Source and Channel Coding Techniques for the MIMO Reverse-link Channel

A Thesis
submitted for the Degree of
Doctor of Philosophy
in the Faculty of Engineering

T. Ganesan
Dept. of ECE, IISc, Bangalore

Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore – 560 012

April 2014

Dedicated to Nithya, Hari, and Hamsini

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deep gratitude to Prof. K.V.S. Hari who initiated me into the world of research and was very supportive till the end of my work. Any countable set of words in the English language cannot adequately express my gratitude towards my thesis advisor, Dr. Chandra Murthy whose *perfection in everything* astonished me many times and motivated me to work towards that goal. He brought rigor and focus in my work which has not only helped me but also has resulted in the existing structure and organization of this thesis.

I would like to thank my internal advisor Dr. Jaiganesh and supervisor Dr. Venugopal at Texas Instruments (TI) India Limited, whose constant support and encouragement made my life easier at work. They gave me the flexibility in managing my work hours and permitted me to publish my research results.

I would also like to thank my ex-supervisors Arvind, C.Srinivasan, Ravi, Subhashish and Nagaraj from TI India for their support during my enrollment. I thank Texas Instruments India Limited for sponsoring my research work as well for providing good facilities for carrying out my research at the office. I also thank the Dean of Engineering at IISc and the IISc administration for supporting the ERP program, without which, I could not have enrolled while working in TI. I would like to thank Dr. Andrew Thangaraj from IITM, Chennai for very insightful discussions on channel coding.

I would like to thank everyone in the Signal Processing for Communications Lab., Department of ECE, IISc for the nice work culture. It is both a fun and a serious research environment. I would like to especially thank Abhay, Bharath, Nandha, Partha, Ranjitha, Sanjeev, Srinivas and Venu for the discussions and good times we had together.

Finally, I would like to thank my family members who had to put up with my extended work hours not only on weekdays but also on the weekends. Most of these hours truly belonged to them.

Abstract

In wireless communication systems, the use of multiple antennas, also known as Multiple-Input Multiple-Output (MIMO) communications, is now a widely accepted and important technology for improving their reliability and throughput performance. However, in order to achieve the performance gains predicted by the theory, the transmitter and receiver need to have accurate and up-to-date Channel State Information (CSI) to overcome the vagaries of the fading environment. Traditionally, the CSI is obtained at the receiver by sending a known training sequence in the forward-link direction. This CSI has to be conveyed to the transmitter via a low-rate, low latency and noisy feedback channel in the reverse-link direction. This thesis addresses three key challenges in sending the CSI to the transmitter of a MIMO communication system over the reverse-link channel, and provides novel solutions to them.

The first issue is that the available CSI at the receiver has to be quantized to a finite number of bits, sent over a noisy feedback channel, reconstructed at the transmitter, and used by the transmitter for precoding its data symbols. In particular, the CSI quantization technique has to be resilient to errors introduced by the noisy reverse-link channel, and it is of interest to design computationally simple, linear filters to mitigate these errors. The second issue addressed is the design of low latency and low decoding complexity error correction codes to provide protection against fading conditions and noise in the reverse-link channel. The third issue is to improve the resilience of the reverse-link channel to fading.

The solution to the first problem is obtained by proposing two classes of *receive filtering techniques*, where the output of the source decoder is passed through a filter designed to reduce the overall distortion including the effect of the channel noise. This work combines the high resolution quantization theory and the optimal Minimum Mean Square Error (MMSE) filtering formulation to analyze, and optimize, the total end-to-end distortion. As a result, analytical expressions for the linear receive filters are obtained that minimize the total end-to-end distortion, given the quantization scheme and source (channel state) distribution. The solution to the second problem is obtained by proposing a new family of error correction codes, termed *trellis coded block codes*, where a trellis code and block code are concatenated in order to provide good coding gain as well as low latency and low complexity decoding. This code construction is made possible due to the existence of a *uniform partitioning* of linear block codes. The solution to the third problem is obtained by proposing three novel *transmit precoding* methods that are applicable to time-division-duplex systems, where the channel reciprocity can be exploited in designing the precoding scheme. The proposed precoding methods convert the Rayleigh fading MIMO channel into parallel Additive White Gaussian Noise (AWGN) channels with fixed gain, while satisfying an average transmit power constraint. Moreover, the receiver does not need to have knowledge of the CSI in order to decode the received data. These precoding methods are also extended to Rayleigh fading multi-user MIMO channels.

Finally, all the above methods are applied to the problem of designing a low-rate, low-latency code for the noisy and fading reverse-link channel that is used for sending the CSI. Simulation results are provided to demonstrate the improvement in the forward-link data rate due to the proposed methods. Note that, although the three solutions are presented in the context of CSI feedback in MIMO communications, their development is fairly general in nature, and, consequently, the solutions are potentially applicable in other communication systems also.

Glossary

Table 1: Abbreviations used in this thesis.

Abbreviation	Description
3GPP	Third generation partnership project
8-PSK	8-ary phase shift keying
downlink	Channel between BS and UT, with BS as the transmitter
forward-link	Same as downlink
i.i.d.	Independent and identically distributed
reverse-link	Channel between UT and BS, with UT as the transmitter
uplink	Same as reverse-link
AWGN	Additive white Gaussian noise
BER	Bit error rate
BLAST code	Bell labs layered Space-time code
BPSK	Binary phase shift keying
BS	Base station (eNodeB as in 3GPP and LTE standards)
CC	Convolutional code
COVQ	Channel optimized VQ
CRC	Cyclic redundancy check
CSI	Channel state information
CSIR	Channel state information at the receiver
CSIT	Channel state information at the transmitter
FDD	Frequency division duplex
FDMA	Frequency division multiple access

Continued on next page

Table 1 – *Continued from previous page*

Notation	Description
IA	Index assignment
LBC	Linear binary code
LDPC code	Low density parity check code
LLR	Log-likelihood Ratio
LTE	Long term evolution
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
MLSD	Maximum likelihood sequence detector
O-STBC	Orthogonal STBC
PAM	Pulse amplitude modulation
QAM	Quadrature amplitude modulation
QPSK	Quadrature phase shift keying
SER	Symbol error rate
SISO	Single-input single-output
SIMO	Single-input multiple-output
SNR	Signal to noise ratio (Normalized for energy per bit)
SOVQ	Source optimized VQ
SQ	Scalar quantization
STC	Space-time code
STBC	Space-time block code
STTC	Space-time trellis code
TCM	Trellis coded modulation
TDD	Time division duplex
TDMA	Time division multiple access
USTC	Unitary STC
UT	User terminal
VQ	Vector quantization

Notation

Table 2: Notation used in this thesis.

Symbol	Description
Capital italic letters, e.g., M	Integers or sets or random variables
Bold small letters, e.g., $\mathbf{g}$	Column vectors
Bold capital letters, e.g, $\mathbf{G}$	Matrices
$\mathbf{g}_i$	i^{th} column of $\mathbf{G}$
g_{ij}	$(i,j)^{\text{th}}$ element of $\mathbf{G}$
$\\|\mathbf{g}\\|_p$	ℓ_p norm of $\mathbf{g}$
$\mathbf{g}^T$	Transpose of $\mathbf{g}$
$\mathbf{g}^H$	Conjugate transpose of $\mathbf{g}$
$\mathbf{G}^H$	Conjugate-transpose of $\mathbf{G}$
$\\|\mathbf{G}\\|_p$	ℓ_p induced norm of $\mathbf{G}$
$\\|\mathbf{G}\\|_F$	Frobenius norm of $\mathbf{G}$
$\vert\mathbf{G}\vert$	Determinant of $\mathbf{G}$
$\text{tr}[\mathbf{G}]$	Trace of matrix $\mathbf{G}$
$diag(x_1, \ldots, x_k)$	Diagonal matrix with $x_1, \ldots, x_k$ as its diagonal elements
$\mathbf{0}$	All zero column vector
$\mathbf{1}$	All ones column vector
$\{\phi\}$	Null set
$\vert X\vert$	Cardinality of the set X
$\lfloor x \rfloor$	Largest integer smaller than or equal to x
$\lceil x \rceil$	Smallest integer larger than or equal to x

Continued on next page

Table 2 – *Continued from previous page*

Notation	Description
$(x)^+$	$\max(0, x)$
$\Re\{x\}$	Real part of the complex number x
$\Im\{x\}$	Imaginary part of the complex number x
$\mathbb{F}(q^p)$	p^{th} extension of $\mathbb{GF}(q)$
$\mathbb{F}_q^n$	An $n-$dimensional vector space in $\mathbb{F}(q)$
$\mathbb{GF}(q)$	Galois field of size q
$\mathbb{N}$	Natural numbers $1, 2, 3, \ldots$
$\mathbb{Z}$	Integer numbers $\ldots, -2, -1, 0, 1, 2, \ldots$
$\mathbb{R}$	Algebraic field of real numbers
$\mathbb{C}$	Algebraic field of complex numbers
$\Pr(X = x)$	Probability of the event $X = x$
$f_X(x)$	Probability Density Function (PDF) of X
$F_X(x)$	Cumulative Distribution Function (CDF) of X
$\mathbb{E}_{X,Y}[f]$	Expectation of function $f(X, Y)$ with respect to
	the joint distribution of X and Y
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{CN}(\mu, \sigma^2)$	circularly symmetric complex Gaussian distribution
	with mean μ and variance σ^2
$\mathcal{C}(n, k, d)_p$	Block code made of $n-$tuples,
	with minimum distance d and cardinality p^k
$\mathcal{C}(n, k)$	Block code made of $n-$tuples
	and with cardinality 2^k in $\mathbb{GF}(2)$
$d_{min}(\mathcal{C})$	Minimum distance of a code $\mathcal{C}$
$\mathcal{W}_H(\mathbf{c}_1)$	The Hamming weight of a code word $\mathbf{c}_1$
$\mathcal{W}_E(\mathbf{c}_1)$	The Euclidean weight of a code word $\mathbf{c}_1$
$(\mathbf{c}_1, \mathbf{c}_2)$	Concatenation of two code words $\mathbf{c}_1$ and $\mathbf{c}_2$
$\mathbf{c}_1 * \mathbf{c}_2$	Element-by-element multiplication of two vectors $\mathbf{c}_1$ and $\mathbf{c}_2$
$\mathcal{D}_H(\mathbf{c}_1, \mathbf{c}_2)$	The Hamming distance between codewords $\mathbf{c}_1$ and $\mathbf{c}_2$

Continued on next page

Table 2 – *Continued from previous page*

Notation	Description
$x \approx y$	x is approximately equal to y
$X_n \doteq X$	X_n is asymptotically equal to X when $n \to \infty$
$X \triangleq Y + Z$	X is defined as $Y + Z$
$f_1(x) \otimes f_2(x)$	Convolution of two functions $f_1(x)$ and $f_2(x)$
$\mathcal{H}_2(x)$	Binary entropy, defined as $-x \log_2 x - (1-x) \log_2(1-x)$
$Q(x)$	Q-function, defined as $\frac{1}{2}\mathrm{erfc}(x/\sqrt{2})$, where $\mathrm{erfc}(x)$ is the complementary error function

Chapter 1

Introduction

"May I speak the truth of Brahman. May I speak the truth. May it protect me. May it protect my teacher. Om. Peace peace peace." -**Aitareya Upanishad**

1.1 MIMO Communication Systems

In wireless communication systems, the use of multiple antennas, also known as Multiple Input Multiple Output (MIMO) communications, is now a widely accepted and important technology for improving their reliability and throughput performance. Following the seminal works of Telatar [1] and Foschini [2], multiple antennas have not only been extensively studied by the academia, but also successfully implemented by the industry. Many current day broadband wireless access systems such as IEEE 802.11n, 3GPP-LTE and LTE-advanced support the use of multiple antennas to improve the spectral efficiency and resilience to signal fading conditions.

Figure 1.1 shows a simple block diagram of a MIMO communication system with N_t transmit antennas and N_r receive antennas. One of the key benefits of using multiple antennas is that the capacity of a multiple antenna link is known to increase linearly with the minimum of the number of antennas at the transmitter and receiver [1,2]. Hence, multiple antennas can be used to increase the number of independent signaling dimensions (also known as the multiplexing gain), and hence, the rate, of the communication link. Another important, and related, feature of MIMO systems is that the use of multiple antennas can provide diversity benefits, thereby

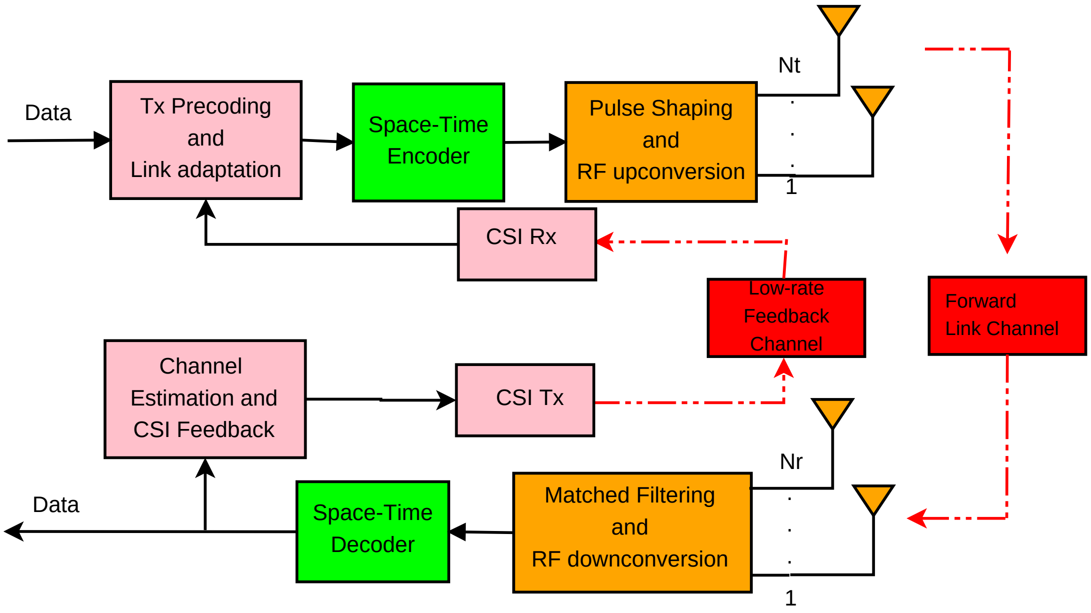
Data
Tx Precoding and Link adaptation
Space-Time Encoder
Pulse Shaping and RF upconversion
Nt
1
CSI Rx
Low-rate Feedback Channel
Forward Link Channel
Channel Estimation and CSI Feedback
CSI Tx
Nr
Data
Space-Time Decoder
Matched Filtering and RF downconversion
1

Figure 1.1: MIMO communication system block diagram.

improving the resilience to fading, since the signal arrives at the different receive antennas through independent paths. Over the past two decades, an enormous amount of research has gone into the design, analysis and optimization of MIMO communication systems, and it remains an active area of research to this day.

It is known that a significant improvement in the capacity of a MIMO communication link is possible when the Channel State Information (CSI) is available at both the transmitter and receiver [1]. When the channel undergoes frequency non-selective fading, as in, for example, narrow-band communication systems, the CSI consists of a complex-valued matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, whose $(i, j)^{\text{th}}$ entry contains the channel gain from the j^{th} transmit antenna to the i^{th} receive antenna, and where N_r and N_t represent the number of antennas at the receiver and transmitter, respectively. The CSI can be estimated at the receiver by sending a known training sequence from the transmitter. To acquire CSI at the transmitter, one has to either send a training sequence in the reverse direction, which is possible in Time Division Duplex (TDD) systems, or send quantized CSI on a low-rate feedback channel from the receiver to the transmitter, which is applicable to both TDD and Frequency Division Duplex (FDD) systems. Both reverse channel training and quantized feedback options are supported in many standards, for example, in the IEEE 802.11n standard [3]. The low-rate reverse-link feedback channel needs to be designed

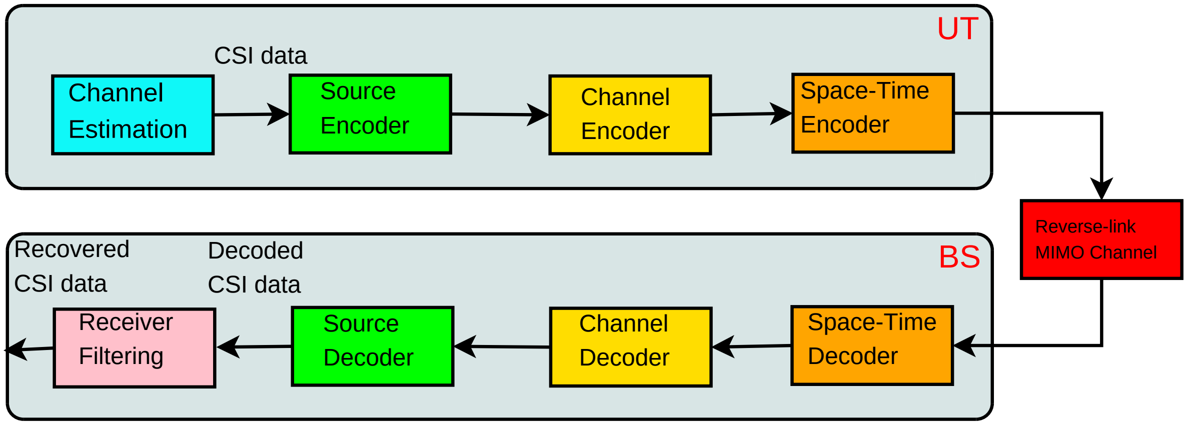

Figure 1.2: Block diagram of the low-rate CSI feedback channel.

with care, as the performance of the MIMO link depends critically on the accurate and timely availability of the CSI at the transmitter. Consequently, the problem of CSI feedback has been studied in detail in recent years (see, for example, [4] for an excellent survey of related literature). The design of the low-rate feedback channel and proposing schemes for improving its reliability and resilience to noise and fading impairments is the focus of this thesis. Figure 1.2 shows a block diagram of the feedback channel for conveying the CSI from the receiver to the transmitter. Designing the feedback channel involves specifying each of the operations in the diagram to obtain as high a quality of received CSI at the base station transmitter as possible. We start with a brief overview of the existing MIMO CSI feedback schemes, and discuss their relative merits and shortcomings.

1.1.1 CSI Feedback in MIMO Systems

In this subsection, we discuss some of the design issues associated with conveying the CSI from the receiver to the transmitter over a low-rate feedback channel that have been addressed in recent literature. The feedback channel design is essentially a problem of communication link design, where the goal is to reconstruct the CSI at the transmitter with as high fidelity and as low latency as possible. The CSI can itself be considered to be a random source that needs to be source-encoded, or quantized, to a finite number of bits prior to transmission. Popular techniques used to compress the CSI include scalar/vector quantization based approaches [4–6] and Grassmannian manifold based approaches [7]. The quantization-based approach has found widespread acceptance in the industry also, and has made its way into present-day communications standards such as the 3GPP [8], LTE-A [9,10], IEEE 802.11n [3]. We now briefly review the codebook based feedback methods supported in these standards.

1.1.1.1 IEEE 802.11n

The IEEE 802.11n standard is an Orthogonal Frequency Division Multiplexing (OFDM) standard, wherein the channel state consists of the channel matrix corresponding to each subcarrier, and there are 108 such subcarriers. Thus, quantization methods need to be penurious with the number of bits allocated for representing the channel matrix on each subcarrier, since the total feedback overhead is the product of the bits used per subcarrier and the number of subcarriers. In the IEEE 802.11n, three different explicit feedback methods are supported:

(a) *CSI matrix feedback method:* In this method, the receiver quantizes the maximum absolute real or imaginary part of the entries of the channel matrix in each subcarrier using 3 bits. Then, the other entries in the channel matrix are normalized by the largest in magnitude entry and the resulting normalized values are quantized using 4 to 8 bits per real and imaginary part of the entry.

(b) *Non-compressed beamforming matrix feedback method:* In this case, the receiver first computes a precoding matrix with orthonormal columns, and then sends a 2 to 8 bit quantized version of the precoding matrix. The quantization is based on a codebook consisting of candidate precoding matrices specified by the standard.

(c) *Compressed beamforming matrix feedback method:* In this case, the receiver sends quantized Givens rotation angles that are used to represent the orthogonal precoding matrix corresponding to the current channel state.

Of the three, the CSI matrix feedback method is the most straightforward approach, as it allows the transmitter to construct the precoding matrix based on its received quantized CSI, but it requires a larger data overhead compared to the other two methods. The latter methods are more efficient in terms of the feedback overhead, but are directly tied to the specific precoding scheme employed by the transmitter.

1.1.1.2 IEEE 802.16e/WiMax

This is also an OFDM based standard. The standard provides several codebooks for 2, 3 and 4 transmit antennas. Precoding matrices using 3 bits per codebook entry are constructed using Grassmannian codebooks. For higher number of transmit antennas, 6 bit codebooks are used, which are obtained using a generator vector that is multiplied by Householder reflection matrices and a diagonal matrix, to save the memory space required for storing the codebooks.

1.1.1.3 3GPP-LTE

The 3GPP LTE is also an OFDM-based standard, which supports codebook based CSI feedback with 2 or 4 transmit antennas. In the 2 transmit antenna case, a codebook with 3 precoding matrices and 2 antenna selection vectors is supported. With 4 transmit antennas, a 4 bit codebook is supported for 1 to 4 spatial streams. The number of spatial streams dictate the number of orthogonal columns in the precoding matrix. To reduce storage space, the codebook entries of various spatial streams are designed to have nested structure, with the codebook for larger number of spatial streams including the codebook with lower number of spatial streams. The codebooks are used by the base station to adapt the rank of the precoding matrix based on the link quality. This standard also supports CSI feedback for multi-user MIMO precoding. An issue with multi-user precoding is that with the limited number of feedback bits, the quantization error creates multi-user interference, which limits the performance especially at high Signal to Noise Ratio (SNR). Hence, the codebook size need to be increased with the SNR. However, designing and practically implementing large codebooks remains a challenge in terms of storage and encoder complexity.

We now discuss several important aspects of MIMO CSI feedback link design.

1.1.2 Source Coding for Noisy Channels

In the MIMO feedback channel, the CSI can be considered as a random source that needs to be compressed/quantized at the user terminal, transmitted over a noisy feedback link, and recovered at the base station, with as low a distortion as possible. This, in turn, can be viewed as a problem of source compression for noisy channels. In particular, when the state of the reverse-link channel is independent of that of the forward-link (for example, in FDD systems), the source encoder is generally unaware of the channel error rate at the time of encoding. This necessitates the study of source coding schemes for noisy channels.

In a typical Vector Quantization (VQ) based source compression system, given a source instantiation, the source encoder looks for the vector in a codebook of vectors that is closest to the source instantiation.[1] The encoder outputs the index of the closest vector, which is then sent to a source decoder over a noisy channel. Due to possible channel errors, the codeword index could be received in error. A conventional source decoder picks the codeword corresponding to the received index as the estimate of the source instantiation. However, due to the channel induced index errors, the overall distortion incurred at the receiver with such a *source optimized VQ* (SOVQ) approach can be considerably higher than the distortion purely due to the source compression (i.e., for error free channels) [11,12].

In the literature, one of the two approaches have been used to mitigate the effect of the noisy channel: optimum Index Assignment (IA) [13–15], and Channel Optimized VQ (COVQ) [16,17]. The former involves mapping codewords to transmit indices such that the most probable error events result in codewords which are close to the codeword corresponding to the correct index. In COVQ, the distortion metric is changed to be the expected distortion after accounting for possible index errors. The codebook is optimized to minimize this expected distortion, resulting in a channel-optimized set of codewords and encoding regions. It is also possible to use IA and COVQ simultaneously to obtain robustness.

It is well-known that optimum index assignment is an NP complete problem and, consequently, sub-optimal methods for IA have been proposed [11, 12, 15, 17–20]. IA and COVQ are techniques that are designed for discrete memoryless channels. Their natural extension for continuous channels (such as additive white Gaussian noise (AWGN) channel) has also been explored, and is termed as Soft-Decision VQ (SDVQ). Here, a soft-metric (such as the Log-Likelihood Ratio (LLR)) is used to estimate the source instantiation, and the expected distortion (after averaging over the noise statistics) is used to define a new set of encoding regions [21,22] at the transmitter. However, these methods cannot be used in the case of the MIMO reverse-link channel, as both the methods require the knowledge of the channel statistics to be available to the encoder in order to be implementable. This is not possible, especially in FDD systems, where the forward and reverse channels are independent.

[1] The notion of "closest vector" will be defined more precisely later in the thesis.

Table 1.1: Examples of transmission schemes for MIMO channels, where CSIR refers to the availability of the perfect CSI at the receiver and CSIT refers to the availability of the perfect CSI at the transmitter.

CSIT, no CSIR	CSIT, CSIR
Zero-forcing Vector perturbation O-STBC and QR-based precoding (see Chapter 4)	BLAST code STBC STTC Spatial multiplexing
no CSIT, no CSIR	**no CSIT, CSIR**
USTC Linear dispersion codes	STBC BLAST code STTC

1.1.3 Channel Coding for MIMO CSI Feedback

Note that the reverse-link channel is also a MIMO fading channel. This naturally raises the question of what technique, or techniques, can offer the best performance for conveying the CSI back to the transmitter. Typically, it is of interest to consider coding and transmission schemes that offer as high a diversity order as possible. The specific scheme used depends on the availability of the state of the reverse-link channel at the transmitter and receiver. The main techniques employed in the literature under different assumptions on the availability of the CSI at the transmitter (user terminal) and receiver (base station) are listed in Table 1.1. Further challenges in the design of diversity transmission schemes for the MIMO reverse-link channel will be discussed in the next section.

Note that the MIMO reverse-link channel is typically a fixed-rate channel, i.e., the channel is quantized using a given, fixed number of bits, and the finite-bit representation of the CSI is sent back to the transmitter. In this scenario, an important metric of the performance of the feedback link is its diversity order. Loosely speaking, the diversity order is the limit of the slope of the probability of error versus SNR curve plotted on a log-log scale, as the SNR goes to infinity.

Next, we discuss the challanges in designing the MIMO CSI feedback system.

1.2 Challenges in MIMO CSI Feedback Design

As mentioned earlier, in typical MIMO communication systems, the CSI is sent from the receiver to the transmitter over a low-rate feedback (control) channel. Hence, one challenge is to compress, or quantize, the available CSI at the receiver using a small number of bits. In the literature, methods based on fixed-rate lossy source coding such as Scalar Quantization (SQ) and Vector Quantization (VQ) have been studied [23]. Different aspects of the CSI feedback such as the impact of errors in the feedback link, delay, etc. have also been considered. However, when the receiver uses a highly efficient source coding technique to compress the available CSI, the resulting data bits could be very sensitive to noise introduced when they are transmitted over the feedback channel. Moreover, the channel error rate may not be known at the receiver prior to transmission. Hence, one challenge in the MIMO CSI feedback design is to come up with techniques to mitigate the errors in the CSI at the transmitter introduced due to the transmission of the quantized CSI data over a noisy feedback link.

One way to handle errors in the feedback link is to use a channel code. In designing such a code, one has to keep in mind the requirements of keeping the code length short and the decoding algorithm simple. Otherwise, the delay in receiving and decoding the CSI at the transmitter would render the feedback practically useless. Hence, a challenge is to come up with low latency codes that offer good coding gain with low decoding complexity.

