VQ and modulator.

Figures 5.6-5.8 differ from the previous three figures only in the rate of the source coder. Here the source rate is $R_s = 2$. We notice that increasing the source rate brings a tremendous improvement in the maximum SSNR. For example, if we compare Figs 5.3 and 5.6, we see the maximum SSNR increases by about 3 dB when the source rate is doubled.

Fig. 5.9-5.11 is similar to Figs 5.6-5.8 except that the signal space dimensions are different. In comparing these two figures, we see that at large SNR, the size of the signal space has no effect on the system performance. However, at lower SNR like 0 dB, we see that the system in Fig. 5.9 performs significantly poorer than that in Fig. 5.6. This is as expected since the signal space dimension is larger in Fig 5.6 than in Fig. 5.9.

In comparing Figs 5.4, 5.7, 5.10 and Figs 5.5, 5.8, 5.11, we can once again confirm the following findings :

1. The most effective way to increase the maximum SSNR is to increase the source rate. Increasing the signal space dimension or the number of codewords has relatively little or no improvement to the maximum SSNR.

2. Large signal space dimension is required for the low channel SNR region. Increasing the source rate has no effect for this operating region.

| Fig. No. | Codebook Size $N$ | Block Length$k$ | Source $\log_2 N/k$ | Dimension Signal Space $M$ | Bit/Dimension | $SNR_{ch}$ For Design | Ref. System |
|---|---|---|---|---|---|---|---|
| Fig. 5.3 | 16 | 4 | 1 | 4 | 1 | 5dB | 2XQPSK |
| Fig. 5.4 | 64 | 6 | 1 | 6 | 1 | 5dB | 3XQPSK |
| Fig. 5.5 | 256 | 8 | 1 | 8 | 1 | 5dB | 4XQPSK |
| Fig. 5.6 | 16 | 2 | 2 | 4 | 1 | 5dB | 2XQPSK |
| Fig. 5.7 | 64 | 3 | 2 | 6 | 1 | 5dB | 3XQPSK |
| Fig. 5.8 | 256 | 4 | 2 | 8 | 1 | 5dB | 4XQPSK |
| Fig. 5.9 | 16 | 2 | 2 | 2 | 2 | 7dB | 16-QAM |
| Fig. 5.10 | 64 | 2 | 3 | 2 | 3 | 12dB | 64-QAM |
| Fig. 5.11 | 256 | 2 | 4 | 2 | 4 | 19dB | 256-QAM |

Table 5.1: Summary of the system parameters used in the different figures. Note that the second last column are the channel SNRs at which optimization are performed.

For illustration purpose, Figs. 5.12 and 5.13 show the VQ constellations and modulation constellations for systems optimized at different channel SNRs. It is interesting to point out that at low channel SNR, the number of distinct VQ codewords in the optimized system may be smaller than the initial codebook size $N$. For example, the number of distinct VQ codewords at a channel SNR of 2 dB is only 11, although the initial codebook contained 16 vectors. This situation may be explained by the fact that the optimal system trades the clean channel performance (SSNR) for better performance in the noisy environment: for a smaller number of centroids the performance in clean channel conditions degrades, but at the same time the distances between the signal points in the modulation space increase and this approach may lead to better performance at low channel SNR. A similar observation was made earlier in [19]. When trying to increase the number of centroids by centroid splitting, the new centroids remain very close to the existing ones and the system tends to converge to the same number of centroids and same average performance as in the initial optimized design.

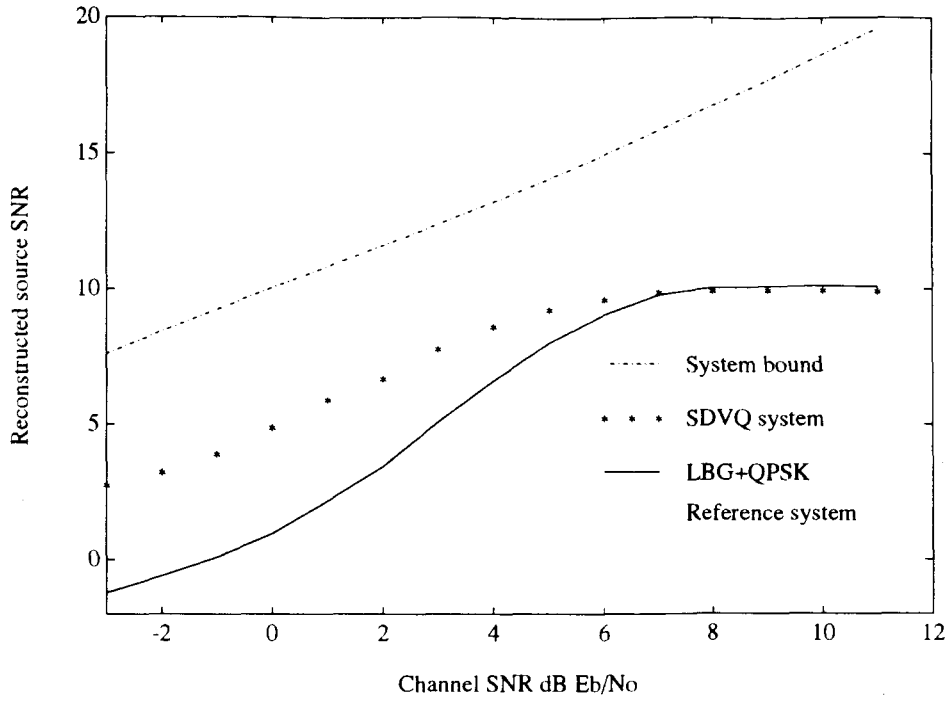


Figure 5.3: Performance of a joint source and channel coding system with $N = 16$ VQ codewords, $k = 4$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 4$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 2 X QPSK is used as the modulation format.

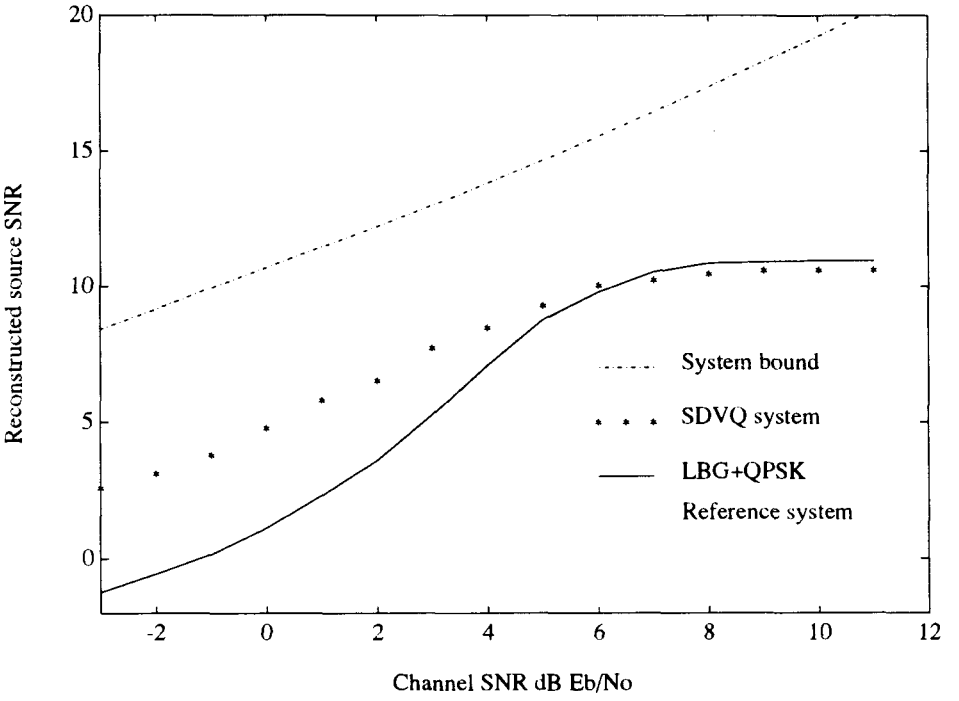


Figure 5.4: Performance of a joint source and channel coding system with $N = 64$ VQ codewords, $k = 6$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 6$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 3 X QPSK is used as the modulation format.

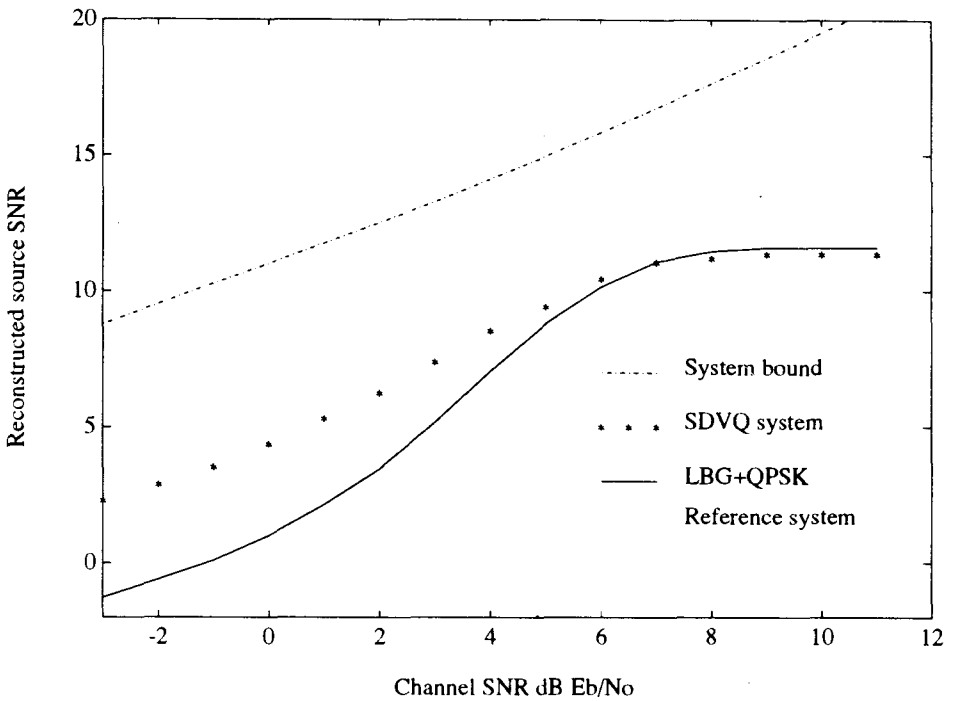


Figure 5.5: Performance of a joint source and channel coding system with $N = 256$ VQ codewords, $k = 8$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 8$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 4 X QPSK is used as the modulation format.

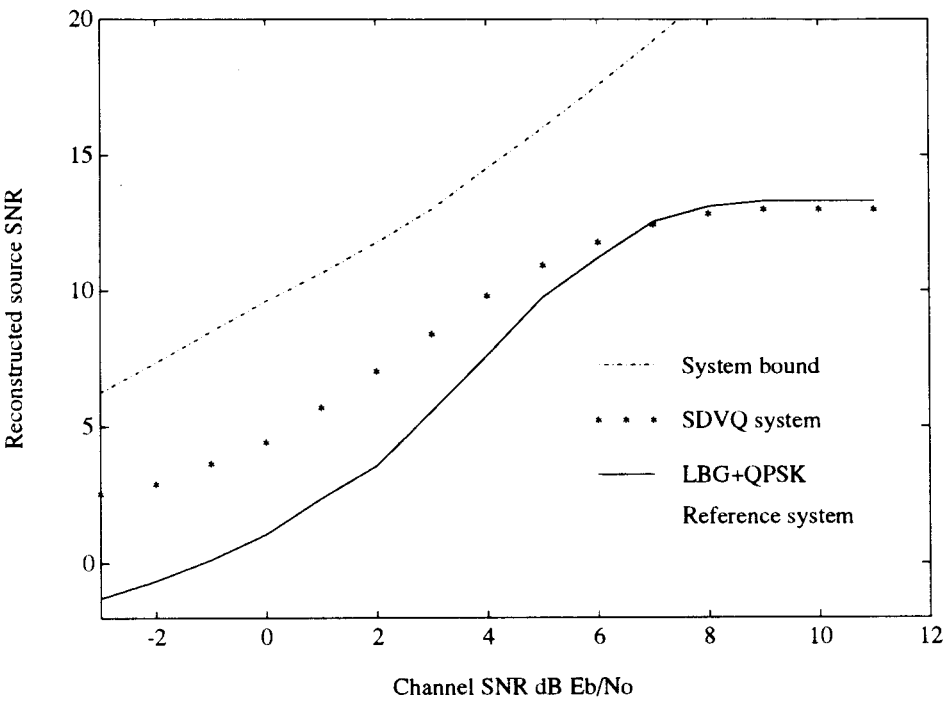


Figure 5.6: Performance of a joint source and channel coding system with $N = 16$ VQ codewords, $k = 2$ VQ dimension, a source rate of $R_s = 2$, and a signal space dimension of $M = 4$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 2 X QPSK is used as the modulation format.

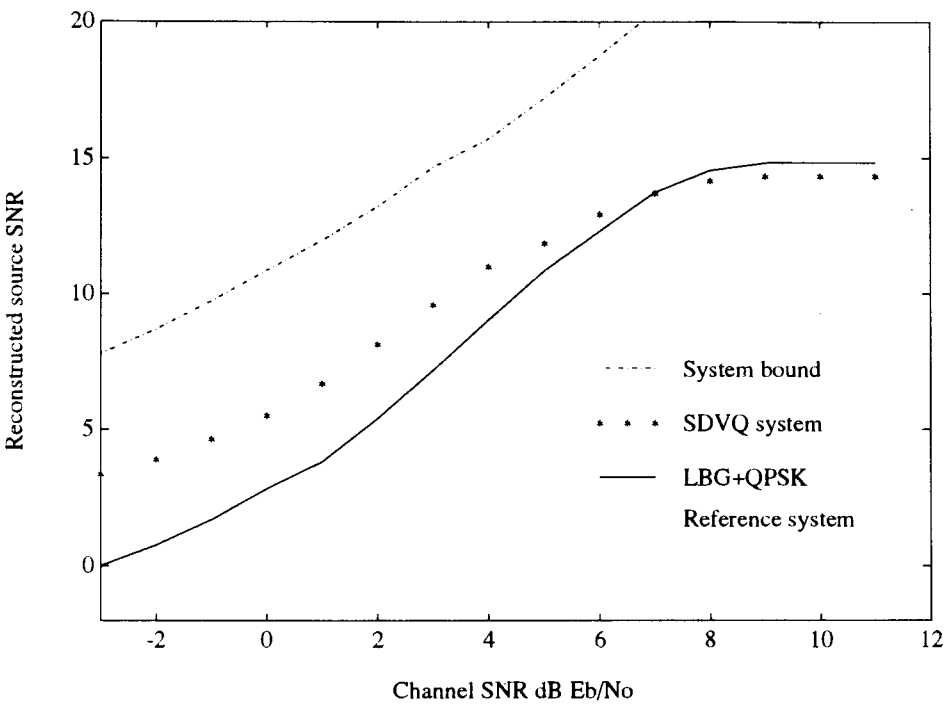


Figure 5.7: Performance of a joint source and channel coding system with $N = 64$ VQ codewords, $k = 3$ VQ dimension, a source rate of $R_s = 2$, and a signal space dimension of $M = 6$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 3 X QPSK is used as the modulation format.

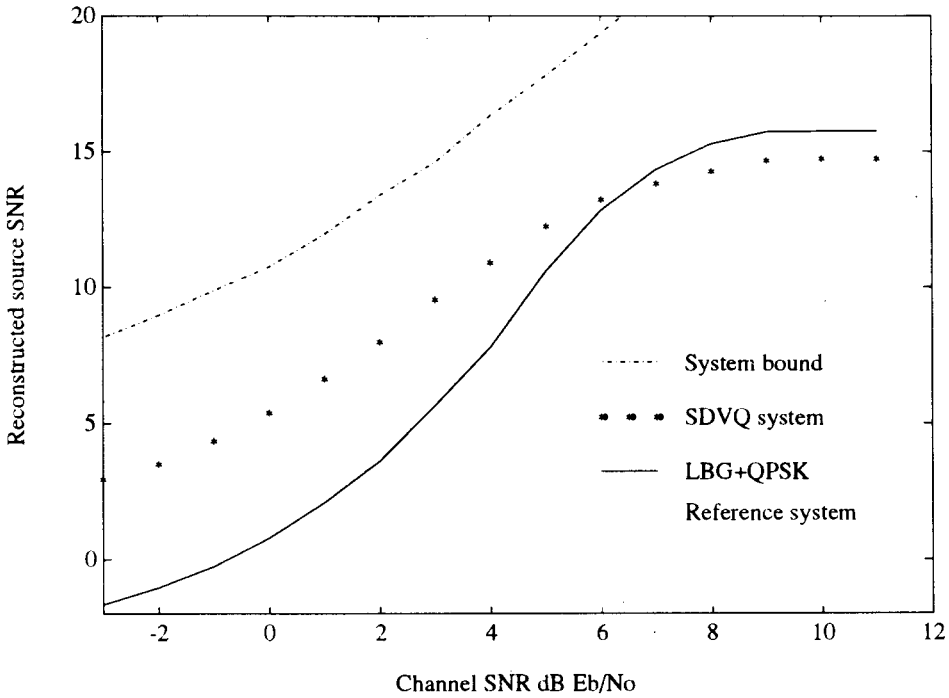


Figure 5.8: Performance of a joint source and channel coding system with $N = 256$ VQ codewords, $k = 4$ VQ dimension, a source rate of $R_s = 2$, and a signal space dimension of $M = 8$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 4 X QPSK is used as the modulation format.

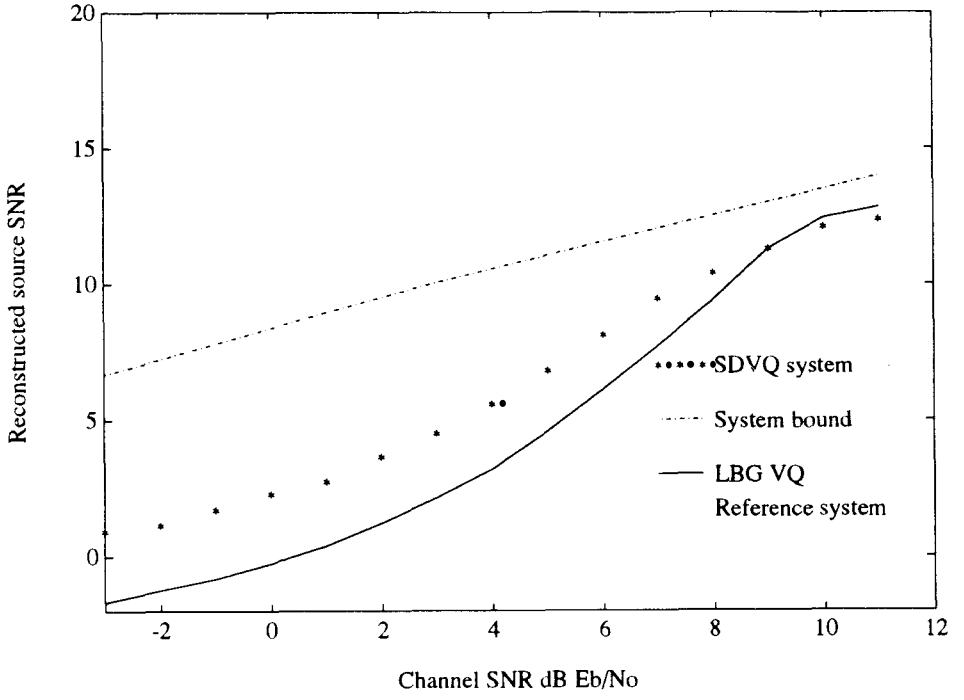


Figure 5.9: Performance of a joint source and channel coding system with $N = 16$ VQ codewords, $k = 2$ VQ dimension, a source rate of $R_s = 2$, and a signal space dimension of $M = 2$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 16-QAM is used as the modulation format.

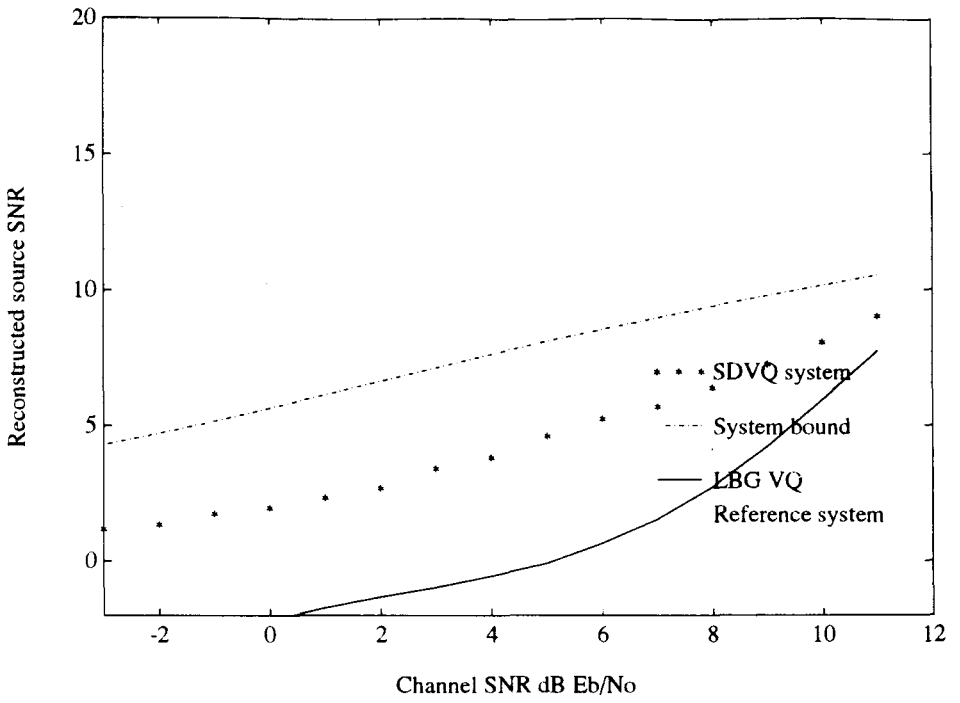


Figure 5.10: Performance of a joint source and channel coding system with $N = 64$ VQ codewords, $k = 2$ VQ dimension, a source rate of $R_s = 3$, and a signal space dimension of $M = 2$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 64-QAM is used as the modulation format.

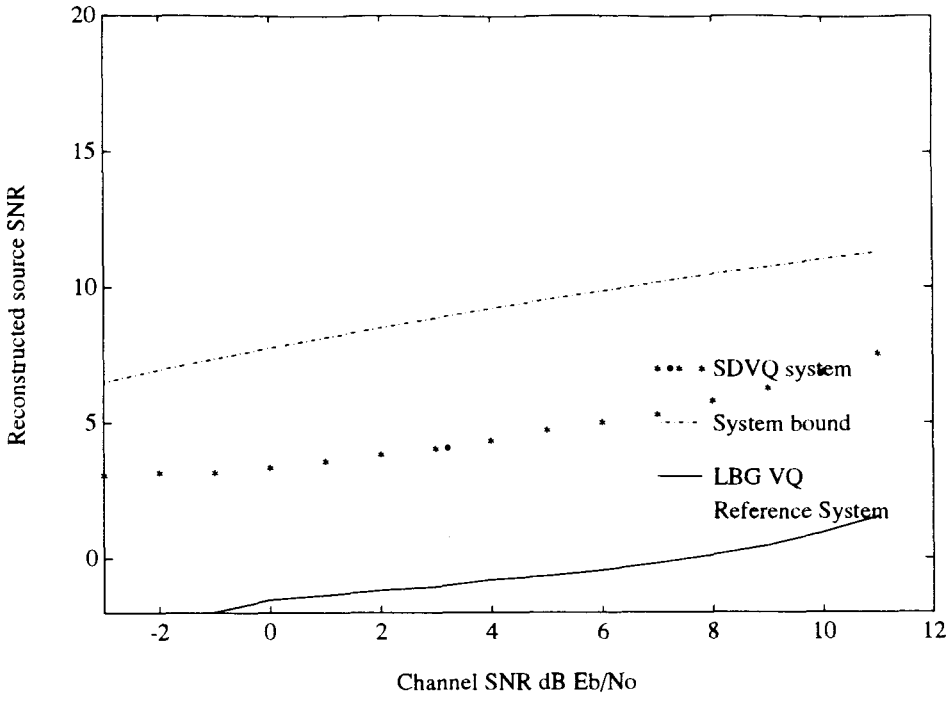


Figure 5.11: Performance of a joint source and channel coding system with $N = 256$ VQ codewords, $k = 2$ VQ dimension, a source rate of $R_s = 4$, and a signal space dimension of $M = 4$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve * * shows the simulation results for the proposed system. The solid curve shows the results for the reference system where 256-QAM is used as the modulation format.

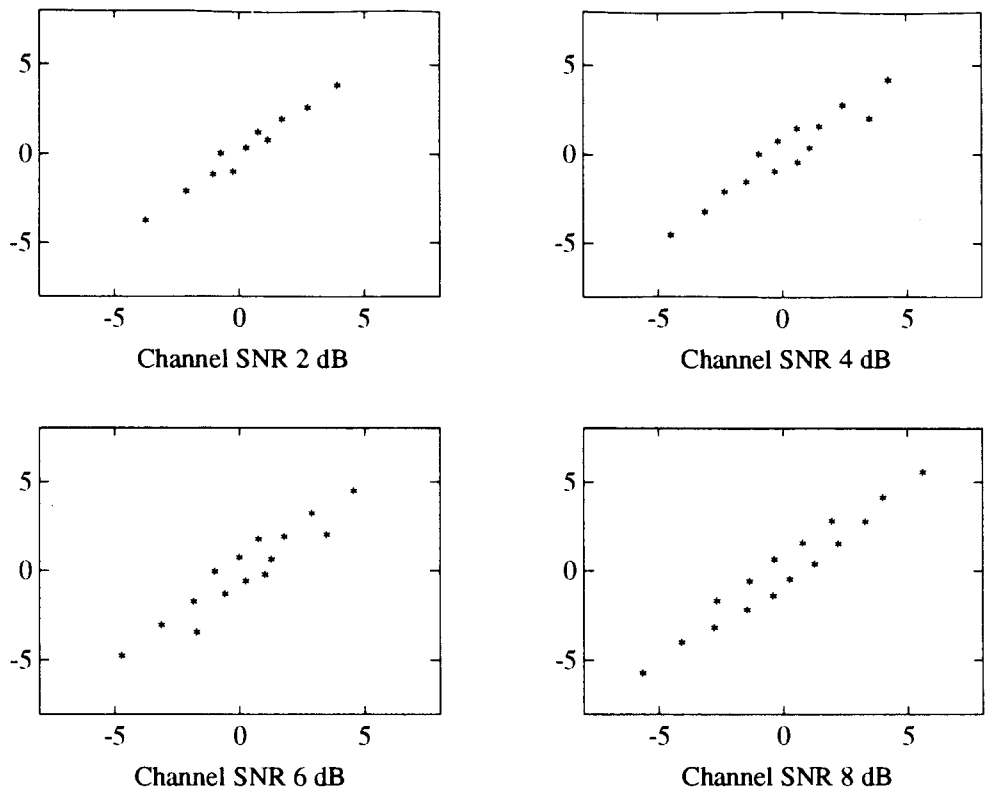


Figure 5.12: VQ codebook constellation at different channel SNR. The source is a first Gauss-Markov with a correlation coefficient of $\rho$=0.9. The various system parameters are $N = 16$, $k = 2$ and $M = 2$.

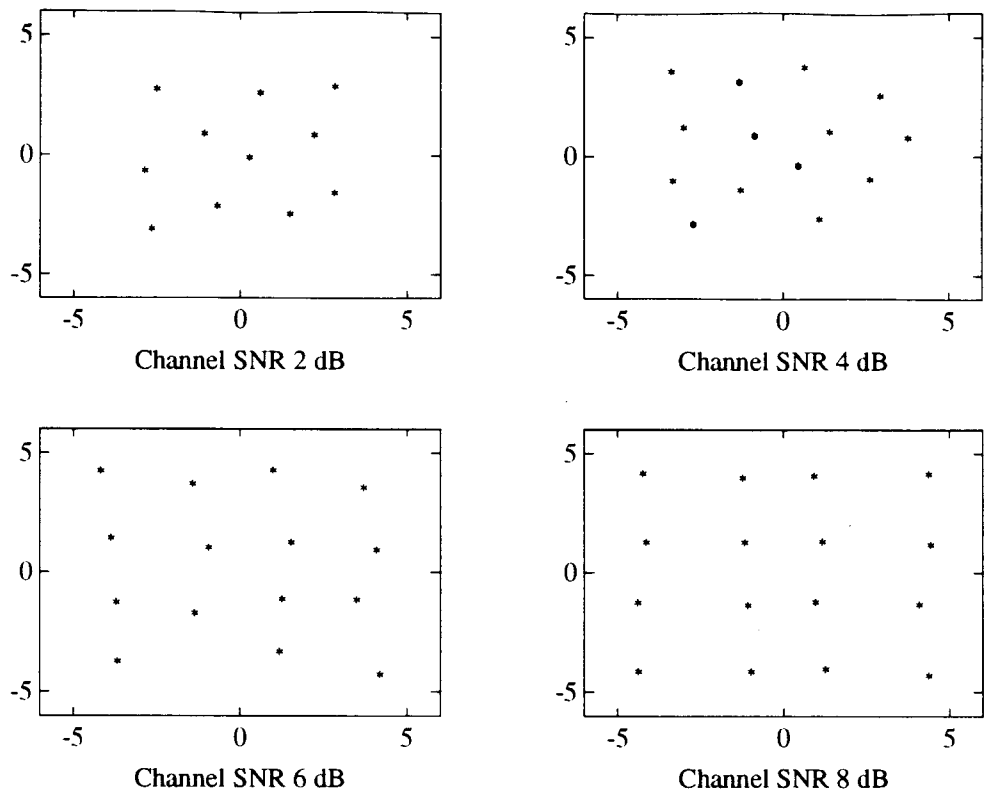


Figure 5.13: Modulation constellations at different channel SNR. The system is the same one in Fig. 5. 12.

We compare the performance of the SDVQ system with the linear receiver system [44] in Figs. 5.14-5.15. It should be pointed out the both two systems are optimized at the exactly channel SNR value. Fig 5.14 is corresponding to the case of the VQ dimension $k$=8, the modulation signal dimension $M$=8, and the source rate $R_s$=1. Fig 5.15 is similar to Fig.14, except that the dimensionality of the modulation signal dimension $M$ is equal to 4. We can see that the performance of our SDVQ system is almost identical to the linear receiver system in the low channel SNR region. However, the SDVQ system is superior to the linear receiver system about 1 dB in the synthesized source SNR in the high channel SNR region. This is due to that the linear receiver is just an approximate implementation of the optimum receiver at the low

SNR channel.

Finally let us look at the performance of the SDVQ and the linear receiver systems in the case of the channel SNR mismatch in Fig. 5.16 and 5.17. Both systems are optimized at a 8 dB channel SNR and the resulting systems operate under different channel SNR conditions. We can see that the SDVQ system is significantly better than the linear receiver system. For example, see Fig. 5.17, when the channel SNR is equal to 10 dB, the SDVQ system is over the linear receiver system about 4 dB in the synthesized source SNR.

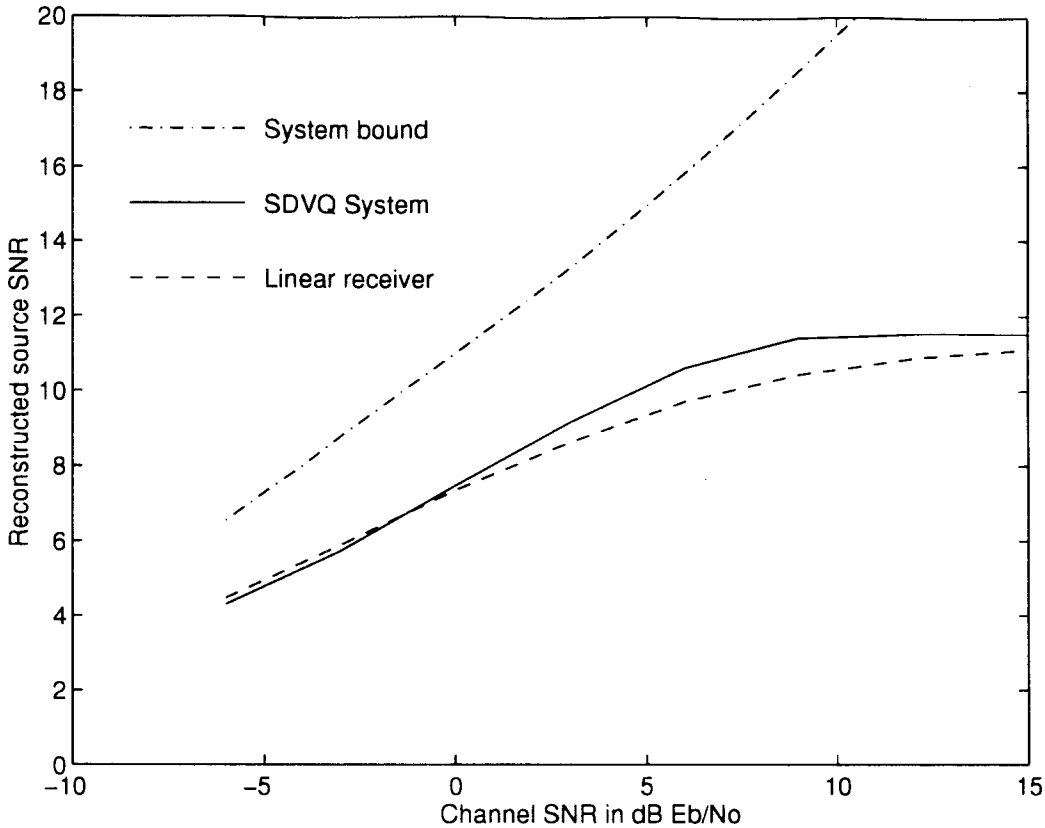


Figure 5.14: Performance comparison of the proposed system with the linear receiver system. Both system are optimized at the exact channel SNR. $N = 256$ VQ codewords, $k = 8$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 8$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve - - shows the simulation results for the proposed system. Solid line is corresponding to the linear receiver system.

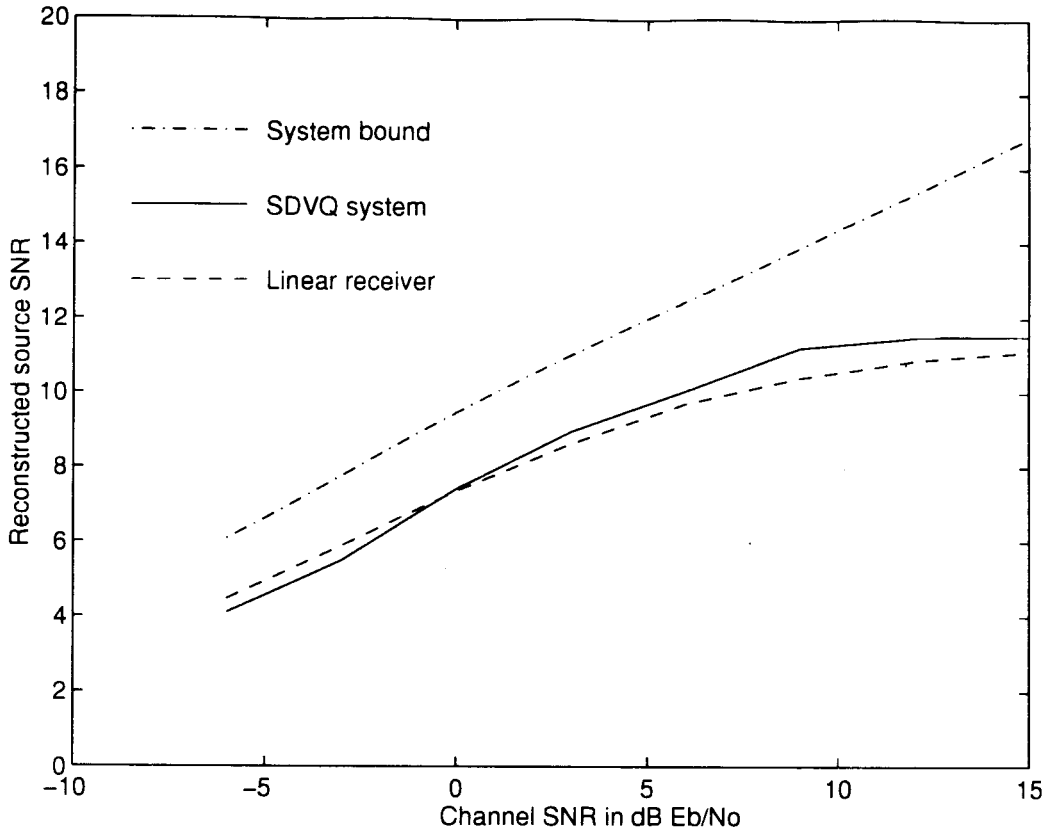


Figure 5.15: Performance comparison of the proposed system with the linear receiver system. Both system are optimized at the exact channel SNR. $N = 256$ VQ codewords, $k = 8$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 4$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve - - shows the simulation results for the proposed system. Solid line is corresponding to the linear receiver system.

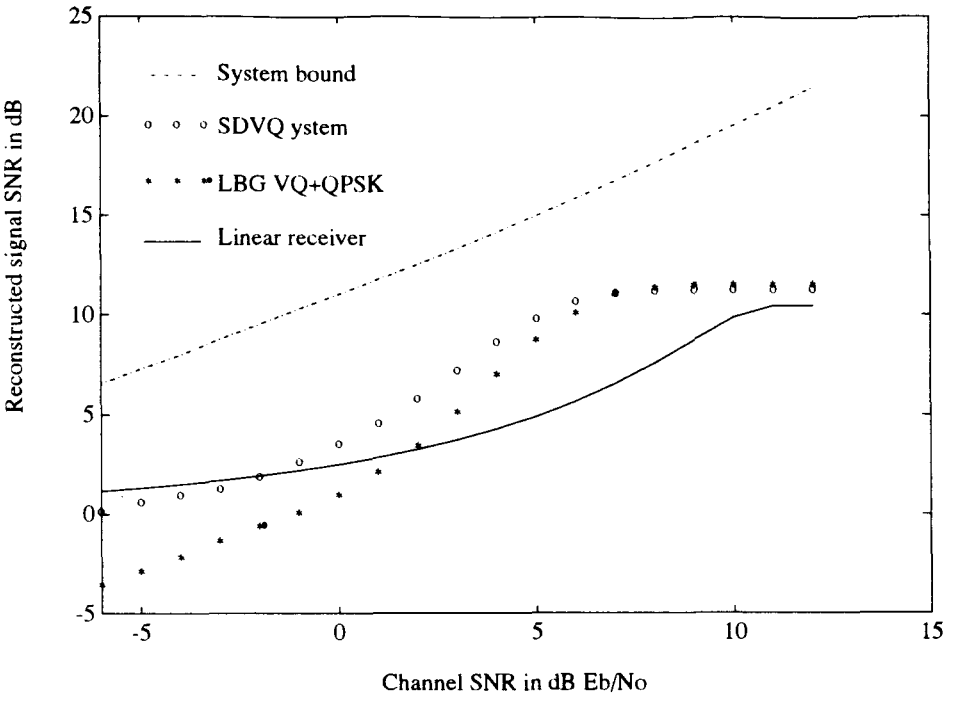


Figure 5.16: Performance comparison of the proposed system with the linear receiver system. Both system are optimized at a particular channel SNR. $N = 256$ VQ codewords, $k = 8$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 8$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve *oo* shows the simulation results for the proposed system. Curve * * represents the performance of the reference system. Solid line is corresponding to the linear receiver system.