Another challenge in the design of MIMO CSI feedback links is that the signal sent over the feedback channel is related to handling the adverse effects of fading and multipath in the wireless link. In particular, in TDD systems, due to channel reciprocity, the forward and reverse-link channel are the same (or very nearly the same). In this case, if the receiver were to acquire CSI through the forward-link training, an interesting challenge is to design the CSI feedback signaling scheme to exploit the receiver's knowledge of the channel, to efficiently feed back the CSI to the transmitter.

In summary, a well-designed MIMO CSI feedback communication scheme should (a) compress the CSI using as few data bits as possible; (b) encode the data bits using a channel code that is of short length and that admits low complexity, fast decoding at the receiver; (c) employ a transmission scheme that offers as large a diversity order; and (d) use filtering or other mechanisms at the receiver to mitigate the excess distortion introduced when a highly compressed source is transmitted over a noisy channel. Meeting these challenges would improve the performance of the reverse-link feedback channel, which would result in accurate and up-to-date availability of CSI at the transmitter. This, in turn, translates to an improved forward-link data rate and/or BER performance in the system. We now briefly present the main contributions of this thesis in light of the above discussion.

1.3 Contributions

In this thesis, we consider the design of the MIMO reverse-link CSI feedback communication channel, and design source coding, channel coding and diversity schemes to improve its performance. Our specific contributions are as follows:

- The first problem we address is the source compression of CSI from the receiver to the transmitter[2] in noisy feedback channel conditions. We propose receiver-only adaptation methods for minimizing the end-to-end average distortion of a fixed-rate source quantization system. For the source encoder, both Scalar and Vector Quantization (SQ and VQ) are considered. The codebook index output by the encoder is sent over a noisy discrete memoryless channel whose statistics are unknown at the transmitter. Due to the latter assumption, the index assignment is considered to be random, which leads to an equivalent symmetric error channel for the index transition probabilities. At the receiver, the code vector corresponding to the received index is passed through a linear receive filter, whose output is an estimate of the source instantiation. Under this setup, an expression for the average Weighted Mean Square Error (WMSE) between the source instantiation and the reconstructed vector at the receiver is derived using high resolution quantization theory. Also, a closed-form expression for the optimum linear receive filter that minimizes the average WMSE is derived. The generality of framework developed is further demonstrated by theoretically analyzing the performance of other adaptation techniques that involve the transmitter also, such as joint transmit-receive linear filtering and codebook scaling. Monte Carlo simulation results validate the theoretical expressions, and show that, depending on the channel statistics, a significant improvement in the WMSE can be

[2] Here, by "receiver" and "transmitter" we refer to the reverse-link receiver (which is the same as the forward-link transmitter) and the reverse-link transmitter (which is the same as the receiver of the forward-link), respectively.

obtained using the proposed linear receive filtering technique.

As an extension of this work for non-symmetric error channels with a given (fixed) index assignment, we present a novel method for analyzing the total distortion as a convex combination of the distortion under random IA and the distortion under what we call *ideal* IA conditions. Using this, we derive expressions for the optimum receive filter that minimizes the end-to-end distortion for the given channel, and analyze its distortion performance.

A third contribution in this context is that we propose a new receive processing scheme that is applicable to continuous channels such as the AWGN channel, which we term semi-hard decision VQ. Here, the receiver makes uses the log likelihood ratios of the received data bits to declare data bits to be in erasure or make hard decisions. The receiver then uses the decoded index, possibly with data bits marked as erased, to estimate the transmitted code vector. We theoretically analyze this scheme for both random IA as well as specific IA, and show how to pick the erasure threshold to minimize the distortion performance. We compare the performance of this scheme with linear receive filtering, conventional source optimized VQ and channel optimized VQ, to demonstrate the performance gains under low SNR conditions. These topics are covered in Chapters 2, 5 and Appendices A, B.

- The second problem we address is the design of low latency error correction codes. We present a new family of block codes referred to as Trellis Coded Block (TCB) codes, which are built using a trellis code and a linear block code (LBC). The TCB code (TCBC) construction is based on an algebraic structure inherent to many LBCs, which allows one to partition an LBC into sub-sets with a constant distance between every pair of code words in the subset. The proposed uniform sub-set partitioning is used to increase the minimum distance of the code, as in trellis coded modulation (TCM). However, unlike conventional TCM, the coding and modulation steps are separated in TCBC. An advantage of this construction is that it can be applied to both discrete as well as continuous channels, while conventional TCM is typically designed for continuous channels. The proposed TCBC is shown to be useful in a variety of applications including forward error correction, low rate quasi-orthogonal sequence generation, lattice code construction, etc. Moreover, the encoder and decoder for TCBC are realized using off-the-shelf trellis and block encoders and trellis decoders. Simulation results demonstrate the performance benefits offered by the TCBC in a variety of applications, and compares them to other existing state of the art codes. These topics are covered in Chapters 3, 5 and Appendices C, D.

- Our third contribution addresses the problem of maximizing the diversity order achievable in the MIMO feedback channel when the transmitter (user terminal) has knowledge of the CSI. Transmission of data when CSI is available at the transmitter arises, for example, in a TDD system, where the forward and reverse channels are the same. Hence, once the receiver acquires the CSI through an initial round of training on the forward-link channel, the reverse-link feedback channel corresponds to a scenario where the CSI is available at the transmitter but not at the receiver. Here, we propose novel transmit diversity techniques for Rayleigh fading MIMO systems when CSI is available at the transmitter, but not at the receiver. Our proposed precoding schemes convert the fading MIMO channel into a fixed-gain AWGN channel, while satisfying an average power constraint. Hence, the proposed precoding schemes achieve an infinite diversity order, which is in sharp contrast with perfect CSIR based schemes, which at best achieve a finite diversity order. Moreover, the proposed schemes are simple-to-implement and facilitate single-symbol maximum likelihood decoding at the receiver. We extend the first precoding scheme to the multiuser Rayleigh fading Multiple Access Channel (MAC) and the third precoding scheme to the multiuser Rayleigh fading MAC, broadcast (BC) and interference channels (IC). We show that the proposed schemes convert the fading MIMO channel into fixed-gain and parallel AWGN channels in all three cases. Monte Carlo simulations illustrate the significant performance improvement obtainable from the proposed precoding schemes compared to existing diversity techniques.

- All the above methods are applied to the MIMO reverse-link CSI feedback channel, to study how the techniques work when implemented in a practical communication system. To this end, we construct an end-to-end simulation platform that includes all the source coding, receive filtering, channel coding and transmit diversity techniques presented in this thesis. This comprehensive setup allows us to evaluate the interoperability and performance of different combinations of the proposed techniques. The quality of the CSI reconstructed at the base station, measured both in terms of the mean squared error and the down-link data rate, are used as performance metrics for numerical evaluation. We demonstrate the performance improvement from the proposed techniques, using Monte Carlo simulations. This topic is covered in Chapter 5.

We note that although we have presented the above techniques in the context of the MIMO reverse-link CSI feedback channel, they are directly applicable to many other communication systems. The transmit diversity schemes we propose can be used whenever the transmitter has channel state information, as in, for example, TDD systems with receiver-initiated training. Also, the receive filtering techniques are applicable, for example, for alleviating the distortion due to channel errors in data storage systems, since the error rate of the storage system is generally not known at the time of encoding. Finally, the proposed TCBC can be used in any application that calls for low latency and low complexity codes with good coding gain.

Publications

The following publications have resulted from the work presented in this thesis:

Journal papers

1. Ganesan Thiagarajan and Chandra R. Murthy, "**Trellis coded block codes: Design and Applications**", *Journal of Communications*, Academic Press, vol. 7, no. 1, pp. 73-85, Jan. 2012.

2. Ganesan Thiagarajan and Chandra R. Murthy, "**Linear Filtering Methods for Fixed Rate Quantization with Noisy Symmetric Error Channels**", *IET Signal Processing*, vol. 7, no. 9, pp. 888-896, Dec. 2013.

3. Ganesan Thiagarajan and Chandra R. Murthy, "**Novel Transmit Precoding Methods for Rayleigh Fading Multiuser TDD-MIMO Systems with CSIT and no CSIR**", *IEEE Transactions on Vehicular Technology*, accepted for publication with minor changes, Mar. 2014.

Conference papers

1. T. Ganesan and Chandra R. Murthy, "**Trellis coded block codes and Applications**", *Proc. National Communication Conference (NCC)*, Guwahati, Jan. 2009.

2. T. Ganesan and Chandra R. Murthy, "**Receiver only Optimized Semi-Hard Decision VQ for Noisy Channels**", *Proc. IEEE Global Communications Conference (Globecom)*, Hawaii, Dec. 2009.

3. Ganesan Thiagarajan and Chandra R. Murthy, "**Novel Precoding Methods for Rayleigh Fading Multiuser TDD-MIMO Systems**", *Proc. IEEE International Conference on Communications (ICC)*, Sydney, Australia, Jun. 2014.

1.4 Organization

This rest of the thesis is organized as follows.

- Chapter 2 describes the design, optimization and analysis of a linear filter at the receiver in order to mitigate the excess distortion incurred in a fixed-rate source compression scheme due to index errors introduced by the noisy channel. This chapter presents two types of techniques: one suitable for hard-decision decoders and another for soft-decision decoders. The necessary material from high resolution analysis and other related derivations for this chapter are given in Appendix A. An alternate method for deriving the optimum receive filter is provided in Appendix B.

- Chapter 3 introduces the low latency and low decoding complexity error correction code we propose, which we term as a *Trellis Coded Block Code* (TCBC). The proofs of the theorems and lemmas in this chapter are covered in Appendix C. Three different applications of TCBC are illustrated in Appendix D.

- Chapter 4 gives three novel transmit diversity schemes which not only employs computationally simple algorithms at the transmitter but also admits low complexity optimal decoding at the receiver. Some of the mathematical derivations pertaining to this chapter are detailed in Appendix E.

- Chapter 5 applies all of the methods discussed in earlier chapters to the MIMO reverse-link channel for designing the low-rate noise-resilient feedback channel for transmitting the channel state information (CSI) to the transmitter. End-to-end simulation results are presented that show the joint benefits of using the new code, the new diversity technique with CSI data available only at the transmitter, and the receive filtering technique to minimize the overall distortion. Further, the benefit of minimizing the distortion on the received CSI on the data rate performance of the forward-link is quantified.

- Finally, Chapter 6 concludes the thesis and lists the possible future extensions.

Chapter 2

Receive Filtering for Source Coding with Noisy Channels

"Lead us from the unreal to the real. Lead us from darkness to light. Lead us from death to immortality. Om Shanthi, Shanthi, Shanthi!" - **Brihadaranyaka Upanishad**

2.1 Introduction

2.1.1 Motivation and Prior work

One of the methods used for reducing the data rate in the CSI feedback channel is to use an efficient source encoder to compress the CSI. Vector Quantization (VQ) based source coding is one of the popular techniques for lossy source compression, for two main reasons: simple off-the-shelf algorithms are available for designing locally optimum code books that minimize the average distortion, and the performance of VQ can be accurately characterized using high rate quantization theory [24, 25]. However, the performance of VQ-based source compression can be very sensitive to errors introduced when the index output by the encoder is transmitted over a noisy channel (compared to the distortion incurred when the channel is error-free) [11, 12].

VQ over discrete memoryless channels have been studied extensively in the literature, and two dominant noise mitigation approaches have emerged: *optimum index assignment* (optimum IA) [13–15], and *channel optimized VQ* (COVQ) [16,17]. The former involves mapping codewords to transmitter indices such that the most probable error events result in codewords which are close to the codeword corresponding to the correct index. It is well-known that optimum IA is an NP complete problem, and several sub-optimal methods have been proposed [11, 12, 15, 17–20]. In COVQ, the distortion metric is changed to be the expected distortion after accounting for possible index errors, and a new set of codewords and encoding regions are designed so that the overall expected distortion is minimized. IA and COVQ are techniques that work well for discrete memoryless channels, and can also be used simultaneously. Their natural extension for continuous channels (such as the additive white Gaussian noise (AWGN) channel) has also been explored, and is termed as *soft-decision VQ* (SDVQ). Here, a soft-metric (such as the bit log-likelihood ratio (LLR)) is used to estimate the source instantiation, and the expected distortion after averaging over the noise statistics is used to define a new set of encoding regions at the transmitter [21, 22]. In [26], a COVQ that exploits the soft-decision information from the channel's output for Rayleigh fading channels was proposed. High-rate analysis of Fixed Rate Quantizer (FRQ) for *noiseless* channels has been studied by many authors [11, 24, 25, 27, 28]. The high-rate analysis has also been extended to the noisy symmetric error channel [29–31]. Similar analysis was extended to a fading symmetric error channel in [32] where numerical evaluations were used for averaging over the channel realizations. However, there is little past work on the performance of receiver-only techniques for FRQ. Also, the aforementioned techniques for improving the distortion performance of FRQ with noisy channels suffer from two main drawbacks. First, they require knowledge of the channel statistics at the transmitter, which may not always be available. Second, they are computationally intensive to optimize (e.g., COVQ or IA) when the number of quantization bits is large and/or the channel statistics change over time. Moreover, in some applications such as recording the compact disc, or the reverse-link feedback of channel state information in multiple antenna systems, the channel statistics are not known at the time of recording/transmission. This motivates the design of techniques for reducing the average distortion that can be implemented solely at the receiver. *Receive filtering*, the focus of our study in this chapter, is a simple adaptation technique that can be implemented at the receiver only, and can help in reducing the average distortion performance of FRQ-based source compression schemes.

2.1.2 Contributions

- In this chapter, we propose, analyze, and optimize two classes of receive filters for minimizing the average end-to-end distortion of both VQ and SQ under various channel models. That is, we obtain a closed form expression for the end-to-end distortion, by approximating the total distortion (average WMSE), when a Linear Receive Filter (LRF) is applied after the source decoder with *random IA* (defined in the next section), using high-rate quantization theory. We then optimize the LRF to minimize the average WMSE. The novelty in our approach lies in combining the minimum mean square error (MMSE) estimation formulation and high resolution analysis for obtaining an analytical expression for the optimum LRF. We derive expressions for the performance of an LRF which minimizes the approximate MSE distortion, with both SQ and VQ. We show that using the proposed LRF results in a lower average WMSE distortion compared to the no-filtering case. Then, we compare the performance of the LRF with combined transmitter based adaptive schemes, especially optimum linear transmit-receive filtering and scaled codebook [33] for showing the benefits of the proposed algorithm. Note that, transmit-receive filtering and scaled codebook method involve adaptation of the encoding scheme at the transmitter. Hence, these schemes require knowledge of the channel statistics at the transmitter also. Through numerical evaluation, it is shown that LRF works as well as COVQ and optimum linear transmit-receiver filter at low and medium SNR where the distortion due to channel errors has large impact on the total distortion. The high rate quantization approximations used in the derivation is given in Appendix A. An alternate derivation for the optimum LRF using a conventional optimization approach [34] is given in Appendix B.

- Another class of simple receive filtering is introduced for continuous memoryless channels. This new method termed semi-hard-decision vector quantization (SHDVQ), is proposed and analyzed for both random IA and good IA. One of the main differences in the proposed method compared to the past work in [21, 22] is that, the soft-metric is not used to recompute a *soft output* codeword (also known as an estimation based decoder, as against a detection based decoder) for every received symbol. Instead the receiver first performs semi-hard-decision decoding and declares some bits as erasures. The other bits may be received correctly or in error. The resulting index is then mapped to a codeword chosen based on IA and the erasure bit locations. The proposed method works for both discrete channels with possible errors and erasures, and for continuous channels, by applying a threshold based hard decision on the LLRs to declare erasures. To keep the receiver simple, a threshold is employed on the LLRs to declare erasures and the decoder outputs a linear combination of codewords based on the erasure location and IA. The erasure threshold that minimizes the expected distortion after accounting for channel errors and erasures is computed from the analytical expressions derived in the sequel.

- A novel performance analysis of the average approximate MSE distortion for any given IA is presented. That is, for any specific IA, we propose to express the total distortion as a convex combination of the distortion with *ideal IA* and *random IA* (defined later). The specific IA employed determines the factor used in the convex combination, and needs to be numerically evaluated only once for a given IA and the number of quantization bits B. The analytical framework and tools presented in this chapter form a powerful technique for studying and optimizing the high rate performance of VQ with a specific IA for noisy channels.

2.2 Problem Setup

We consider a random n-dimensional source vector $\mathbf{x}$ with zero mean, variance σ_x^2 and continuous probability density function (pdf) $f_{\mathbf{x}}(\mathbf{x})$ over a compact support $\mathcal{D}_{\mathbf{x}} \subset \mathbb{R}^n$. The source encoder maps $\mathbf{x}$ to the closest vector $\mathbf{y}$ in a codebook $\mathcal{C}$ of cardinality N, with respect to the WMSE distortion $d(\mathbf{x}, \mathbf{y}) \triangleq (\mathbf{x}-\mathbf{y})^T \mathbf{W} (\mathbf{x}-\mathbf{y})$, where $\mathbf{W}$ is an $n \times n$ symmetric positive definite matrix. Also, for SQ, the matrix $\mathbf{W}$ is assumed to be diagonal. For both SQ and VQ, the Lloyd-Max algorithm [35] is used to design codebooks that are source-optimized for a *noiseless* channel. We denote the source-optimized codebook by the set $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_N\}$, and the corresponding *point density function* [25] by $\lambda(\mathbf{x})$, i.e., there are roughly $N\lambda(\mathbf{x})\,\mathrm{d}\mathbf{x}$ code points in a small volume $\mathrm{d}\mathbf{x}$ containing $\mathbf{x} \in \mathcal{D}_{\mathbf{x}}$. When a linear transmit filter $\mathbf{T}$ is employed, the encoder uses the transformed codebook $\mathcal{C} \triangleq \{\mathbf{T}\hat{\mathbf{x}}_1, \mathbf{T}\hat{\mathbf{x}}_2, \ldots, \mathbf{T}\hat{\mathbf{x}}_N\}$ for nearest-neighbor based quantization. That is, the encoder outputs index i whenever $\mathbf{x} \in \tilde{\mathcal{R}}_i \triangleq \{\mathbf{x} : d(\mathbf{x}, \mathbf{T}\hat{\mathbf{x}}_i) \leq d(\mathbf{x}, \mathbf{T}\hat{\mathbf{x}}_j), 1 \leq j \leq N\}$. We use the notation $\mathcal{R}_i$ to represent $\tilde{\mathcal{R}}_i$ with $\mathbf{T} = \mathbf{I}$, i.e., without transmit filtering. Now, whenever $\mathbf{x} \in \mathcal{R}_i \triangleq \{\mathbf{x} : d(\mathbf{x}, \hat{\mathbf{x}}_i) \leq d(\mathbf{x}, \hat{\mathbf{x}}_j), 1 \leq j \leq N, j \neq i\}$, the encoder outputs index i, which is sent over a noisy Discrete Memoryless Channel (DMC), and is received as a possibly different index

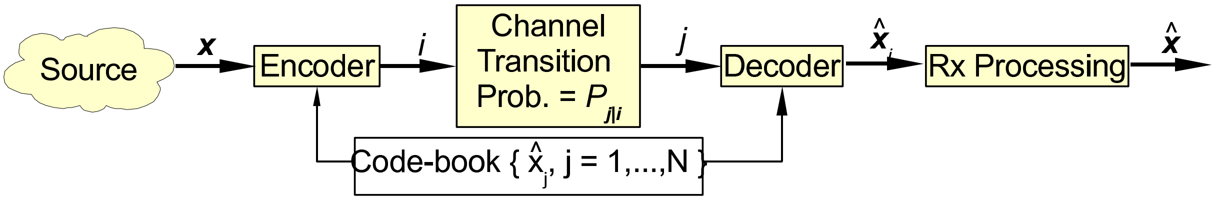
Source
x
Encoder
i
Channel Transition Prob. = P_{j|i}
j
Decoder
x̂_j
Rx Processing
x̂
Code-book { x̂_j, j = 1,...,N }

Figure 2.1: Block diagram of the system model.

j at the receiver. The system model is illustrated in Fig. 2.1.

At the receiver, the VQ decoder outputs the codeword $\hat{\mathbf{x}}_j$ corresponding to the received index j (which may not equal i, due to channel-induced index errors), as shown in the Fig. 2.1. Then, a linear filter $\mathbf{R}$ at the receiver outputs $\mathbf{R}\hat{\mathbf{x}}_j$ as an estimate of the source instantiation $\mathbf{x}$. Thus, the end-to-end distortion in the source vector $\mathbf{x} \in \mathcal{R}_i$ is $d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j)$.

2.2.1 Index Assignment

When the index output by the VQ encoder is transmitted over a noisy channel, the index assignment (IA), i.e., the mapping of codewords to the indices that are transmitted over the channel, plays a role in determining the overall distortion performance. When the bits corresponding to a given index are transmitted over noiseless channel, all IAs are equivalant. That is, we can assign any unique $B = \log_2 N$ bit pattern to the codebook vectors. On the other hand, when the channel is noisy, IA does affect the average distortion, and a good IA can achieve a lower distortion than a bad IA. However, the number of possible IAs is $N!$, and the problem of finding the optimum index assignment is known to be NP complete.

In this chapter, we study two approaches to circumvent the problem of finding the optimum index assignment. In the first approach, we consider *random IA*, where the IA is chosen uniformly at random from all possible IAs. This can be realized in practice by periodically and synchronously changing the IA at both the transmitter and receiver, and sequentially employing all possible IAs. This results in a symmetric error channel as stated earlier [29]. In the second approach, we implement the Linearity Increasing Swap Algorithm (LISA) , a Hadamard transform based tool, for finding a good IA [20]. Mathematically, the operation of index assignment results in a non-linear mapping between indices and the corresponding codewords that are output by the receiver (and also between the source instantiation and the corresponding index output by the encoder). The LISA tool tries to linearize this non-linear map. The performance of VQ for noisy channels with a *good IA* obtained from the LISA is analyzed by modeling the IA as a convex combination of an *Ideal IA* and a *Random IA*. An IA is said to be “ideal” if any single bit error on any transmitted index results in an index of one of the codewords which is closest in distortion sense to the codeword corresponding to the transmitted index. Note that, it is not guaranteed that such an index assignment is possible for all values of B and dimension n. Extending this to multi-bit errors, ideal IA ensures that for a given channel error rate, the distortion caused is smaller than any other IA. However, we assume ideal IA for two reasons. One, it gives mathematical tractability for analyzing the total distortion as a function of channel transition probabilities. Two, the distortion computed for ideal IA gives the lowest bound for the total distortion among all possible IAs. Moreover, for any given values of B and n, ideal IA condition can be satisfied for a certain fraction of the codewords. This gives the motivation for modeling the good IA as a convex combination of ideal IA and random IA where a certain fraction of the codewords have index assignments close to that of ideal IA. This will be discussed in more detail in the sections to follow. We now describe the channel model used in the study.

2.2.2 Channel Model

For simplicity, we assume that the codebook index i is mapped to a binary sequence and transmitted over a (possibly fading) AWGN channel using uncoded BPSK modulation. However, the framework presented here can be easily extended to other modulation and coding schemes, as long as the index transition probability due to channel errors can be computed. We assume that the channel SNR is known at the receiver, using which, it can compute the probability of error q given by

$$q = Q(\sqrt{2\gamma_b}), \tag{2.1}$$

where γ_b is the receiver SNR. When the channel is flat-fading, the SNR γ_b depends on the random channel instantiation. The pdf of γ_b with L^{th} order diversity is given by

$$p(\gamma_b) = \frac{1}{(L-1)!\overline{\gamma}_b^L}\gamma_b^{L-1}e^{-\gamma_b/\overline{\gamma}_b}, \tag{2.2}$$

where $\overline{\gamma}_b$ denotes the average SNR at the receiver. The probability of error for BPSK transmission in a fading channel can be calculated by averaging over the distribution of γ_b as

$$\overline{q} = \int_0^\infty Q(\sqrt{2\gamma_b})p(\gamma_b)d\gamma_b. \tag{2.3}$$

We consider the following two models for the BPSK demodulator implemented at the receiver.

2.2.2.1 Hard-Decision Decoding

When the bits are decoded using hard decision decoding, the codebook index transitions can be modeled as a DMC, with the transition probability being dependent on the IA and the cross-over probability of the underlying binary symmetric channel. When the IA is random, it is easy to see that the DMC is equivalent to a symmetric error channel with $N \times N$ transition probability matrix [29] whose $(i, j)^{\text{th}}$ element $P_{j|i}$ represents the probability that the index i is received as index j, given by

$$P_{j|i} = \epsilon_N + (1 - N\epsilon_N)\delta(i, j), \tag{2.4}$$

where $\delta(i, j) = 1$ when $i = j$ and 0 otherwise; and $\epsilon_N = (1 - (1 - q)^B)/(N - 1)$. For any given specific IA, the transition probability is given by

$$P_{\pi(j)|\pi(i)} = q^{\mathcal{W}_H(\pi(j)\oplus\pi(i))}(1 - q)^{B-\mathcal{W}_H(\pi(j)\oplus\pi(i))},$$

where $\pi : [1, 2, \ldots, N] \rightarrow [1, 2, \ldots, N]$ denotes the IA (a bijective map), $\mathbf{a} \oplus \mathbf{b}$ denotes the XOR operation between the binary vectors $\mathbf{a}$ and $\mathbf{b}$ and $\mathcal{W}_H(\mathbf{b})$ denotes the Hamming weight of $\mathbf{b}$.

2.2.2.2 Channel Model for Vector Quantization

In this work, we abstract the combined channel[1] as a DMC, and parameterize the index transition probability using the channel bit error rate. With random IA, the $N \times N$ transition probability matrix has its $(i, j)^{\text{th}}$ element $P_{j|i}$ given by (2.4), where $N = N_v$ is the number of codebook vectors used.

[1] The effective channel comprising the channel encoder, the noisy channel, and channel decoder is referred to as the combined channel.

2.2.2.3 Channel Model for Scalar Quantization

When the bits are encoded using a scalar quantizer, the codebook index transitions for random IA can again be modeled as a DMC. The $N \times N$ transition probability matrix is given by (2.4), where, now, $N = N_s^n$, and N_s is the number of quantization levels per dimension.