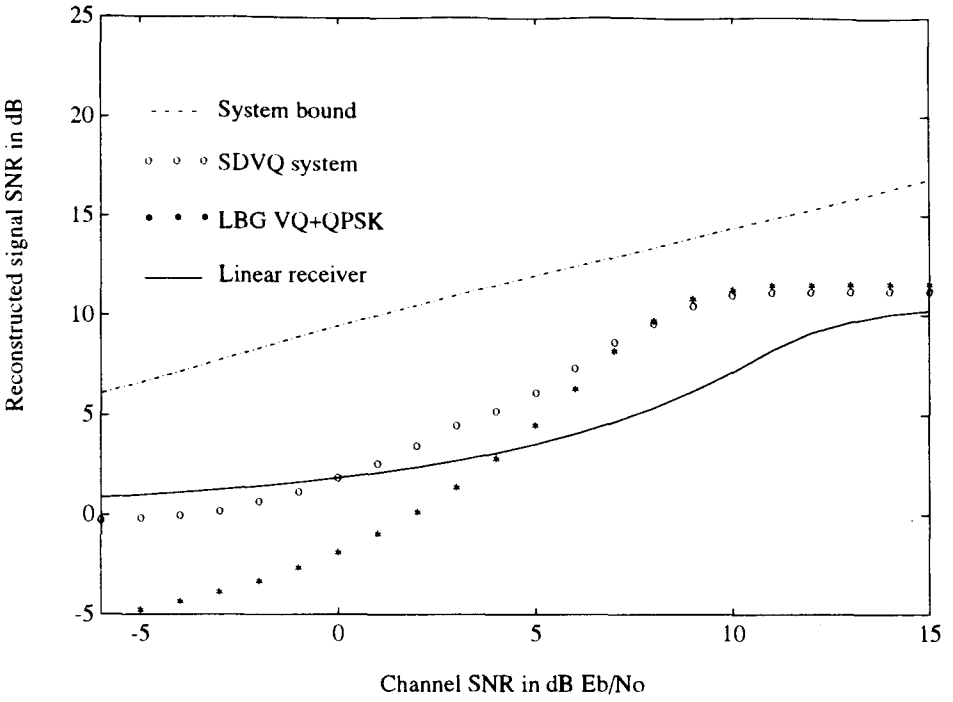


Figure 5.17: Performance comparison of the proposed system with the linear receiver system. Both system are optimized at a particular channel SNR. $N = 256$ VQ codewords, $k = 8$ VQ dimensions, a source rate of $R_s = 1$, and a signal space dimension of $M = 4$. Curve $- \cdot -$ represents the upper bound predicted by the rate distortion theory. Curve *oo* shows the simulation results for the proposed system. Curve * * represents the performance of the reference system. Solid line is corresponding to the linear receiver system.

## 5.5 Summary

We have presented in this chapter an iterative procedure to joint optimize source and channel coding in terms of the criterion of the minimum mean-square error (MSE) between the original and the reconstructed source signals. We applied this procedure to design a combined codec for a first order Gauss-Markov source. The results indicate that our codec significantly outperforms a conventional system. In addition, it is observed that at low channel SNR, a modulation signal space with a large signal

dimension is needed for satisfactory performance. On the other hand for high channel signal-to-noise ratio, the source rate should be increased for better performance.

# Chapter 6

# Sequential Reconstruction of Vector Quantized Signals

We described a symbol by symbol soft decision decoding algorithm that jointly optimizes the VQ and the modulation signal set in Chapter 5. However, if the VQ output indices are correlated, a block decoding technique can be used to further improve the system performance.

Ideally, a VQ or source encoder should remove all the redundancy in the source. However, due to the constraints such as the encoding delay and VQ complexity, successive VQ indices are often correlated. In particular, if the source already has a certain redundancy, the receiver could take advantage of this residual redundancy to reduce the effect of the channel noise. The problem thus is how to design such a receiver.

This chapter presents an optimal sequential decoding scheme for a joint source and channel coding system operating in the AWGN channel. The minimum mean-square error (MSE) between the original and the reconstructed source signals is used as the optimality criterion. The system being investigated consists of a vector quantizer (VQ) whose output indices are mapped directly into points in the modulation signal

space. The modulation signal is then transmitted over an AWGN channel. A sequential decoder based on the Bayesian estimate is used to reconstruct the source signal from the received signal samples. A recursive algorithm for implementing the block soft decision receiver is introduced. Compared to the symbol by symbol decoding technique presented in Chapter 5, it is found that sequential decoding significantly improves the system performance when transmitting from a correlated source.

This chapter is organized into five sections. Section 5.1 provides a brief review of previous work on detecting a Markov source based on the hard decision receiver, and a study of the residual redundancy at the output of the source coder. By modeling the VQ output as a first order Markov process, we derive in Section 5.2 a block decoder under the MSE criterion. In section 5.3, a recursive signal reconstruction procedure is presented to implement block decoding. The simulation results on the performance of the sequential decoding technique are presented in Section 5.4. Finally, a summary of this chapter is given in Section 5.5.

## 6.1 Residual Redundancy

The decoder described in chapter 5 actually performs symbol by symbol soft decision decoding, i.e., as soon as a signal sample is received, the decoder would reconstruct a corresponding source signal sample, based solely on that particular received sample. Such a decoder works well for a memoryless source; however, for sources with memory, the performance can be further improved. Even if a very sophisticated compression technique is used to try to remove the source redundancy, there will be some residual correlation at the output of the source coder at different time instances. This residual redundancy can be used by the receiver to provide error protection without bandwidth expansion.

By modeling the residual correlation as a first order Markov process, several authors have considered the problem of detecting a correlated source over a noisy channel. Devore [78] considered detecting a binary Markov source through a binary symmetrical channel and gave necessary conditions for the optimal detectors. Sayood and Borkenhagen [79] considered the problem of detecting a discrete Markov source over a memoryless noisy channel in a joint source-channel DPCM image coding system . Phamdo and Farvardin [80] considered the instantaneous maximum a posteriori (MAP) detection of a discrete Markov source over noisy channels and its application to the combined source-channel coding system. The main difference between [79] and [80] is that the detector in [79] is suboptimum, hence, it is not a MAP detector. In [79] only sequential detection was considered, i.e., the determination of the most probably transmitted sequence given an observed sequence, while the objective of instantaneous MAP detection is to determine the most probably transmitted symbol given the observations up to the current time. More recently, Gerlach developed a hard decision receiver for combined speech extrapolation and error detection based on maximum a posteriori probability of the source parameters [81]. The optimal receiver structures resulting from the above papers were developed based on the signal detection theory. However, as we proved in chapter 4, the optimal receiver under the MSE criterion should make the conditional estimate. We first present a block decoding algorithm to exploit the residual redundancy, then develope a recursive algorithm to implement the optimal soft decision block decoding, which constitutes the main contribution presented in this chapter.

Let us first review briefly the structure of the optimal symbol by symbol decoder for a combined source and channel coding system. The communication system model is shown in Fig. 6.1. Assume that $\mathbf{x}$ is a $k$-dimensional vector derived from a random source. The vector $\mathbf{x}$ is mapped into a VQ codeword chosen from the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$.

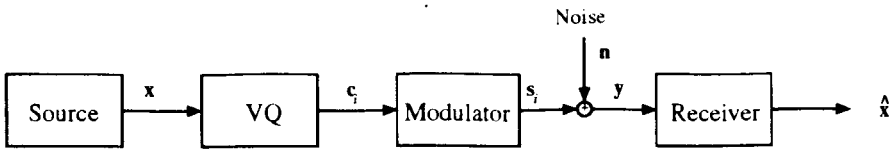


Figure 6.1: System model for symbol by symbol decoding

Assuming $\mathbf{c}_i$ was chosen, then its index $i$ is mapped directly into the modulation signal point $\mathbf{s}_i$. The set $\{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ forms the signal constellation.

In the channel, the transmitted signal is combined with additive Gaussian noise $\mathbf{n}$. Given the received signal $\mathbf{y} = \mathbf{s}_i + \mathbf{n}$, the decoder should provide the "best" estimate of the original signal $\mathbf{x}$. The soft-decision detector provides the estimate

$$\hat{\mathbf{x}}(\mathbf{y}) = \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p(\mathbf{y}|\mathbf{s}_i)}{\sum_{i=1}^{N} P_i p(\mathbf{y}|\mathbf{s}_i)} \tag{6.1}$$

The decoder described by (6.1) actually performs a symbol by symbol soft decision decoding. Due to practical constraints such as the encoding delay and the VQ complexity, the VQ indices at different time intervals are often correlated. This redundancy can be modeled as a first order Markov process. For demonstration purposes, we investigate the transition probability

$$P(\mathbf{c}_i|\mathbf{c}_j) = Prob(\mathbf{c}_i \text{ at time m} \,|\, \mathbf{c}_j \text{ at time m-1}) \tag{6.2}$$

of the VQ codebook. The VQ codebook is designed by the SDVQ, see Section 5.2, which is a modified version of the Generalized Llyod Algorithm (GLA). As input data we used a first order Gauss-Markov source as well as the line spectral pair (LSP) parameters of a speech signal. The simulation results are shown in Fig.6.2 and 6.3 for different VQ parameters.

The two planar axes represent the VQ's indices and the vertical axis represents the transition probability. If the source is memoryless, the transition probability should

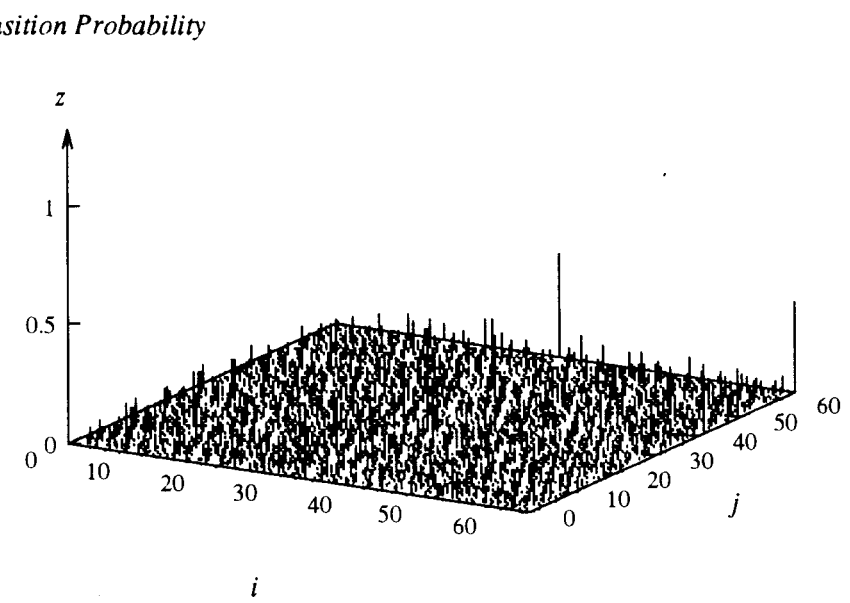


Figure 6.2: Transition probability of a VQ codebook for an AR(1) source with coefficient $\rho$=0.9. Codebook size is $N$=64, and the vector dimension $k$=6.

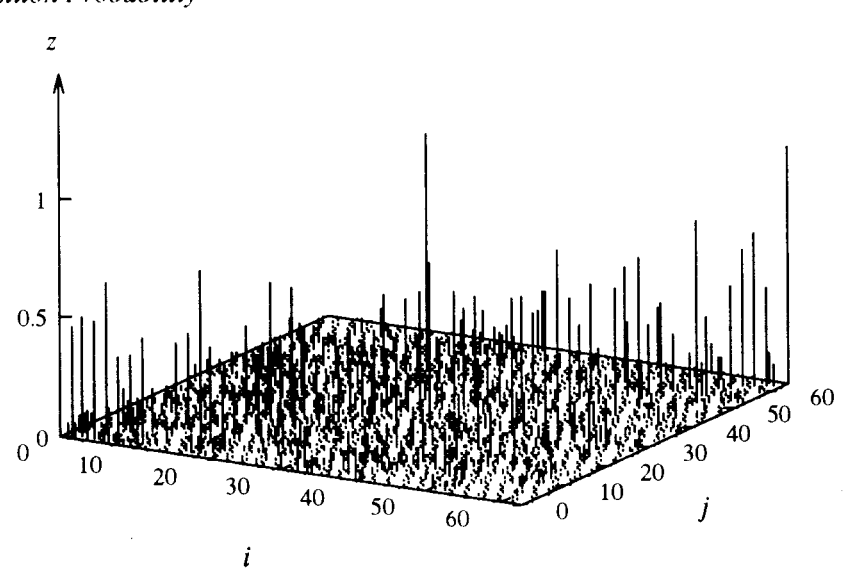


Figure 6.3: Transition probabilities of the first stage in a 4-stage VQ codebook for the LSP parameters of speech. Number of stages is 4. The codebook size is 64 for each stage, and the vector dimension is 10.

not depend on the previous index, i.e,

$$P(\mathbf{c}_i|\mathbf{c}_j) = P(\mathbf{c}_i) \tag{6.3}$$

which means that the transition probabilities along the index $j$ axis should be equal to each other. It is clear from the figures that the vertical readings are not uniform and this implies that redundancy is present at the VQ's output. The next problem is how to model this redundancy. In the previous studies [79] and [80], the VQ output was usually modeled as a first order Markov process, since it is a good representation of the source model and easy to perform mathematic analysis. In this thesis, we adopt the same assumption for the VQ source output. It will be demonstrated later in this chapter that this redundancy can be exploited at the receiver to provide error protection against the channel noise.

## 6.2 The Block Decoder

We will derive the optimal block decoding algorithm under the MSE criterion. Referring

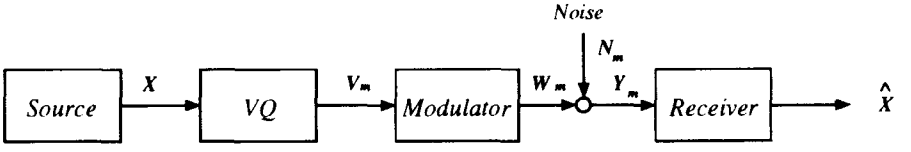


Figure 6.4: System model for block decoding

to Figure 6.4, in order to describe the sequential decoding process, we need to introduce a time index to describe the received signals and the reconstructed source signals at different instants. Let us denote the modulator output at time $m$ by $\mathbf{w}_m$, and the corresponding noise vector, and received vector by $\mathbf{n}_m$, and $\mathbf{y}_m$ respectively. In addition, let the VQ codeword and the reconstructed source signal corresponding to $\mathbf{w}_m$ be denoted by $\mathbf{v}_m$ and $\hat{\mathbf{x}}_m$. It should be clear that $\mathbf{v}_m \in \{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$. and

$\mathbf{w}_m \in \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$. Now if we define

$$\begin{aligned} \mathbf{W_L} &= (\mathbf{w}_1, \ldots, \mathbf{w}_L), \\ \mathbf{N_L} &= (\mathbf{n}_1, \ldots, \mathbf{n}_L), \\ \mathbf{Y_L} &= (\mathbf{y}_1, \ldots, \mathbf{y}_L), \\ \mathbf{V_L} &= (\mathbf{v}_1, \ldots, \mathbf{v}_L), \\ \hat{\mathbf{X}}_\mathbf{L} &= (\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L), \end{aligned} \tag{6.4}$$

where $L$ is the block length, then the optimal block decoder computes the conditional mean of the source given the received signal, i.e.,

$$\hat{\mathbf{X}}_L(\mathbf{Y}_L) = \frac{\sum_{\mathbf{V}_L} P(\mathbf{V}_L) p(\mathbf{Y}_L|\mathbf{V}_L)\mathbf{V}_L}{\sum_{\mathbf{V}_L} P(\mathbf{V}_L)\, p(\mathbf{Y}_L|\mathbf{V}_L)} \tag{6.5}$$

where $P(\mathbf{V}_L)$ is the probability of the sequence of centroids $(\mathbf{v}_1, \ldots, \mathbf{v}_L)$, and $p(\mathbf{Y}_L|\mathbf{V}_L)$ is the joint pdf of all the $\mathbf{y}_L$'s conditioned on the source sequence $(\mathbf{v}_1, \ldots, \mathbf{v}_L)$. Note that the summations in the above equation are taken over all possible $L$-length source sequences. Comparing with the symbol by symbol decoder in (6.1), we see immediately that (6.5) is simply a more general formulation of (6.1).

Because of the summations in (6.5), signal reconstruction by the block decoder appears to be very complex. A brute force approach would require the decoder to compute $N^L$ terms for each of the summations, where $N$ is the codebook size. Clearly this is not feasible for any VQ of practical interest. Fortunately, by modeling the VQ's output as a first order Markov process, the computation of (6.5) can be significantly simplified. To develop a sequential decoding algorithm, let us first express the probability $P(\mathbf{V_L})$ as

$$P(\mathbf{V}_L) = P(\mathbf{v}_1) \prod_{i=2}^{L} P(\mathbf{v}_i|\mathbf{v}_{i-1}) \tag{6.6}$$

where $P(\mathbf{v}_i|\mathbf{v}_{i-1})$ is the transition probability from the previous state $i-1$ to the

current state $i$. The conditional pdf $p(\mathbf{Y}_L|\mathbf{V}_L)$ can be expressed as

$$p(\mathbf{Y}_L|\mathbf{V}_L) = p(\mathbf{y}_1|\mathbf{v}_1)\prod_{i=2}^{L} p(\mathbf{y}_i|\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i), \tag{6.7}$$

If no coding is introduced in the modulation mapping,

$$p(\mathbf{y}_i|\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = p(\mathbf{y}_i|\mathbf{v}_i)$$

we obtain a common modulation mapping strategy. But we can introduce a first order conditional mapping for the modulation

$$p(\mathbf{y}_i|\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i) = p(\mathbf{y}_i|\mathbf{v}_{i-1}, \mathbf{v}_i) \tag{6.8}$$

which means that the received signal is related not only to the current input $\mathbf{v}_i$ but also to the previous input $\mathbf{v}_{i-1}$. Equation (6.8) describes an optimal modulation mapping rule for the Markov source, which suggests that a coded modulation should be adopted for the source with memory. In a traditional modulation system, a fixed modulation or mapping is used to transmit a VQ index. This strategy is fine if the VQ indices are independent to each other, i.e., no redundancy exits at the output of source coder. However, as shown in Fig. 6.2 and Fig. 6.3, redundancy does exist even though a sophisticated source compression scheme such as VQ is used. The redundancy can be used to design a coded modulation scheme. Therefore, the optimal modulation scheme should be a conditional mapping. The optimal mapping rule would thus have a trellis structure. It is quite interesting to note that we could use the trellis coded modulation (TCM) for a first order Markov source. It is quite clear that the TCM can be derived from the transition probability of the Markov source.

## 6.3 The Sequential Decoding Algorithm

Now we want to develop a sequential decoding algorithm starting with equation (6.5) by exploiting the properties of a first order Markov source. The combination of

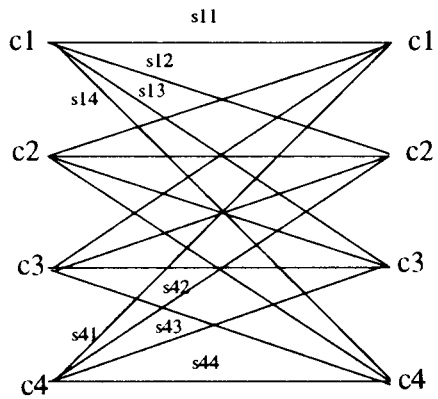


Figure 6.5: Trellis Coded Modulation for a first order Markov process

the block decoder and the recursive procedure leads to a sequential decoder. Since equation (6.5) is quite complicated, we will develop a recursive algorithm for the numerator and denominator separately. Note that the denominator of equation (6.5) can be written as

$$\begin{aligned}\alpha_m &= \sum_{\mathbf{V}_m} P(\mathbf{V}_m)p(\mathbf{Y}_m|\mathbf{V}_m) \\ &= \sum_{\mathbf{v}_m}\sum_{\mathbf{V}_{m-1}} P(\mathbf{v}_{m-1})p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1})P(\mathbf{V}_{m-1})p(\mathbf{Y}_{m-1}|\mathbf{V}_{m-1})\end{aligned} \tag{6.9}$$

Furthermore, let us define

$$f(\mathbf{v}_m) = \sum_{\mathbf{V}_{m-1}} P(\mathbf{v}_{m-1})p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1})P(\mathbf{V}_{m-1})p(\mathbf{Y}_{m-1}|\mathbf{V}_{m-1})$$

then we have

$$\begin{aligned}f(\mathbf{v}_m) &= \sum_{\mathbf{v}_{m-1}} P(\mathbf{v}_m)p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1}) \\ &\quad \sum_{\mathbf{V}_{m-2}} P(\mathbf{v}_{m-1}|\mathbf{v}_{m-2})p(\mathbf{y}_{m-1}|\mathbf{v}_{m-1},\mathbf{v}_{m-2})P(\mathbf{V}_{m-2})p(\mathbf{Y}_{m-2}|\mathbf{V}_{m-2}) \\ &= \sum_{\mathbf{v}_{m-1}} f(\mathbf{v}_{m-1})P(\mathbf{v}_m|\mathbf{v}_{m-1})p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1})\end{aligned} \tag{6.10}$$

and

$$\alpha_m = \sum_{\mathbf{v}_m} f(\mathbf{v}_m), \tag{6.11}$$

Note that the numerator of equation (6.5) is the vector:

$$\begin{aligned}\mathbf{b}(\mathbf{V}_m) &= \sum_{\mathbf{V}_m} P(\mathbf{V}_m)p(\mathbf{Y}_m|\mathbf{V}_m)\mathbf{V_m} \\ &= \sum_{\mathbf{v}_m}\sum_{\mathbf{V}_{m-1}} P(\mathbf{v}_{m-1})p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1})P(\mathbf{V}_{m-1})p(\mathbf{Y}_{m-1}|\mathbf{V}_{m-1})(\mathbf{V}_{m-1},\mathbf{v}_m)\end{aligned} \tag{6.12}$$

Decomposing the equation (6.12), we get

$$\mathbf{b}_j(\mathbf{v}_m) = \begin{cases} \sum_{\mathbf{V}_{m-1}} \mathbf{v}_j P(\mathbf{v}_m|\mathbf{v}_{m-1})p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1})P(\mathbf{V}_{m-1})p(\mathbf{Y}_{m-1}|\mathbf{V}_{m-1}) & j < m \\ \mathbf{v}_m f(\mathbf{v}_m) & j = m \end{cases} \tag{6.13}$$

In the following, we develop a recursive formula for $\mathbf{b}_j(\mathbf{v}_m)$. As we know

$$\begin{aligned}\mathbf{b}_j(\mathbf{v}_m) = &\sum_{\mathbf{v}_{m-1}} P(\mathbf{v}_m|\mathbf{v}_{m-1})p(\mathbf{y}_m|\mathbf{v}_m,\mathbf{v}_{m-1}) \\ &\sum_{\mathbf{V}_{m-2}} \mathbf{v}_j P(\mathbf{v}_{m-1}|\mathbf{v}_2)p(\mathbf{y}_{m-1}|\mathbf{v}_{m-1},\mathbf{v}_{m-2})P(\mathbf{V}_{m-2})p(\mathbf{Y}_{m-2}|\mathbf{V}_{m-2})\end{aligned}$$

Realizing that the last term is equal to $\mathbf{b}_j(\mathbf{v}_{m-1})$ for $j = 1, 2, \cdots, m-2$, we obtain the following recursive formula

$$\mathbf{b}_j(\mathbf{v}_m) = \begin{cases} \sum_{\mathbf{v}_{m-1}} \mathbf{b}_m(\mathbf{v}_{m-1}) P(\mathbf{v}_m|\mathbf{v}_{m-1}) p(\mathbf{y}_m|\mathbf{v}_m, \mathbf{v}_{m-1}) & j < m \\ \mathbf{v}_m f(\mathbf{v}_m) & j = m \end{cases} \tag{6.14}$$

Subsequently, we can express the components of the vector $\mathbf{b}(\mathbf{V}_m)$ as

$$\mathbf{b}_{j,m} = \sum_{\mathbf{v}_j} \mathbf{b}_j(\mathbf{v}_m) \qquad j = 1, 2, \ldots, m \tag{6.15}$$

Furthermore, it is straightforward to express the reconstructed signal vectors $\hat{x}_1, \ldots, \hat{x}_L$ for an observation window of $L$ in the form:

$$\hat{\mathbf{x}}_m^{(L)} = \frac{\mathbf{b}_{L,m}}{\alpha_m} \tag{6.16}$$

where the superscript $L$ is deliberately introduced to emphasize the dependency of the reconstructed signal vectors on the block length $L$.

The VQ output $\mathbf{v}_m$ represents a state in the Markov process, there are $N$ such state variables at any time $m$. Correspondingly, the term $f(\mathbf{v}_m)$ is a function of $N$ state variables $\mathbf{v}_m$. The update of $f(\mathbf{v}_m)$ at any time $m > 1$ requires altogether $2N^2$ multiplications; see second line in (6.10). At time $m = 1$, there are only $N$ multiplications required. Similarly, the term $h_j(\mathbf{v}_m)$ is also a function of the state variable $\mathbf{v}_m$. Since the index $j$ can vary from 1 to $m$, this means at time $m$, there are altogether $N$ values for $\mathbf{v}_m$. The updating of $f(\mathbf{v}_m)$ and $h_j(\mathbf{v}_m)$ at time $m$ requires altogether $(m-1)N^2 + N$ multiplications. Thus in summary, for a block length of $L$, the updating of the $f(\mathbf{v}_m)$'s and the $h_j(\mathbf{v}_m)$'s requires a total of $N(L+1) + (L-1)(L+4)N^2/2$ multiplications. For large $N$, this is approximately equal to $(NL)^2/2$ multiplications for decoding $L$ symbols, or equivalently a complexity of $LN^2/2$ per symbol. For the direct block decoding in terms of the equation (6.5), the VQ index $\mathbf{v}_m$ has $N$ states for each time interval $m$, therefore, the total number of states for $L$ time

intervals is $N^L$. The number of the multiplications for brute-force block decoding is about $2N^L/L$. We can see that a significant reduction in complexity is obtained by applying the sequential decoding algorithm: the complexity is reduced from $2N^L/L$ to $LN^2/2$.

Since both the decoding delay and the complexity increase linearly with $L$, an excessively large block size is clearly not desirable, nor is it necessary. Recall that the VQ output can be modeled as a first order Markov process and each VQ index represents a state of the Markov processes. According to the property of the Markov processes, the next state is dependent only on the present state. On other hand, for a homogeneous Markov process, the transition probability $P(\mathbf{c}_i/\mathbf{c}_j)$ is independent of the time in which the transition occurs. Therefore the state transition matrix at time $m$, $\mathbf{P}(m)$, can be expressed as

$$\mathbf{P}(m) = \mathbf{P}(1)^m \tag{6.17}$$

where $\mathbf{P}(1)$ is the single-step (conditional) transition probability matrix, defined as

$$\mathbf{P}(1) = \begin{vmatrix} P(1/1) & P(2/1) & \cdots & P(N/1) \\ P(1/2) & P(2/2) & \cdots & P(N/2) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ P(1/N) & P(2/N) & \cdots & P(N/N) \end{vmatrix}$$

and the $i,j$-th element of $\mathbf{P}(1)$ represents the probability of transferring at any time $m$ from state $i$ to state $j$. As we know, a homogeneous Markov chain can reach a steady-state probability matrix $\mathcal{P}$ after many transitions [82].

$$\lim_{m\to\infty} \mathbf{P}(1)^m = \mathcal{P} \tag{6.18}$$

For our case, we can compare the norm of the difference of two consecutive transition probability matrices and determine how long time the system needs to enter the steady state. Fig. 6.6 shows the norm of the difference of two consecutive transition

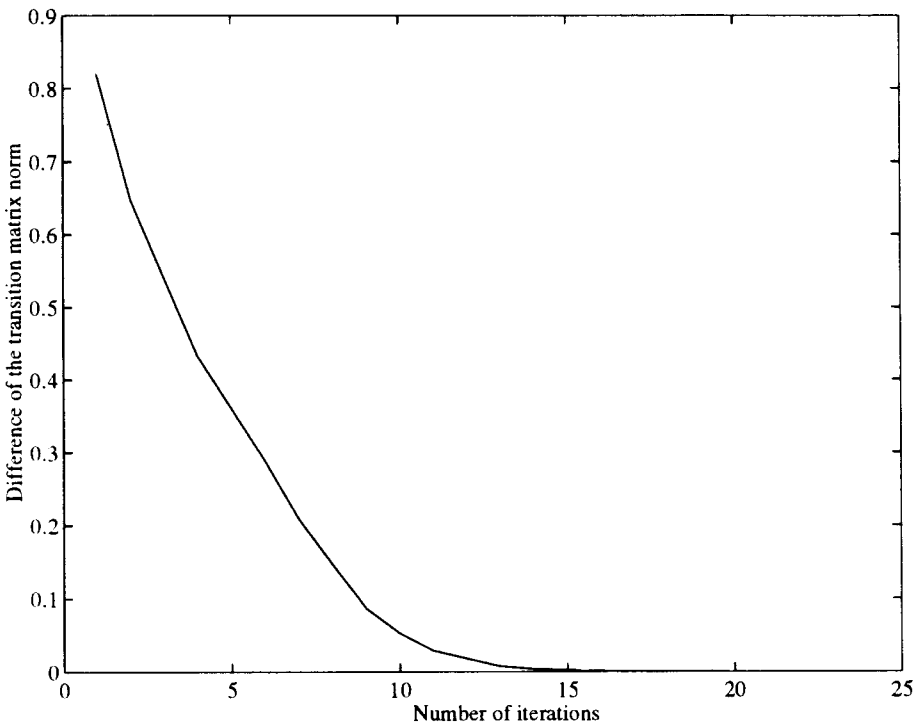


Figure 6.6: The norm of the difference of consecutive transition matrix v.s. iteration number

probability matrices as a function of the iteration number, where the norm is defined as

$$e(m) = \sum_{i=1}^{N} \sum_{j=1}^{N} \|\mathbf{P}_m(i,j) - \mathbf{P}_{m-1}(i,j)\|^2 \tag{6.19}$$

Our experiments show that after 16 transitions, the transition matrix enters steady-state. In other words, the Markov process has a memory span of about 16 symbols. If a source has a memory span of $L_o$ symbols, then the block length $L$ should be at least $L_o$. In our case, the reconstructed signal vectors $\hat{\mathbf{x}}_m^{(L)}$ in (6.16) will satisfy

$$\hat{\mathbf{x}}_m^{(n)} \approx \hat{\mathbf{x}}_m^{(\infty)} \qquad n > L_o + m - 1 \tag{6.20}$$

The above equation implies that except for an initial delay of $L_o$ symbol time, decoding can be performed continuously, one symbol at a time, by only keeping and extending all those state variables $h_m(\mathbf{v}_m)$ with an index m greater than $L - L_o$. This is the basis of our sequential decoder.

## 6.4 Simulation Results

In our experiments, the vector quantizer used in the sequential decoding soft decision vector quantizer (SD-SDVQ) is the same as that used in the SDVQ system based on symbol-by-symbol decoding (i.e., vector quantizer codebook optimization does not account for sequential decoding). Obviously, the system is a sub-optimum SD-SDVQ, but it is easy to implement and gives a good indication of the coding gain that can be obtained by using the sequential decoding technique.

To demonstrate the advantages of sequential decoding over symbol-by-symbol decoding, such as the COVQ, SDVQ and SOVQ(LBG), Figs. 6.7-6.10 compare SD-SDVQ to the COVQ, SDVQ and SOVQ(LBG) for different VQ source dimensions and codebook sizes at rates of 1 bit/sample (Fig. 6.7-6.8) and 2 bits/sample (Fig.

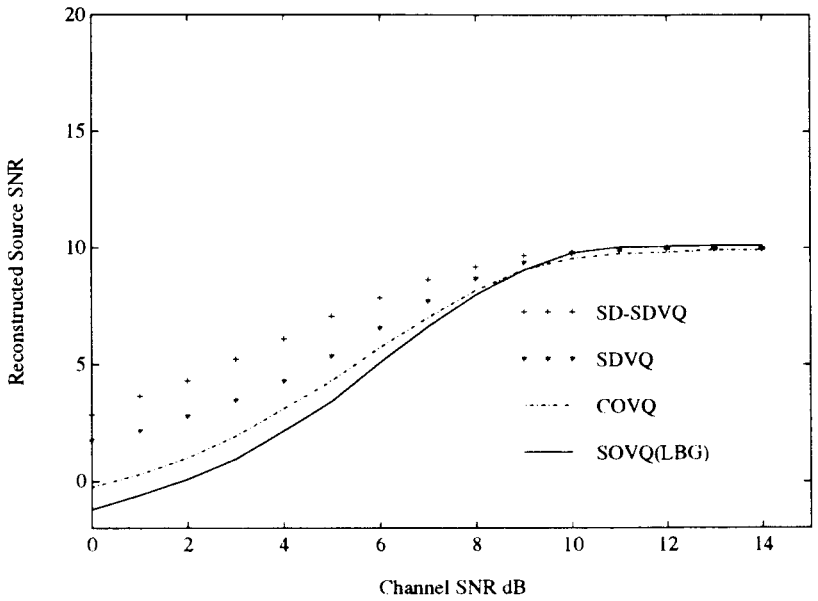


Figure 6.7: Performance of COVQ, SD-SDVQ, SDVQ and SOVQ(LBG) systems. The rate is 1 bit/symbol. VQ codebook size $M = 16$, block length $k$=4. Source **AR**-1 with $\rho$=0.9. The design channel SNR for SDVQ and COVQ is 5 dB.

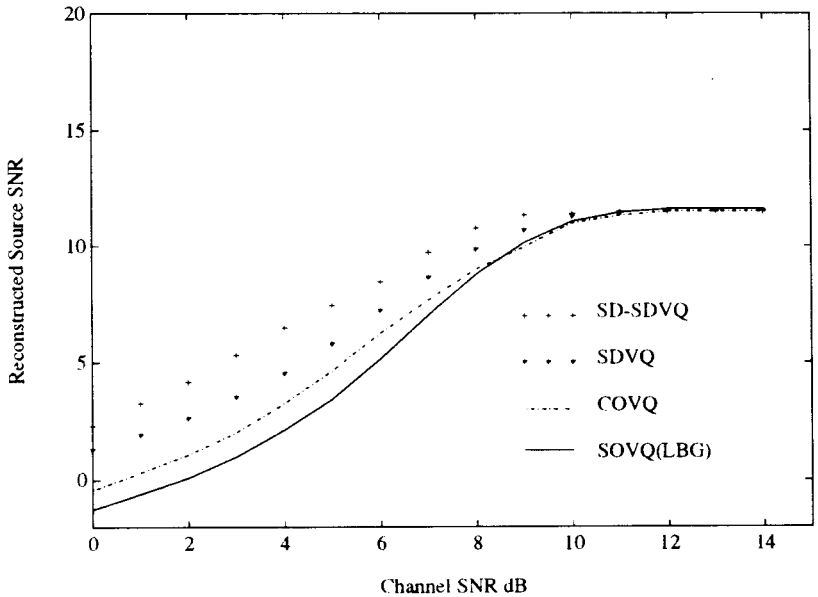


Figure 6.8: Performance of COVQ, SD-SDVQ, SDVQ and SOVQ(LBG) systems. The rate is 1 bit/symbol. VQ codebook size $N = 256$, block length $k$=6. Source **AR**-1 with $\rho$=0.9. The design channel SNR for SDVQ and COVQ is 5 dB.

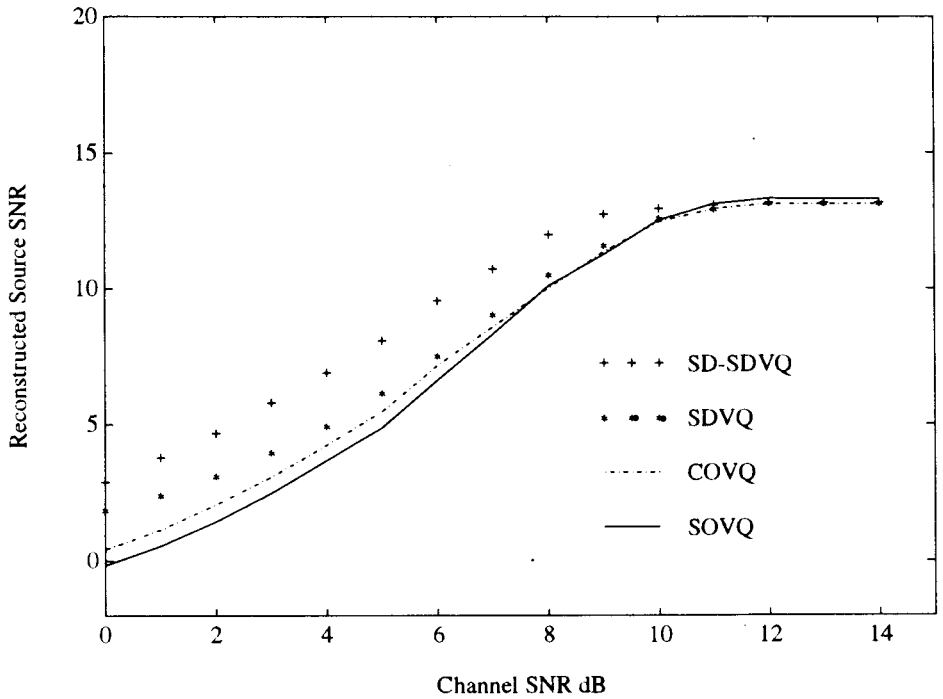


Figure 6.9: Performance of COVQ, SD-SDVQ, SDVQ and SOVQ(LBG) systems. The rate is 2 bit/symbol. VQ codebook size $M = 16$, block length $k$=2. Source **AR**-1 with $\rho$=0.9. The design channel SNR for SDVQ and COVQ is 5 dB.

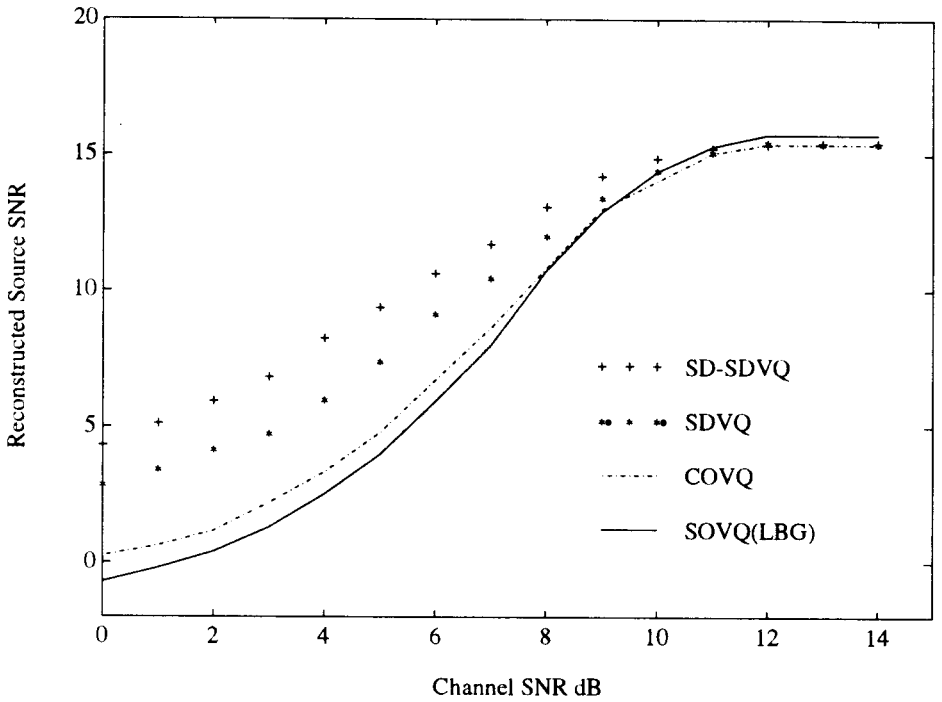


Figure 6.10: Performance of SD-SDVQ, SDVQ, COVQ and SOVQ(LBG) systems. The rate is 2 bit/symbol. VQ codebook size $N = 256$, block length $k$=3. Source **AR**-1 with $\rho$=0.9. The design channel SNR for SDVQ and COVQ is 5 dB.

6.9-6.10), where a first-order Gauss-Markov source is used as input. The COVQ and SDVQ systems were designed for an average SNR of 5 dB and the optimal modulation constellation, described in section 5.5, is used to transmit the VQ indices. Simulated annealing was used to optimize the index assignment for the COVQ, SDVQ and SOVQ(LBG).