2.2.2.4 Semi-Hard Decision Decoding

Here, it is assumed that the receiver first computes the Log-Likelihood Ratio (LLR) from the received symbol, and either outputs a bit or declares an erasure, depending on whether the magnitude of the LLR exceeds or falls below a threshold. The threshold at which the receiver declares an erasure is a design parameter, that will be optimized later.[2] In this case, the AWGN channel is converted into a binary symmetric error and erasure channel, where a bit is erased with probability ρ and toggled with probability α. For a B bit transmission, the channel transition probability matrix is now given by its $(i,j)^{th}$ element $P_{j|i} = \rho^{B_1}\alpha^{B_2}(1-\rho-\alpha)^{B_3}$, where B_1 is the number of erasures in received index j, B_2 is the number of incorrectly received bits, and $B_3 = B - B_1 - B_2$ is the number of bits received correctly. For example, the $\mathbf{P}$ matrix (for $B = 2$) can be written as

$$\left[\begin{array}{c|cccc}
 & 00 & 01 & 10 & 11 \\
\hline
00 & (1-\alpha-\rho)^2 & \alpha(1-\alpha-\rho) & \alpha(1-\alpha-\rho) & \alpha^2 \\
01 & \alpha(1-\alpha-\rho) & (1-\alpha-\rho)^2 & \alpha^2 & \alpha(1-\alpha-\rho) \\
10 & \alpha(1-\alpha-\rho) & \alpha^2 & (1-\alpha-\rho)^2 & \alpha(1-\alpha-\rho) \\
11 & \alpha^2 & \alpha(1-\alpha-\rho) & (1-\alpha-\rho)^2 & \alpha(1-\alpha-\rho)
\end{array}\right.$$

$$\left.\begin{array}{ccccc}
0E & 1E & E0 & E1 & EE \\
\hline
\rho(1-\alpha-\rho) & \alpha\rho & \rho(1-\alpha-\rho) & \alpha\rho & \rho^2 \\
\rho(1-\alpha-\rho) & \alpha\rho & \alpha\rho & \rho(1-\alpha-\rho) & \rho^2 \\
\alpha\rho & \rho(1-\alpha-\rho) & \rho(1-\alpha-\rho) & \alpha\rho & \rho^2 \\
\alpha\rho & \rho(1-\alpha-\rho) & \alpha\rho & \rho(1-\alpha-\rho) & \rho^2
\end{array}\right]$$

where E denotes a bit erasure.

Depending on the decoded and erased bit positions, the decoder chooses one of the codewords

[2] If the threshold is set to zero, this channel defaults to the hard-decision decoding channel.

(which are precomputed depending on the erasure location and IA) as the estimate of the transmitted vector. Moreover, to keep the storage requirements minimal, additional codewords are stored only for all 1 bit erasure cases. Hence, there are $B\ 2^{B-1}$ additional codewords at the receiver apart from the 2^B codewords used for no-erasure cases. In the sequel, the Sec. 2.4 describes the design of the new codewords used at the receiver.

2.2.3 Optimization Problem

At the receiver, upon receiving j, the decoder outputs the corresponding codebook entry $\mathbf{y} \triangleq \hat{\mathbf{x}}_j$, which is multiplied by a linear receive filter $\mathbf{R}$ to obtain $\mathbf{Ry}$ as an estimate of $\mathbf{x}$. Thus, the end-to-end distortion in the source vector $\mathbf{x} \in \mathcal{R}_i$ is $d(\mathbf{x}, \mathbf{Ry})$. The average distortion is given by

$$J = \mathbb{E}[(\mathbf{x} - \mathbf{Ry})^T \mathbf{W} (\mathbf{x} - \mathbf{Ry})] = \sum_{i,j=1}^{N} P_{j|i} \int_{\mathcal{R}_i} (\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_j)^T \mathbf{W} (\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_j) f_{\mathbf{x}}(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{2.5}$$

where the expectation is taken over both the source distribution and the channel transition probabilities. Hence, our goal is to analyze (2.5) and design the LRF, $\mathbf{R}$, to minimize the average WMSE distortion. The next section presents the derivation of the optimal LRF.

2.3 Optimum Linear Receive Filters

2.3.1 Receive Filter for Random IA

Let $\mathbf{x} \in \mathcal{R}_i$ be the source instantiation and $\mathbf{y}$ be the corresponding codeword at the receiver, which could be different from $\hat{\mathbf{x}}_i$ due to channel errors. Then, the vector $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{2.6}$$

where $\mathbf{n}$ is an additive noise vector. For the development to follow, it will be necessary to consider two cases for $\mathbf{y}$ separately: $\mathbf{y} = \hat{\mathbf{x}}_i$ when the correct index is received, and $\mathbf{y} = \hat{\mathbf{y}}$ when an incorrect index is received. Hence, when the channel makes no error, $\mathbf{n} = \hat{\mathbf{x}}_i - \mathbf{x}$. When the channel does make an error, $\mathbf{n} = \hat{\mathbf{y}} - \mathbf{x}$, where $\hat{\mathbf{y}}$ is chosen uniformly among all other codewords $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{i-1}, \hat{\mathbf{x}}_{i+1}, \ldots, \hat{\mathbf{x}}_N\}$ due to the structure of the index transition probability matrix.

Equivalently, $\mathbf{n}$ is distributed as

$$f_{\mathbf{n}}(\mathbf{n}) = \begin{cases} (\hat{\mathbf{x}}_i - \mathbf{x}), & \text{with probability } 1-(N-1)\epsilon_N \\ (\hat{\mathbf{y}} - \mathbf{x}), & \text{with probability } (N-1)\epsilon_N \end{cases}, \tag{2.7}$$

where $\hat{\mathbf{y}}$ is spatially distributed according to $\tilde{\lambda}_{\mathbf{y}}(\mathbf{y}) = c\lambda_{\mathbf{x}}(\mathbf{y}), \mathbf{y} \notin \mathcal{R}_i$ and 0 otherwise, where c is a normalization constant chosen such that $\int_{\mathcal{D}_{\mathbf{x}}} \tilde{\lambda}_{\mathbf{y}}(\mathbf{y}) d\mathbf{y} = 1$. Exactly which $\hat{\mathbf{y}}$, among the $(N-1)$ possible *error codewords* is received, depends on the assumption made on the type of index assignment (IA). This is considered in two separate cases: (i) the random IA case, and (ii) the good IA case, in the sequel.

2.3.1.1 Linear Receive Filter for VQ

Let $\Sigma_{\mathbf{xy}} \triangleq \mathbb{E}[\mathbf{x}\mathbf{y}^T]$ and $\Sigma_{\mathbf{yy}} \triangleq \mathbb{E}[\mathbf{y}\mathbf{y}^T]$. The optimal LRF, $\mathbf{R}_{\text{opt}}$, that minimizes the WMSE cost function (2.5) is given by the well-known expression [36]

$$\mathbf{R}_{\text{opt}} = \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1}. \tag{2.8}$$

The main result of this chapter, stated as the following theorem, is valid under standard high-rate approximations [11, 27, 28]. The key results from high rate quantization theory that are relevant to our work are briefly reviewed in Appendix A.

Theorem 1. *The LRF that minimizes an approximation of the average WMSE distortion in (2.5) when the VQ index is transmitted over an SEC with index error rate ϵ_N is given by*

$$\mathbf{R}_{opt} = \tilde{\Sigma}_{\mathbf{xx}} \left[\frac{N\epsilon_N}{1-N\epsilon_N} \Sigma_\lambda + \tilde{\Sigma}_{\mathbf{xx}} \right]^{-1}, \tag{2.9}$$

where $\tilde{\Sigma}_{\mathbf{xx}} \triangleq \Sigma_{\mathbf{xx}} - \mathbf{\Theta}$, *with* $\mathbf{\Theta} \triangleq \mathbf{\Phi}_n \Gamma_n N^{\frac{-2}{n}}$, $\mathbf{\Phi}_n \triangleq \frac{\kappa_n^{\frac{-2}{n}} |\mathbf{W}|^{\frac{1}{n}}}{n+2} \mathbf{W}^{-1}$, $\Gamma_n \triangleq \left[\int_{\mathcal{D}_{\mathbf{x}}} f_{\mathbf{x}}^{\frac{n}{n+2}}(\mathbf{x})\, \mathrm{d}\mathbf{x} \right]^{\frac{n+2}{n}}$, *and* $\Sigma_\lambda \triangleq \int_{\mathcal{D}_y} \mathbf{y}\mathbf{y}^T \lambda(\mathbf{y})\, \mathrm{d}\mathbf{y}$. *The corresponding minimum average WMSE is given by*

$$E_{d,VQ}^{R_{opt}} \;=\; tr \left(\mathbf{W} \left[\Sigma_{\mathbf{xx}} - (1-N\epsilon_N)\tilde{\Sigma}_{\mathbf{xx}} \left(\frac{N\epsilon_N}{1-N\epsilon_N} \Sigma_\lambda + \tilde{\Sigma}_{\mathbf{xx}} \right)^{-1} \tilde{\Sigma}_{\mathbf{xx}}^T \right] \right), \tag{2.10}$$

Proof. In order to derive the optimum receive filter, one needs to derive expressions for the covariance matrices $\Sigma_{\mathbf{xy}}$ and $\Sigma_{\mathbf{yy}}$. Now, from (2.6), $\Sigma_{\mathbf{xy}} = \mathbb{E}[\mathbf{x}(\mathbf{x}+\mathbf{n})^T] = \Sigma_{\mathbf{xx}} + \mathbb{E}[\mathbf{x}\mathbf{n}^T]$. Using the distribution of $\mathbf{n}$ in (2.7), $\mathbb{E}[\mathbf{x}\mathbf{n}^T]$ can be written as

$$\mathbb{E}[\mathbf{x}\mathbf{n}^T] = (1-(N-1)\epsilon_N)\mathbb{E}[\mathbf{x}(\hat{\mathbf{x}}_i - \mathbf{x})^T] + (N-1)\epsilon_N \mathbb{E}[\mathbf{x}(\hat{\mathbf{y}} - \mathbf{x})^T]. \tag{2.11}$$

The expectation in the first term above is approximated as

$$\begin{aligned} \mathbb{E}[\mathbf{x}(\hat{\mathbf{x}}_i - \mathbf{x})^T] &= \sum_{i=1}^{N} \int_{\mathcal{R}_i} \mathbf{x}(\hat{\mathbf{x}}_i - \mathbf{x})^T f_{\mathbf{x}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} & (2.12) \\ &\approx -\sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i) \int_{\mathcal{E}_i} (\hat{\mathbf{x}}_i + \mathbf{e})\mathbf{e}^T \, \mathrm{d}\mathbf{e} = -\mathbf{\Theta}, & (2.13) \end{aligned}$$

where $\mathbf{e} \triangleq (\mathbf{x} - \hat{\mathbf{x}}_i)$ and $\mathcal{E}_i \triangleq \{\mathbf{e} : \mathbf{e} + \hat{\mathbf{x}}_i \in \mathcal{R}_i\}$. The above is obtained using the approximations that the polytope generating the Voronoi regions is geometrically centered about the origin [25] and the expression for $\mathbb{E}[\mathbf{e}\mathbf{e}^T]$ from (A.9) in Appendix A.2.

Similarly, $\mathbb{E}[\mathbf{x}(\hat{\mathbf{y}} - \mathbf{x})^T]$ can be shown to be

$$\begin{aligned} \mathbb{E}[\mathbf{x}(\hat{\mathbf{y}} - \mathbf{x})^T] &= \sum_{i=1}^{N} \int_{\mathcal{R}_i} f_{\mathbf{x}}(\mathbf{x}) \left[\frac{1}{N-1} \left\{ \left(\sum_{j=1}^{N} \mathbf{x}(\hat{\mathbf{x}}_j - \mathbf{x})^T \right) - \mathbf{x}(\hat{\mathbf{x}}_i - \mathbf{x})^T \right\} \right] \mathrm{d}\mathbf{x} & (2.14) \\ &\approx -\frac{N}{N-1}\Sigma_{\mathbf{xx}} + \frac{1}{N-1}\mathbf{\Theta}. & (2.15) \end{aligned}$$

The fact that $\mathbb{E}[\mathbf{x}] = \mathbf{0}$ and the approximation in (2.13) has been used to obtain the above.

$$\Sigma_{\mathbf{xy}} = (1 - N\epsilon_N)\tilde{\Sigma}_{\mathbf{xx}}. \tag{2.16}$$

Now, $\Sigma_{\mathbf{yy}} = \mathbb{E}[(\mathbf{x}+\mathbf{n})\mathbf{y}^T] = \mathbb{E}[\mathbf{x}\mathbf{y}^T] + \mathbb{E}[\mathbf{x}\mathbf{n}^T]^T + \mathbb{E}[\mathbf{n}\mathbf{n}^T]$, and note that the first two terms are as derived above. Using (2.7), $\Sigma_{\mathbf{nn}} \triangleq \mathbb{E}[\mathbf{n}\mathbf{n}^T]$ is given by

$$\Sigma_{\mathbf{nn}} = (1-(N-1)\epsilon_N)\mathbb{E}[(\hat{\mathbf{x}}_i - \mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x})^T] + (N-1)\epsilon_N \mathbb{E}[(\hat{\mathbf{y}} - \mathbf{x})(\hat{\mathbf{y}} - \mathbf{x})^T], \tag{2.17}$$

with $\hat{\mathbf{x}}_i$ being the codeword corresponding to the source instantiation $\mathbf{x}$ and with $\hat{\mathbf{y}}$ being chosen uniformly among all other codewords $\{\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_{i-1}, \hat{\mathbf{x}}_{i+1}, \ldots, \hat{\mathbf{x}}_n\}$ due to the structure of the index transition probability matrix. The expectation in the first term is given by (A.9) in Appendix A as $\mathbb{E}[(\hat{\mathbf{x}}_i - \mathbf{x})(\hat{\mathbf{x}}_i - \mathbf{x})^T] = \mathbf{\Theta}$. To compute the expectation in the second term, note that $\mathbb{E}[(\hat{\mathbf{y}}-\mathbf{x})(\hat{\mathbf{y}}-\mathbf{x})^T] = \sum_{i=1}^{N}\int_{\mathcal{R}_i}\mathbb{E}[(\hat{\mathbf{y}}-\mathbf{x})(\hat{\mathbf{y}}-\mathbf{x})^T|\mathbf{x}]f_{\mathbf{x}}(\mathbf{x})\,\mathrm{d}\mathbf{x}$. Now, if $\mathbf{x}\in\mathcal{R}_i$, we have

$$\begin{aligned}\mathbb{E}[(\hat{\mathbf{y}}-\mathbf{x})(\hat{\mathbf{y}}-\mathbf{x})^T|\mathbf{x}] &= \frac{1}{N-1}\left(\sum_{j=1}^{N}(\hat{\mathbf{x}}_j-\mathbf{x})(\hat{\mathbf{x}}_j-\mathbf{x})^T-(\hat{\mathbf{x}}_i-\mathbf{x})(\hat{\mathbf{x}}_i-\mathbf{x})\right) \qquad (2.18)\\ &\approx \frac{N}{N-1}\left(\Sigma_\lambda-\mathbf{x}\mu_\lambda^T-\mu_\lambda\mathbf{x}^T+\mathbf{x}\mathbf{x}^T\right)-\frac{1}{N-1}(\hat{\mathbf{x}}_i-\mathbf{x})(\hat{\mathbf{x}}_i-\mathbf{x})^T,\end{aligned}$$

where $\mu_\lambda \triangleq \frac{1}{N}\sum_{j=1}^{N}\hat{\mathbf{x}}_j \approx \int_{\mathcal{D}_y}\mathbf{y}\lambda(\mathbf{y})\,\mathrm{d}\mathbf{y}$ and $\Sigma_\lambda \triangleq \frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T \approx \int_{\mathcal{D}_y}\mathbf{y}\mathbf{y}^T\lambda(\mathbf{y})\,\mathrm{d}\mathbf{y}$ from the Monte Carlo approximation [29], and can be computed in closed-form given the source-optimized point density of the codebook. Hence,

$$\mathbb{E}[(\hat{\mathbf{y}}-\mathbf{x})(\hat{\mathbf{y}}-\mathbf{x})^T] \approx \frac{N}{N-1}\left(\Sigma_\lambda+\Sigma_{\mathbf{xx}}\right)-\frac{1}{N-1}\boldsymbol{\Theta}. \qquad (2.19)$$

Again, the fact that $\mathbb{E}[\mathbf{x}]=\mathbf{0}$ and (A.9) in Appendix A have been used to obtain the above expression. Substituting the above into (2.17) and using the result to evaluate $\Sigma_{\mathbf{yy}}$, we obtain

$$\Sigma_{\mathbf{yy}} = (1-N\epsilon_N)\tilde{\Sigma}_{\mathbf{xx}}+N\epsilon_N\Sigma_\lambda. \qquad (2.20)$$

Substituting (2.16) and (2.20) in (2.8), the optimum LRF can now be obtained as given in (2.9). Finally, substituting for $\mathbf{R}_{\mathrm{opt}}$ in (2.5) and simplifying the resulting expression yields the following:

$$\begin{aligned}E_{d,VQ}^{Ropt} &= tr\left(\mathbf{W}\left[\Sigma_{\mathbf{xx}}-\Sigma_{\mathbf{xy}}\mathbf{R}_{\mathrm{opt}}^T-\mathbf{R}_{\mathrm{opt}}\Sigma_{\mathbf{yx}}+\mathbf{R}_{\mathrm{opt}}\Sigma_{\mathbf{yy}}\mathbf{R}_{\mathrm{opt}}^T\right]\right) \qquad (2.21)\\ &= tr\left(\mathbf{W}\left[\Sigma_{\mathbf{xx}}-(1-N\epsilon_N)\mathbf{R}_{\mathrm{opt}}\tilde{\Sigma}_{\mathbf{xx}}^T\right]\right). \qquad (2.22)\end{aligned}$$

The expression in (2.10) follows by substituting for R_{opt} from (2.9) in the above, completing the proof. □

Remark 1. The covariance of the output of the decoder, denoted $\Sigma_{\mathbf{yy}}$, given by (2.20), is valid for $0\leq\epsilon_N<\frac{1}{N}$. Note that, when $\epsilon_N=0$, we have $\Sigma_{\mathbf{yy}}\triangleq\mathbb{E}\{\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\}=\Sigma_{\mathbf{xx}}-\boldsymbol{\Theta}\neq\frac{1}{N}\sum_{i=1}^{N}\hat{\mathbf{x}}_i\hat{\mathbf{x}}_i^T\triangleq\Sigma_\lambda$.

2.3.1.2 Comparison with No Filtering

The expected WMSE without the LRF is obtained by substituting $\mathbf{R} = \mathbf{I}$ in (2.5) and simplifying, as

$$E_{d,VQ}^{\text{conv}} = tr\left(\mathbf{W}\left[\Sigma_{\mathbf{xx}} - (1 - N\epsilon_N)\tilde{\Sigma}_{\mathbf{xx}} + N\epsilon_N\Sigma_\lambda\right]\right). \tag{2.23}$$

The above equation is interesting, as it provides an accurate and easy-to-evaluate expression for the expected WMSE of VQ with a noisy DMC and random IA. With some manipulation, it can be shown to be equivalent to the expected distortion expression in the literature [30] (see Theorem 1), and hence, it can be viewed as an alternative and perhaps simpler derivation of that result. Now, substituting for $E_{d,VQ}^{R_{\text{opt}}}$ from (2.22), the reduction in the WMSE distortion due to the LRF is given by

$$E_{d,VQ}^{\text{conv}} - E_{d,VQ}^{R_{\text{opt}}} = tr\left(\mathbf{W}\left[(1 - N\epsilon_N)\mathbf{R}_{\text{opt}}\underbrace{\tilde{\Sigma}_{\mathbf{xx}}^T}_{a} - (1 - N\epsilon_N)\tilde{\Sigma}_{\mathbf{xx}} + N\epsilon_N\Sigma_\lambda\right]\right). \tag{2.24}$$

Substituting for R_{opt} from (2.9) and replacing $\tilde{\Sigma}_{\mathbf{xx}}$ with $\left(\tilde{\Sigma}_{\mathbf{xx}} + \frac{N\epsilon_N}{1-N\epsilon_N}\Sigma_\lambda\right) - \frac{N\epsilon_N}{1-N\epsilon_N}\Sigma_\lambda$ for the term marked a in the above equation, since $\tilde{\Sigma}_{\mathbf{xx}}$ is a symmetric matrix, we get

$$E_{d,VQ}^{\text{conv}} - E_{d,VQ}^{R_{\text{opt}}} = tr\left(\mathbf{W}\left[N\epsilon_N\left(\mathbf{I} - \underbrace{\tilde{\Sigma}_{\mathbf{xx}}}_{b}\left(\tilde{\Sigma}_{\mathbf{xx}} + \frac{N\epsilon_N}{1 - N\epsilon_N}\Sigma_\lambda\right)^{-1}\right)\Sigma_\lambda\right]\right). \tag{2.25}$$

Applying the same substitution for the term marked b, we get

$$E_{d,VQ}^{\text{conv}} - E_{d,VQ}^{R_{\text{opt}}} = tr\left(\mathbf{W}\left[\frac{N^2\epsilon_N^2}{1 - N\epsilon_N}\underbrace{\Sigma_\lambda\left(\tilde{\Sigma}_{\mathbf{xx}} + \frac{N\epsilon_N}{1 - N\epsilon_N}\Sigma_\lambda\right)^{-1}\Sigma_\lambda}_{c}\right]\right). \tag{2.26}$$

Since $\tilde{\Sigma}_{\mathbf{xx}} = \Sigma_{\mathbf{xx}} - \mathbf{\Theta}$ and $\mathbf{\Theta}$ decreases as $N^{-2/n}$, $\tilde{\Sigma}_{\mathbf{xx}}$ is positive definite for sufficiently large N. Due to the positive definiteness of Σ_λ, the matrix marked c is positive definite, and hence, the LRF offers a positive improvement performance over no filtering for all channel conditions.

2.3.1.3 Linear Receive Filter for SQ

In this section, we derive the optimum LRF for SQ of a vector source, and characterize the improvement in the performance from receive filtering. In practice, it is typical to use SQ when the source dimensions are independent and the distortion function is separable, e.g., when $\mathbf{W}$ is a non-negative diagonal matrix. Hence, in this section, attention is restricted to independent random variables in each of the dimensions of the vector source and the weighting matrix $\mathbf{W}$ is set as the identity matrix.

Let N_s be the number of quantization levels per dimension for SQ, so that $N = N_s^n$ is the size of the overall n-dimensional codebook. Within a given dimension, the probability that a codebook index i is *incorrectly* received as an index $j \neq i$ is $N_s^{(n-1)}\epsilon_N$. This is because the n-dimensional codebook has $N_s^{(n-1)}$ indices with j as the component of the index in the given dimension. Due to this, we obtain the following marginal index transition probability matrix for each dimension: $P_{j|i}^{(SQ)} \triangleq N_s^{n-1}\epsilon_N + (1 - N_s^n\epsilon_N)\delta(i,j)$, where $1 \leq i \leq N_s$ and $1 \leq j \leq N_s$ are per-dimensional codebook indices and ϵ_N is as defined in the previous section. It is interesting to note that the equivalent pairwise index error rate of SQ, $\epsilon_{N_s} \triangleq N_s^{n-1}\epsilon_N$ satisfies $N_s\epsilon_{N_s} = N\epsilon_N$. Now, the total distortion with SQ is simply the sum of distortions incurred in each of the dimensions. Hence, the optimum LRF in the i-th dimension, which is a scaling factor denoted $r_{\text{opt}}(i)$, can be obtained by setting $n = 1$ in (2.9) and substituting the above index transition probability, as follows:

$$r_{\text{opt}}(i) = \left(\sigma_{x_i}^2 - \Gamma_i \frac{N_s^{-2}}{12}\right)\left[\frac{N\epsilon_N}{1 - N\epsilon_N}\sigma_{\lambda_i}^2 + \sigma_{x_i}^2 - \Gamma_i\frac{N_s^{-2}}{12}\right]^{-1}, \tag{2.27}$$

with $\sigma_{x_i}^2$ and $\sigma_{\lambda_i}^2$ being the i-th diagonal components of $\Sigma_{\mathbf{xx}}$ and Σ_λ, respectively, and $\Gamma_i \triangleq \left[\int_{\mathcal{D}_i} f_{X_i}^{\frac{1}{3}}(x)\,\mathrm{d}x\right]^3$. Here, $f_{X_i}(x)$ is the pdf and $\mathcal{D}_i$ is the domain, of the i-th component of $\mathbf{x}$. Using (2.27) and (2.22), the expected MSE after LRF for SQ can be obtained as

$$E_{d,SQ}^{R_{\text{opt}}} = \sum_{i=1}^{n}\sigma_{x_i}^2 - (1 - N\epsilon_N)\left(\sigma_{x_i}^2 - \Gamma_i\frac{N_s^{-2}}{12}\right)^2\left[\frac{N\epsilon_N}{1 - N\epsilon_N}\sigma_{\lambda_i}^2 + \sigma_{x_i}^2 - \Gamma_i\frac{N_s^{-2}}{12}\right]^{-1}. \tag{2.28}$$

2.3.1.4 Comparison with No Filtering

Without receive filtering, we can obtain an expression for the expected distortion of SQ for noisy channels by substituting $n = 1$ in (2.23) and simplifying, to get:

$$E_{d,SQ}^{\text{conv}} = \sum_{i=1}^{n} \left[(1 - N\epsilon_N)\Gamma_i \frac{N_s^{-2}}{12} + N\epsilon_N \left(\sigma_{x_i}^2 + \sigma_{\lambda_i}^2 \right) \right]. \tag{2.29}$$

The reduction in the MSE distortion due to the LRF can be obtained from (2.26) as

$$E_{d,SQ}^{\text{conv}} - E_{d,SQ}^{R_{\text{opt}}} = \sum_{i=1}^{n} \frac{N^2\epsilon_N^2}{1 - N\epsilon_N} \sigma_{\lambda_i}^4 \left[\frac{N\epsilon_N}{1 - N\epsilon_N} \sigma_{\lambda_i}^2 + \sigma_{x_i}^2 - \Gamma_i \frac{N_s^{-2}}{12} \right]^{-1}, \tag{2.30}$$

and the terms in the summation above are all positive for reasonably large N. Hence, for high-rate quantization, the LRF obtains a lower MSE distortion compared to the no filtering case. It is instructive to compare the above expressions with the average distortion for VQ obtained in the previous subsection. The simulation results in the next section validate the above analysis and quantify the relative performance of SQ and VQ with and without receive filtering, for noisy DMCs.

2.3.1.5 Comparison with Transmitter Adaptation Methods

Now, we derive the optimum linear transmit and receive filter for minimizing the end-to-end distortion and compare its performance with the receive-only filter. The system model is shown in Figure 2.2. The past literature relevant for this work include COVQ, where the codebook is adapted as per the channel statistics [12], and the scheme where a scaled codebook used in both the transmitter and receiver [33]. In [33], the scaling factor is numerically optimized to minimize the end-to-end distortion. The optimal scaling factor is depends on the channel SNR, and can be precomputed and stored as a lookup table.

For the system shown in Figure 2.2, the source instantiation is quantized using the scaled codebook, where the transmit linear filter (scaling matrix) $\mathbf{T}$ is set as $\mathbf{T} = \alpha\mathbf{I}$ for analytical tractability and ease of comparison with existing literature. Using such a scaled codebook is also equivalent to applying a scalar α^{-1} to the source instantiation prior to quantization, and applying another scalar α at the decoder output. The model we use here is thus fairly general, and by choosing $\tilde{\mathbf{R}} = \mathbf{R}\mathbf{T}^{-1}$, where $\mathbf{R}$ is the optimized Rx filter for the given Transmit-only

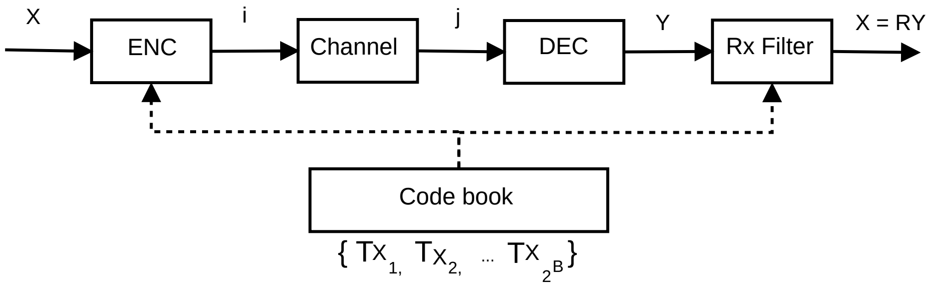

Figure 2.2: Combined transmit and receive filtering system model.

filter $\mathbf{T}$, we can obtain a joint transmit-receive filter. Hence, the following special cases can be handled in this approach: (a) Rx only filter, when $\mathbf{T} = \mathbf{I}$, (b) Transmit-only filter, when $\mathbf{R} = \mathbf{I}$ and (c) Transmit-Receive filter, when the optimum $\mathbf{R}$ is computed for the given $\mathbf{T}$.