Figures 6.7-6.8 show that SD-SDVQ provides significant performance improvement over other symbol-by-symbol decoding scheme, particularly in low channel SNR. For example, in Fig. 6.8, at a channel SNR of 3 dB and a codebook size of 256, a vector dimension $k = 8$, SD-SDVQ performance is better than SDVQ by about 1.2 dB, than COVQ by about 2 dB and than SOVQ(LBG) by about 2.5 dB. We can see that the SD-SDVQ provide the best performance in the all experiments.

## 6.5 Conclusions

We have derived in this chapter a sequential decoder for reconstructing vector-quantized and digitally modulated signals transmitted over an AWGN channel. This decoding algorithm has a computational complexity of $LN^2/2$, where $L$ is the observation window size and $N$ is the size of the VQ codebook. The simulation results show that the sequential decoding can improve the system performance over symbol-by-symbol decoding, especially at low channel signal-to-noise ratio.

# Chapter 7

# Application of the Joint Source and Channel Coding Algorithm

As we see from the previous chapters, joint optimization of source and channel coding can bring about a significant improvement in the system performance compared with the traditional design method. Here in this chapter, we consider the application of this coding philosophy to the mobile channel, which is characterized by the presence of the Rayleigh fading. In addition, we will look at the the performance of the joint codec in transmitting the line spectral pair (LSP) parameters of the speech signal over the AWGN channel. The solutions to these two relatively more practical problems constitute the main subjects of this chapter.

On the subject of joint optimization of source and channel coding over a Rayleigh fading channel, once again, a decoder based on the conditional estimate is used to reconstruct the source signal from the received signal and the channel state information (CSI). We will present an iterative algorithm that jointly optimizes the VQ and the modulation signal set in terms of the minimum MSE criterion. It will be shown that a jointly optimized system based on average channel characteristics significantly outperforms a reference system based on a VQ designed for the given source and

a standard QAM modulation signal set. We also compared the performance of the symbol-by-symbol decoding with the sequential decoding. It is found that the distortion can be further reduced by the sequential decoding algorithm, especially in the low SNR channel.

On the subject of transmitting LSPs over the AWGN channel, in a view of the real-time implementation, a structured VQ is used to reduce the computational complexity and storage capacity. We use a multi-stage VQ (MSVQ) to represent the LSP. To take account of the properties of the human ear, a weighted mean squared-error (WMSE) is used as the optimization criterion. The conditions for optimality under the WMSE criterion are derived. It is found that the optimal receiver is still a conditional estimator and the jointly optimized system can reduce the spectral distortion in the noisy channel significantly.

## 7.1 Optimized Decoder Over the Rayleigh Fading Channel

It is well known that the land/mobile radio channel exhibits Rayleigh fading. The block diagram of the communication system over the Rayleigh fading channel is shown in Figure 7.1. The source is a zero-mean stationary random process, and $\mathbf{x}$ is a $k$-

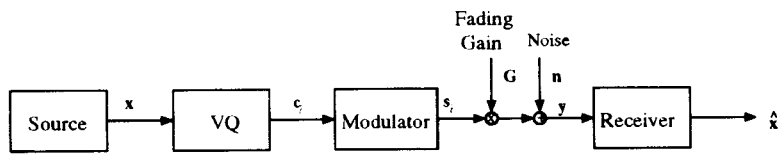


Figure 7.1: System model over Rayleigh fading channel

dimensional vector derived from the random source with the pdf $p(\mathbf{x})$. The vector $\mathbf{x}$ is mapped to a VQ codeword chosen from the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$ according to a given optimization criterion. Assuming the source vector is represented by $\mathbf{c}_i$, then its index

$i$ is mapped directly into the modulation signal point. Since every two dimensions of $\mathbf{s}_i^*$ can be viewed as a QAM symbol, it is thus most convenient to treat $\mathbf{s}_i^*$ as a complex vector with $M/2$ components. Let $s_{ij}^*$, $j = 1, 2, \ldots, M/2$, be the $j^{th}$ component of $\mathbf{s}_i^*$, then after transmission over the Rayleigh channel with additive white Gaussian noise, the correspondingly received signal is

$$y_j^* = g_j^* s_{ij}^* + n_j^*, \tag{7.1}$$

where $n_j^*$ and $g_j^*$ are zero mean, complex Gaussian random variables representing the fading gain and the additive Gaussian noise respectively, experienced by $s_{ij}^*$. Since interleaving/deinterleaving is usually used to combat fading, it is reasonable to assume that $g_i^*$'s are independent and identically distributed (i.i.d). The same is true for the $n_i^*$'s. Note that if the white Gaussian noise in the channel has a two-sided power spectral density of $N_o/2$, then the variance of the $n_i^*$'s is $N_o/2$. In addition, if the variance of the fading gain $g_j$ is $\sigma_g^2$, then the symbol signal energy is

$$E_{avg} = \sum_{i=1}^{N} P_i \|\mathbf{s}_i^*\|^2 \sigma_g^2 \tag{7.2}$$

where $P_i$ is the probability that $\mathbf{s}_i^*$ is transmitted. The channel signal-to-noise ratio (SNR) is simply $E_{avg}/\frac{N_o}{2}$. As in many other studies of the fading channel, we assume perfect channel state information (CSI) is available at the receiver, i.e., the $g_j$'s are known at the receiver. It is not difficult to estimate these complex fading gains, especially at slow fading rates. This may be done, for example, by inserting known training symbols at regularly spaced intervals [83]. If the fading rate is, say f Hz, and the baud rate is B Hz, then the spacing of training symbols should be at least B/(2f) symbols. The fade rate is usually small compared with the baud rate (around half a percent in Digital Cellular, for example) and hence the overhead associated with channel estimation is very small. Interpolation may be used to determine the complex gains that affect the data symbols between training symbols.

The received signal can be written in a matrix form as

$$\mathbf{y}^* = \mathbf{G}^*\mathbf{s}_i^* + \mathbf{n}^* \tag{7.3}$$

where $\mathbf{G}^* = diag(g_1^*, g_2^*, \cdots, g_{M/2}^*)$. For convenience, let us define a fading gain vector

$$\mathbf{g}^* = [g_1^*, g_2^*, ..g_{M/2}^*]^T \tag{7.4}$$

Given the received signal $\mathbf{y}^*$ and the fading gain $\mathbf{g}^*$, the decoder should provide the "best" estimate of the original signal $\mathbf{x}$.

The MSE criterion is used to measure the distortion between the original and the reconstructed source signals. Let $p(\mathbf{x}, \mathbf{y}^*|\mathbf{g}^*)$ denote the conditional joint density function of $\mathbf{x}$ and $\mathbf{y}^*$ given the fading vector $\mathbf{g}^*$. Then the average distortion can be written as

$$D = \frac{1}{k}\int\int\int \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y}^*, \mathbf{g}^*)\|^2 p(\mathbf{x}, \mathbf{y}^*|\mathbf{g}^*)d\mathbf{x}d\mathbf{y}^* p(\mathbf{g}^*)d\mathbf{g}^* \tag{7.5}$$

where $p(\mathbf{g}^*)$ denotes the pdf of the fading gain vector $\mathbf{g}^*$.

As shown in Chapter 5, the optimum decoder under the minimum MSE criterion is one that computes the conditional expectation of the signal $\mathbf{x}$ given the received signal $\mathbf{y}^*$ and the fading gain $\mathbf{g}^*$. In other words

$$\hat{\mathbf{x}}(\mathbf{y}^*, \mathbf{g}^*) = E[\mathbf{x}|\mathbf{y}^*, \mathbf{g}^*] = \frac{1}{p(\mathbf{y}^*|\mathbf{g}^*)}\int \mathbf{x}p(\mathbf{x})p(\mathbf{y}^*|\mathbf{x}, \mathbf{g}^*)d\mathbf{x} \tag{7.6}$$

where $p(\mathbf{x})$ is the pdf of $\mathbf{x}$, $p(\mathbf{y}^*/\mathbf{x}, \mathbf{g}^*)$ is the conditional pdf of $\mathbf{y}^*$ given $\mathbf{x}$ and $\mathbf{g}^*$. Let $\Omega_i$, $i = 1, 2, \cdots, N$ denote the $i$th partition in the VQ, and let $p_{\mathbf{n}^*}()$ be the pdf of the complex Gaussian noise vector $\mathbf{n}^*$. Since we assume that perfect CSI is available at the receiver, it can be shown that (7.6) can be rewritten as

$$\hat{\mathbf{x}}(\mathbf{y}^*, \mathbf{g}^*) = \frac{\sum_{i=1}^N P_i \mathbf{c}_i p(\mathbf{y}^*|\mathbf{s}_i^*, \mathbf{g}^*)}{\sum_{i=1}^N P_i p(\mathbf{y}^*|\mathbf{s}_i^*, \mathbf{g}^*)}, \tag{7.7}$$

where

$$p(\mathbf{y}^*|\mathbf{s}_i^*, \mathbf{g}^*) = p_n(\mathbf{y}^* - \mathbf{G}^*\mathbf{s}_i^*) = \frac{1}{(N_0\pi)^{M/2}} exp(-\frac{\|\mathbf{y}^* - \mathbf{G}^*\mathbf{s}_i^*\|^2}{N_0}) \tag{7.8}$$

is the conditional pdf of $\mathbf{y}^*$ given that $\mathbf{s}_i^*$ is transmitted and the fading gain is $\mathbf{g}^*$,

$$\mathbf{c}_i = \frac{\int_{\Omega_i} \mathbf{x}p(\mathbf{x})d\mathbf{x}}{\int_{\Omega_i}(\mathbf{x})d\mathbf{x}} \tag{7.9}$$

is the VQ's clean channel centroid associate with the partition region $\Omega_i$, and

$$P_i = \int_{\Omega_i} p(\mathbf{x})d\mathbf{x} \tag{7.10}$$

is the probability of occurrence of the $i$th codevector $\mathbf{c}_i$ .

## 7.2 Optimization of VQ and Modulation Signal Set for the Rayleigh Fading Channel

We first consider the problem of optimizing the VQ partitions for a given modulation constellation and Rayleigh fading gain.

As in the AWGN channel, the GLA procedure is used again to find the optimal VQ. Assumed that an initial VQ codebook, $\mathbf{c}_i, i = 1, \cdots, N$, is given. Furthermore, assume that the codeword $\mathbf{c}_i$ is transmitted by the signal vector $\mathbf{s}_i^*$. Therefore after substituting (7.7) into (7.5), the average distortion $D$ can be rewritten as

$$D = \frac{1}{k}\sum_{m=1}^{N}\int_{\Omega_m}\int_{-\infty}^{+\infty}\int_{\mathbf{g}^*}\left\|\mathbf{x} - \frac{\sum_{i=1}^{N} P_i P\mathbf{c}_i p_{\mathbf{n}^*}(\mathbf{G}^*(\mathbf{s}_m^* - \mathbf{s}_i^*) + \mathbf{n}^*)}{\sum_{i=1}^{N} P_i p_n(\mathbf{G}^*(\mathbf{s}_m^* - \mathbf{s}_i^*) + \mathbf{n}^*)}\right\|^2 \tag{7.11}$$

$$p_{\mathbf{n}^*}(\mathbf{n}^*)p(\mathbf{g}^*)d\mathbf{g}^* p(\mathbf{x})d\mathbf{x}$$

where $\Omega_i$, $i = 1, \cdots, N$, are the various partitions. Since each inner integral term

$$I_m = \int_{\mathbf{n}^*}\int_{\mathbf{g}^*}\left\|\mathbf{x} - \frac{\sum_{i=1}^{N} P_i\mathbf{c}_i p_{\mathbf{n}^*}(\mathbf{G}^*(\mathbf{s}_m^* - \mathbf{s}_i^*) + \mathbf{n}^*)}{\sum_{i=1}^{N} P_i p_{\mathbf{n}^*}(\mathbf{G}^*(\mathbf{s}_m^* - \mathbf{s}_i^*) + \mathbf{n}^*)}\right\|^2 p_{\mathbf{n}^*}(\mathbf{n}^*)p(\mathbf{g}^*)d\mathbf{g}^* d\mathbf{n}^* \tag{7.12}$$

in equation (7.12) is non-negative, the average distortion will be minimized if the following partitioning rule is adopted for the VQ:

$$\mathbf{x} \in \Omega_m \quad \text{if} \quad I_m \leq I_j \quad \text{for all} \quad j \neq m \tag{7.13}$$

The partitions defined by (7.12) and (7.13) depend on the VQ centroids $\mathbf{c}_1, \ldots, \mathbf{c}_N$, which in turn depend on the partitions according to the following

$$\mathbf{c}_i = \frac{\int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_i} p(\mathbf{x}) d\mathbf{x}} \tag{7.14}$$

Therefore the optimal centroids and partitions can be obtained by iterating (7.12) and (7.14) until convergence occurs. Convergence is guaranteed since the distortion can either decrease or remain the same at each step of the iteration.

The mapping of the VQ centroids into the modulation signal points (index assignment) has an important impact on the system performance. The simulated annealing algorithm is used for optimizing the index assignment [23].

The distortion in (7.11) depends on the VQ partitions $\{\Omega_m, m = 1, \cdots, N\}$ and the modulation constellation $\{\mathbf{s}_m^*, m = 1, \cdots, N\}$. Therefore, the average distortion could be further reduced by optimizing the modulation constellation for a given VQ partition.

Due to the inherent property of the nonlinear receiver it is quite difficulty to find an analytical solution for the optimal modulation constellation, As in Chapter 5, we still adopt the constraint gradient-search algorithm to find the optimal constellations.

## 7.3 Sequential Decoding over the Rayleigh Fading Channel

As we discussed in Chapter 6, the redundancy exits at the VQ output due to the constraints on the VQ complexity, therefore a block decoding technique can be used

to exploit the redundancy.

The optimal block decoder for the Rayleigh fading channel is given by

$$\hat{\mathbf{X}}_L(\mathbf{Y}_L^*, \mathbf{G}_L^*) = \frac{\sum_{\mathbf{V}_L} P(\mathbf{V}_L) p(\mathbf{Y}_L^* | \mathbf{V}_L, \mathbf{G}_L^*) \mathbf{V}_L}{\sum_{\mathbf{V}_L} P(\mathbf{V}_L)\, p(\mathbf{Y}_L^* | \mathbf{V}_L, \mathbf{G}_L^*)} \tag{7.15}$$

where $P(\mathbf{V}_L)$ is the probability of the sequence of centroids $(\mathbf{v}_1, \ldots, \mathbf{v}_L)$, and $p(\mathbf{Y}_L^* | \mathbf{V}_L, \mathbf{G}_L^*)$ is the joint pdf of all the $\mathbf{Y}_L^*$'s conditioned on the source sequence $(\mathbf{v}_1, \ldots, \mathbf{v}_L)$ and the fading gain $(\mathbf{g}_1^*, \ldots, \mathbf{g}_L^*)$. Note that the summations in the above equation are taken over all possible $L$-length source sequences.

Following a similar development of the sequential decoding algorithm for the AWGN channel, we can express the reconstructed source vectors $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L$ for an observation window of $L$ in the form:

$$\hat{\mathbf{x}}_m^{(L)} = \frac{\mathbf{b}_{L,m}}{\alpha_m} \tag{7.16}$$

where the superscript $(L)$ is deliberately introduced to emphasize the dependency of the reconstructed signal vectors on the block length $L$. Note the recursive formula for $\mathbf{b}_{L,m}$ and $\alpha_m$ are related to the given Rayleigh fading gain matrix $\mathbf{G}_\mathbf{L}^*$.

## 7.4 Joint Optimization of Source and Channel Coding over Rayleigh Fading Channel

A strategy of the generalized Lloyd's algorithm (GLA) can be used to jointly optimize the VQ and modulation constellations over the Rayleigh fading channel. This strategy first optimizes the VQ for the given modulation constellation, then optimizes the modulation constellation for the given VQ. By iterating these two steps until the convergence, a jointly optimized system is obtained.

The design procedure described above was applied to a first order Gauss-Markov source signal. The correlation coefficient between successive samples is set to 0.9. The

corresponding numerical results are shown in Figs 7.2-7.5 for the different VQ size $N$, the signal space dimension $M$, and the VQ source rate $R_s$, defined as

$$R_s = \frac{log_2(N)}{k} \tag{7.17}$$

The reference systems used for comparison in Figs. 7.2-7.5 are constructed as follows. For given $N$, $M$, and $R_s$, we first design the optimal VQ for the AR-1 source in a noiseless channel, using the LBG algorithm [43]. Then from the value of $N$ and $M$, determine the most appropriate basic QAM constellation. The decoder in the reference system is assumed to perform hard decision decoding, i.e. it determines the most likely transmitted VQ codeword from the received vector $\mathbf{y}^*$ and the channel state information $\mathbf{g}^*$. The simulated annealing algorithm is used to find the optimal VQ-modulation mapping for the reference system. As for sequential decoding, the same VQ and modulation constellation designed for symbol by symbol decoding is used. Neither VQ nor modulation constellation has been re-optimized for the sequential decoding strategy. Due to the homogeneous Markov property of the particular VQ output, the decoding delay $L_o$ is equal to 16.

We show in Figs 7.2-7.8 the performance of our systems as a function of the channel signal to noise ratio, $E_{av}/N_0$, in dB. Figs 7.2-7.4 are for a source rate of 1 bit/sample while Figs. 7.5-7.7 are for a source rate of 2 bit/sample. The difference among Figs 7.2-7.4 (or Figs. 7.5-7.7) is the number of codewords N. The performance measure is the reconstructed source signal-to-noise ratio (SSNR), defined as

$$SSNR = \frac{E(|\mathbf{x}|^2)}{E(|\mathbf{x} - \hat{\mathbf{x}}|^2)} \tag{7.18}$$

It should be pointed out that system optimization were performed at a channel SNR of 17 dB. The resultant systems were then used over the entire SNR range. The optimization point (at 17 dB) gives the best compromise between the performance

in the high SNR and the low SNR regions. Note that in principle we can optimize the system at every channel SNR. However this would require an adaptive system which coordinates between the transmitter and the receiver. As observed from these figures, the sequential decoder significantly outperforms the two reference systems. For example, it can be seen from that, see fig. 7.2 and 7.4, with $N = 16$ and a channel SNR of 10 dB, the gain in SSNR provided by sequential decoding is roughly 5 dB compared to the hard decision decoder (conventional system) and 1.5 dB compared to the symbol by symbol decoder. When $N$ increases to 256, the performance advantage (in SSNR) the sequential decoder has over the symbol by symbol decoder increases to 2.4 dB at the same channel SNR. When the channel SNR decreases, the advantage of using a sequential decoder becomes even more obvious.

To put the Rayleigh channel results in perspective, we show in Fig. 7.8 the performance comparison with the pure AWGN channel. It is found that Rayleigh fading causes a 15 dB degradation in the channel SNR.

## 7.5 Linear Prediction and Spectrum Information

In the earlier chapters, we evaluate the performance of the joint source and channel coding through the use of a first order Gauss-Markov source. Now, we consider a more practical application of our SDVQ technique in transmitting the line spectral pairs (LSP) of a digitized speech signal.