Now, the noise $\mathbf{n}$ can be written as $\mathbf{n} = \mathbf{y} - \mathbf{x}$. Its PDF given by the following, when $\mathbf{x} \in \tilde{\mathcal{R}}_i$

$$f_{\mathbf{n}}(\mathbf{n}) = \begin{cases} \tilde{\mathbf{x}}_i - \mathbf{x} & \text{, with probability } (1 - (N-1)\epsilon_N) \\ \tilde{\mathbf{x}}_j - \mathbf{x} & \text{, with probability } (N-1)\epsilon_N \end{cases} , \tag{2.31}$$

where $\tilde{\mathcal{R}}_i$ denotes the modified Voronoi region according to the scaled codebook $\{\mathbf{T}\hat{\mathbf{x}}_i\}$.

$$\tilde{\mathcal{R}}_i \triangleq \left\{\mathbf{x} : (\tilde{\mathbf{x}}_i - \mathbf{x})^T \mathbf{W} (\tilde{\mathbf{x}}_i - \mathbf{x}) < (\tilde{\mathbf{x}}_j - \mathbf{x})^T \mathbf{W} (\tilde{\mathbf{x}}_j - \mathbf{x}), j \neq i \right\}. \tag{2.32}$$

Also, define $\tilde{\mathcal{E}}_i$ by shifting $\tilde{\mathcal{R}}_i$ to the origin by subtracting $\tilde{\mathbf{x}}_i$ from all $\mathbf{x} \in \tilde{\mathcal{R}}_i$.

The end-to-end distortion for W-MSE distortion measure can be written as

$$\begin{aligned} J &= \mathbb{E}\left\{(\mathbf{x} - \tilde{\mathbf{R}}\mathbf{y})^T \mathbf{W} (\mathbf{x} - \tilde{\mathbf{R}}\mathbf{y})\right\} \\ &= tr\left[\mathbf{W}\left(\Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{xy}}\tilde{\mathbf{R}}^T - \tilde{\mathbf{R}}\Sigma_{\mathbf{yx}} + \tilde{\mathbf{R}}\Sigma_{\mathbf{yy}}\tilde{\mathbf{R}}^T\right)\right]. \end{aligned} \tag{2.33}$$

Using high rate quantization cell approximation, one can compute the matrices $\Sigma_{\mathbf{xy}}, \Sigma_{\mathbf{yy}}$ as shown in the Rx filter derivation.

$$\Sigma_{\mathbf{xy}} = \mathbb{E}\left\{\mathbf{x}(\mathbf{x} + \mathbf{n})^T\right\} = \Sigma_{\mathbf{xx}} + \mathbb{E}\left\{\mathbf{x}\mathbf{n}^T\right\} \tag{2.34}$$

$$\mathbb{E}\left\{\mathbf{x}\mathbf{n}^T\right\} = (1 - (N-1)\epsilon_N)\mathbb{E}\{\mathbf{x}(\tilde{\mathbf{x}}_i - \mathbf{x})^T\} + (N-1)\epsilon_N \mathbb{E}\{\mathbf{x}(\tilde{\mathbf{y}} - \mathbf{x})^T\}, \tag{2.35}$$

where $\tilde{\mathbf{x}}_i = \mathbf{T}\hat{\mathbf{x}}_i$, $\tilde{\mathbf{y}} = \mathbf{T}\hat{\mathbf{x}}_j, j \neq i$. Now, $\mathbb{E}\{\mathbf{x}(\tilde{\mathbf{x}}_i - \mathbf{x})^T\}$ can be written as

$$\begin{aligned}
\mathbb{E}\{\mathbf{x}(\tilde{\mathbf{x}}_i - \mathbf{x})^T\} &= \sum_{i=1}^{N} \int_{\mathbf{x}\in\tilde{\mathcal{R}}_i} \mathbf{x}(\tilde{\mathbf{x}}_i - \mathbf{x})^T f_{\mathbf{x}}(\mathbf{x})\mathrm{d}\mathbf{x}, \\
&\overset{(a)}{\approx} -\sum_{i=1}^{N} f_{\mathbf{x}}(\tilde{\mathbf{x}}_i) \int_{\mathbf{e}\in\tilde{\mathcal{E}}_i} (\tilde{\mathbf{x}}_i + \mathbf{e})\mathbf{e}^T \mathrm{d}\mathbf{e}, \\
&\overset{(b)}{\approx} -\sum_{i=1}^{N} f_{\mathbf{x}}(\tilde{\mathbf{x}}_i) \int_{\mathbf{e}\in\tilde{\mathcal{E}}_i} \mathbf{e}\mathbf{e}^T \mathrm{d}\mathbf{e}, \\
&\overset{(c)}{\approx} -\sum_{i=1}^{N} f_{\mathbf{x}}(\tilde{\mathbf{x}}_i) \frac{\tilde{V}_i}{n+2} \left(\frac{\tilde{V}_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}} \mathbf{W}^{-1}, \\
&\overset{(d)}{\approx} -\frac{N^{\frac{-2}{n}}\kappa_n^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{n+2}\mathbf{W}^{-1} \int_{\mathbf{x}\in\mathcal{D}_{\mathbf{x}}} \lambda_T^{\frac{-2}{n}}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x})\mathrm{d}\mathbf{x}, \qquad (2.36)
\end{aligned}$$

where we have used the following approximations: (a) $f_{\mathbf{x}}(\mathbf{x}) \approx f_{\mathbf{x}}(\tilde{\mathbf{x}}_i), \mathbf{x} \in \tilde{\mathcal{R}}_i$, (b) $\int_{\mathbf{e}\in\tilde{\mathcal{E}}_i} \tilde{\mathbf{x}}_i\mathbf{e}^T \approx 0$ due to the geometric centroid nature of the codebook and (c) quantization cell approximation as an ellipsoid, (d) Monte Carlo integration approximation for large N, and $\tilde{V}_i$ denotes the volume of the new quantization cell $\tilde{\mathcal{R}}_i$.

To simplify the analytical expressions further, we restrict ourselves to the special case where $x_i \sim \mathcal{N}(0,1)$ and $\mathbf{T} = t\mathbf{I}, 0 \leq t \leq 1$. For an n-dimensional Gaussian vector with i.i.d. elements according to $\mathcal{N}(0,1)$, it can be shown that

$$c = \int f_{\mathbf{x}}^{\frac{n}{n+2}}(\mathbf{x})\mathrm{d}\mathbf{x} = (2\pi)^{\frac{-n^2}{2(n+2)}} \left(\frac{n}{n+2}\right)^{\frac{n}{2}}.$$

To compute the point density of the scaled codebook, one can use the fact that

$$\lambda_{\mathbf{T}}(\mathbf{y}) \quad = \quad \frac{1}{|\mathbf{T}|}\lambda_{\mathbf{x}}(\mathbf{T}^{-1}\mathbf{y}) = \frac{1}{t^n}\frac{f_{\mathbf{x}}^{\frac{n}{n+2}}\left(\frac{\mathbf{x}}{t}\right)}{c},$$

which can be shown to be an i.i.d. Gaussian vector with elements distributed as $\mathcal{N}\left(0, t^2\left(\frac{n+2}{n}\right)\right)$. Now, substituting the above in (2.36), we get

$$\begin{aligned}
\mathbb{E}\{\mathbf{x}(\tilde{\mathbf{x}}_i - \mathbf{x})^T\} = -\frac{N^{\frac{-2}{n}}\kappa_n^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{n+2}\mathbf{W}^{-1}2\pi t^2\left(\frac{n+2}{n}\right)(2\pi)^{\frac{-n}{2}} \int_{\mathbf{x}\in\mathcal{D}_{\mathbf{x}}} e^{-\frac{\mathbf{x}^T\mathbf{x}}{2}\left(1-\frac{2n}{n(n+2)t^2}\right)}\mathrm{d}\mathbf{x} \\
= -\frac{2\pi N^{\frac{-2}{n}}\kappa_n^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}t^2}{n\left(1-\frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}}\mathbf{W}^{-1} = -\mathbf{\Theta}_t, (2.37)
\end{aligned}$$

where t is constrained to lie in $\left[\sqrt{\frac{2}{n+2}}, 1\right)$ for the above expression to be meaningful. Note that, this reduces to $-\mathbf{\Theta}$ in (A.9) when $t = 1$, as expected. Next, one can compute $\mathbb{E}\{\mathbf{x}(\tilde{\mathbf{y}} - \mathbf{x})^T\}$ along similar lines. That is,

$$\begin{aligned} \mathbb{E}\{\mathbf{x}(\tilde{\mathbf{y}} - \mathbf{x})^T\} = \sum_{i=1}^{N} \int_{\mathbf{x} \in \tilde{\mathcal{R}}_i} f_{\mathbf{x}}(\mathbf{x}) \frac{1}{N-1} \left[\left(\sum_{j=1}^{N} \mathbf{x}(\tilde{\mathbf{x}}_j - \mathbf{x})^T \right) - \mathbf{x}(\tilde{\mathbf{x}}_i - \mathbf{x})^T \right] \mathrm{d}\mathbf{x} \\ \approx -\frac{N}{N-1} \Sigma_{\mathbf{xx}} + \frac{1}{N-1} \mathbf{\Theta}_t, \end{aligned} \tag{2.38}$$

since $\sum_{j=1}^{N} \tilde{\mathbf{x}}_j \approx 0$. Substituting (2.37) and (2.38) in (2.35), we get

$$\mathbb{E}\{\mathbf{x}\mathbf{n}^T\} = -N\epsilon_N \Sigma_{\mathbf{xx}} - (1 - N\epsilon_N)\mathbf{\Theta}_t \tag{2.39}$$

$$\Sigma_{\mathbf{xy}} = \Sigma_{\mathbf{xx}} + \mathbb{E}\{\mathbf{x}\mathbf{n}^T\} = (1 - N\epsilon_N)(\Sigma_{\mathbf{xx}} - \mathbf{\Theta}_t). \tag{2.40}$$

Now, consider $\mathbb{E}\{\mathbf{y}\mathbf{y}^T\}$.

$$\mathbb{E}\{\mathbf{y}\mathbf{y}^T\} = \mathbb{E}\{(\mathbf{x} + \mathbf{n})\mathbf{y}^T\} = \mathbb{E}\{\mathbf{x}\mathbf{y}^T\} + \mathbb{E}\{\mathbf{n}\mathbf{x}^T\} + \mathbb{E}\{\mathbf{n}\mathbf{n}^T\}.$$

Since $\mathbb{E}\{\mathbf{x}\mathbf{y}^T\}$ and $\mathbb{E}\{\mathbf{x}\mathbf{n}^T\}$ are computed already, we need to compute only $\mathbb{E}\{\mathbf{n}\mathbf{n}^T\}$.

$$\begin{aligned} \mathbb{E}\{\mathbf{n}\mathbf{n}^T\} &= (1 - (N-1)\epsilon_N)\mathbb{E}\{(\tilde{\mathbf{x}}_i - \mathbf{x})(\tilde{\mathbf{x}}_i - \mathbf{x})^T\} + (N-1)\epsilon_N \mathbb{E}\{(\tilde{\mathbf{y}} - \mathbf{x})(\tilde{\mathbf{y}} - \mathbf{x})^T\} \quad (2.41) \\ &= (1 - (N-1)\epsilon_N)\mathbf{\Theta}_t + (N-1)\epsilon_N \sum_{i=1}^{N} \int_{\mathbf{x} \in \tilde{\mathcal{R}}_i} \mathbb{E}\{(\tilde{\mathbf{y}} - \mathbf{x})(\tilde{\mathbf{y}} - \mathbf{x})^T | \mathbf{x}\} f_{\mathbf{x}}(\mathbf{x}) \mathrm{d}\mathbf{x}, \end{aligned}$$

$$\begin{aligned} \mathbb{E}\{(\tilde{\mathbf{y}} - \mathbf{x})(\tilde{\mathbf{y}} - \mathbf{x})^T | \mathbf{x}\} &= \frac{1}{N-1} \left(\sum_{j=1}^{N} (\tilde{\mathbf{x}}_j - \mathbf{x})(\tilde{\mathbf{x}}_j - \mathbf{x})^T (\tilde{\mathbf{x}}_i - \mathbf{x})(\tilde{\mathbf{x}}_i - \mathbf{x})^T \right) \quad (2.42) \\ &= \frac{N}{N-1} \left(\Sigma_{\lambda_t} - \mathbf{x} \left(\sum_j \tilde{\mathbf{x}}_j \right) + \mathbf{x}\mathbf{x}^T \right) - \frac{1}{N-1} (\tilde{\mathbf{x}}_j - \mathbf{x})(\tilde{\mathbf{x}}_j - \mathbf{x})^T \\ \mathbb{E}\{(\tilde{\mathbf{y}} - \mathbf{x})(\tilde{\mathbf{y}} - \mathbf{x})^T\} &= \frac{N}{N-1} (\Sigma_{\lambda_t} + \Sigma_{\mathbf{xx}}) - \frac{1}{N-1} \mathbf{\Theta}_t, \quad (2.43) \\ \mathbb{E}\{\mathbf{n}\mathbf{n}^T\} &= (1 - N\epsilon_N)\mathbf{\Theta}_t + N\epsilon_N \left(\Sigma_{\mathbf{xx}} + \Sigma_{\lambda_t} \right). \quad (2.44) \end{aligned}$$

Therefore, $\Sigma_{\mathbf{yy}}$ can be written as

$$\Sigma_{\mathbf{yy}} \quad = \quad (1 - N\epsilon_N)(\Sigma_{\mathbf{xx}} - \mathbf{\Theta}_t) + N\epsilon_N \Sigma_{\lambda_t}. \tag{2.45}$$

By substituting for $\Sigma_{\mathbf{xy}}$, $\Sigma_{\mathbf{yy}}$ and Σ_{λ_t} in (2.33), one can compute the end-to-end distortion for three special cases of $\tilde{\mathbf{R}}$: (a) Transmit-only filter case when $\tilde{\mathbf{R}} = \frac{1}{t}\mathbf{I}$, (b) Optimized Rx filter when $\tilde{\mathbf{R}} = \mathbf{R}_{opt}$ and (c) Scaled codebook when $\tilde{\mathbf{R}} = \mathbf{I}$ (as in [33]).

For the special case of $\Sigma_{\mathbf{xx}} = \mathbf{W} = \mathbf{I}$, one can show that E_d^{TxFilt} can be simplified to

$$\begin{aligned} E_d^{\text{TxFilt}}(t) = n - \frac{2n(1 - N\epsilon_N)}{t} \left(1 - \frac{2\pi N^{\frac{-2}{n}} \kappa_n^{\frac{-2}{n}} t^2}{n \left(1 - \frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}} \right) \\ + \frac{n}{t^2} \left[(1 - N\epsilon_N) \left(1 - \frac{2\pi N^{\frac{-2}{n}} \kappa_n^{\frac{-2}{n}} t^2}{n \left(1 - \frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}} \right) \right] + N\epsilon_N(n+2), \end{aligned}$$

where $t \in \left[\sqrt{\frac{2}{n+2}}, 1\right)$. Differentiating the above with respect to t and equating to zero does not give any insight into the relationship between the t and the other parameters such as $N\epsilon_N$. Hence, we numerically evaluate the function for various t for the given $N\epsilon_N$ and find the optimal value of t which minimizes the total distortion. Now, we compute the distortion with the optimal receive filter and with the transmit filter $\mathbf{T} = t\mathbf{I}$. Let $\tilde{\mathbf{R}} = r\mathbf{I}$. Hence, the total distortion E_d^{TxRxFilt} can be written as

$$\begin{aligned} E_d^{\text{TxRxFilt}}(t) = n - 2nr(1 - N\epsilon_N) \left(1 - \frac{2\pi N^{\frac{-2}{n}} \kappa_n^{\frac{-2}{n}} t^2}{n \left(1 - \frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}} \right) \\ + nr^2(1 - N\epsilon_N) \left(1 - \frac{2\pi N^{\frac{-2}{n}} \kappa_n^{\frac{-2}{n}} t^2}{n \left(1 - \frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}} \right) + N\epsilon_N r^2 t^2 (n+2), \end{aligned}$$

Now, taking a derivative with respect to r and equating to zero gives the optimal value r_{opt} for the given value of t. That is,

$$r_{opt} = \frac{(1 - N\epsilon_N)tr(\mathbf{I} - \mathbf{\Theta}_t)}{(1 - N\epsilon_N)tr(\mathbf{I} - \mathbf{\Theta}_t) + N\epsilon_N tr(\Sigma_{\lambda_t})}. \tag{2.46}$$

Note the similarity of the expression for the optimum receive filter obtained earlier for the receive-only filtering case. Substituting this into the total distortion gives the minimum end-to-end distortion for the combined transmit and receive filtering. Similarly, one can calculate the total distortion for the scaled codebook given in [33] as follows.

$$E_d^{\mathrm{scaledCB}}(t) = n - 2n(1 - N\epsilon_N)\left(1 - \frac{2\pi N^{\frac{-2}{n}}\kappa_n^{\frac{-2}{n}}t^2}{n\left(1 - \frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}}\right)$$
$$+n(1 - N\epsilon_N)\left(1 - \frac{2\pi N^{\frac{-2}{n}}\kappa_n^{\frac{-2}{n}}t^2}{n\left(1 - \frac{2}{(n+2)t^2}\right)^{\frac{n}{2}}}\right) + N\epsilon_N t^2(n+2),$$

which also can be evaluated for the optimum value t. Note that, all the above three distortions can be compared with the total distortion without any filtering. The latter can be obtained by setting $t = 1$ in the above the expression for $E_d^{\mathrm{scaledCB}}(t)$.

2.3.2 Receive Filter for Ideal IA

For simplicity of presentation, we restrict the analysis to 1 bit errors, which dominates the performance at moderate-to-high SNRs. The extension of our method to multi-bit errors is straightforward. The expected distortion can be written as

$$E_d = \sum_{i=1}^{N} f_X(\hat{\mathbf{x}}_i) \sum_{j=1}^{N} P_{j|i} \int_{\mathbf{x}\in\mathcal{R}_i} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j)\mathrm{d}\mathbf{x}. \tag{2.47}$$

If at most one bit errors occur, the channel transition probabilities can be written as follows:

$$\begin{aligned} P_{i|i} &= Q \triangleq (1-q)^B \\ P_{j|i} &= \frac{1-Q}{B} = \frac{1-(1-q)^B}{B} \ \forall j \in S_i, \end{aligned} \tag{2.48}$$

where S_i denotes the set of neighbors for the i^{th} codeword considering all 1 bit errors in the index corresponding to the i^{th} codeword. In Appendix A, it is shown that (2.47) can be written as (A.17) which is reproduced here for convenience.

$$\begin{aligned} E_d^{\text{ideal}} &\doteq \int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x}) \left[\mathbf{x}^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\mathbf{x}\right] d\mathbf{x} + \frac{nN^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{(n+2)\kappa_n^{\frac{2}{n}}} \left[\int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})d\mathbf{x}\right] \\ &+\frac{4(1-Q)N^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{n\ \kappa_n^{\frac{2}{n}}}\text{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right)\left[\int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})d\mathbf{x}\right]. \end{aligned}$$

After some manipulations, the average distortion can be written as

$$\begin{aligned} E_d^{\text{ideal}} &\doteq tr\left(\mathbf{W}(\mathbf{I}-\mathbf{R})\Sigma_{\mathbf{x}}(\mathbf{I}-\mathbf{R})^T\right) \\ &+ E_d^{\text{SO}}\left[1+\frac{4(1-Q)(n+2)}{n^2}tr\left(\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\mathbf{R}^T\right)\right]. \end{aligned} \tag{2.49}$$

Now, straightforward differentiation of the above with respect to $\mathbf{R}$ results in an equation for the optimum receive filter matrix $\mathbf{R}_{\text{opt}}$, as follows:

$$\left[\mathbf{W}^T\mathbf{R}\Sigma_{\mathbf{x}}^T+\mathbf{W}\mathbf{R}\Sigma_{\mathbf{x}}-\mathbf{W}^T\Sigma_{\mathbf{x}}^T-\mathbf{W}\Sigma_{\mathbf{x}}\right]+\frac{4(1-Q)(n+2)E_d^{\text{SO}}}{n^2}\left[\mathbf{W}^T\mathbf{R}\mathbf{W}^{-T}+\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\right]=0.$$

For symmetric $\mathbf{W}$ and $\Sigma_{\mathbf{x}}$, a closed form expression for $\mathbf{R}_{\text{opt}}$ can be obtained as

$$\mathbf{R}_{\text{opt}}=\left(\Sigma_{\mathbf{x}}+\frac{4(1-Q)(n+2)E_d^{\text{SO}}}{n^2}\mathbf{W}^{-1}\right)^{-1}\Sigma_{\mathbf{x}}. \tag{2.50}$$

Clearly, when the channel is error free, i.e., $Q=1$, the optimum filter matrix turns out to be the identity matrix, as expected.

2.3.3 Receive Filter for Specific IA

In this section, we model the expected distortion for a given IA as a convex combination of the expected distortion for the ideal IA and expected distortion for the random IA. The weighting constant η used in the combination is determined later using computer simulations. That is, the expected distortion can be written as

$$
\begin{aligned}
E_d^{\mathrm{IA}} &\doteq \eta E_d^{\mathrm{Ideal}} + (1-\eta) E_d^{\mathrm{random}} \\
&= \eta \left\{ tr\left(\mathbf{W}(\mathbf{I}-\mathbf{R})\Sigma_{\mathbf{x}}(\mathbf{I}-\mathbf{R})^T\right) + E_d^{\mathrm{SO}}\left[1 + \frac{4(1-Q)(n+2)}{n^2} tr\left(\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\mathbf{R}^T\right)\right]\right\} \\
&+ (1-\eta)\left\{N\epsilon_N\left[\mathrm{tr}(\mathbf{W}\Sigma_{\mathbf{x}}) + \mathrm{tr}(\mathbf{W}\mathbf{R}\Sigma_\lambda\mathbf{R}^T)\right]\right. \\
&+ \left.(1-N\epsilon_N)\,\mathrm{tr}\left(\mathbf{W}\left(\mathbf{I}-\mathbf{R}\right)\Sigma_{\mathbf{x}}\left(\mathbf{I}-\mathbf{R}\right)^T\right)\right\} + (1-\eta)E_d^{\mathrm{SO}} \\
&= \left[1-(1-\eta)N\epsilon_N\right] tr\left(\mathbf{W}(\mathbf{I}-\mathbf{R})\Sigma_{\mathbf{x}}(\mathbf{I}-\mathbf{R})^T\right) \\
&+ E_d^{\mathrm{SO}}\left(1 + \eta\left[\frac{4(1-Q)(n+2)}{n^2} tr\left(\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\mathbf{R}^T\right)\right]\right) \\
&+ (1-\eta)N\epsilon_N\left[\mathrm{tr}(\mathbf{W}\Sigma_{\mathbf{x}}) + \mathrm{tr}(\mathbf{W}\mathbf{R}\Sigma_\lambda\mathbf{R}^T)\right]. \qquad (2.51)
\end{aligned}
$$

Using straightforward differentiation with respect to $\mathbf{R}$ and equating to zero, we get

$$
\mathbf{R}_{\mathrm{opt}} = \Sigma_{\mathbf{x}}\left[\Sigma_{\mathbf{x}} + \frac{(1-\eta)N\epsilon_N}{1-(1-\eta)N\epsilon_N}\Sigma_\lambda + \frac{4\eta(1-Q)(n+2)E_d^{\mathrm{SO}}}{\left[1-(1-\eta)N\epsilon_N\right]n^2}\mathbf{W}^{-1}\right]^{-1}. \qquad (2.52)
$$

The proportionality constant η can be obtained by simulations[3] for the given IA. For a Gaussian i.i.d. source with variance per dimension $\sigma^2 = 1$, the optimum receive filter simplifies to

$$
\mathbf{R}_{\mathrm{opt}} = \left[\frac{n+2(1-\eta)N\epsilon_N}{n(1-(1-\eta)N\epsilon_N)}\mathbf{I} + \frac{4\eta(1-Q)(n+2)E_d^{SO}}{\left[1-(1-\eta)N\epsilon_N\right]n^2}\mathbf{W}^{-1}\right]^{-1}.
$$

It is interesting to note that even if the weight matrix is identity, $\mathbf{W} = \mathbf{I}$, there still is a correction term corresponding to contribution from the ideal IA. That is, the output of the receive filter is a scaled version of the received codeword. The scale value reduces as E_d^{SO} increases (i.e., less bits used in the quantization) or as η increases. In presenting simulation results, we evaluate (2.51) numerically for the computed $\mathbf{R}_{\mathrm{opt}}$ in (2.52).

Next, we describe and explore another receive filtering technique for mitigating the channel noise.

[3] The parameter η roughly measures what percentage of the IA mapping behaves like an ideal IA. From our numerical simulations, we have found that $\eta = 0.6$ works well for most IAs, and for wide range of B. When $\eta = 0$, the solution in (2.52) converges to (2.9) in Theorem 1, except for some loss in the accuracy of the filter computed using (2.52). This loss is due to the slightly different approximations used to get the closed form expression.

2.4 Semi-Hard Decision VQ for Noisy Channel

In this section, we assume that the channel is a binary symmetric error and erasure channel as described in Section 2.2. Since the receiver uses semi-hard decisions on LLRs (for declaring erasures), the technique is termed semi-hard-decision VQ.

Recall that the average distortion of a B bit quantizer, for an n-dimensional source with source distribution $f_{\mathbf{X}}(\mathbf{x})$ and an error free channel, can be written as

$$E_d^{\text{SO}} = \sum_{i=1}^{2^B} \int_{\mathbf{x}\in\mathcal{R}_i} f_{\mathbf{X}}(\mathbf{x})\ d(\mathbf{x},\mathbf{x}_i)\ \mathrm{d}\mathbf{x}. \tag{2.53}$$

For SOVQ, with MSE as the distortion metric, it is known that [37]

$$E_d^{\text{SO}} = \frac{n}{n+2}\kappa_n^{\frac{-2}{n}}\ 2^{\frac{-2B}{n}} \int_{\mathbf{x}} \lambda^{\frac{-2}{n}}(x)\ f_{\mathbf{X}}(\mathbf{x})\ \mathrm{d}\mathbf{x}, \tag{2.54}$$

where $\lambda(\mathbf{x})$ is the “point density function”. The source optimized point density function $\lambda(\mathbf{x}) = cf_{\mathbf{X}}^{\frac{n}{n+2}}(\mathbf{x})$ where c is a normalization constant [11]. When the index is transmitted via an error free channel, any IA is optimum. However, if the channel can be modeled as a noisy discrete memoryless channel with index transition probability matrix $\mathbf{P}$, then the average distortion can be written as

$$E_d = \sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i) \sum_{j=1}^{2^B} p_{\pi(i)\pi(j)} \int_{\mathbf{x}\in\mathcal{R}_i} \|\mathbf{x} - \hat{\mathbf{x}}_j\|_2^2\ \mathrm{d}\mathbf{x}, \tag{2.55}$$

where $\pi(i)$ represents the IA, which is a bijective map $\pi : [1:2^B] \rightarrow [1:2^B]$ and $\hat{\mathbf{x}}_j$ represents the codeword vectors used in the receiver decoder. As mentioned earlier, IA is an NP complete problem and often sub-optimal methods such as simulated annealing [17], Hadamard transform tool based mapping [20], etc. are used to arrive at a good IA. Thus, in the rest of this chapter, the notation $\pi(i)$ is replaced by i for simplicity under the assumption that a good index assignment has been chosen. The following subsection describes the design of new codewords used at the *receiver*. Then, we analytically derive the distortion for the proposed semi-hard decision VQ scheme. Unlike the previous section, where the receive filter is linear, the receive filter applied here is non-linear, in the sense that it is a function of the erasure threshold γ that is applied to the received LLRs to declare erasures (a design parameter, explained later in this section).

2.4.1 New Codewords

It is known that, for MSE distortion, the optimal codewords $\hat{\mathbf{x}}_j$ at the *receiver* for a discrete error and erasure channel, described by the $\mathbf{P}$ matrix, can be expressed as a linear combination of the codewords $\mathbf{x}_i$ at the *encoder* (i.e., $\mathbf{x}_i$ is the representation vector for $\mathcal{R}_i$, the i^{th} quantization region, in the nearest neighbor-based quantizer) as [11]

$$\hat{\mathbf{x}}_j = \frac{\sum_{i=1}^{2^B} p_{ij}\mathbf{x}_i}{\sum_{i=1}^{2^B} p_{ij}}, \quad 1 \leq j \leq 3^B, \tag{2.56}$$

where there are 3^B codewords at the receiver because each of the B bits could be received correctly, erased, or flipped. For a continuous channel, the optimum decoder output for SDVQ, given the channel statistics $\Pr(R = r|I = i)$, can be written as [22]

$$\hat{\mathbf{x}}_R = \sum_{i=1}^{2^B} \Pr(I = i|R = r)\ \mathbf{x}_i, \quad R \in \mathcal{D}_R, \tag{2.57}$$

where $\mathcal{D}_R$ is the domain of the received signal vector. For example, if BPSK transmission is used for transmitting the indices, then $\mathcal{D}_R = \mathbb{R}^B$. However, defining the codewords as in (2.57) has the drawback that it requires the computation of the probability terms for each continuous-valued received symbol, r. Here, it is proposed to employ the new set of codewords given below for a given $\mathbf{P}$ matrix. The use of the proposed codewords allows one to pre-compute them given the channel statistics, thereby making them more suitable for real-time applications. That is, in the proposed SHDVQ, the following code points are output by the decoder:

(i) when an index j, $1 \leq j \leq 2^B$ is received, $\hat{\mathbf{x}}_j = \mathbf{x}_j$ (i.e., the codeword employed at the transmitter corresponding to index j),

(ii) when a single bit is erased, i.e., for $2^B + 1 \leq j \leq 2^B + B\ 2^{B-1}$, $\hat{\mathbf{x}}_j = \frac{\mathbf{x}_{j,0}+\mathbf{x}_{j,1}}{2}$ where $\mathbf{x}_{j,0}$ (similarly, $\mathbf{x}_{j,1}$) is the codeword at the transmitter with index j, but the erasure bit location is replaced by a bit 0 (similarly, bit 1)[4], and finally

(iii) when multiple bits are erased, the decoder outputs the all zero codeword, i.e., for $2^B + B\ 2^{B-1} + 1 \leq j \leq 3^B$, $\hat{\mathbf{x}}_j = \mathbf{0}$.

[4] Note that, this does not require the receiver to know j, but only the erasure location. Given the erasure location, the receiver can compute the mean of the two codewords of the original codebook that could have resulted in the received bit and erasure sequence.

The next subsection analytically computes the distortion for such a receiver. In the following, we consider different cases for the possible errors and erasures in the codeword indices. We start by computing the expected distortion when either no errors or exactly 1 bit erasures occur. Then, we analyze the expected distortion when multi-bit erasures occur, and when 1 bit errors occur. As mentioned earlier, we ignore the events where multi-bit errors occur, as this is a low-probability event for moderate-to-high SNRs. Finally, we obtain the overall average distortion by multiplying these average distortion expressions with their corresponding probabilities of occurrence.

2.4.2 Average Distortion with 1 Bit Erasure and Ideal IA

Consider the case where at most 1 bit erasures occur, and the IA is ideal. Then, the average distortion, given by (2.55), can be expressed as the sum of the contribution from the correct index reception and an erroneous index reception with 1 bit erasures. Let $\phi_E = (1-\rho)^B$ represent the probability of correct index reception. Then,

$$E_{d,I}^{\mathrm{1E}} = \phi_E E_d^{\mathrm{SO}} + \tfrac{1-\phi_E}{B} \textstyle\sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i) \sum_{j \in S(i)} \int_{\mathbf{x} \in \mathcal{R}_i} \|\mathbf{x} - \hat{\mathbf{x}}_j\|_2^2 \, \mathrm{d}\mathbf{x}, \tag{2.58}$$

where the subscript 'I' is used to emphasize that ideal IA is assumed, E_d^{SO} is the average distortion due to source quantization (given by (2.53)), and $S(i)$ is the set of B indices with 1 bit erasures which have a non-zero probability in i^{th} row of $\mathbf{P}$.

Assuming ideal IA, the closest codewords differ in their indices by 1 bit. For high rate coding, when the shape of the source Voronoi regions are similar, the distortion between the new codewords $\hat{\mathbf{x}}_j$ and the signal vectors in region $\mathcal{R}_i$ can be upper bounded by sum of distortion between the codeword of $\mathcal{R}_i$ and an offset vector $\mathcal{E}$. That is,

$$d(\mathbf{x}, \hat{\mathbf{x}}_j) \leq d(\mathbf{x}, \mathbf{x}_i) + \mathcal{E}_{i,j}, \tag{2.59}$$

where $\mathcal{E}_{i,j} = d(\mathbf{x}_i, \hat{\mathbf{x}}_j)$. Now, $\frac{1}{B}\sum_{j \in S(i)} \mathcal{E}_{i,j}$ can be approximated by the average distortion between the codeword $\mathbf{x}_i$ and the boundary of the Voronoi region $\mathcal{R}_i$. [5] With high rate quantization, it is known that, the Voronoi regions $\mathcal{R}_i$ can be well approximated by using hyper-ellipsoids [27]. That is, the hyper-ellipsoid is the set of $\mathbf{x}$ satisfying the condition

$$(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i) \leq \left(\frac{\upsilon_i^2}{\kappa_n^2}\right)^{\frac{1}{n}}, \tag{2.60}$$

where κ_n is the volume of an n-dimensional sphere of unit radius and υ_i is the volume of the region $\mathcal{R}_i$. With the above approximation, it can be written that

$$\overline{\mathcal{E}}_i \triangleq \frac{1}{B}\sum_{j\in S(i)} \mathcal{E}_{i,j} \quad \approx \left(\tfrac{\upsilon_i^2}{\kappa_n^2}\right)^{\frac{1}{n}}. \tag{2.61}$$

Substituting (2.61) and (2.59) in (2.58), it can be shown that

$$E_{d,I}^{\mathrm{1E}} \approx E_d^{\mathrm{SO}} + (1-\phi_E)\ \kappa_n^{\frac{-2}{n}} \sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i)\upsilon_i^{\frac{2}{n}}\upsilon_i. \tag{2.62}$$

In the above, the volume υ_i can be approximated as $\upsilon_i \approx 1/(2^B\lambda(\mathbf{x}_i))$, and hence, we get

$$E_{d,I}^{1E} \approx E_d^{\mathrm{SO}} + \frac{(1-\phi_E)\ \kappa_n^{\frac{-2}{n}}}{2^{\frac{(n+2)B}{n}}} \sum_{i=1}^{2^B} \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{\lambda^{\frac{n+2}{n}}(\mathbf{x}_i)}.$$

Using the Monte Carlo integration formula,

$$\frac{1}{N}\sum_{i=1}^{N}\beta(\mathbf{y}_i) \doteq \int_{\mathbf{y}} \beta(y)\lambda(\mathbf{y})\mathrm{d}\mathbf{y}, \tag{2.63}$$

where $\doteq$ represents the asymptotic equality when $N \to \infty$, and $\lambda(\mathbf{y})$ represents the point density of the codepoints, $E_{d,I}^{\mathrm{1E}}$ can be written as

$$E_{d,I}^{\mathrm{1E}} \approx E_d^{\mathrm{SO}} + \frac{(1-\phi_E)\ \kappa_n^{\frac{-2}{n}}}{2^{\frac{2B}{n}}} \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})\ \mathrm{d}\mathbf{x}. \tag{2.64}$$

[5] This is valid since the new codewords for the erasure case are approximately at the boundary of the i^{th} Voronoi region.

Also, recognizing that the integral term in (2.64) is proportional to the E_d^{SO} in (2.55),

$$E_{d,I}^{\text{1E}} \approx E_d^{\text{SO}} \left[1 + (1-\phi_E) \left(\frac{n+2}{n} \right) \right]. \tag{2.65}$$

For an n-dimensional Gaussian distributed independent random variables (with zero mean and unit variance in each dimension), the average distortion can be shown to be

$$E_{d,I}^{\text{1E}} \approx \left[1 + (1-\phi_E) \left(\frac{n+2}{n} \right) \right] \frac{2\pi\kappa_n^{\frac{-2}{n}}}{2^{\frac{2B}{n}}} \left(\frac{n+2}{n} \right)^{\frac{n}{2}}. \tag{2.66}$$

The significance of the above equation is that it shows that the average distortion with ideal IA decreases at the *same rate* $2^{\frac{-2B}{n}}$ as that for the average distortion of SOVQ, albeit with larger coefficient. This is in contrast with the high rate distortion for random IA, which is discussed next.

2.4.3 Average Distortion with 1 Bit Erasure and Random IA

When the IA is random, the above development will not work, as 1 bit erasures need not result in codewords corresponding to neighboring cells being received. In this case, the average MSE distortion can be derived from (2.55) as follows. Note that, the codewords at the decoder corresponding to indices with a single bit erasure and random IA when index i is sent, are computed by the receiver as $\hat{\mathbf{x}}_j = \frac{\mathbf{x}_i + \mathbf{x}_l}{2}$ where l is some random index. Then, it follows that

$$E_{d,R}^{\text{1E}} = E_d^{\text{SO}} + \sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i) \sum_{j \in S'(i)} \frac{1-\phi_E}{B} \|\mathbf{x}_i - \hat{\mathbf{x}}_j\|^2 \upsilon_i,$$

where the subscript 'R' is used to emphasize that random IA is assumed, $j > 2^B$ and for these values of $j \in S'(i)$, $\hat{\mathbf{x}}_j = \frac{\mathbf{x}_i + \mathbf{x}_l}{2}$ (for some random l). That is, $S'(i)$ now contains B indices corresponding to codewords of the form $\frac{\mathbf{x}_i + \mathbf{x}_l}{2}$, where the indices l are randomly chosen. The above expression simplifies to

$$E_{d,R}^{\text{1E}} = E_d^{\text{SO}} + \frac{1-\phi_E}{4} \sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i) \left[\frac{1}{B} \sum_{j \in S'(i)} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right] \upsilon_i.$$

Using the Monte Carlo integration formula (2.63), we get

$$\tfrac{1-\phi_E}{4}\sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i)\left[\tfrac{1}{B}\sum_{j\in S'(j)}\|\mathbf{x}_i-\mathbf{x}_j\|^2\right] \approx \sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i)\int_{\mathbf{z}}\|\mathbf{x}_i-\mathbf{z}\|^2\lambda(\mathbf{z})d\mathbf{z} = \tfrac{1-\phi_E}{4}(\sigma_x^2+\sigma_\lambda^2),$$

where σ_x^2 and σ_λ^2 are the variances of the source distribution and point density function, respectively. For the n-dimensional i.i.d. Gaussian example, the total distortion can be simplified to

$$E_{d,R}^{\text{1E}} \approx E_d^{\text{SO}} + \left(\frac{1-\phi_E}{2}\right)\left(\frac{n+1}{n}\right)\sigma_x^2 \tag{2.67}$$

The significance of the above expression is that it clearly shows that the high rate distortion with random IA always floors – the distortion is eventually dominated by the second term, since the E_d^{SO} term reduces as $2^{\frac{-2B}{n}}$. Thus, if a particular practical IA is non-ideal, we would expect its performance would be in between the performance with ideal IA and random IA, which is explored next.

2.4.4 Average Distortion for a given IA

As mentioned earlier, the average distortion analysis is done in two parts: with ideal IA and with random IA. Then, the overall distortion of any given IA is modeled as a convex combination of the distortion with the ideal IA and random IA. The convex combination factor can be seen as a single parameter that measures the "goodness" of the IA, and can be obtained experimentally using simple measurements. Thus, the expression for total distortion for i.i.d. Gaussian source due to 1 bit erasure and a given IA can be modified as

$$E_{d,c}^{\text{1E}} \approx \quad E_d^{\text{SO}}\left[1+\eta_1(1-\phi_E)\left(\tfrac{n+2}{n}\right)\right] + (1-\eta_1)\left(\tfrac{1-\phi_E}{2}\right)\left(\tfrac{n+1}{n}\right)\sigma_{\mathbf{x}}^2, \tag{2.68}$$

where $\eta_1 \in (0,1]$ depends on n, B and the IA. Our simulation results have shown that 60% of the codewords meet the ideal IA condition for the natural IA, and hence, we set $\eta_1 = 0.6$ in evaluating the above expression. This will be demonstrated in the simulation results section.

2.4.5 Average Distortion with Zero Output for Erasures

Consider the case when the decoder outputs zero codeword (the mean of the source) whenever an erasure occurs. In the above analysis, if we replace $\hat{\mathbf{x}}_j$ with all zero codeword $\mathbf{0}$, we can show that

$$E_d^{\text{all zero}} \approx E_d^{\text{SO}} + (1-\phi_E)\sigma_x^2. \tag{2.69}$$

Thus, whenever an erasure occurs, the decoder introduces a maximum error equal to the variance of the source. Comparing (2.67) and (2.69), it can be concluded that using the average between the two codewords always pays-off in reducing the total distortion even when the IA is random. That is, the scheme in (2.67) is better than the scheme in (2.69) in terms of total distortion. This justifies the use of the additional code vectors as proposed in this section, even when the IA is random.

2.4.6 Average Distortion for 1 bit Errors

Along similar lines as the analysis for 1 bit erasures, the 1 bit error case can also be analyzed under the assumptions of ideal IA and random IA. It can be shown that

$$E_{d,c}^{1\text{e}} \approx \quad E_d^{\text{SO}}\left[1 + 4\eta_2(1-\phi_e)\left(\tfrac{n+2}{n}\right)\right] + 2(1-\eta_2)(1-\phi_e)\left(\tfrac{n+1}{n}\right)\sigma_{\mathbf{x}}^2, \tag{2.70}$$

where ϕ_e is the probability of correct reception and η_2, the convex combination, equals η_1 in (2.68), since the IA is the same for both cases.

2.4.7 Average Distortion for the Proposed Receiver

For a given channel, there is a non-zero probability that multi-bit erasures and errors can occur. Here, we consider a decoder that outputs the receiver optimized codewords described in Section 2.4.1 whenever a single bit erasure occurs and outputs the all-zero codeword whenever multi-bit erasures occur. Thus, the total distortion can be written as

$$E_d^{\text{total}} = p_0\ E_d^{\text{SO}} + p_{1e}\ E_{d,c}^{1\text{e}} + p_{1E}\ E_{d,c}^{1\text{E}} + p_{\text{rest}}\ E_d^{\text{all zero}}, \tag{2.71}$$

where $p_{\text{rest}} = 1-(p_0+p_{1e}+p_{1E})$, p_0 is the probability that no error has occurred, p_{1e} is the probability that 1 bit error has occurred and p_{1E} is the probability of 1 bit erasure. For the given SNR and erasure threshold γ, one can compute these probabilities easily using the Gaussian error function. In the following section, we determine the optimal threshold for minimizing the total distortion.

2.4.8 Optimum Threshold Selection

The tradeoff involved in selecting the optimal threshold is as follows. The average distortion due to erroneously received bits exceeds that obtained by declaring the error-bit as an erasure. This is because erasures lead to codewords that are closer to the transmitted codeword, compared to the codeword output by the receiver when an error occurs. Hence, if a bit is likely to be in error, it is better to declare it as an erasure, but if it is likely to be received correctly, it is better accept the hard-decisions. So, if the erasure threshold (i.e., the threshold on the LLRs below which the bit is declared to be in erasure) is too high, there will be many erasures (and even bits that could be decoded correctly will be declared as erasures), leading to a larger average distortion. A similar argument applies when the threshold is too low, leading to large distortion because many bits will be received incorrectly. Thus, there exists an optimal value of the threshold for a given SNR. In order to compute the optimum threshold, one needs to consider the distortion incurred due to error cases also. Let γ be the threshold on the LLR below which a bit is declared to be in erasure. Then, the probability of correct index reception is given by

$$1 - \alpha - \rho = Q\left(\frac{\gamma - 2\text{SNR}}{2\sqrt{\text{SNR}}}\right),$$

where $Q(.)$ is the standard Gaussian tail function for computing Gaussian tail error probability. [6] Along similar lines, the probability of receiving with 1 bit error is given by

$$\alpha = Q\left(\frac{\gamma + 2\text{SNR}}{2\sqrt{\text{SNR}}}\right).$$

From the above equations, ρ can be found and used for computing the $\mathbf{P}$ matrix. The optimum γ^* can be found by differentiating E_d^{tot} with respect to γ and equating to zero. However, since this is mathematically cumbersome, numerical computation is used to find the optimum γ^* for the given values of SNR, B and n. Figure 3.3 shows the value optimum threshold γ for various values of SNR for a 2-dimensional Gaussian source, with $B = 6$. It can be noticed that γ gets smaller as the SNR increases. This is expected, since, as the channel approaches an error free channel, the receiver should declare erasures less frequently.

[6] $Q(x)$ is defined as $\frac{1}{2}\text{erfc}\left(\frac{x}{\sqrt{2}}\right)$, where erfc() is the standard Gaussian complementary cumulative distribution function.

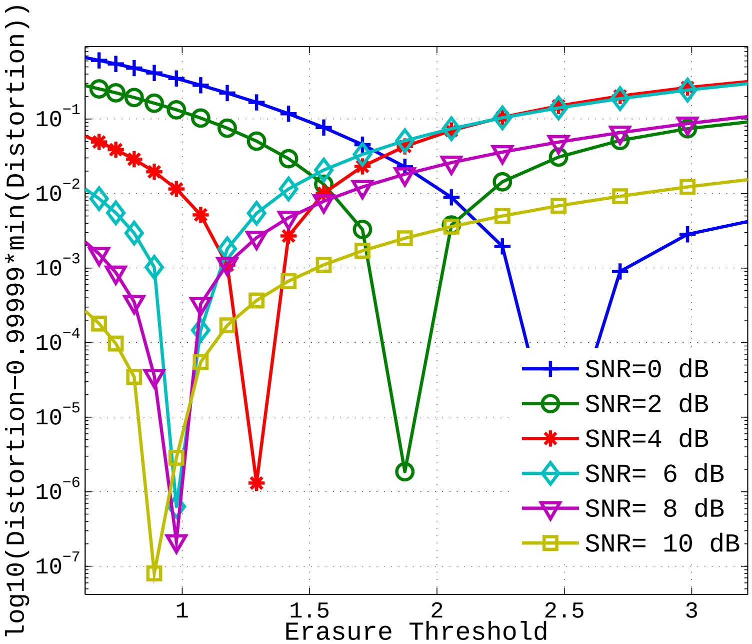

Figure 2.3: Optimum erasure threshold as function of SNR for 2-dimensional Gaussian i.i.d. source and $B = 6$.

2.5 Simulation Results

In this section, we validate the analytical expressions derived above and illustrate the improvement in the average distortion that can be obtained through the above LRF and SHDVQ, using Monte Carlo simulations. An n-dimensional i.i.d. zero mean Gaussian distributed vector with unit variance per dimension is used as the source and $50,000$ instantiations are used for generating the optimal encoder codebook using the Lloyd-Max algorithm [35]. The covariance of this source-optimized codebook is used as the covariance of the point density, for evaluating the theoretical expressions. Another set of $50,000$ instantiations are used for encoding the source using the above computed codebook. The index from the encoder is sent over a noisy channel and the optimal LRF is applied at the decoder before computing the end-to-end distortion. The noisy channel is modeled as a binary symmetric channel (BSC) with transition probability $q = Q(\sqrt{2\ \mathrm{SNR}})$ [38] that depends on the SNR per bit, and $B = \log_2 N$ bits are employed for source quantization.

2.5.1 Receive Filter for Non-Fading Channel

Figure 2.4 shows the average MSE and WMSE distortion for the source-optimized VQ with $n = 3$ and $B = 9$ bits. In the WMSE case, the matrix:

$$\mathbf{W} = \begin{bmatrix} 0.69 & 0.26 & 0.03 \\ 0.26 & 0.44 & 0.43 \\ 0.03 & 0.43 & 1.87 \end{bmatrix},$$

accurate to 2 decimals, was used. The matrix was generated using a random unitary matrix as eigenvectors and eigenvalues equal to $[2.0, 0.8, 0.2]$, which ensures that $\mathrm{tr}(\mathbf{W})$ is the same as in the MSE case. The excellent match between the simulation results and the theoretical expression is clear from the figure. At low SNR, the distortion with and without filtering are found to approach n and $2(n+1)$, which matches with (2.10) and (2.23), respectively. Also, at high SNR, all schemes converge to the high-rate distortion of VQ for noiseless channels, as expected.

Figure 2.5 compares the MSE performance of SQ and VQ based LRF for reducing the total distortion, with $n = 2$ and $B = 8$. It is interesting to note that the LRF greatly diminishes the performance difference between SQ and VQ. This corroborates with the theoretical expressions, since, at high rate, neglecting the terms of order $N^{-2/n}$, we have

$$E_{d,SQ}^{R_{\mathrm{opt}}} - E_{d,VQ}^{R_{\mathrm{opt}}} = \frac{2N\epsilon_N(1 - N\epsilon_N)^2(n-1)}{(1 + 2N\epsilon_N)(1 + 2N\epsilon_N/n)}.$$

Figure 2.6 plots the ratio of the average MSE of the LRF, joint transmit-receive linear filter, the scaled codebook [33] and COVQ [12], to the average MSE of SQ with no filtering. For obtaining this plot, $n = 4$ and $B = 48$ bits were used. The COVQ provides the best performance for all SNRs, at the cost of a computationally expensive reoptimization of the codebook and feedback of the optimal codebook from the receiver to the transmitter for every SNR. This is followed by the performance of the joint transmit-receive linear filter, which also requires the feedback of the optimal transmit filter. The joint transmit-receive linear filter outperforms the scaled codebook at all SNRs. The scaled codebook outperforms the LRF at high SNR, while the LRF performs better for SNR ≤ 4 dB. However, the scaled codebook requires numerical computation of the optimal scale factor and its adaptation at both the transmitter and the

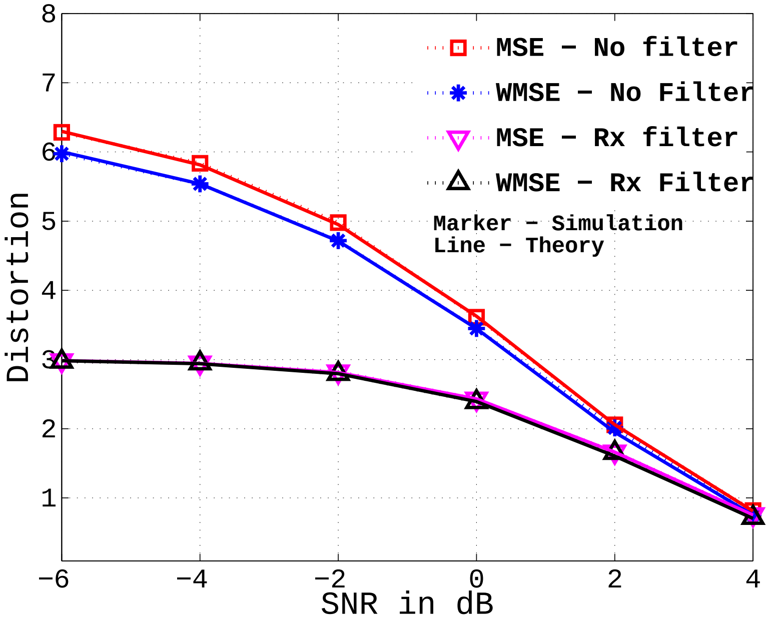

Figure 2.4: MSE and WMSE *vs.* SNR for a 3-D Gaussian i.i.d. random source with $B = 9$ bit VQ.

receiver for every SNR. The LRF is the least computationally expensive and the simplest to implement. It always performs better than no filtering, and offers several dBs of improvement in the average MSE for practical SNR values.

Finally, Table 2.1 lists the percentage improvement in MSE distortion from the LRF compared to the no-filtering case. We observe that the percentage improvement is the highest for small N and n, and is higher for larger $N\epsilon_N$. Moreover, in all three cases of $N\epsilon_N$, comparing the percentage improvement for $N = 64$ with $N = 512$, we observe that it has decreased for $n = 1$ and 2, but increased for $n = 3$. That is, as N is increased, the percentage improvement increases till about 4 bits per dimension, after which it starts to decrease. These observations agree with the theoretical expressions obtained above.

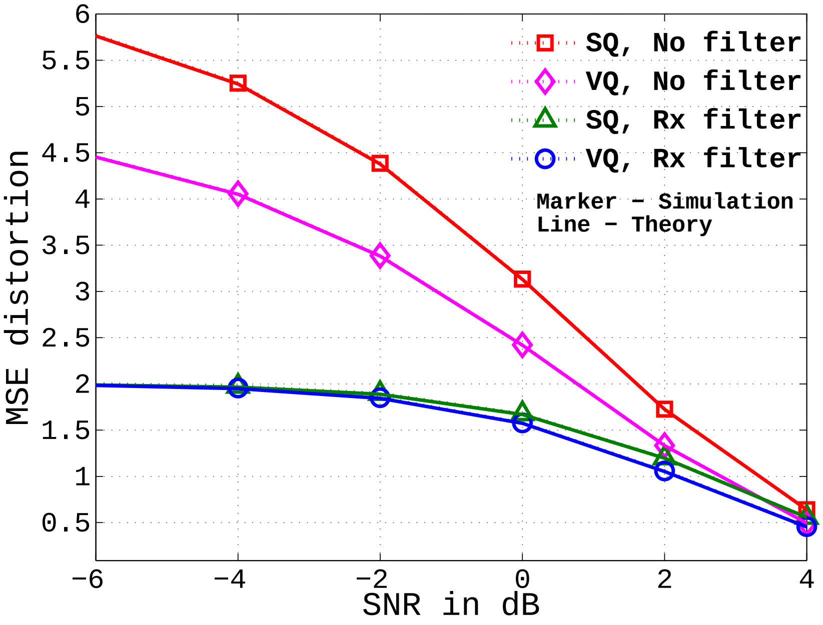

Figure 2.5: Performance comparison of the LRF designed for SQ and VQ with $n = 2$ and $B = 8$ bits.