LSP is a representation form of the linear prediction parameter of the speech signal. Linear prediction is a very important and powerful signal processing technique used in the speech analysis and speech coding. The basic idea behind the method is that each input speech sample can be predicted by a linear combination of a finite number of past input speech samples. Mathematically, the linear predictor is described by the

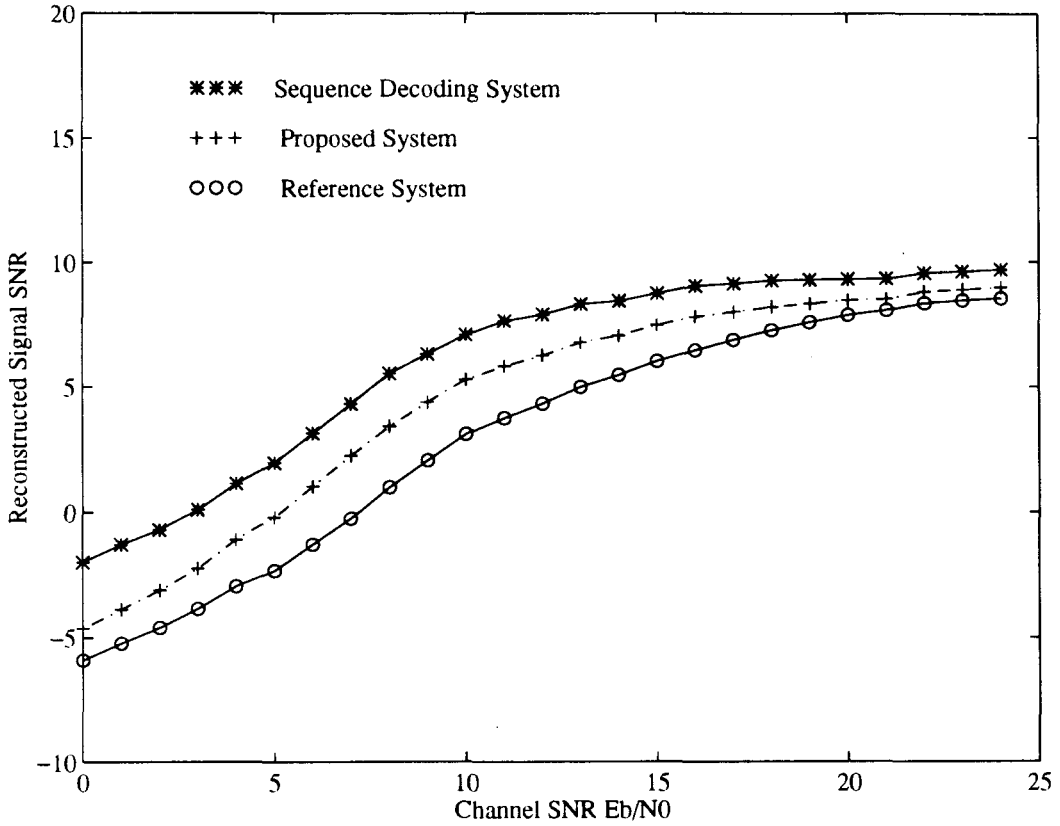


Figure 7.2: System Performance for vector dimension $k$=4, rate 1 bit/sample, and signal space dimension M=4. Curve *** denotes the performance of sequential decoding, Curve +++ is the performance of the optimized system for symbol by symbol decoding. Curve o o o is for the reference system where 2XQPSK symbols is used as the modulation format.

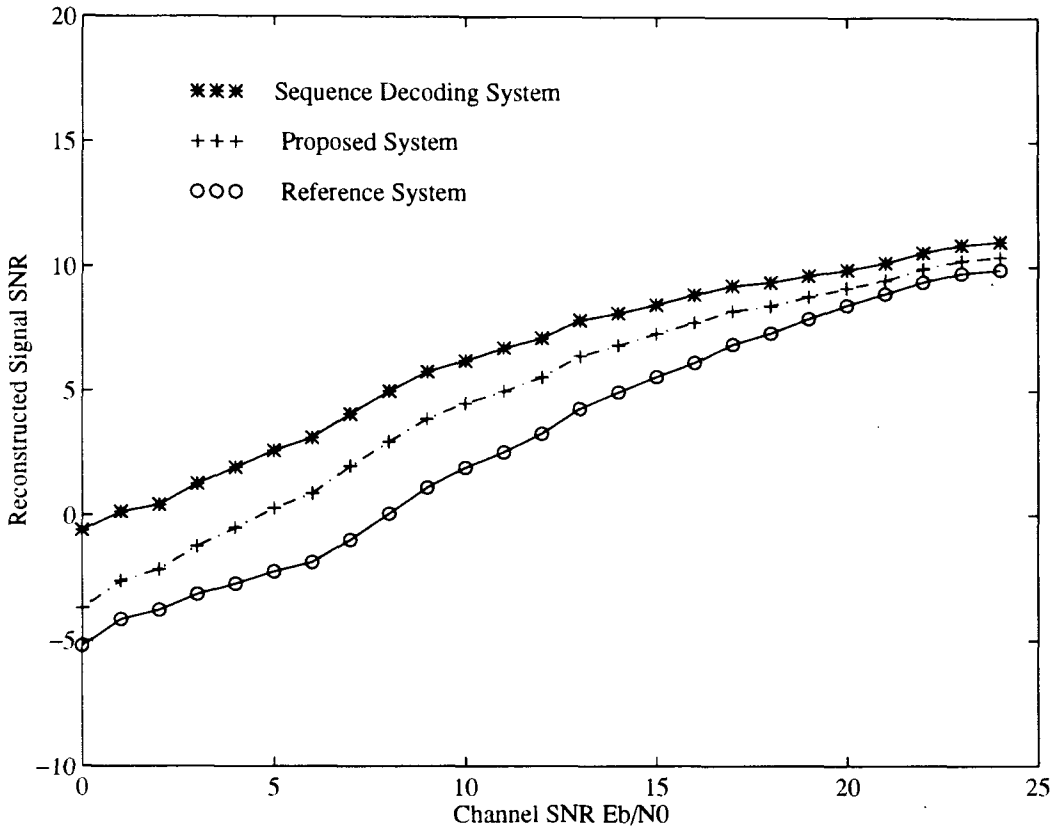


Figure 7.3: System Performance for vector dimension $k$=6, rate 1 bit/sample, and signal space dimension M=6. Curve *** denotes the performance of sequential decoding, Curve +++ is the performance of the optimized system for symbol by symbol decoding. Curve o o o is for the reference system where 3XQPSK symbols is used as the modulation format.

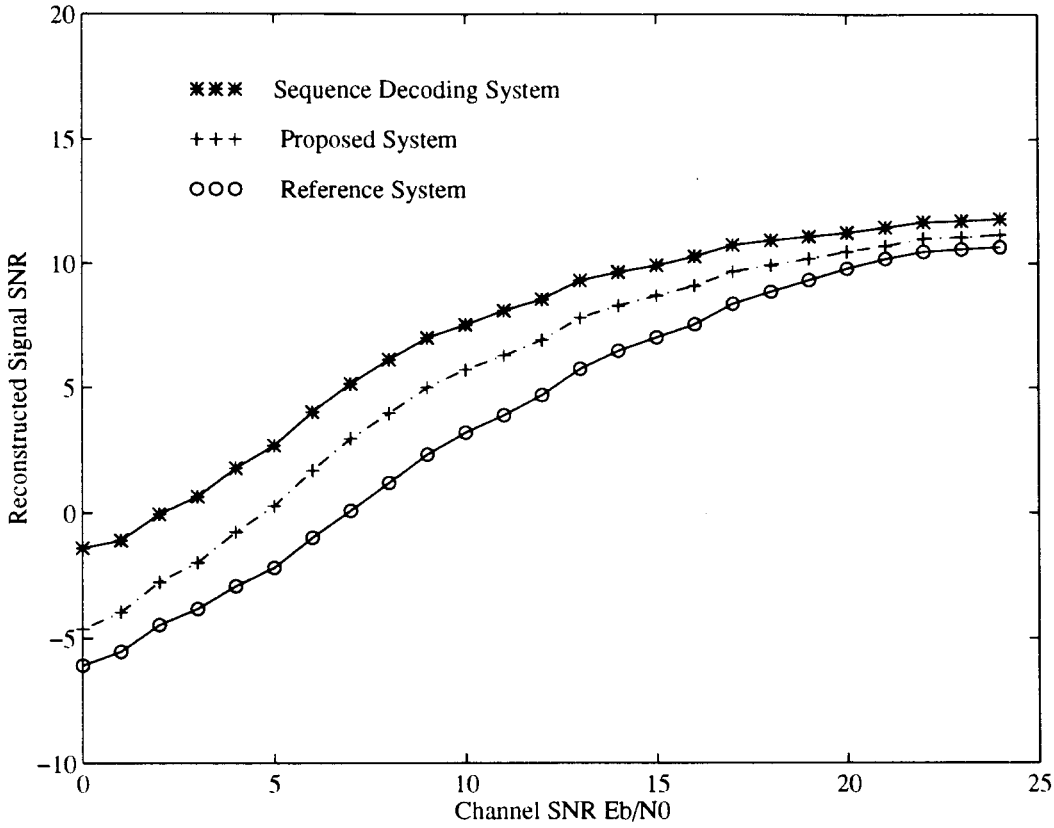


Figure 7.4: System Performance for vector dimension $k$=8, rate 1 bit/sample, and signal space dimension M=8. Curve *** denotes the performance of sequential decoding, Curve +++ is the performance of the optimized system for symbol by symbol decoding. Curve o o o is for the reference system where 4XQPSK symbols is used as the modulation format.

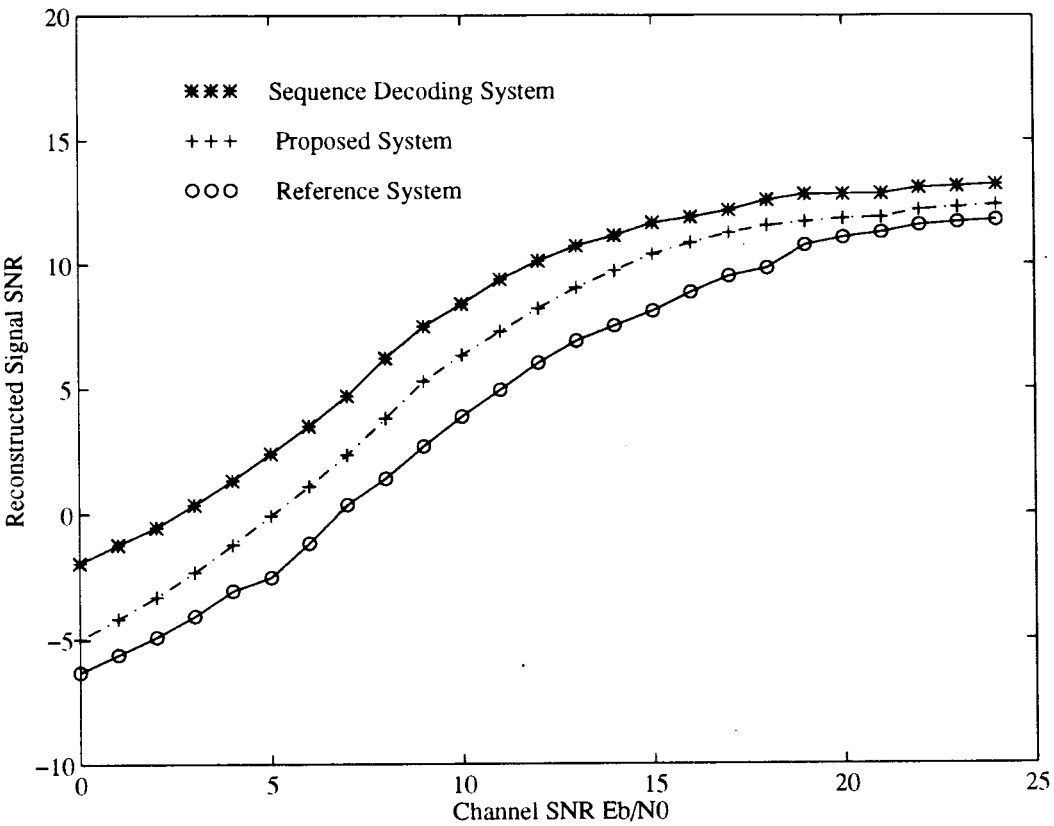


Figure 7.5: System Performance for vector dimension $k$=2, rate 2 bit/sample, and signal space dimension M=4. Curve *** denotes the performance of sequential decoding, Curve +++ is the performance of the optimized system for symbol by symbol decoding. Curve o o o is for the reference system where 2XQPSK symbols is used as the modulation format.

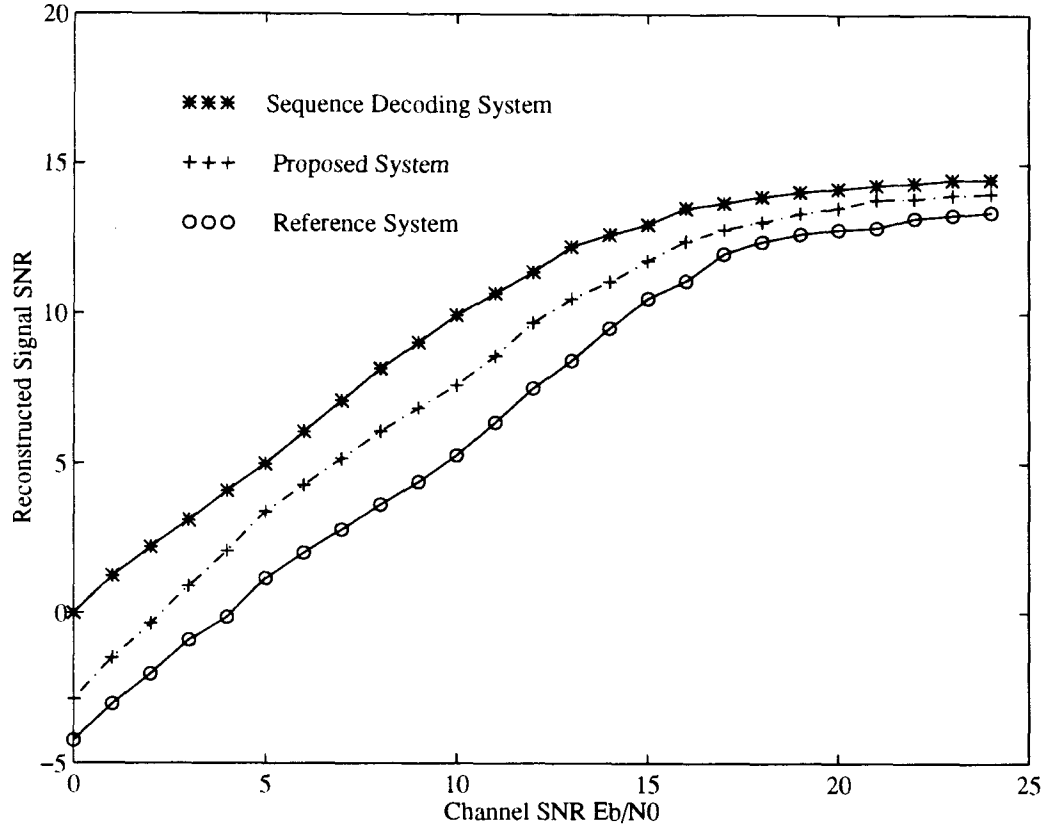


Figure 7.6: System Performance for vector dimension $k$=3, rate 2 bit/sample, and signal space dimension M=6. Curve *** denotes the performance of sequential decoding, Curve +++ is the performance of the optimized system for symbol by symbol decoding. Curve o o o is for the reference system where 3XQPSK symbols is used as the modulation format.

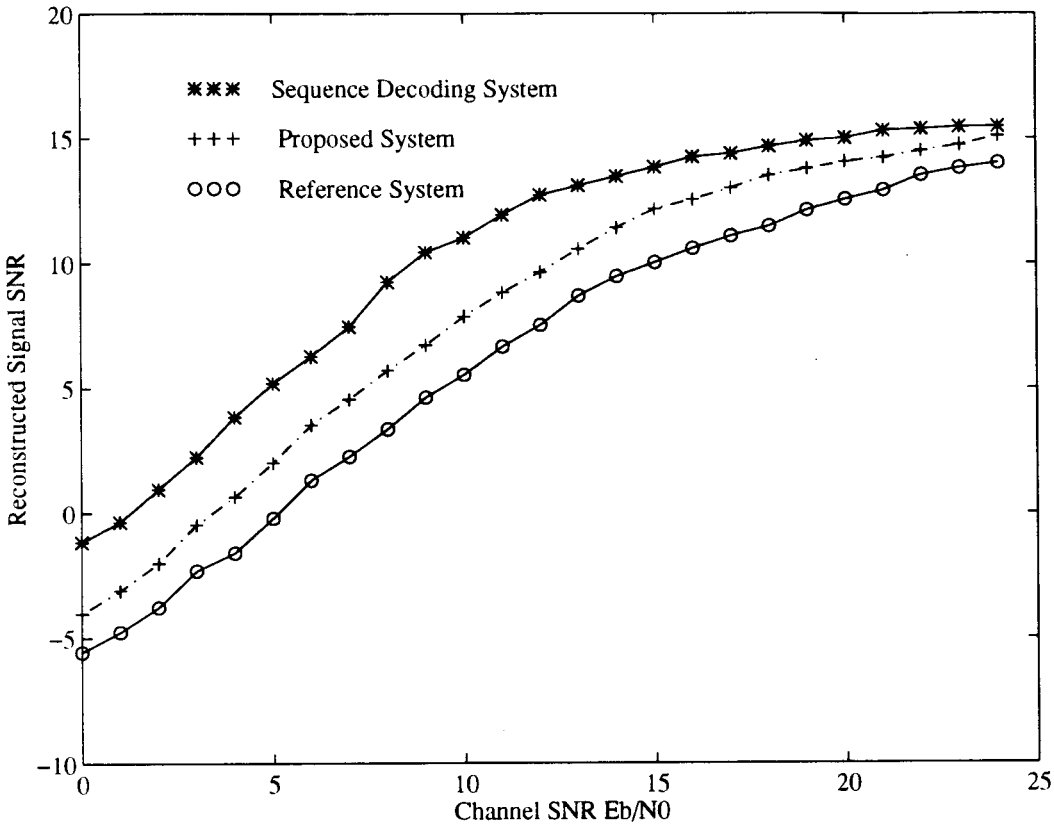


Figure 7.7: System Performance for vector dimension $k$=2, rate 2 bit/sample, and signal space dimension M=4. Curve *** denotes the performance of sequential decoding, Curve +++ is the performance of the optimized system for symbol by symbol decoding. Curve o o o is for the reference system where 4XQPSK symbols is used as the modulation format.

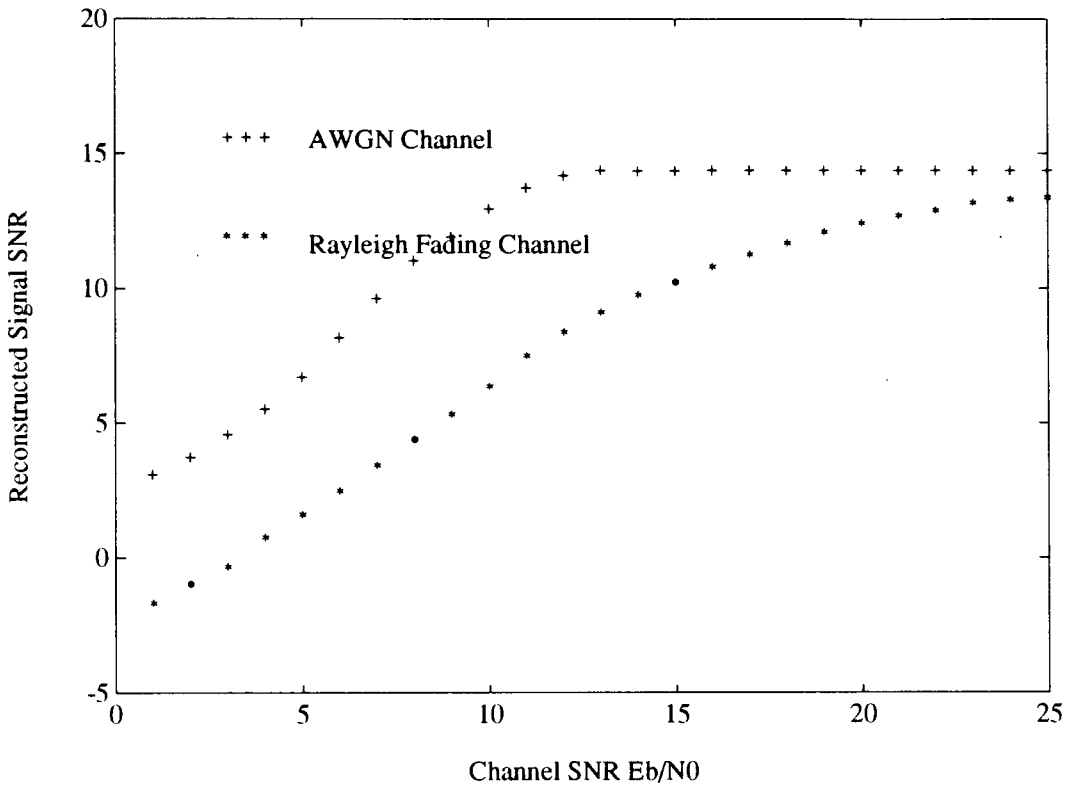


Figure 7.8: Effect of Rayleigh fading on the system performance. The VQ codebook size $N = 64$, block length $k = 3$, and a signal space dimension $M = 6$. The source is AR-1 with $\rho = 0.9$. Curve +++ is the AWGN channel result. Curve *** is the Rayleigh fading channel results.

equation

$$\hat{u}(n) = \sum_{k=1}^{K} a_k u(n-k) \tag{7.19}$$

By subtracting the predicted signal, $\hat{u}(n)$, from the current value, $u(n)$, we obtain a residual (or error) signal,

$$e(n) = u(n) - \hat{u}(n) \tag{7.20}$$

The coefficients is chosen to minimize the MSE of the prediction error, therefore the partial derivative of the MSE with respect to each coefficient $a_i$ $(i = 1, 2, \cdots, K)$ should be zero. Then, the following condition for optimality is found:

$$E\{e(n)u(n-i)\} = 0 \tag{7.21}$$

This equation is called the *orthogonal principle* in the linear prediction theory. It implies that the prediction error should be orthogonal to the input data. By submitting equ. (7.19), (7.20) into (7.21) we obtain the following linear equation

$$\mathbf{Ra} = \mathbf{r}_u \tag{7.22}$$

where $\mathbf{R}$ is the autocorrelation matrix,

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & \cdots & r(K-1) \\ r(1) & r(0) & \cdots & r(K-2) \\ r(2) & r(1) & \cdots & r(K-3) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ r(K-1) & r(K-2) & \cdots & r(0) \end{bmatrix}$$

$\mathbf{a} = [a_1, a_2, \cdots, a_K]^T$ is the linear prediction coefficients and $\mathbf{r}_u = [r(1), r(2), \cdots, r(K)]^T$, where $r(i), i = 1, 2, , \cdots, K$, is the autocorrelation function of the signal $u(n)$. The equation (7.22) is called Yule-Walker equation. The autocorrelation matrix $\mathbf{R}$ is a

positive, Toeplitz matrix. Due to its special structure a computational efficient procedure, Levison-Durbin algorithm [59]-[60], can be used to find the optimal prediction coefficients. If the speech signal is passed through a filter with the optimal prediction coefficients, the residual sequence needs fewer bits of information to achieve a satisfactory signal-to-noise ratio. Note the prediction coefficients are needed be quantized and coded only at a rather slow frame rate, typically about 10-30 ms to reflect the changing nature of the speech signal.