2.5.2 Receive Filter for Fading Symmetric Error Channel

In this simulation, we consider a Rayleigh fading channel to first compute the BSC bit transition probability q for each channel instantiation, and then average the expected distortion performance over 500 channel instantiations. The indices obtained after source compression are transmitted over the channel and the optimal receive filter computed under two scenarios: (a) optimal receive filtering for each channel instantiation, and (b) optimal receive filtering for the "average channel", i.e., where the optimal LRF is computed for the average q for the given SNR. The simulation results are compared for no receiver filtering, receive filter for *Random IA* (i.e., symmetric error channel) as well as for receive filter using specific IA (obtained using the LISA [20]). The convex combination parameter $\eta = 0.6$ is used for this simulation also.

Figure 2.7 plots the average MSE versus SNR for a 2-dimensional i.i.d. zero mean Gaussian source with unit variance per dimension, and with 6-bit quantization. It can be observed that the total distortion in the fading channel is improved with LRF compared to no filtering. Moreover, with only a small loss in the performance, one can also use the single receive filter computed using the average channel cross-over probability. The advantage of using the LRF is also shown

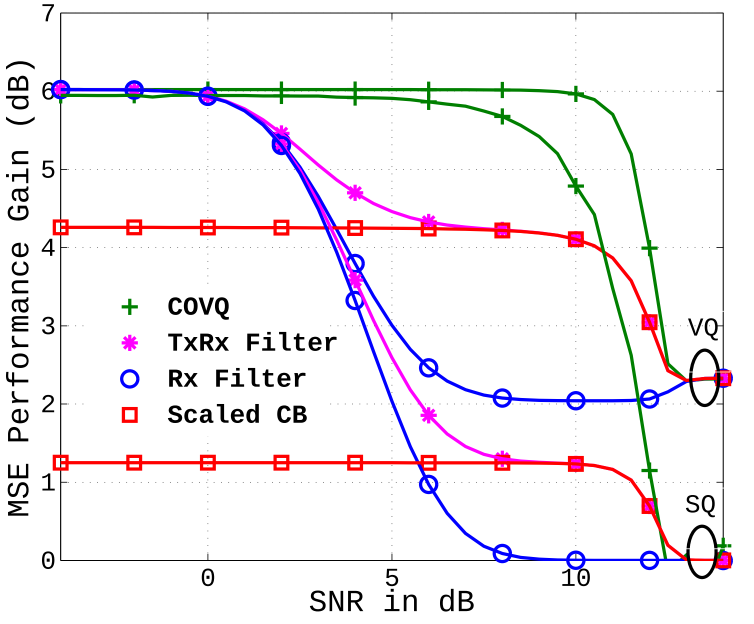

Figure 2.6: Comparison of the LRF, joint transmit-receive filter, scaled codebook and COVQ for $n = 4$ and $B = 48$ bits.

in terms of an improvement in the fade margin required to achieve a given target distortion in Table 2.2. Here, the target distortion is set as 6 and 12 dB above the minimum distortion that can be achieved in the absence of channel errors for the given number of quantization bits. It can be observed that there is about 2 dB improvement due to *good IA* and another 2 dB improvement due to receive filtering for a *specific IA* (highlighted in the table in bold font). Moreover, there is 1 dB difference in the performance of receive filter designed for average q and receive filter designed on a per-channel instantiation basis.

2.5.3 SHDVQ for AWGN

Here, comparison is done between the average distortion for SOVQ with both noiseless and noisy channels, receive filter and the proposed SHDVQ. The encoder output is sent via an AWGN channel for with noise variance depending on the SNR condition. At the receiver, the LLR values of the bits comprising the received indices are computed, and used to declare erasures, by comparing the LLR to a threshold. The value of the threshold is chosen by numerically minimizing E_d^{tot} in (2.71).

Thus, the total distortion was computed for various SNRs, and the results for a 2-dimensional

Table 2.1: Percentage performance improvement due to receive filtering.

$\frac{E^{\text{conv}}_{d,VQ}-E^{R_{\text{opt}}}_{d,VQ}}{E^{R_{\text{opt}}}_{d,VQ}} \times 100$	$N\epsilon_N = 0.05$		$N\epsilon_N = 0.1$		$N\epsilon_N = 0.25$	
	Sim.	Th.	Sim.	Th.	Sim.	Th.
$n = 1, N = 64$	6.65	6.72	13.63	13.72	36.15	36.32
$n = 1, N = 512$	2.77	2.82	5.72	5.77	15.55	15.65
$n = 2, N = 64$	3.97	4.13	9.12	9.33	26.20	26.53
$n = 2, N = 512$	3.09	3.54	6.89	7.36	19.47	19.96
$n = 3, N = 64$	2.11	1.89	5.45	5.15	17.57	17.21
$n = 3, N = 512$	2.85	3.19	6.88	7.21	20.41	20.60

Table 2.2: Improvement in the link margin under Rayleigh fading channels. The table lists the channel SNR needed to achieve a target distortion of 4 E_d^{SO} and a target distortion 16 E_d^{SO}, for $B = 8$ and $B = 6$, and for a 2-dimensional standard Gaussian source.

No. of Quant. bits	Target distortion (dB)	Channel SNR needed (dB)					
		No Filter		Rx Filter (avgCh)		Rx Filter	
		Random IA	Specific IA	Random IA	Specific IA	Random IA	Specific IA
$B = 8$	4 E_d^{SO}	**24**	**21**	24	21	21.5	**19.5**
	16 E_d^{SO}	**17**	**13.7**	16.5	13.5	15	**12.3**
$B = 6$	4 E_d^{SO}	**17.2**	**15**	17	14.7	15.1	**13.5**
	16 E_d^{SO}	**9.5**	**7**	8.5	5.5	6.5	**4.5**

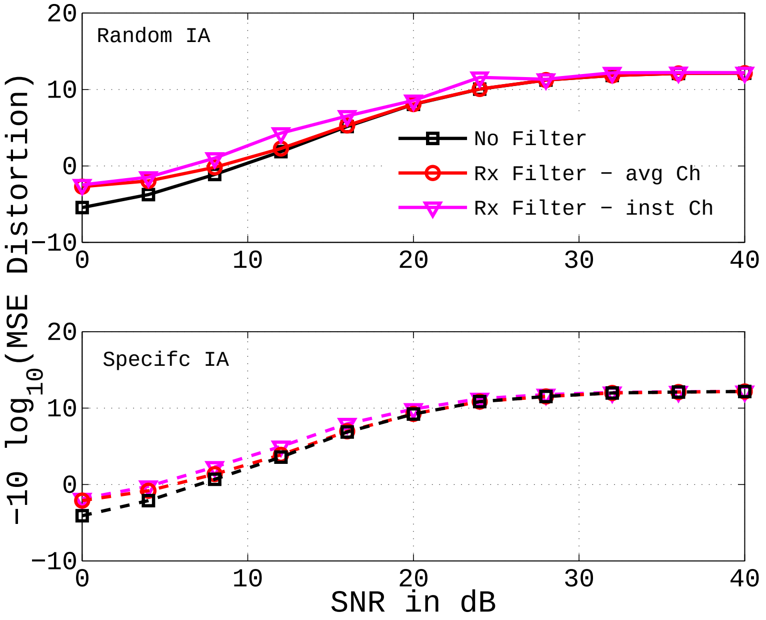

Figure 2.7: MSE *vs.* SNR for a 2-dimensional Gaussian i.i.d. random source with $B = 6$. The index is sent over a Fading Binary Symmetric channel.

Gaussian vector with $B = 6$ is given in Figure 2.8. The figure also compares the performance with COVQ [29] and receive filtering [34] under *Random IA* as well as *specific IA*. It can be observed that the SHDVQ method performs as well as or better than COVQ and receiver filtering method at all SNRs for *Random IA*. However, for specific IA, receive filtering outperforms SHDVQ. Figures 2.9 and 2.10 demonstrate the accuracy of the analytical expressions for the performance of SHDVQ with various number of quantization bits and with $\eta = 0.6$, as before.

2.6 Summary

In this chapter, we presented two receiver-only adaptation techniques for source quantization over noisy channels, that help to mitigate the excess distortion incurred due to index errors caused by the noisy channel. The advantage of these methods is that encoder remains agnostic to the channel, and hence require no feedback from the source decoder to the source encoder. Analytical expressions were derived for the average distortion and the optimal linear receive filter was designed for an symmetric error channel by optimizing the approximate total distortion with

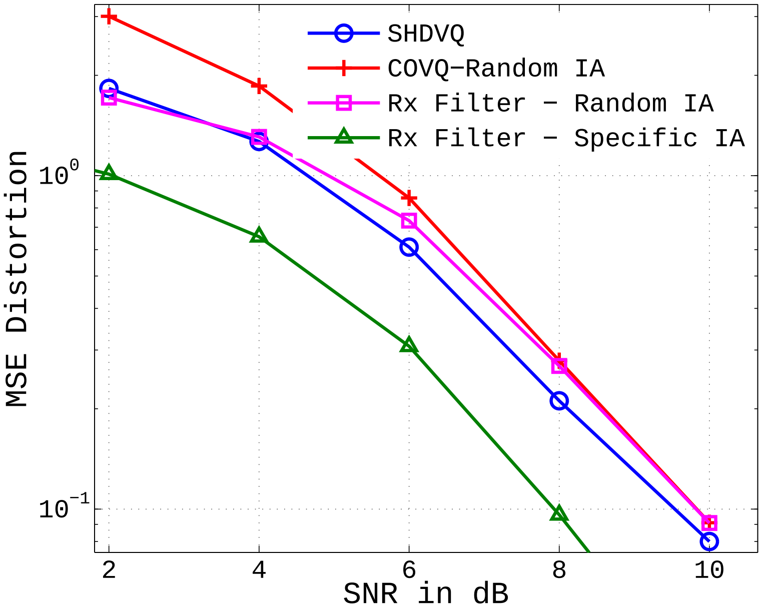

Figure 2.8: Comparison of distortion *vs.* SNR for 2-dimensional Gaussian i.i.d. Source with $B = 6$.

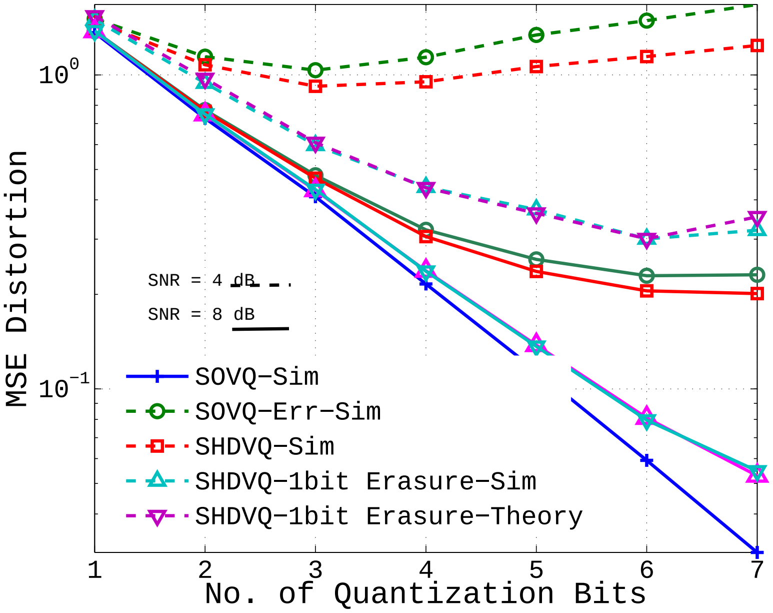

Figure 2.9: MSE performance for a 2-dimensional Gaussian i.i.d. source with a fixed IA and different SNRs.

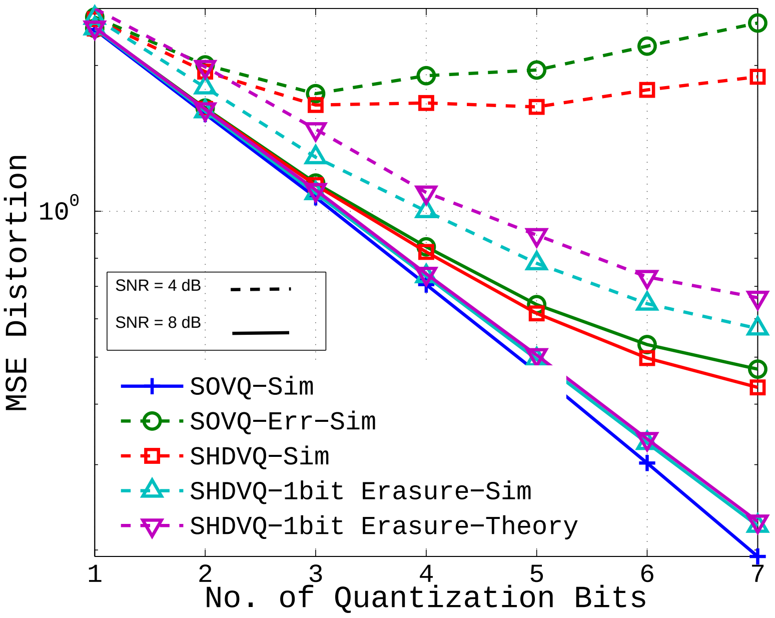

Figure 2.10: MSE performance for a 3-dimensional Gaussian i.i.d. source with a fixed IA and different SNRs.

the weighted-MSE distortion metric. We also designed the receive filter for a specific IA, by modeling the total distortion as convex combination of distortion for random IA and ideal IA. The performance of the receive filter-based approach was compared with no filtering, COVQ and transmit filter based methods. It was shown that the receive filter provides an improvement in the average distortion compared to no filtering, and it performs as well as COVQ at low SNR. The performance under a fading symmetric error channel was evaluated numerically and improvement in average distortion due to the receive filter was illustrated in this case also. Then, an another receive processing technique for continuous channel such as the AWGN channel was presented with semi-hard decisions at the receiver. The end-to-end distortion for the proposed receiver codebook was analyzed for both ideal IA and random IA. Simulation results were presented to demonstrate the accuracy of the analytical expressions as well as illustrate the performance improvement under noisy channel conditions. In conclusion, the receive processing techniques proposed in this chapter are simple, easy to implement, and help in improving the average distortion performance when the output of the source encoder is transmitted over a noisy channel whose statistics are available at the receiver. In Chapter 5, we apply these techniques to the problem of reverse-link CSI feedback, and show the improvement in the forward-link throughput that is attainable by using these techniques for mitigating the distortion caused due to feedback channel errors.

Chapter 3

Trellis Coded Block Codes

"Om! That (Brahman) is infinite, and this (universe) is infinite. The infinite proceeds from the infinite. (Then) taking the infinitude of the infinite (universe), It remains as the infinite (Brahman) alone." - **Brihadaranyaka Upanishad**

3.1 Introduction

3.1.1 Motivation and Prior work

Recall that an important factor that determines the performance of a MIMO wireless link is the availability of accurate and up-to-date CSI at the transmitter. Hence, it is desirable to have low complexity and low latency channel codes which can help in improving the quality of the CSI estimated at the transmitter. Most of the known block codes which have good distance properties, unfortunately, also have code latencies proportional to their code length (e.g., RS or BCH codes [39]). On the other hand, Convolutional Codes (CC) have low latency, but designing high rate codes with good distance property is computationally very intensive and there is no systematic design procedure available. In this chapter, we design *Trellis Coded Block Codes* (TCBC), which combine the benefits from both the block codes and trellis codes. Moreover, the TCB codes are applicable to both discrete channels and continuous channels. The TCB code is obtained by concatenation of a block codes and a convolutional codes in a novel way.

Combining the block codes and trellis codes has been done in multiple ways. For example,

(i) Trellis Coded Modulation (TCM) [40, 41], (ii) Concatenated codes [42], and (iii) Multi-level coding (MLC) [42] are some of the popular techniques. But, each of these techniques have their own shortcomings. For example, TCM depends on constellation expansion and it is not applicable for discrete channels, while concatenated codes have large decoding latency. The MLC also has large decoding latency, and the multi-level decoder is sub-optimal.[1] Our approach of concatenation is different from all the above in the sense that, we select one of the sub-codes of the block code depending on the trellis code output. Although this is similar to the TCM, the difference lies in the application of uniform code partitioning theorem derived in this thesis to generate the sub-codes, and in the procedure for selection of the codeword within a sub-code by the rest of the data bits.

To put the uniform code partitioning that we explore in this work in context, there are four known ways of partitioning linear block codes. These are: (i) the sub-codes obtained by a coset decomposition [43], (ii) sub-codes of smaller length which constitute the mother code by concatenation [44], (iii) an association scheme to group code-pairs with a given Hamming distance between them [39], and (iv) decomposing N-tuple codes into cosets using finite groups that exploit the algebraic structure available in Cartesian product space [45]. In this work, we describe another algebraic structure, which we call uniform sub-code partitioning, that allows one to partition code words into disjoint sub-code sets with a certain uniform distance property. We provide the details in the later sections.

3.1.2 Contributions

In this chapter, TCBC are designed along the similar principle of TCM, but exploiting the underlying algebraic structure in all linear block codes rather than constellation expansion.

The main contributions in this chapter are:

- A new algebraic result describing partitioning of linear block codes into uniform sub-codes.
- An encoder/decoder structure which utilizes the uniform sub-set partitioning in Hamming space (and hence is usable in both discrete and continuous channels) is proposed. The codes so constructed are referred to as Trellis Coded Block Codes (TCBC). The proposed decoder is shown to be a Maximum Likelihood Sequence Detection (MLSD) decoder.

[1] The ML decoder for MLC is computationally very complex.

- Analytical expressions for an upper bound on the BER performance of the TCBC are derived.

- Perhaps the most practically useful implication of this work is that it provides a systematic method of constructing fixed length codes with a desired minimum distance property, using off-the-shelf block and convolutional codes as building blocks. These provide a system designer a convenient way to custom design codes with desired distance properties while also meeting a given decoding latency requirement.

3.2 Uniform Code Partitioning

The following definitions are used in the description of the new algebraic structure of Linear Block Code (LBC) in the binary field $\mathbb{F}_2^n$.

Definition 1. Uniform set: *A set in $\mathbb{F}_2^n$ is said to be* **uniform** *if the distance d_u between any pair of elements is a constant.*

Note that the above definition differs from that of the equi-distant codes used in the coding theory literature [39] in that a uniform set need not be closed under addition.

Definition 2. Maximal uniform set: *A uniform set U is said to be* **maximal** *if it is the largest possible set in terms of cardinality, for the given length n and the uniform distance d_u.*

Definition 3. Non-trivial set: *A set U is said to be* **non-trivial** *if it contains atleast* 3 *non-zero elements.*

If the uniform code is linear, it is a constant weight code, except for the all zero code-word. Hadamard codes [39] are an example of such a uniform code. One straightforward method of partitioning a given code into uniform sub-sets is pair-wise partitioning,[2] where elements are grouped into pairs, such that the distance between the pairs is constant, as explained in the lemma below.

Lemma 1. *For any LBC, there exists a disjoint code-word pair set (partitioning) such that distance between the code-word pairs is constant. In fact, there exists at least one code-word pair partition for every Hamming weight in the code's distance spectrum.*

[2] This pair-wise partitioning is unrelated to the association scheme [46]. Association schemes find all pairs of code-words that are at various Hamming distances whereas pairwise partitioning finds disjoint pairwise subsets of the parent code.

Proof. Let d be a distance in the distance spectrum of $\mathcal{C}$. Then, there exists atleast one code-word $\mathbf{c}_1$ such that $\mathcal{D}_H(\mathbf{c}_1, \mathbf{0}) = d$. Now, add any other code-word $\mathbf{c}_2$ to both $\mathbf{0}$ and $\mathbf{c}_1$ to get the codeword pair $(\mathbf{c}_2, \mathbf{c}_1 + \mathbf{c}_2)$ satisfying $\mathcal{D}_H(\mathbf{c}_2, \mathbf{c}_1 + \mathbf{c}_2) = d$. Proceeding this way, we can create disjoint code pairs with distance d for every Hamming weight in the distance spectrum of $\mathcal{C}$. □

While the above lemma is useful, it is desirable to have partitions that have more than two code-words in each subset. This is because having fewer subsets helps reduce the complexity of the outer trellis code that will be introduced later. That is, we seek to find a partition

$$\mathcal{C} = \bigcup_{i=1}^{L} \mathbf{C}_i, \tag{3.1}$$

such that $\mathbf{C}_i \cap \mathbf{C}_j = \{\phi\}$, $1 \leq i, j \leq L$, $i \neq j$, where L is the number of constituent uniform sub-sets and $\mathbf{C}_i$, $i = 1, 2, \ldots, L$ are *non-trivial* uniform sub-codes.

In the sequel, Theorem 2 asserts that such partitioning exists for many binary LBCs. Now, we state and prove some useful properties of uniform linear sub-codes which will set the stage for stating Theorem 2.

Lemma 2. *The distance d_u for any non-trivial uniform linear code $\mathbf{C}_0$ is even. Moreover, the uniform code is linear if and only if $d_u = 2\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1)$ for any two non-zero $\mathbf{c}_0, \mathbf{c}_1 \in \mathbf{C}_0$.*

Proof: See Appendix C.1.

Remark 2. An immediate consequence of the above Lemma is that the uniform distance of a non-trivial linear uniform code and its even parity extension code are the same. This property will be used in the derivations to follow.

Lemma 3. *Let $\mathbf{c_0}, \mathbf{c}_1, \mathbf{c}_2$ belong to a uniform linear code with distance d_u and $\mathbf{c}_i \neq \mathbf{0}$, for $i = 0, 1, 2$. Then, $\mathbf{c_0} = \mathbf{c}_1 + \mathbf{c}_2$ if and only if $\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \mathbf{c}_2) = 0$.*

Proof: See Appendix C.2.

Remark 3. Combining Lemmas 2 and 3, it is immediate to see that if $\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \mathbf{c}_2) = d_u/4$, then $\mathcal{W}_H(\mathbf{c}_0 + \mathbf{c}_1 + \mathbf{c}_2) = d_u$, where $\mathbf{c}_0$, $\mathbf{c}_1$ and $\mathbf{c}_2$ belong to a uniform linear code $\mathbf{C}_0$ with atleast 8 code-words.

Lemma 4. *There exists a non-trivial uniform sub-code* $\mathbf{C}_u$ *of the rate-1 code* $\mathbb{F}_2^n$ *with*

$$d_u = \begin{cases} \frac{n}{2} & \textit{if} \quad n = 4k \\ \frac{n-1}{2} & \textit{if} \quad n = 4k+1 \\ \frac{n+2}{2} & \textit{if} \quad n = 4k+2 \\ \frac{n+1}{2} & \textit{if} \quad n = 4k+3 \end{cases}, \tag{3.2}$$

where k *is an integer* ≥ 1*. Moreover, a uniform* linear *subset* $\mathbf{C}_0^F$ *which spans a vector space with dimension at least* 2 *can be constructed from* $\mathbf{C}_u$*.*

Proof: See Appendix C.3.

We can now state the main theorem of this section.

Theorem 2. *For a binary LBC* $\mathcal{C}$*, if* $\mathbf{C}_0 \triangleq \mathbf{C}_0^F \cap \mathcal{C}$ *is a non-trivial uniform set for some* $\mathbf{C}_0^F$ *satisfying the properties in Lemma 4, the following hold:*

(i) Tiling property: $\mathbf{C}_0$ *and its cosets tile* $\mathcal{C}$ *and one can build* $\mathbf{C}_0^{\max}$*, a linear maximal uniform sub-code of* $\mathcal{C}$*, from the cosets of* $\mathbf{C}_0$*,*

(ii) Cardinality bounds: *The cardinality of* $\mathbf{C}_0$ *is bounded as* $2^2 \leq |\mathbf{C}_0| \leq 2^{\lfloor \log_2 n+1 \rfloor}$*, and*

(iii) Cardinality of the maximal linear uniform set: $|\mathbf{C}_0| = 2^{j^*+1}$ *where* $j^* \geq 1$ *is the largest integer such that* (a) $\mathbf{C}_0$ *has a subset* $\mathbf{C}_{j^*}$ *with cardinality* j^*+1 *and non-zero entries such that*

$$\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \ldots * \mathbf{c}_{j^*}) = \frac{d_u}{2^{j^*}}, \tag{3.3}$$

where $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{j^*} \in \mathbf{C}_{j^*}$*, and* (b) *For* $l = 1, 2, \ldots, j^*-1$*, for all subsets* $\mathbf{C}_l$ *of* $\mathbf{C}_{j^*}$ *with cardinality* $l+1$*,*

$$\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \ldots * \mathbf{c}_l) = \frac{d_u}{2^l}, \tag{3.4}$$

where, with a slight abuse of notation, $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_l \in \mathbf{C}_l$*.*

Proof: See Appendix C.4.

Discussion: The above theorem suggests the following procedure for obtaining uniform sub-codes for a given code $\mathcal{C}$:

(i) Find code-words in $\mathcal{C}$ with Hamming weight d_u given by (3.2). Denote this sub-set as $\mathbf{C}_u$.

(ii) Find a sub-set of $\mathbf{C}_u$ that is closed by using the linearity conditions (3.3) and (3.4) in case (iii) of Theorem 2. Denote this sub-set as $\mathbf{C}_0$.

(iii) Now, $\mathbf{C}_0$ and its cosets form a uniform partitioning of $\mathcal{C}$.

If no non-trivial $\mathbf{C}_0$ can be found in step (ii), then a non-trivial uniform sub-code partitioning does not exist for the code $\mathcal{C}$. In this case, we partition the code into subsets with pairs of elements. The sub-sets are constructed from the pair of code-words formed by the all-zero code-word and one of the code-words in $\mathbf{C}_u$, and its cosets.

Note that the partitioning of $\mathcal{C}$ into its maximal uniform linear sub-codes is not unique. However, the number of partitions L is unique due to the linearity of the sub-codes. Moreover, for a given number of subsets, the uniform subset partitioning results in the maximum possible $d_{\min}$ among all possible partitions into non-trivial subsets. The utility of finding uniform distance sub-codes with large cardinality, as given by Theorem 2, will be seen in the TCBC construction, presented in the next section. Specifically, the larger the cardinality of the uniform subsets, the lesser the complexity of the CC associated with constructing the TCB code. In general, the uniform partitioning theorem reveals the fundamental structure in the LBCs that is used to construct the code. This idea is schematically illustrated in Figure 3.1. That is, in the case of LBCs, for the given length, the code is built using the one underlying uniform coset, by taking union of the shifted versions of the coset.

Examples: The following examples illustrate the uniform partitioning given in Theorem 2. The Hamming $(7,4,3)_2$ code can be partitioned into two *maximal uniform sub-sets*[3] as follows: $\mathbf{C}_0 = \{0,1,6,7,10,11,12,13\}$ and $\mathbf{C}_1 = \{2,3,4,5,8,9,14,15\}$. All the code words in each sub-set are at an equal Hamming distance ($d_u = \frac{n+1}{2} = 4$) from each other. Another example is the Maximum Length Shift Register (MLSR) $(6,3,3)_2$ code[4], which can be partitioned into $\mathbf{C}_0 = \{0,3,4,7\}$ and $\mathbf{C}_1 = \{1,2,5,6\}$. The Hamming distance between any pair of elements in both sub-sets $\mathbf{C}_0$ and $\mathbf{C}_1$ is $d_u = \frac{n+2}{2} = 4$. Hadamard codes are themselves maximal uniform codes. Hence, any partition would give a non-maximal, but uniform sub-sets with $d_u = \frac{n}{2}$. MLSR $(9,4,3)_2$ code[5] can be partitioned into uniform cosets with $\mathbf{C}_0 = \{0,2,9,11\}$ and uniform

[3] The code-words in the sub-set are denoted by indices which are the decimal equivalent of the binary data vectors in the code $\mathcal{C}$. The binary data vectors are the binary k-tuples which get multiplied by the generator matrix $\mathbf{G}$ to generate the code-words in $\mathcal{C}$.

[4] $\mathcal{C} = \{00,16,23,35,47,51,64,72\}$ in octal notation.

[5] $\mathcal{C} = \{000,075,107,172,217,262,310,365,436,443,531,544,621,654,726,753\}$ in octal notation.

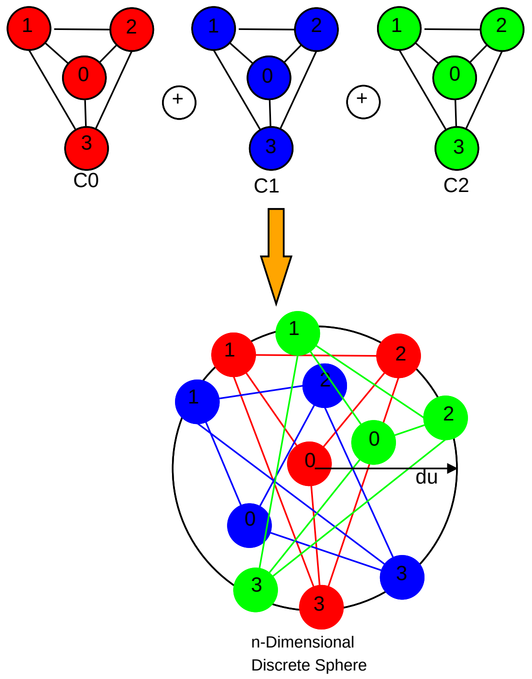

Figure 3.1: Structure of LBC as union of cosets.

distance $d_u = \frac{n-1}{2} = 4$. The binary Golay $(23, 12, 7)_2$ code can be partitioned into 512 cosets of 8 elements each. Elements in the uniform linear coset is listed here as an illustration: $\mathbf{C}_0$ $= \{0,\ 120,\ 79,\ 929,\ 1434,\ 1764,\ 1739,\ 2508\}$. The uniform distance of this sub-set $\mathbf{C}_1$ is 12. Similarly, the Bose-Chaudhuri-Hocquenghem (BCH) $(31, 11, 11)_2$ code can be partitioned into 128 cosets with 16 elements each. The generator polynomial for this cyclic code is given by $g(x) = x^{20} + x^{18} + x^{17} + x^{13} + x^{10} + x^9 + x^7 + x^6 + x^6 + x^4 + x^2 + 1$. The elements of a uniform coset, of the code generated by the non-systematic generator matrix built from $g(x)$, is given here for illustration: $\mathbf{C}_1 = \{4,\ 8,\ 59,\ 69,\ 256,\ 422,\ 487,\ 491,\ 542,\ 545,\ 962,\ 1010,\ 1084,\ 1193,$ $1561,\ 1627\}$. The uniform distance for this sub-set is 16, which is again greater than 11, the $d_{\min}$ of the parent BCH code. It can be noticed that for these binary LBCs,the uniform distance of the sub-sets is $\frac{n}{2}$, $\frac{n-1}{2}$, $\frac{n+2}{2}$ or $\frac{n+1}{2}$, as given in (3.2). Finally, an example of a code without a non-trivial uniform sub-code partitioning is the MLSR $(9, 3, 4)$ code[6] with $\mathbf{C}_0 = \{0, 1\}$ and $d_u = \frac{n-1}{2} = 4$.

[6] $\mathcal{C} = \{000, 164, 235, 351, 472, 516, 647, 723\}$ in octal notation.

3.3 Trellis Coded Block Codes

We now introduce TCBC, a new family of codes based on the aforementioned uniform subset partitioning. We construct the (n, k) TCB code with k input bits and n-output bits as follows. Let l of the k bits be input into a rate $\frac{l}{m}$ trellis code, whose output selects one of 2^m sub-sets (partitions) of an $(n, k-l+m)$ LBC, and the $k-l$ additional input bits select one of the code-words in the selected sub-set, resulting in an (n, k) TCB code. Note that, n is the length of the LBC $\mathcal{C}$ with 2^{k-l+m} code-words, which is partitioned into 2^m uniform sub-sets with 2^{k-l} code-words in each sub-set.

The minimum distance of this hybrid code is determined by the smaller of the (constant) minimum distance between code-words in each sub-set and $\frac{K-1}{l}d_{\min}$, where K is the constraint length of the CC and $d_{\min}$ is the minimum distance of the parent LBC. [7] We choose K such that $\frac{K-1}{l}d_{\min} > \min_{1\leq i\leq 2^m} d_{\min}(\mathbf{C}_i)$, and hence the minimum distance of the TCB code is determined by the uniform distance in any sub-set (assigned to parallel transitions of the trellis of the CC). Thus, by choosing the sub-set partitioning appropriately, one can obtain a coding gain in a conceptually similar manner to TCM/coset codes. Hence, the TCBC can be viewed as a generalization of TCM, where the modulation is not integrated with the codeword construction. This enables the use of TCBC for both continuous (e.g., the AWGN) as well discrete (e.g., the binary symmetric) channels. Also, as the modulation scheme is not coupled with the code design, the system designer gets more flexibility in choosing communication sub-system parameters such as signal constellation, constellation shaping, etc.

3.3.1 Encoder and Decoder Structure

We now describe the encoding and decoding operations in the proposed TCBC. For simplicity of exposition, we first illustrate the central idea through an example construction, and then present a general procedure for code construction. Figure 3.2 shows an example of a TCB encoder that uses the Hamming $(7, 4)$ code as the parent LBC. The encoder comprises a trellis code, which selects the sub-set indices $(\mathbf{v}_0, \mathbf{v}_1)$, a look-up-table (LUT) which selects the data word based on sub-set indices and the remaining input data bits $(\mathbf{v}_2, \mathbf{v}_3)$, followed by the LBC

[7] When one out of l input bits is nonzero, it takes $\frac{K-1}{l}$ transitions before the trellis path merges with the all zero path. Thus, the minimum distance in this path is given by $d_{\min}\frac{(K-1)}{l}$. On the other hand, the minimum distance in the parallel transitions in the trellis equals the uniform distance, d_u, by construction.

Table 3.1: LUT mapping for TCB (7,3) code.

$\mathbf{v_0v_1v_2v_3}$	$\mathbf{u_0u_1u_2u_3}$	$\mathbf{v_0v_1v_2v_3}$	$\mathbf{u_0u_1u_2u_3}$
0000	0000	1000	1100
0001	0001	1001	1101
0010	0110	1010	1010
0011	0111	1011	1011
0100	0100	1100	1000
0101	0101	1101	1001
0110	0010	1110	1110
0111	0011	1111	1111

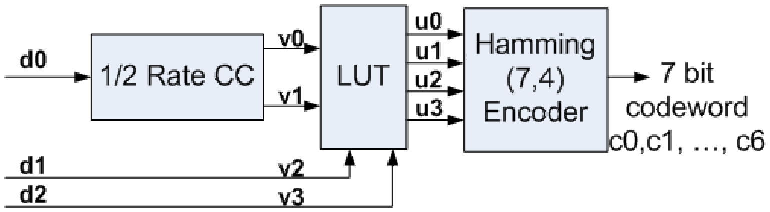

Figure 3.2: An example of a TCBC encoder.

encoder. The entries of the LUT enforce the sub-set partition structure. For example, for a 4 sub-set partitioning of the Hamming $(7,4)$ code as $\mathbf{C}_0 = \{0,1,6,7\}$, $\mathbf{C}_1 = \{10,11,12,13\}$, $\mathbf{C}_2 = \{2,3,4,5\}$ and $\mathbf{C}_3 = \{8,9,14,15\}$, if $(\mathbf{v}_0\ \mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3) = (0000)$, the LUT will select the first code-word in $\mathbf{C}_0$, if $(\mathbf{v}_0\ \mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3) = (1111)$, the LUT will select the fourth code-word in $\mathbf{C}_3$, and so on. The complete LUT for the TCBC $(7,3)$ using the Hamming $(7,4)$ code is given in Table 3.1 as an illustration. Once the data-words to be encoded are determined, the output of the LUT denoted as $(\mathbf{u}_0,\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3)$ is sent to a conventional Hamming $(7,4)$ block encoder which generates the 7-bit code-word. These coded bits can be transmitted on a binary symmetric channel (BSC) as is, or can be mapped into any digital modulation constellation symbols such as binary phase shift keying (BPSK) and transmitted on an additive white Gaussian noise (AWGN) channel.

The decoder structure is shown in Figure 3.3. A trellis decoder (e.g., Viterbi decoder, Fano

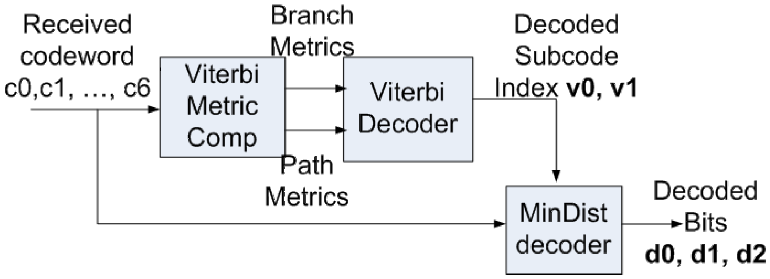
Received codeword c0,c1, ..., c6
Viterbi Metric Comp
Branch Metrics
Path Metrics
Viterbi Decoder
Decoded Subcode Index **v0, v1**
MinDist decoder
Decoded Bits **d0, d1, d2**

Figure 3.3: An example of a TCBC decoder.

decoder, etc.) is used for detecting the sub-set index. The branch metric for each transition in the trellis is chosen to be the smallest distance between the received word and the code-words in the sub-code selected by the output of the trellis assigned to that transition. Once sub-set indices are estimated (after tracing the data bits back as in the Viterbi decoder), this information is used to compute the distance (Hamming/Euclidean) between the transmitted (block) code word and the code-words in the selected sub-set/coset. The code-word with the least distance is used to estimate the remaining data bits. Thus, the decoder can be constructed using off-the-shelf Viterbi and (block) sub-set minimum distance decoders as its components. This enables an efficient hardware implementation of the decoder compared to a brute force minimum distance decoder.

Now, we describe a design procedure for constructing a TCBC with a desired $d_{\min}^{\mathrm{TCBC}}$ and rate $\frac{k}{n}$.

Design Procedure

(i) For the given $d_{\min}^{\mathrm{TCBC}}$, compute the code length n such that $d_u = d_{\min}^{\mathrm{TCBC}}$, where d_u is given by (3.2).

(ii) Choose any LBC of length n and rate $r_L > \frac{k}{n}$. Denote the minimum distance of the LBC by $d_{\min}^{\mathrm{LBC}}$.

(iii) Find a maximal linear uniform partitioning of the chosen LBC. In case non-trivial partitioning is not possible, choose a different LBC. Let the number of sub-codes be $L = 2^m$. Note that L is a power of 2 since the sub-codes are cosets of a linear sub-code.

(iv) Compute $l = m + k - n\ r_L$. Choose a rate $\frac{l}{m}$ CC with constraint length K satisfying

$\frac{K-1}{l} d_{\min}^{\text{LBC}} > d_u$.

(v) Send l out of the k input bits into the $\frac{l}{m}$ rate CC. Select one of the 2^m sub-codes of the LBC based on the m output bits of the CC. Use the remaining $k-l$ bits to choose one of the codewords from the selected sub-code. That $k-l>0$ can be seen from the following argument. Since $l = m+k-nr_L$, we have $n\, r_L - m = k-l$. The fact that $r_L > \frac{m}{n}$ implies that $k-l>0$. Thus, the n-bit TCBC codeword is completely determined for every k input bits.

The n bit codeword sequence is now transmitted over the channel after suitable constellation mapping and modulation. At the receiver, the process is reversed: the output of the channel is used to compute the metrics of a Viterbi decoder, which detects the l input bits of the CC. The remaining $k-l$ bits are detected using a minimum distance decoder for the coset chosen by the output of the Viterbi decoder. The minimum distance decoder does not impose a significant computational burden on the receiver since it computes distances only to codewords within the sub-code. Moreover, the following Lemma asserts that the proposed decoder is a maximum likelihood sequence detector (MLSD).

Lemma 5. *The proposed TCB decoder is an MLSD decoder for both binary symmetric and AWGN channels.*

Proof: See Appendix C.5.

3.3.2 Rate and Coding Gain

Recall that the parent LBC used to generate the TCB code has a rate $r_L = \frac{k-l+m}{n}$. The TCBC itself has a rate $r_T = \frac{k}{n}$. Hence, the rate loss associated with TCBC is $\frac{m-l}{n}$. By choosing $m = l+1$, one can restrict the loss in TCB encoder to be $\frac{1}{n}$. In many practical codes, the improvement due to increased minimum distance compensates for this small loss $(\frac{1}{n})$ in the coding rate, and hence the proposed TCBC offers an overall coding gain improvement compared to the parent LBC. Here, the coding gain of a code is defined as the product of its minimum distance and rate. The goal of the TCBC construction prescribed above is to ensure that the gain in the minimum distance due to the uniform distance partitioning offsets the small rate loss incurred in the encoding process. This is analogous to the gain obtained in conventional TCM

Table 3.2: Ratio of coding gains.

$(n, k, d_{\min})$	d_u	$\frac{d_u}{d_{\min}}$	G_T/G_L for $(m-l)$=	
			$\lfloor(k-1)/2\rfloor$	1
$(7, 4, 3)$	4	4/3	12/12	1
$(15, 11, 3)$	8	8/3	48/33	80/33
$(23, 12, 7)$	12	12/7	84/84	132/84

via first doubling the constellation size followed by set partitioning to increase the minimum distance.

Discussion: Suppose the LBC $\mathcal{C}$ with minimum distance $d_{\min}$ can be partitioned into the union of uniform cosets with d_u given by (3.2). Then, for the rate $\frac{k}{n}$ TCBC designed using the procedure mentioned above with $\mathcal{C}$ as the parent LBC, if $d_{\min} < d_u \leq \frac{K-1}{l} d_{\min}$, we have $G_T > G_L$, i.e., the coding gain of TCBC, denoted G_T, is higher than that $\mathcal{C}$, denoted G_L. To illustrate this, consider a modified notation where we use (n, k, d) LBC to design $(n, k-(m-l))$ TCB code. Table 3.2 shows bounds on the ratio of coding gains G_T/G_L for three choices of the parent $(n, k, d_{\min})$ code. It can be seen that for $(m-l) \leq \lfloor\frac{k-1}{2}\rfloor$, the G_T/G_L is greater than or equal to 1. i.e., there is a coding gain improvement for the TCBC especially when $d_u/d_{\min}$ is close to 2. Here, we have used the fact that in the designed $(n, k-(m-l))$ TCB code, the number of bits $(m-l)$ is at most $(k-2)$. It also can be noticed that $d_u/d_{\min}$ is closer to 2 which compensates for small rate associated with various values of $(m-l) \geq 1$.

3.3.3 Latency and Complexity

The TCBC encoder builds its codewords by concatenating short length codewords. Moreover, the short length codes are connected via a trellis to enable MLSD at the receiver. Thus, the length of a TCB code-word is an integer multiple of the length of the parent code used in the encoder, and can be chosen by the system designer depending on the latency constraints. Also, from the TCBC decoder structure, it can be observed that one need not wait till the entire TCB code-word is received, in order to decode the message bits. In fact, the maximum delay incurred in the receiver is the *trace-back length* used by the Viterbi decoder, which is typically set at about 10 times the constraint length of the CC. Thus, independent of the length of the TCBC, the decoder can start decoding the message bits after receiving a few parent code-words, which will be much smaller than the actual length of the code.

From the proposed encoder and decoder structures for TCBC, it can be seen that the complexity of encoding and decoding is significantly smaller than traditional LBC encoders and decoders with the same rate and overall length. If one has to encode and decode large length block codes without any hierarchical structure, it would involve large generator matrices and look-up tables. The optimal ML decoder for such a code would clearly have formidable computational requirements. Even for codes with structure, the complexity of encoding and decoding in extended Galois-field arithmetic and root-finding algorithms is non-trivial. Hence, constructing large length block codes with structure and good distance property is not an easy task. In this context, TCBCs offer the advantage that they can be used to systematically build good, large length block codes using existing LBCs, and their minimum distance can be easily characterized.

The TCBC encoder has an additional CC code (compared to an LBC) whose complexity is negligible when compared to the complexity of an LBC encoder of the same overall length. Moreover, the LBC encoder needs to work only at the sub-set level, which is a significant computational advantage. The TCBC decoder also implements a Viterbi decoder whose trellis complexity is 2^{K+l}, where K is the constraint length of the CC and l is the number of input bits given to the rate $\frac{l}{m}$ CC encoder. Although the bit metric computation for each of the transitions in the trellis of the CC requires 2^{k-l} distance computations, total distance computation is only 2^k per trellis stage. After the Viterbi decoding, a minimum distance decoder is used for decoding the remaining $k-l$ bits at the sub-set level, which requires only 2^{k-l} distance computations. Note that the minimum distance decoder can reuse the information available in the branch metrics assigned to the trellis, thus further saving on the computational cost. Thus, the TCBC decoder performs MLSD decoding of large length block codes at very low complexity, without constructing the trellis structure of the concatenated parent LBC.

3.3.4 Further Properties and Performance Bounds

We now make the following observations regarding the proposed TCB code.

(i) *TCBC as a generalized TCM code:* The TCBC can be considered as a generalized version of the classical TCM code, since it can separate the coding and modulation, as opposed to TCM. Some key similarities and differences between the TCB and TCM codes are listed

Table 3.3: Comparison of TCB and TCM codes.

Property	**TCM Code**	**TCB Code**
Coding and Modulation	Combined	Separate
Rate	$\frac{l}{m}$ & constellation	$\frac{l}{m}$ & partitioning
$d_{\min}$	$f(d_{\text{free}}^{CC}, d_{\min}^{\text{CONST}})$	$f(d_{\text{free}}^{CC}, K, d_u, l)$
Parallel transition bits	uncoded	coded
Channel supported	Continuous	Continuous or discrete
Decoding	MLSD	MLSD and min. distance decoder

in Table 3.3. In Table 3.3, $d_{\min}^{\text{CONST}}$ represents the minimum distance of the constellation used in TCM. Another difference between TCBC and TCM is that the parallel transition bits in TCBC are coded using codewords within the sub-block code.

(ii) From Theorem 2, the maximum uniform partitioning results in partitions with large cardinality for the given d_u, leading to small m in the rate $\frac{l}{m}$ CC used to build the code. Codes with small m and high rate are readily available from off-the-shelf designs [47], which simplifies the TCB code construction.

(iii) The uniform distance helps in bounding the Bit Error Rate (BER) of the code, since all error events for parallel transitions have the same effect on the overall BER. For example, using the union bound and the theory of Multi-Level Coding (MLC) [42], it can be shown that

$$P_b^{\text{BSC}} \approx \frac{1}{k}\left[2^{d_{\text{free}}^{CC}}\left[p(1-p)\right]^{\frac{d_{\text{free}}^{CC}}{2}} + (|\mathbf{C}_0|-1)\binom{n}{\tilde{d}}p^{\tilde{d}}(1-p)^{(n-\tilde{d})}\right], \qquad (3.5)$$

$$P_b^{\text{AWGN}} \approx \frac{1}{k}\left[B_{d_{\text{free}}^{CC}}\, Q\left(\sqrt{\frac{2d_{\text{free}}^{CC}R_1\mathcal{E}_b}{\sigma_n^2}}\right) + (|\mathbf{C}_0|-1)\, Q\left(\sqrt{\frac{2\tilde{d}R_2\mathcal{E}_b}{\sigma_n^2}}\right)\right], \qquad (3.6)$$

Table 3.4: Parameters for the SER performance study.

$d_{\text{free}}^{\text{TCBC}}$	$d_H(\mathbf{C}_i)$	*CC Rate*	*CC Poly.*	K
4	4	1/2	$[3; 7]$	3
6	7	2/3	$[7\ 4\ 1;\ 1\ 2\ 3]$	3
7	7	2/3	$[13\ 6\ 13;\ 6\ 13\ 17]$	4

where $Q(x)$ is the standard Gaussian tail function, $B_{d_{\text{free}}^{CC}}$ is the number of neighbors of the trellis code at distance d_{free}^{CC}, $\tilde{d} = \min_i d_{\min}(\mathbf{C}_i)$ and $\mathbf{C}_0$ corresponds to the coset with smallest uniform distance $\tilde{d}$.

3.4 Simulation Results

In this section, Monte Carlo simulation results are presented to illustrate the performance gains due to the proposed TCB codes.

3.4.1 Binary Symmetric Channel

To demonstrate the coding gain in a BSC, Monte Carlo simulations are performed with TCBC-based on the Hamming $(7,4)$ code as the underlying LBC. The encoder and decoder structures shown in Figures 3.2 and 3.3 are used to obtain a TCB $(7,3)$ code for various values of $d_{\text{free}}^{\text{TCBC}}$. The various LBC partitioning and CC codes used for the simulations are listed in Table 3.4. Note that the rate of all three codes is the same $\left(\frac{3}{7}\right)$, but higher coding gain is obtained by the codes that have higher trellis complexity. The SER curves are plotted in Fig. 3.4. The SER in these plots are computed using 10^8 data bits with 10^3 bits in each symbol.

The SER performance of the TCB $(7,3)$ with $d_{\text{free}}^{\text{TCBC}} = 4$ is similar to that of the Hamming $(7,4)$ code since a $d_{\text{free}}^{\text{TCBC}}$ of 3 or 4 will have roughly the same SER. It can be observed that the slope of the SER curve matches that of the theoretical curve (of a hypothetical[8] code of the same length and chosen minimum distance) more closely at low channel cross over probability (equivalently high SNR). A higher $d_H(\mathbf{C}_i)$ improves the performance, as expected. For higher

[8] Note that such code does not exist. i.e., no code exists with $n = 7$ and $d_{\min} = 5$. We assume such a code exists and compute its SER as per theoretical formula to show that TCBC with $d_{\min} = 5$ achieves similar performance (slope of the SER curve).

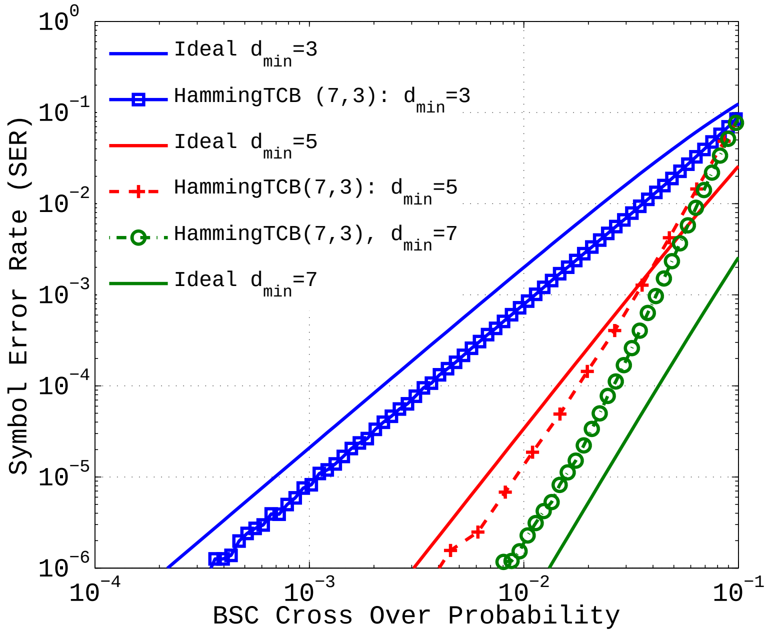

Figure 3.4: SER of TCBC $(7, 3)$ using the Hamming $(7, 4)$ code in BSC.

channel cross-over probability (equivalently low SNR), the SER is worse than the theoretical number due to error propagation in the Multi-Stage Decoding (MSD). Finally, the several orders of magnitude improvement in SER relative to the parent LBC obtained using the proposed TCBC is clear from the graph.

3.4.2 AWGN Channel

The performance gain from TCBC in the AWGN channel is now demonstrated through Monte Carlo simulations. The TCB and CC code-words are mapped suitably to BPSK or QPSK symbols, and sent over the AWGN channel. The performance of the proposed codes are compared with a CC code of the same rate and trellis complexity. For benchmarking, the BER performance is also compared with Turbo codes of the same length and rate for two different number of iterations in the Turbo decoder.

3.4.2.1 TCBC-FEC code and Comparison with CC Codes

The example TCBC-FEC construction given in the Appendix D.1 is numerically evaluated here. The output of the TCBC is encoded using BPSK symbols and transmitted over AWGN

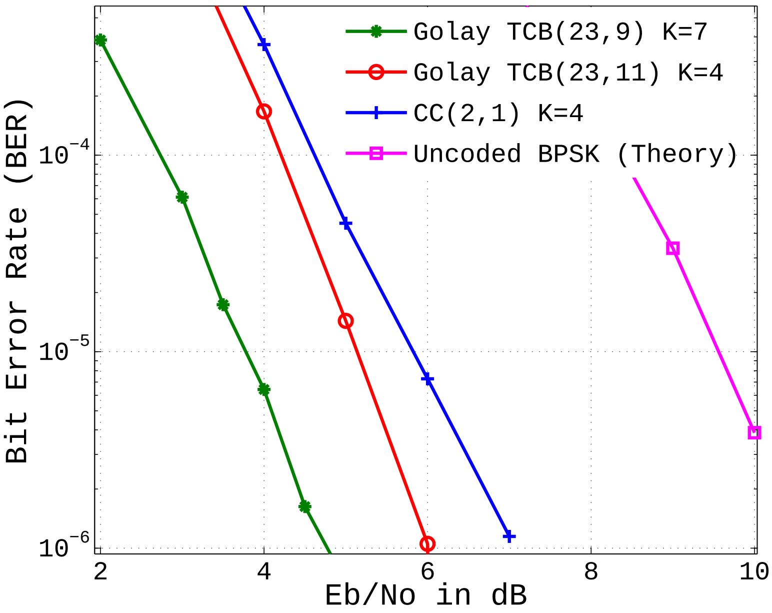

Figure 3.5: Comparison of CC and TCBC for $R = 0.5, K = 4$ with example TCBC-FEC construction.

channel. The received code sequences are decoded using the TCBC decoder. It can be noticed in Figure 3.5 that the code achieves a coding gain for 6 dB for a block length of 1000 uncoded bits. To highlight the performance of this code this plot is overlaid with BER performance of other codes described next.