On the other hand, we can look at the LPC analysis of speech signal from the view of signal spectrum. LPC analysis is to extract a set of parameters from the speech signal which specifies the filter transfer function giving the best spectral match to the signal being encodered. The spectral density function of the speech signal is given by

$$P(\omega) = \frac{\sigma^2}{\|A(e^{j\omega})\|}$$

where the $\sigma^2$ is the variance of the prediction error signal, and $A(e^{j\omega})$ is the spectral response of the linear filter. Such a filtering process is often referred to as spectral flatting, since regardless of the spectral shape of the input speech signal, the spectrum of residual sequence is substantially flat. Practically, a good estimate of the signal spectral density can be obtained by the so-called model-based approach. For the speech signal analysis 10 order predictor is usually adopted, such as VSELP [5], DOD FS-1016 [7]. Figure 7.9 shows the linear prediction spectrum of a 20 ms segment of voiced speech. A Hamming window was used to weight the speech data prior to the linear prediction analysis. We can see from the figure that the spectral envelope of the LPC filter frequency response contains a number of peaks at frequencies closely related to the formant frequencies, the resonant frequencies of the vocal tract.

In low-bit rate speech coding algorithms such as CELP, the input speech spectrum information is quantized and transmitted, along with other parameters to the receiver.

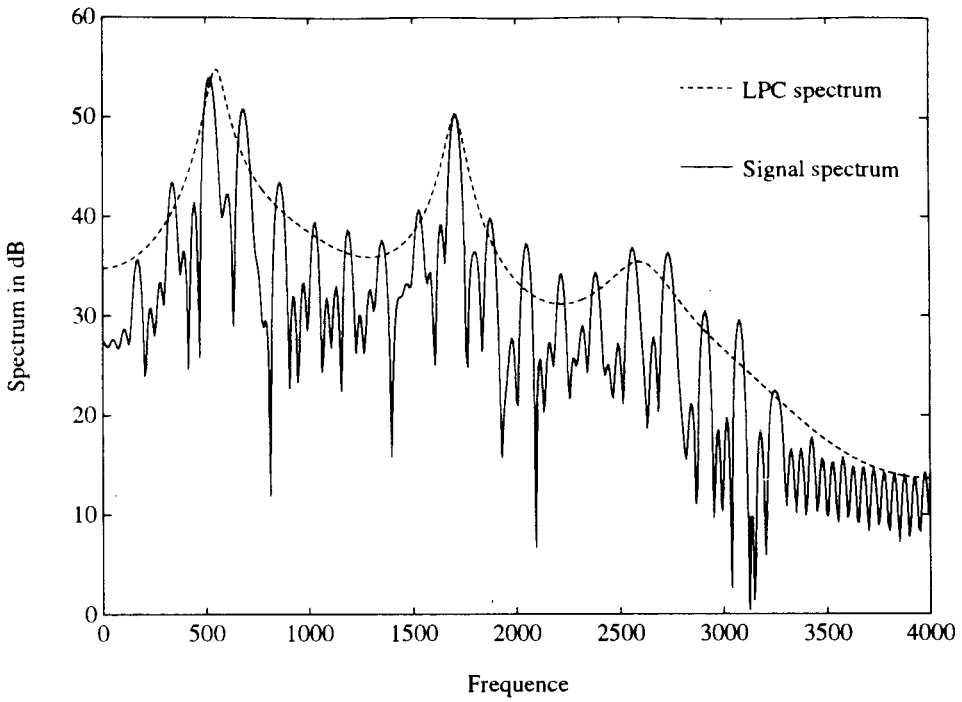


Figure 7.9: Linear prediction spectrum and signal spectrum

Vector quantization of spectral parameters has attracted considerable attention for the low bit rate speech coding. However, the application of spectral parameter vector quantization has been limited in actual speech coding system due to the high VQ computational complexity and the large distortion over noisy channels. The optimal $K$-th order linear predictor is represented by:

$$A(z) = \sum_{i=1}^{K} a_i z^i$$

where $a_i$ are the coefficients of the linear predictor. And the same predictor with quantized coefficients $\hat{a}_i$ is denoted by $\hat{A}(z)$. The line spectrum pairs (LSP), historically, is a popular one-to-one transformation of the LPC parameters that result in a set of parameters which can be efficiently quantized while maintaining stability [84]-[86]. Many good properties of LSP make it a good candidate for the presentation of the speech signal spectral information. The first such property is monotonicity, which

restricts each LSP in a vector to be larger than the previous LSP. Given the monotonicity property, errors can be easily detected and corrected. The second desirable property is its robustness towards error, which means that an error in any single LSP do not propagate to other LSP.

By introducing a suitable structure in the VQ codebook, both the memory and computational complexity required by a VQ can be reduced dramatically. These techniques range from the multi-stage codebooks and group codebooks to overlaying non-uniform binary tree structures on the codebook. In a multi-stage VQ (MSVQ) system, an input LSP vector $\mathbf{x}$ is approximated by:

$$\hat{\mathbf{x}} = \mathbf{c}_0^{i_0} + \mathbf{c}_1^{i_1} + \cdots + \mathbf{c}_{K-1}^{i_{K-1}} \tag{7.23}$$

where $\hat{\mathbf{x}}$ is the quantization presentation of $\mathbf{x}$, $K$ is the number of stages, and $\mathbf{c}_j^m$ is the $m$-th codevector from the $j$-th stage. The $j$-th stage codebook comprises of $L = 2^b$ codevectors. The total bits for the MSVQ system are equal to $b \times K$ bits. A sequential search procedure is traditionally used in MSVQ by selecting the $j$-th vector such that the mean squared error

$$\epsilon_j = \|\mathbf{e}_{j-1} - \mathbf{c}_j\|^2 \tag{7.24}$$

is minimized over all possible choices of $j$-th stage codebook $\mathbf{c}_j$, where $\mathbf{e}_{j-1}$ is the residual error up to $j$-1th stage. Since the minimization of MSE for overall stages is not equal to the minimization of MSE for each stage, the sequential search procedure is not the optimal search procedure. The sub-optimal search procedure will result in the performance degradation.

Another answer to reduce the VQ complexity is to use group vector quantization (GVQ). In GVQ, the vector to be quantized is divided into several sub-vectors where an independent VQ is applied to each sub -vector. The only constraint placed on the

individual code book search is that the reconstructed vector for LSP is required to be monotonically increasing from the first element to the last. Recently, the spectral distortion near 1 dB has been obtained by using GVQ technique in a 24 bits/frame system, where 12 bits were used to encode the the first 4 LSP parameters and the other 6 LSP parameters respectively [89]. As we see, a group VQ codebook is equivalent to a multi-stage VQ where some vector components are forced to zero. The MSVQ performance can be improved by iteratively optimizing the MSVQ codebook, as reported by Chan, Gupta and Gersho [90]. MSVQ was previously studied in [91]. In [91] several measures were utilized to improve the performance of the traditional MSVQ, which include using a multiple search procedure in each stage search to approximate the full-search codebook search, and jointly optimizing the MSVQ codebooks of total stages.

The performance of traditional MSVQ degrades significantly in the presence of transmission errors. In noisy channels, the received index may differ from the transmitted index and as a result the reproduction vector may be a poor representation of the original source vector. Due to the inherent structure of MSVQ codebook the errors in the first several stages are very sensitive to the channel error, which leads to very large distortion if an error occurs in the demodulator. As we see in chapter 5, the optimal decoder should make a linear combination of the clean channel centroids with weights depending on the received signal and the channel statistics.

## 7.6 Weighting Mean Squared-Error Criterion for the LSP

Because of the different sensitivities of the human ear to various frequencies, a

weighted mean-squared error (WMSE) distortion measure is usually used as an optimization criterion in training the VQ for the LSP. The WMSE between the original and the reconstructed signal is given by:

$$\epsilon = \|(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{W}(\mathbf{x} - \hat{\mathbf{x}})\| \tag{7.25}$$

where $\mathbf{W}$ is a diagonal matrix whose elements may depend on $\mathbf{x}$. For the LSP, the perceptually based weighted matrix introduced in [92] is used. The weighted matrix elements are given by:

$$w_i = \begin{cases} u(f_i)\sqrt{D_i/D_{max}} & 1.375 \leq D_i \leq D_{max} \\ u(f_i)D_i/\sqrt{1.375 D_{max}} & D_i < 1.375 \end{cases} \tag{7.26}$$

where

$$u(f_i) = \begin{cases} 1, & f_i < 1000Hz \\ \frac{-0.5}{3000}(f_i - 1000) + 1 & 1000 \leq f_i \leq 4000Hz \end{cases} \tag{7.27}$$

$f_i$ denotes the $i$-th components of the LSP, $D_i$ denotes the group delay for $f_i$, and $D_{max}$ is the maximum group delay. The function $u(D_i)$ accounts for the specific spectral sensitivity of each frequency $f_i$. The group delay can be computed as the gradient of the phase angle of the filter at a frequency corresponding to the $i$-th LSP with the sampling rate being 8000 Hz.

The following spectral distortion measure is used to measure the performance of the resulting LSP MSVQ codebook:

$$d_{SD}(A(z), \hat{A}(z)) = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} 10 \log_{10} \left( \frac{|A(e^{j2\pi\omega})|^2}{|\hat{A}(e^{j2\pi\omega})|^2} \right) \right\}^{1/2} \tag{7.28}$$

As for the optimization of the joint codec under the WMSE criterion, it can be shown that the optimal decoder is once again the one that computes the conditional expectation of the source given the received signal

$$\hat{\mathbf{x}} = E(\mathbf{x}/\mathbf{y}) \tag{7.29}$$

It is well known that there are two conditions that must be satisfied for the optimal VQ. One is the partitions, and the other is the centroids. In the following, we will address these two issues.

We first consider the optimal partition condition. For a given source, the decoder in (7.29) and the given channel statistics, the overall WMSE can be expressed as

$$D = \sum_{i=1}^{M} \int_{\Omega_i} \int_n \left[\mathbf{x} - E(\mathbf{x}/\mathbf{y} = \mathbf{s}_i + n)\right]^T \mathbf{W} \left[\mathbf{x} - E(\mathbf{x}/\mathbf{y} = \mathbf{s}_i + \mathbf{n})\right] p(\mathbf{x}) p(\mathbf{n}) d\mathbf{n} d\mathbf{x}$$

It can be proved that the optimal partitions are given by

$$\mathbf{x} \in \Omega_i \quad \text{if} \quad I_i \leq I_m \quad \text{for all} \quad i \neq m \tag{7.30}$$

where $I_m$ denotes a scalar

$$I_m = \int_n \left[\mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p_n(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}{\sum_{i=1}^{N} P_i p_n(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}\right]^T W \left[\mathbf{x} - \frac{\sum_{i=1}^{N} P_i \mathbf{c}_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}{\sum_{i=1}^{N} P_i p_{\mathbf{n}}(\mathbf{s}_m - \mathbf{s}_i + \mathbf{n})}\right] p(\mathbf{n}) d\mathbf{n}$$

On the other hand, the optimal centroids are given by

$$\mathbf{c}_i = \frac{\int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}}{\int_{\Omega_i} \mathbf{x} p(\mathbf{x}) d\mathbf{x}} \tag{7.31}$$

The optimal partitions specified by (7.30) depend on the VQ centroids $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N$, which in turn are related to the partitions, see (7.31). Therefore the GLA procedure can be used to design the optimal VQ until convergence occurs.

## 7.7 MSVQ LSP Codebook Design

There are several approaches to design the MSVQ codebook, such as sequential [57], iterative [92] and joint optimization of the MSVQ [91]. In the sequential MSVQ design algorithm, each stage VQ codebook is designed using a training sequence consisting of quantization error vectors resulting from the previous stages. This training procedure implicitly assumes that all the VQ codebooks in the the following stages are zero.

Optimizing in this manner is clearly sub-optimal due to the correlation property of the different stage errors. The iterative sequential design and simultaneous joint design algorithms are concerned with designing MSVQ to minimize the overall WMSE. In the iterative design, the other stage codebooks are assumed fixed and known, and the objective is to design the current stage VQ codebook to minimize the overall WMSE. Once an initial set of VQ codebooks are obtained, each stage can be re-optimized given the other stage VQ codebooks. Thus, the overall weighted mean square error (WMSE) is minimized rather than the WMSE for encoding the residual vector resulting from the previous stage VQ. Generally speaking, the MSVQ codebook obtained from the sequential design can be used as an initial codebook for the iterative optimization.

The MSVQ codebooks are designed by using the GLA algorithm to minimize average WMSE based on a training sequence. The GLA consists of two steps, one is to cluster the training sequence according to a given set of centroids (or codevectors), and the other is to choose the centroids to minimize the distortion over the particular partitions. We iterate these two steps until the convergency condition is satisfied. For a sequential design algorithm, the current stage residual vector is obtained by subtracting the previous stage reconstruction vectors from the original vector. The residual vector is then quantized by the current stage VQ. A soft decision VQ design procedure based on the equation (7.30) and (7.31) is used to design the current stage VQ starting with the first stage. A condition estimator is used as a source decoder to obtain a reconstructed signal from the received channel signal.

## 7.8 Experimental Results

The first step toward MSVQ design is to collect a set of representative speech vectors. This training set should include 50 to 1000 training vectors per code word. Here

400,000 training vectors were used for the codebook design, and another set of 256,000 vectors were used for testing the performance of the resulting VQ codebook.

The use of soft decision VQ requires the knowledge of the noise pdf. For the AWGN channel this pdf is completely determined by the channel SNR. Once again the SDVQ system is designed for an average expected value of the channel SNR. When used on a given channel, such a SDVQ may suffer some performance degradation due to the noise level mismatch, still it will offer a significant advantage over other procedures.

The vector quantizer design used in sequential decoding soft decision vector quantizer (SD-SDVQ) is the same as that used in the SDVQ system based on symbol-by-symbol decoding (i.e., the vector quantizer codebook optimization does not account for sequential decoding). Although this method will not lead to an optimum SD-SDVQ, the system is easy to implement and gives a good indication of the coding gain that can be obtained by using the sequential decoding technique.

The soft decision vector quantizer (SDVQ) and the channel optimized vector quantizer (COVQ), discussed in Chapter 4, were trained by using the same database as that used for the source optimized vector quantizer (SOVQ) discussed in Chapter 2. The standard QPSK modulation signal set was used to transmit the MSVQ indices. Each VQ index was mapped into a signal formed by concatenating a number of QPSK symbols. For example for codebook size $N = 256$, each index was mapped into a signal obtained by concatenating four QPSK symbols.

Figures 7.9 and 7.10 show the average spectral distortion comparison among SOVQ, COVQ, SDVQ and SD-SDVQ for LSP quantization using two different MSVQ structures. Both MSVQ systems use 24 bits for the quantization of speech LSP parameters. Fig.7.9 corresponds to the case of a 4 stage MSVQ with 6 bits per stage, while Fig. 7.10 corresponds to the case of a 3 stage MSVQ with 8 bits per stage. SDVQ shows the performance improvement over SOVQ and COVQ, particularly for

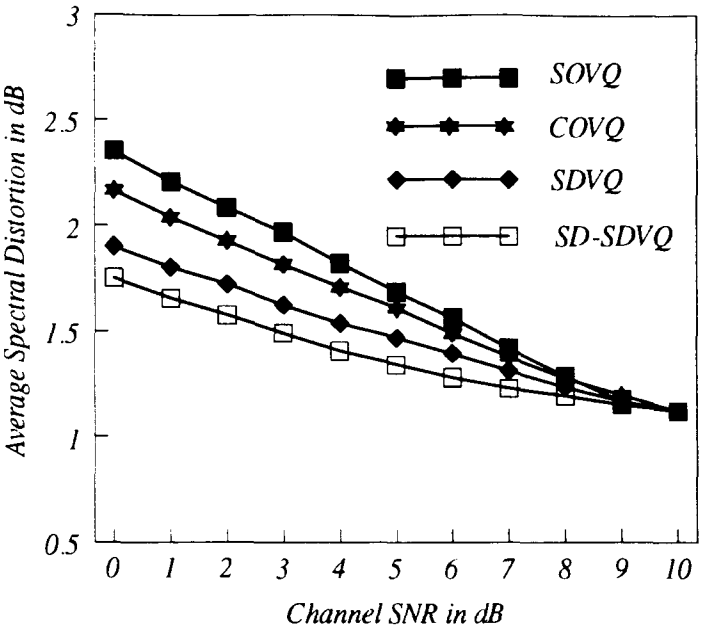


Figure 7.10: Performance of SD-SDVQ, SDVQ, COVQ, and SOVQ (GLA) for quantization of speech line spectral pairs (LSP). Four stages with 6 bits/stage were used. The design SNR is 8 dB for SDVQ and COVQ.

the low channel SNR region. For example, at a channel SNR of 5 dB and for the 3-stage MSVQ, the average spectral distortion of SDVQ is about 0.35 dB less than SOVQ and 0.15 dB less than COVQ. SD-SDVQ further improves the performance by reducing the SD by about 0.15 dB with respect to SDVQ. For example, at a channel SNR of 5 dB and for the 3-stage MSVQ, the average spectral distortion of SDVQ is about 0.35 dB less than SOVQ and 0.15 dB less than COVQ. SD-SDVQ further improves the performance by reducing the SD by about 0.15 dB with respect to SDVQ.

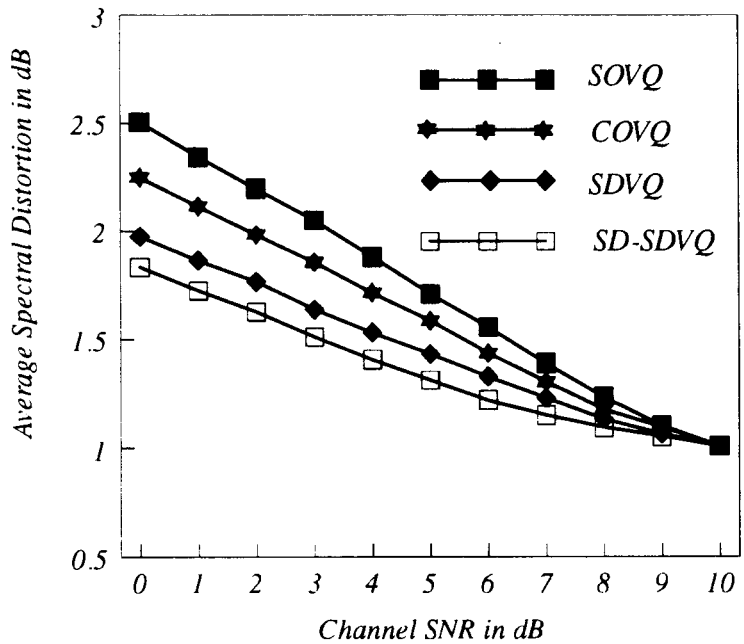


Figure 7.11: Performance of SD-SDVQ, SDVQ, COVQ, and SOVQ (GLA) for quantization of speech line spectral pairs (LSP). Three stages with 8 bits/stage were used. The design SNR is 8 dB for SDVQ and COVQ.

## 7.9 Conclusions

We have presented in this chapter an iterative procedure for joint source and channel coding optimization over a Rayleigh fading channel. We also developed a sequential decoding algorithm for the fading channel. These approaches were applied to design a combined source and channel coding system for a first order Gauss-Markov source. The simulation results show that a significant improvement can be obtained compared with the separately designed system, and that the sequential decoder can further reduce the average distortion by exploiting redundancy in the VQ indices.

We also compared the performance of different multi-stage vector quantizer used to represent the line spectral pair parameters of the speech signal. The multi-stages vector quantizers were designed by SOVQ, COVQ, SDVQ, and SD-SDVQ. A weighted mean square error (WMSE) was used as the optimization criterion to design the SDVQ. The simulation results show that soft decision vector quantizer has a smallest spectral distortion among the source-optimized, the channel-optimized and the soft decision vector quantizer for the LSP parameters.

# Chapter 8

# Summary and Review

## 8.1 Summary

We have considered the problem of joint optimization of source and channel coding for the AWGN and Rayleigh fading channels. In our system, the source signal is compressed by a vector quantizer (VQ), the output of the VQ ,i.e., its index, is mapped directly into a signal vector in the modulation signal space and then it is transmitted over a noisy channel. The optimization variables in the joint source and channel coder are the VQ encoder, the mapping from the VQ index to the modulation constellation, the modulation constellation and the receiver structure. The objective is to minimize the mean-square error (MSE) between the original and the reconstructed source signal, subject to constrains on the average energy and bandwidth.

Based on Bayesian estimation theory, a soft decision vector quantizer (SDVQ) was developed. It is shown that the optimal receiver should calculate the conditional mean of the source signal under the received signal. The output of the receiver is a linear combination of the VQ centroids and the weighting coefficients are nonlinear function of the received signal. Several approximations at various channel SNR were discussed. An iterative algorithm is presented to jointly design the VQ and the modulation signal

set. The algorithm first optimizes the VQ codebook for a fixed signal set, and then optimizes the signal set for a fixed VQ codebook. Iterating these two steps until convergence occurs will provide at least a local optimum solution to the problem. The algorithm has been used to design the VQ and the signal constellation for a first order Gauss-Markov source operating in the AWGN and Rayleigh fading channels. The simulation results indicate that system performance benefits significantly from the joint design, especially for the low channel SNR.

Due to the constraints on the VQ encoder delay and complexity, the source coder can not remove all the redundancy in the source. The residual redundancy is modeled as a first order Markov process. We further developed a sequential decoder algorithm to exploit the residual redundancy to combat the channel noise without the bandwidth expansion. The simulation results show that further improvement can be obtained by using sequential decoding strategy, especially for the Rayleigh fading channel.

## 8.2 Critical Review of the Results and Further Research

We have shown that the soft decision receiver based on the conditional estimate is quite robust in the noisy channel and can improve the system performance when compared to traditional receivers. However, the implementation of the conditional estimate requires more computations than the traditional receivers since it deals typically with the computation of an exponential function. Complexity reduction for the soft decision receiver based on conditional estimate may be a subject of further research.

In this study, the receiver is optimized at the nominal channel SNR and the resulting receiver is then tested in all channel conditions. By using this approach, we

considered the channel mismatch introduced by the channel SNR. In practical systems one may encounter other types of mismatch, for example, mismatch of the noise probability distribution (PDF), when a system designed for a given PDF (say AWGN) is used on a channel with a different noise PDF. The efect of other types of mismatch remains a subject for further study.

The next generation of mobile or personal communication system (PCS) will not only provide voice, but also services such as fax and video. Obviously from a transmission and transport point of view, we want to use a common modulation scheme to transmit different kinds of information. This poses some difficulties to our existing joint source and channel coding design approach, which is intended for a single service. However, one possible remedy is to use the optimization procedure with a proper cost function derived from the traffic intensity of the different services and their respective distortion measures.

Part of the experimental results presented in the thesis were obtained on a first-order Markov source and the typical source SNRs achieved by the proposed system is about 5 dB. Such an SNR is too low for practical applications. There are two reasons why the SNR is so low. First, the Markov source is only a theoretical model and it is more difficult to encode than practical signals. Second, due to the limited computing resources available during the thesis work, relatively low-rate vector quantizers were used (the amount of computations required by a VQ increases exponentially with the rate).

On the other hand, the results obtained for the quantization of LPC parameters show spectral distortions in the range of 1-2 dB. These are practical values and the improvements obtained by the proposed system are significant from practical point of view. LPC quantization is an essential part of most modern speech coders. Extending the work to include a full speech coder is a possible subject for future research. Modern

speech coders use extensively vector quantization for encoding the excitation and the gains, hence extending the work to include these blocks is relatively straightforward.

# References


[1] D.M. Balston and R.C.V. Macario *Ed*, *Cellular Radio Systems*, Artech House, Boston, 1993

[2] D.C. Cox, "Wireless Network Access for Personal Communications", *IEEE Communication Magazine*, vol. 30, No. 12, Dec., 1992

[3] D. M. Balston, "The Pan-European Cellular Technology", IEE Conference Publication, 1988

[4] GSM/ETSI Recommendations, European Telecommunications Standards Institute, Sophia Antipolis, France. 06.10 GSM Full Rate Speech Tanscoding, 1991

[5] I. Gerson and M. Jaswick, "Vector Sum Excited Linear Prediction(VSELP) Speech Coding at 8 kbit/s", Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp 461-464, 1990

[6] Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction, ITU-T Recommendation G.728, Geneva, Switzland, Sept. 1992

[7] Telecommunications: Analog to Digital Conversion of Radio Voice by 4800 bit/s: Code Excited Linear Prediction(CELP) FED-STD-1016, Office of Technical Standards, National Communications System, Washington DC, Nov. 1989

[8] B.S. Atal, V. Cuperman and A. Gersho Ed., *Advance in Speech Coding*, Klumer Academic Publishers, Massachusetts, 1991

[9] B.S. Atal, V. Cuperman and A. Gersho Ed., *Speech and Audio Coding for Wireless and Network Applications*, Klumer Academic Publishers, Massachusetts, 1993

[10] M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP):High Quality Speech at Very Low Bit Rates", IEEE Inter. Conf. on Acoustics, Speech and Signal Processing, Vol.1 25.1.1., 1985

[11] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill Inc., New York, 1979

[12] R.E. Blahut, *Theory and Practice of Error Control Coders*, Addison-Wiely, Mass., 1983

[13] C.E. Shannon, "A Mathematical Theory of Communication", *Bell System Tech. Journal*, Vol. 27, pp.379-423, 623-656, 1948

[14] C.E. Shannon, "Coding Theorem for a Discrete Source with a Fidelity Criterion", IRE Nat. Conf. Record, pp.142-163, 1953

[15] Wozencraft and Jacobs, *Principles of Communication Engineering*, John Wiley and Sons, New York, 1965

[16] T. Fine,"Properties of an optimum digital system and applications", *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 287-296, Oct. 1964.

[17] A. E. Kurtenbach and P. A. Wintz,"Quantizing for noisy channels," *IEEE Trans. Commun.* , vol. COM-17, pp. 291-302, April 1969.

[18] J. K. Wolf and J. Ziv, "Transmission of Noisy Information to a Noisy Receiver With Minimum Distortion," *IEEE Trans. Inform. Theory*, vol. IT-16, No.4, pp. 344-351, July 1970

[19] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 827-838, Nov. 1987

[20] J.D. Gibson and T.R. Fischer,"Alphabet-Constrained Data Compression", *IEEE Trans. Inform. Theory*, IT-28, No. 1, pp. 291-302, April, 1982.

[21] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. COM-27, No. 9, pp. 1644-1659, Nov. 1979

[22] H. Kumazawa, M. Kasahara and T. Namekawa, "A Construction of Vector Quantizers for Noisy Channels," *Electro. Eng. Japan* , Vol.67-B, No.4, pp. 39-45, 1984

[23] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Trans Inform. Theory* , vol. IT-36, No.4, pp. 799-809, July, 1990.

[24] N. Rydbeck and C. W. Sundberg, "Analysis of Digital Errors in Nonlinear PCM", *IEEE Trans. Commun.*, vol. COM-24, Jan. 1976, pp. 59-65.

[25] J. R. B. De Marca, N. Jayant, "An Algorithm for Assigning Binary Indices to the Codevectors of a Multi-Dimensional Quantizer", in Proc. IEEE Int. Conf. Comm., Seattle, June 1987, pp. 1128-1132.

[26] J. R. B. De Marca, N. Farvardin, Y. Shoam, "Robust Vector Quantization for Noisy Channels", in Proc. Mobile Satellite Conf., May 1988, pp. 515-520, JPL Publ. 88-9.

[27] K. Zeger and A. Gersho, "Zero redundancy Channel Coding in Vector Quantization, *IEE Electron. Lett.*, Vol.23, 1987, pp.654-655.

[28] K. Zeger and A. Gersho, "Pseudo-Gray Coding", *IEEE Trans. on Comm.*, vol. 38, no. 12, December 1990, pp. 2147-2158.

[29] P. Knagenhjelm, "How Good Is Your Index Assignment", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis, Minnesota, USA, April, 1993, pp. II 423-426.

[30] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing", *Science*, vol. 220, 1983, pp. 671-680,

[31] W. B. Kleijn and R. A. Sukkar, "Efficient Channel Coding for CELP Using Source Information," *Speech Communication*, vol-12, 1992, pp. 547-566.

[32] J. Hagenauer, W. Seahardri and C.E.W. Sundberg, "The Performance of Rate-Compatible Punctured Convolutional Codes for Digital Radio", *IEEE Trans. Commun.*, vol-38, No.7, 1990, pp. 966-980

[33] J. G. Duanham and R. M. Gray, "Joint Source and Channel Trellis Encoding", *IEEE Trans. Inf. Theory*, vol. IT-27, July 1981, pp. 516-519.

[34] E. Ayanoglu and R. M. Gray, "The Design of Joint Source and Channel Trellis Waveform Coders," *IEEE Trans. Inform. Theory*, vol. IT-33, Nov. 1987, pp. 855-865.

[35] R. Skinnemoen, "A Modulation-Organized Vector Quantizer(MOR-VQ) for joint source and channel coding", Globelcom94, San Francisco, U.S.A, 1994

[36] E. Bedrosian, "Weighted PCM," *IRE Trans. Inform. Theory,* vol. IT-4, March 1958

[37] C. L. Weber, *Elements of Detection and Signal Design*, New York, Springer-Verlag, 1987

[38] M. K. Simon and J. G. Smith, "Hexagonal Multiple Phase-and-amplitude-shift-keyed Signal sets," *IEEE Trans. Commun.*, vol. COM-21, Oct. 1973

[39] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellation in the presence of Gaussian noise," *IEEE Trans. Commun.*, vol. COM-22, January 1974

[40] M. L. Honig, K. Steiglitz and S. A.Norman, "Optimization of Signal Sets for Partial-Response Channels: Part I: Numerical Techniques," *IEEE Trans. Inform. Theory*, Vol. IT-37, No.5, Sept. 1991

[41] R. Steel, C. E. W. Sundberg and W. C. Wong, "Transmission of Log-PCM via QAM over Gaussian and Rayleigh Fading Channels", , *IEE Proc., Vol.134, Pt. F*, No. 6, Oct. 1987

[42] S. P. Lloyd, "Least squares quantization in PCM," unpublished Bell Laboratories Technical Note, 1957. Portions presented at Institute of Mathematical Statistics Meeting, Atlantic City, HJ, Sept. 1957. Published in Special Issue on Quantization, *IEEE Trans. Inform. Theory*, vol. IT-28, pp. Mar. 1982

[43] Y. Linde, A. Buzo, and R. M. Gray, " An algorithm for Vector Quantization Design," *IEEE Trans. Commun.* vol. COM-28, Dec. 1980, pp.84-95.

[44] V. Vaishampayan and N. Farvardin, "Joint Design of Block Source Codes and Modulation Signal Sets," *Computer Science Technical Report*, University of Maryland, Feb. 1990

[45] Feng-Hua Liu, Paul Ho and Vladimir Cuperman, "Joint Source and Channel Coding Using a Non-Linear Receiver", *Proceeding of IEEE ICC'93*, Geneva, Switzerland, June 1993

[46] V. Cuperman, F-H. Liu and P. Ho, " Soft Decision Vector Quantization for Noisy Channels", *IEEE Speech Workshop for Telecommunications*, Saint-Adele, Quebec, Canada, 1993

[47] V. Cuperman, F-H. Liu and P. Ho, "Robust Vector Quantization for Noisy Channel by Soft Decision and Sequential Decoding, *European Trans. Telecommunication*, Oct., 1994

[48] F-H. Liu, P. Ho and V. Cuperman, "Joint Vector Quantization-Modulation Optimization for Noisy Channel", Submitting to IEEE Trans. Communication, 1993

[49] F-H. Liu, P. Ho and V. Cuperman, "Joint Source and Channel Coding Using a Non-Linear Receiver over Rayleigh Fading Channel", Proc. of IEEE Globecom-93, Mini-Communication Theory Conference, Houston, Texas, 1993, U.S.A

[50] F-H. Liu, P. Ho and V. Cuperman, " Sequential Reconstruction of Vector Quantized Signals Transmitted Over Rayleigh Fading Channel, IEEE ICC-94, New Orleans, U.S.A

[51] Gallager, *Information Theory and Reliable Communication*. New York, McGraw-Hill, 1968

[52] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiely & Sons, Inc., New York, 1991

[53] T. Berger, *Rate Distortion Theory: A Mathematical Basis For Data Compression*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971

[54] U. Grenander and G. Szegö, *Toeplitz Forms And Their Applications*. Berkerly, University of California Press, 1958

[55] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, Englewood Cliffs, New Jerser, 1984

[56] J. Makhoul, S. Roucos and H. Gish, "Vector Quantization in Speech Coding", *Proceedings of the IEEE*, Vol. 73, pp. 1551-1588, 1985

[57] A. Gersho and G.W. Gray, *Vector Quantization and Signal Processing*, Klumer Academic Publishers, Boston, 1991

[58] J. Makhoul, "Linear Prediction: A Tutorial Review," *Proceedings of the IEEE*, Vol.63, pp. 561-580, Apr. 1975

[59] J. D. Markel, and A. H. Gray, *Linear Prediction of Speech*, New York, Spring-Verlag, 1976

[60] N. Levinson, "The Weiner RMS Error Criterion in Filter Design and Prediction," *Journal of Mathematics and Physic*, Vol. 25, pp 261-278, 1946

[61] J.R. Deller, J. G. Proakis and J.H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, New York, 1993

[62] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, New Jersey, 1989

[63] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals",*IEEE Trans. Inform. Theory*, vol. IT-28, pp.55-67, 1982

[64] E. Biglieri, D. Divsalar, P. J. McLane and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*, Macmillan Publishing Company, New York, 1991

[65] Van Trees, H. L., *Detection, Estimation and Modulation Theory: Part I*, New York, John Wiley and Sons, 1968

[66] J.G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1983.

[67] N.D. Yacoub, *Foundation of Mobile Radio Engineering*, CRC Press, London, 1993

[68] R.C.V. Macario (Ed.), *Personal & Mobile Radio Systems* , Peter Peregrinus Ltd., England, 1991

[69] R.S. Kennedy, *Fading Dispersive Communication Channels*, John Wiley Inc., New York, 1969

[70] N. Farvardin and V. Vaishampayan, " On the Performance and Complexity of Channel-Optimized Vector Quantizers," *IEEE Trans. Inform. Theory*, vol. IT-37, No.1, pp. 134-137, January 1991.

[71] W.A. Gardner,"Structural Characterization of Locally Optimum Detectors in Terms of Locally Optimum Estimators and Correlators," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 924-932, Nov. 1982

[72] M. V. Eyuboğlu and G. D. Forney, Jr."Trellis Precoding: Combined Coding, Precoding and Shaping for Interference Channels", *IEEE Trans Inform. Theory* , vol. IT-38, No.2, , pp. 105-111, March, 1992.

[73] E. Aarts and J. Korst *Simulated Annealing and Boltzmann Machines*, New York, John Wiley & Sons Ltd., 1989

[74] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing", *Science* vol. 220, pp 671-680,

[75] M.C. Jeruchim, P. Balaban and K.S. Shanmugan, *Simulation of Communication Systems*, Plenum Publishing, New York, 1992

[76] W. H. Press et al., *Numerical Recipes: the Art of Scientific Computing*, Cambaridge, Cambaridge University Press, 1989

[77] S.H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*, Klumer Academic Publisher, Boston, 1994

[78] J.L. Devore, "A Note on the Observation of a Markov Source Through a Noisy Channel," *IEEE Trans. Inform. Theory*, Vol. 20, N0v. 1974, pp. 726-764

[79] Sayood and Borkenhagen,"Use of Residual Redundancy in the Design of Joint Source/Channel Coders", *IEEE Trans. on Communication*, Vol.39, Jun. 1991

[80] N. Phamdo and N. Farvardin, "Optimal Detection of Discrete Markov Sources Over Discrete Memoryless Channels-Applications to Combined Source-Channel Coding," *Systems Research Center Technical Report*, University of Maryland, Feb. 1992

[81] C. Gerlach, "A Probabilistic Framework for Optimum Speech Extrapolation in Digital Mobile Radio", Proceedings of IEEE IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis, Minnesota, USA, April, 1993, pp. II 419-422.

[82] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd Ed., McGraw-Hill Company, New York, 1990

[83] M. Moher and J. Lodge, "TCMP-A Modulation and Coding Strategy for Rician Fading Channels", *IEEE J. Select. Areas Commu.* Vol. 7, No. 91, pp. 1347-1355, Dec., 1989

[84] Y. Tohkura and F. Itakura, "Spectral Sensitivity Analysis of PARCOR Parameters for Speech Data Compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, pp. 273-280, June 1979.

[85] F. Soong and B. Juang, Line Spectrum Pair (LSP) and Speech Data Compression," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1983.

[86] G. Kang and L. Fransen, "Application of Line-Spectrum Pairs To Low-Bit-Rate Speech Encoders," IEEE International Conference on Acoustics, Speech, and Signal Processing , 1985.

[87] G. Kang and L. Fransen, "Experimentation with Synthesized Speech Generated From Line-Spectrum Pairs," *IEEE Transactions on Acoustics, Speech, and Signal Processing* , vol. ASSP-35, April 87.

[88] P. Kabal and R. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, December 1986.

[89] K.K. Paliwal and B. Atal, Efficient Vector Quantization of LPC Parameters at 24 bits/frame," IEEE International Conference on Acoustics, Speech, and Signal Processing , pp. 661-664, March 1991.

[90] W.-Y. Chan, S. Gupta, and A. Gersho, "Enhanced Multistage Vector Quantization by Joint Codebook Design," *IEEE Trans. on Communications*, Vol-COM40, No. 11, pp. 1693-1697, Nov. 1992.

[91] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, V. Cuperman, "Efficient Search and Design Procedures for Robust Multi-Stage VQ of LPC Parameters for 4 kb/s Speech Coding", *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 4, October 1993 pp. 1-13.

[92] F. Tzeng, "Analysis-By-Synthesis Linear Predictive Speech Coding at 2.4 kbit/s," *Proc. Globecom* 89 , pp. 1253-1257, 1989.