Figure 3.5 also compares the BER of a rate 0.47 TCBC against a rate $\frac{1}{2}$ CC with $K = 4$. The rate $\frac{1}{2}$ CC is generated using the polynomials [17 13]. The Golay-TCB $(23q, 11q, 12)$ code is generated using the rate $\frac{8}{9}$ CC with $K = 4$ given in [47]. It can be seen that the proposed code outperforms the CC by 0.5 dB at a BER of 10^{-5} for a block length of 1000 bits. Moreover, this code and CC achieve a coding gain of about 4.25 dB and 3.5 dB respectively, related to uncoded transmission for a 1000 bit block length.

3.4.2.2 Comparison with Turbo codes

For comparison purposes, a rate $\frac{1}{5}$ parallel concatenated turbo code with a block length of 175 bits is constructed using a rate $\frac{1}{3}$ CC generated by the polynomials [7 5 3] in the systematic form. This rate $\frac{1}{5}$ code is decoded using an iterative Soft-Output-Viterbi-Algorithm (SOVA) decoder

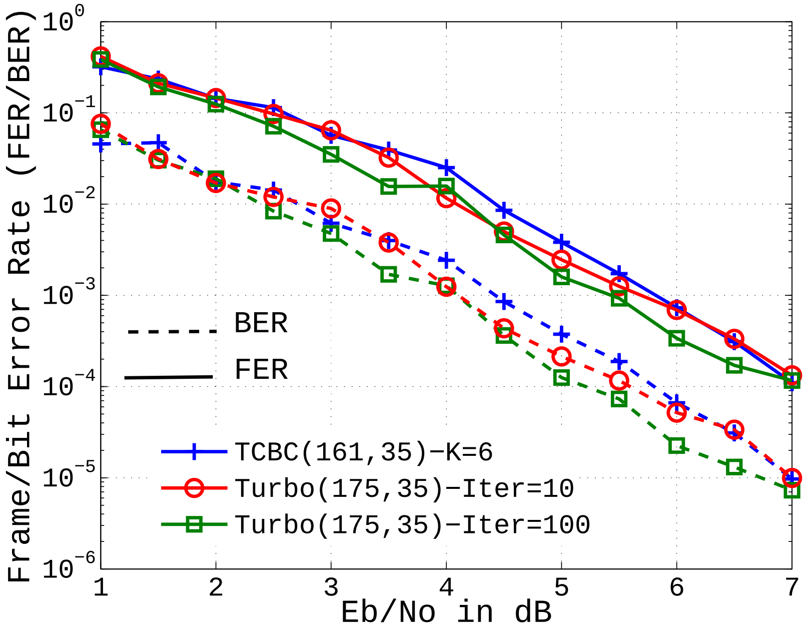

Figure 3.6: Comparison of Turbo code $(175, 35)$ and Golay-TCBC $(161, 35)$.

[48]. Note that, this turbo code can be considered as a $(175, 35)$ block code. For comparison, TCBC $(161, 35)$ is constructed using $q = 7$ consecutive symbols of a TCBC $(23q, 5q, 12)$ (whose rate is slightly higher than that of the turbo code) which is built based on the binary Golay $(23, 12, 7)$ code. The Golay-TCBC is constructed using the rate $\frac{2}{9}$ CC[9]. The turbo decoder is run with 10 and 100 iterations. The BER and SER of the two decoders are plotted in Figure 3.6. It can be observed that there is little difference in performance of the turbo code and the TCB code for 10 iterations in the turbo decoder at an SER of 10^{-3}. Also, the TCBC performs less than 0.5 dB worse than the turbo code for 100 iterations in the turbo decoder at the same SER. Note that, the decoder of TCBC is non-iterative and hence can be parallelized to decode with low latency, which is not possible in the case of iterative decoders. Thus, TCB provides an alternate coding and decoding method with low latency which can perform nearly as well as turbo codes in AWGN, when the symbol size is of the order of a few hundred bits. Figure 3.7 shows the performance of the turbo code $(1000, 200)$ for 10 iterations and Golay-TCBC $(920, 200)$. It can be observed that the performance gap is about 1 dB.

[9] CC(9,2) is defined by the polynomials $[6, 3, 7, 6, 3, 7, 6, 3, 7; 3, 10, 17, 3, 10, 17, 3, 10, 17]$.

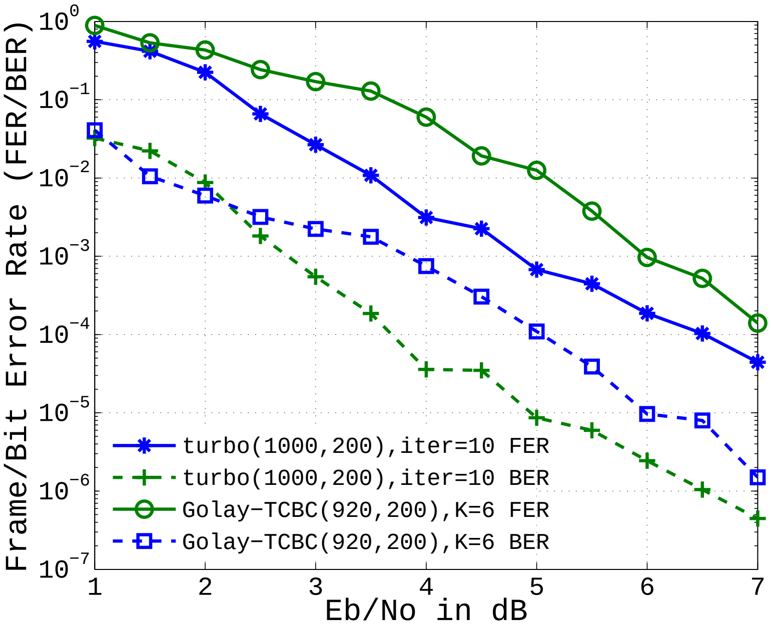

Figure 3.7: Comparison of the Turbo code $(1000, 200)$ and Golay-TCBC $(920, 200)$.

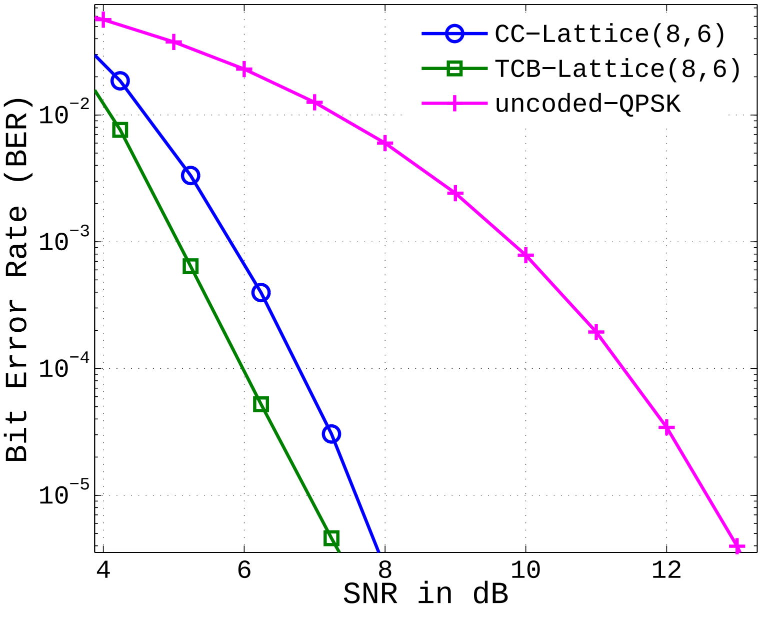

Figure 3.8: Comparison of the CC-based and TCBC lattice codes $(8, 6)$.

3.4.2.3 Comparison with CC-based lattice codes

Figure 3.8 compares the BER performance of a CC-based lattice code and a TCB based lattice code constructed in section D.3. Both codes are rate $\frac{6}{8}$ codes, and the BER of uncoded-QPSK is also plotted for comparison. The TCB-based code outperforms the CC-based lattice code, for the same trellis complexity and using the same lattice partition in $\mathbb{R}^8$. The CC-based lattice code comprises a $(4, 3)$ CC generated using polynomials [0 2 1 3; 2 1 1 3; 3 2 2 2] with $K = 6$, $d_{\text{free}}^{CC} = 4$ and TCB-based lattice code comprises a $(3, 2)$ CC generated using polynomials [3 6 7; 14 1 17] with $K = 6$, $d_{\text{free}}^{CC} = 6$. The lattice code is generated in $\mathbb{R}^8$ using QPSK constellation points. The lattice code-words are transmitted over an AWGN channel and decoded using the CC-based lattice decoder and TCB-based lattice decoder, respectively. In both decoders, the coded bits are recovered using a standard Viterbi decoder and the uncoded bits are recovered using a minimum distance decoder.

3.5 Summary

In this chapter, a new algebraic structure of binary linear block codes (LBC) was presented and a new family of codes referred to as trellis coded block codes was introduced, which can be used in discrete as well as continuous channels. The procedure developed here can be used to obtain a coding gain starting from any LBC. It was shown that, at a small loss in the rate $(\frac{1}{n})$, the BER performance can be improved relative to the parent code. Such codes could come in handy when one needs to design codes with short length and low decoding latency. This is made possible via the uniform sub-set partitioning of block codes. A simple encoder/decoder structure which uses an off-the-shelf block encoder and a Viterbi decoder was proposed. It was shown that the proposed decoder is a maximum likelihood sequence detector. Analytical expressions for bounds on the BER performance of the TCB code were obtained based on the theory of multi-stage decoding. It was also shown that using the proposed decoder structure, the latency of decoding can be made significantly smaller than the length of the block code. Moreover, the non-iterative decoder for the TCBC enables parallel hardware implementation. Three different applications of the proposed TCBCs were described in Appendix D, to illustrate their utility in practical systems. Finally, Monte Carlo simulation results demonstrated the performance gains in BSC and AWGN channels. In Chapter 5, we use the proposed TCBC for encoding the quantized

CSI for feedback to the transmitter in a multiple antenna communication system with a noisy feedback link, and illustrate the positive impact of TCBC on the forward-link average data rate.

Chapter 4

Transmit Diversity Techniques with CSIT and no CSIR

"As rivers flow into the sea and in doing so lose name and form, even so the wise man, freed from name and form, attains the Supreme Being, the Self-luminous, the Infinite." -**Mundaka Upanishad**.

4.1 Introduction

One of the seminal results in fading Multiple-Input Multiple-Output (MIMO) communication with Channel State Information (CSI) at the receiver (CSIR) is the exact characterization of the trade-off between the diversity and multiplexing gain (the so-called Diversity Multiplexing gain Trade-off (DMT)) [49]. A key finding in this work is that, for Rayleigh fading MIMO channels with perfect CSIR, the maximum diversity order can at most be $N_r N_t$, where N_r and N_t denote the number of antennas at the receiver and transmitter, respectively. In turn, this has the important implication that, as a function of the SNR expressed in dB, the logarithm of the probability of error in communication can at best decrease linearly with a slope of $N_r N_t$, even with constant-rate transmission, as the SNR goes to infinity. Since that early result, the DMT has been extended to various cases with full/partial knowledge of CSI at the transmitter (CSIT) and CSIR.

However, to the best of our knowledge, diversity transmission schemes, and the corresponding achievable DMT, of a fading MIMO channel with CSI available only at the transmitter and no CSIR has not been studied in the literature. This is perhaps because the acquisition of CSIT has typically been viewed as a two-stage process: CSI is first acquired at the receiver, and then fed back from the receiver to the transmitter in a quantized or analog fashion. Thus, the existing studies inherently assume an initial estimation of CSI at the receiver. However, when the channel is *reciprocal*, i.e., when the forward and reverse channels are the same, CSI can be directly acquired at the transmitter by sending a known training sequence in the reverse-link direction. The channel can be modeled as being reciprocal, for example, in Time Division Duplex (TDD) communication systems [50–54].[1] It is hence possible to acquire CSIT without first acquiring CSIR. In this context, some important questions that we seek to answer in this work are: If perfect CSI is available at only the transmitter, what is the best diversity order that can be achieved? How does it compare with the diversity order that can be obtained when perfect CSI is available only at the receiver? We answer these questions by proposing novel transmission schemes based on CSIT and no CSIR. The schemes are fundamentally different from straightforward techniques such as zero-forcing transmission in that zero-forcing does not satisfy an average power constraint at the transmitter under Rayleigh fading [55]. Further, since our techniques require CSI only at the transmitter, there is no need for forward-link training and channel estimation at the receiver. We show that our proposed schemes can convert a Rayleigh fading MIMO channel into fixed-gain parallel AWGN channels, while simultaneously satisfying an average power constraint at the transmitter. The proposed schemes thus equalize the fading channel, and achieve an infinite diversity order. Also, and perhaps more significantly, we show that our proposed precoding schemes extend elegantly to the fading multi-user multiple access, broadcast and interference channels, thereby achieving an infinite diversity order in these cases also.

4.1.1 Motivation and Prior work

The importance of spatial diversity for reliable communication in a wireless communication system with multiple antennas is now very well understood [56, 64]. As already mentioned,

[1] Note that channel reciprocity also requires that the transmit and receive radio-frequency (RF) chains are well-calibrated, which is assumed here [3].

Table 4.1: Summary of maximum diversity order achievable in Rayleigh fading Single-Input Multiple-Output (SIMO), Multiple-Input Single Output (MISO), and MIMO channels.

Scenario	CSI Condition	Maximum Diversity Order	References
SIMO	CSIR	N_r	[56]
SIMO	CSIRT̂	$2N_t, \infty$	[57, 58]
MISO	CSIR	N_t	[59, 60]
MISO	CSIR̂T̂	$N_t^2(N_t^2 + N_t + 1) + N_t$	[61]
MIMO	CSIR	$N_t N_r$	[49]
MIMO	CSIR̂	$N_t N_r \left[\frac{rT_c}{T_c - L_{tr}}\right]$	[62]
MIMO	CSIRT̂	$N_r N_t (N_r N_t + 2)$	[63]
MIMO	CSIR̂T̂	$2N_r N_t$	[63]
MIMO	CSIT	∞	This work

a diversity order of $N_t N_r$ can be achieved with perfect CSIR [49]. On the other hand, an exponential diversity order can be achieved when perfect CSI is available at both transmitter and receiver [65]. It is also known that under partial CSIT and perfect CSIR, a diversity order greater than $N_t N_r$ can be achieved (See [57], [66], and [7-18] in [63]). In the above papers, the available CSIT is exploited either for inverting dominant modes, or for power control, to improve the diversity order, under the assumption of perfect CSIR. In [55, 67–69], channel-inversion based power control and precoding was considered. However, channel inversion fails to satisfy the average power constraint [55], requiring the use of regularization [67] or computationally intensive sphere encoding schemes [68, 69]. Table 4.1 lists the key results from the literature on the maximum diversity order that is achievable under various assumptions on the availability of CSI at the receiver and transmitter. In the second column of the table, the CSI condition is represented as CSIR (or CSIT), when the CSI at the receiver (or transmitter) is perfect, and no CSI is available at the transmitter (or receiver). Also, a hat over a letter, e.g., CSIR̂T̂, represents the system with estimated CSI at the receiver and transmitter, where the estimation error variance is assumed to be inversely proportional to the training SNR. Thus, to the best of our knowledge, none of the existing studies analyze the diversity order that is achievable when

CSI is available only at the transmitter, which is our focus in this Chapter.

4.1.2 Contributions

- We propose three novel and simple-to-implement transmit precoding schemes which require CSI only at the transmitter. Our first proposed technique uses a real Orthogonal Space-Time Block Code (O-STBC) based signaling with power control, and can be used with 3 or more transmit antennas. With 2 transmit antennas, the same scheme is also applicable with the complex Alamouti code based signaling and an appropriate power control. The second technique we propose is a modification of the existing Maximum Ratio Transmission (MRT) scheme [70]. The third technique we propose uses a more general complex signaling scheme, thereby recovering the spectral efficiency lost due to the real-valued signaling, but is applicable when $N_t \geq 2N_r$. We show that our proposed transmit precoding schemes achieve an infinite diversity order. Added benefits of our proposed approach are that forward-link training is not required, and optimal decoding at the receiver is very simple.
- We extend the first transmit precoding scheme to the Rayleigh fading Multiple Access Channel (MAC). We show that an infinite diversity order is achieved in this case as well.
- We extend the second transmit precoding scheme to three kinds of multi-user Rayleigh fading channels: the MAC, the Broadcast Channel (BC), and the Interference Channel (IC). We show that an infinite diversity order is achieved in all three cases.
- We illustrate the performance of the proposed schemes via Monte Carlo simulations. We show the AWGN-like waterfall behavior of the probability of error versus SNR curves in the single-user and multi-user cases. We also present simulation results with imperfect CSIT obtained using reverse-link training, as well as with practical peak-to-average power constraints, and show that, for practical SNRs, the waterfall behavior is still retained.

The three precoding schemes we propose are quite different from each other. The first is based on O-STBC signaling, second is based on the MRT scheme and the third is based on the use of the QR-decomposition of the channel to achieve a form of active interference cancellation at the transmitter. The real O-STBC based scheme is simpler to implement than the QR-based scheme, but the latter allows the use of complex-valued signaling, and extends elegantly to the multi-user MAC, BC and IC. Since our proposed precoding scheme converts the Rayleigh fading

MIMO MAC, BC and IC with CSIT and no CSIR into fixed-gain AWGN channels, one can use existing results for the Gaussian channel [71] to immediately obtain achievable rate regions for all three cases.

4.2 Transmit Precoding Based on Real O-STBC Signaling

4.2.1 Equivalent Channel Model

For our first proposed scheme, we consider real O-STBC signaling. At the receiver, we consider the real part of the baseband received signal, and, hence, we can consider both the baseband equivalent channel as well as the additive noise as having real-valued components. Before presenting the proposed transmit precoding scheme, we start with the following mathematical model for the received signal at the i^{th} receive antenna:

$$\mathbf{y}_i = \sqrt{\frac{k\rho}{N_t}}\mathbf{X}\mathbf{h}_i + \mathbf{n}_i, \tag{4.1}$$

where $\mathbf{y}_i \in \mathbb{R}^L$ denotes the received signal vector for $L \geq N_t$ consecutive symbols. The channel vector between the transmit antennas and the i^{th} receive antenna is denoted by $\mathbf{h}_i \in \mathbb{R}^{N_t}$, and is assumed to have Gaussian, independent and identically distributed (i.i.d.) entries with zero mean and unit variance, denoted by $\mathcal{N}(0,1)$. The real O-STBC codeword is denoted by $\mathbf{X} \in \mathbb{R}^{L\times N_t}$. The noise vector is denoted by $\mathbf{n}_i \in \mathbb{R}^L$, and is assumed to have i.i.d. $\mathcal{N}(0, \sigma_n^2)$ entries. Also, ρ is the total transmit power available across the N_t antennas per channel use, and k is a constant used to meet the average transmit power constraint. Using the equivalent representation of the codeword matrix $\mathbf{X}$ in terms of its constituent Hurwitz-Radon matrices [60], it is shown in [72] that (4.1) can be written as

$$\mathbf{y}_i = \sqrt{\frac{k\rho}{N_t}}\tilde{\mathbf{H}}_i\mathbf{x} + \mathbf{n}_i, \tag{4.2}$$

where $\tilde{\mathbf{H}}_i \in \mathbb{R}^{L\times L}$ denotes the equivalent channel matrix and the vector $\mathbf{x} \in \mathbb{R}^L$ contains the symbols used to construct $\mathbf{X}$. Note that, the matrix $\tilde{\mathbf{H}}_i$ is obtained from $\mathbf{h}_i$ using a simple mapping $\pi : \mathbb{R}^{N_t} \to \mathbb{R}^{L\times L}$ [72].

Examples: Consider the 4 × 4 real O-STBC code designed in [60]. In this case, it can be shown that

$$\mathbf{X}^T = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 \\ s_2 & s_1 & s_4 & -s_3 \\ s_3 & -s_4 & s_1 & s_2 \\ s_4 & s_3 & -s_2 & s_1 \end{bmatrix}, \quad \tilde{\mathbf{H}} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \\ h_2 & -h_1 & h_4 & -h_3 \\ h_3 & -h_4 & -h_1 & h_2 \\ h_4 & h_3 & -h_2 & -h_1 \end{bmatrix}, \tag{4.3}$$

where s_j denotes the j^{th} data symbol drawn from a finite size constellation set, $\mathbf{x} = [s_1\ s_2\ s_3\ s_4]^T$, and $h_j = h_{ji}, j = 1, 2, \ldots, N_t$ are the channel coefficients between the N_t transmit antennas and i^{th} receive antenna. Note that, for simplicity, we have omitted the receive antenna index i in writing the expression for $\tilde{\mathbf{H}}$.

As an example with a non-square O-STBC, for the $\mathcal{G}_3$ code in [60], the codeword and the equivalent channel matrix can be written, respectively, as

$$\mathbf{X}^T = \begin{bmatrix} s_1 & -s_2 & -s_3 & -s_4 \\ s_2 & s_1 & s_4 & -s_3 \\ s_3 & -s_4 & s_1 & s_2 \end{bmatrix}, \tilde{\mathbf{H}} = \begin{bmatrix} h_1 & h_2 & h_3 & 0 \\ h_2 & -h_1 & 0 & -h_3 \\ h_3 & 0 & -h_1 & h_2 \\ 0 & h_3 & -h_2 & -h_1 \end{bmatrix}. \tag{4.4}$$

It can be verified that, in both cases, the equivalent channel matrices $\tilde{\mathbf{H}}$ are orthogonal. In fact, this property is true for all real O-STBCs, as we show next.

By the equivalence of the two representations, we have $\mathbf{Xh} = \tilde{\mathbf{H}}\mathbf{x}$. Multiplying by $\mathbf{X}^T$ on both sides, we get

$$\beta\mathbf{h} = \mathbf{X}^T\tilde{\mathbf{H}}\mathbf{x}, \tag{4.5}$$

where $\mathbf{X}^T\mathbf{X} = \beta\mathbf{I}_{N_t}$, and $\beta = \sum_{i=1}^{N_t} s_i^2 > 0$, since $\mathbf{X}$ is a real O-STBC codeword. Also, $\mathbf{I}_{N_t}$ represents the $N_t \times N_t$ identity matrix. Now, suppose $\mathbf{h} \neq \mathbf{0}$, but the columns of $\tilde{\mathbf{H}}$ are linearly dependent. Then, there exists a nonzero $\mathbf{x}$ that lies in the null space of $\tilde{\mathbf{H}}$, and substituting such an $\mathbf{x}$ in the above leads to $\mathbf{h} = \mathbf{0}$, i.e., a contradiction. Hence, any nonzero channel vector $\mathbf{h}$ leads to an $\tilde{\mathbf{H}}$ with full column rank. Next, we show that $\tilde{\mathbf{H}}$ is orthogonal.

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ denote two data vectors and $\mathbf{X}_1$ and $\mathbf{X}_2$ denote their corresponding O-STBC matrices. Further, let $\mathbf{x}_{k,j}$ denote the j^{th} column of $\mathbf{X}_k$, for $k = 1, 2$. Due to the structure of

O-STBC codes, $\mathbf{x}_{2,j}^T\mathbf{x}_{1,i} = -\mathbf{x}_{2,i}^T\mathbf{x}_{1,j}$ for $j \neq i$, and $\mathbf{x}_{1,i}^T\mathbf{x}_{2,i} = \mathbf{x}_{1,j}^T\mathbf{x}_{2,j} = \mathbf{x}_1^T\mathbf{x}_2$ [73]. Hence,

$$\mathbf{h}^T\mathbf{X}_1^T\mathbf{X}_2\mathbf{h} = \sum_i\sum_j h_ih_j\mathbf{x}_{1,i}^T\mathbf{x}_{2,j} = \sum_i h_i^2\mathbf{x}_{1,i}^T\mathbf{x}_{2,i} = \mathbf{x}_1^T\mathbf{x}_2\sum_i h_i^2. \tag{4.6}$$

Using the fact that $\mathbf{Xh} = \tilde{\mathbf{H}}\mathbf{x}$, we get

$$\mathbf{x}_1^T\left(\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\right)\mathbf{x}_2 = \mathbf{x}_1^T\mathbf{x}_2\sum_i h_i^2. \tag{4.7}$$

The above equation holds for any pair of vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, if and only if $\tilde{\mathbf{H}}$ is orthogonal and $\tilde{\mathbf{H}}^T\tilde{\mathbf{H}} = \tilde{\mathbf{H}}\tilde{\mathbf{H}}^T = (\sum_{i=1}^{N_t} h_i^2)\mathbf{I}_L$. Thus, the equivalent channel matrix $\tilde{\mathbf{H}}$ is an orthogonal matrix.

Remark 4. The equivalent channel representation in (4.2) and the orthogonality property in (4.7) also hold for the complex 2×2 Alamouti code (see Exercise 9.4 of [74]). However, it does not necessarily hold for other complex O-STBCs. Due to this, it is not possible to directly extend the real O-STBC based transmit precoding scheme proposed in the next subsection to complex O-STBC signaling, except when $N_t = 2$.

4.2.2 Proposed Transmit Precoding Scheme

4.2.2.1 Single Receive Antenna Case

For the ease of the presentation, we first consider the single receive antenna case. We premultiply the data vector $\mathbf{x}$ with the matrix $\mathbf{P} \triangleq \frac{1}{\alpha}\tilde{\mathbf{H}}^T$ where $\alpha = h_1^2 + h_2^2 + \ldots + h_{N_t}^2$ is a scalar. Then, we use the vector $\mathbf{Px}$ to generate the real O-STBC codeword $\mathbf{X}$. Since the channel matrix $\tilde{\mathbf{H}}$ is orthogonal, such a precoding equalizes the effective channel, i.e.,

$$\mathbf{y} = \sqrt{\frac{k\rho}{N_t}}\tilde{\mathbf{H}}\mathbf{Px} + \mathbf{n} = \sqrt{\frac{k\rho}{N_t}}\mathbf{x} + \mathbf{n}. \tag{4.8}$$

In the above, the constant k is used to satisfy the transmit power constraint; we derive its value below. Note that, with the aforementioned precoding, optimal data decoding at the receiver is very simple,[2] as the equivalent channel consists of L parallel Single-Input Single-Output (SISO) AWGN channels with their gain independent of the channel instantiation. Since the effect of fading has been perfectly equalized at the transmitter, the proposed scheme achieves an infinite

[2] The data decoding is even simpler than that of O-STBC with perfect CSIR.

diversity order. Moreover, as the equivalent channel has a fixed gain, channel estimation is not required at the receiver.

Satisfying the Average Transmit Power Constraint

In the proposed scheme, the columns of actual O-STBC matrix transmitted, $\mathbf{X}$, are constructed using permutations and sign-inversions of the entries of the precoded vector $\mathbf{Px}$. Hence, the average transmit power over L channel uses, which is given by $\mathrm{tr}(\mathbf{X}^T\mathbf{X})$, can be written as

$$P_{\text{avg}} = \frac{k\rho L}{N_t}\left(\mathbb{E}_{\mathbf{h},\mathbf{x}}\left[\mathbf{x}^T\mathbf{P}^T\mathbf{P}\mathbf{x}\right]\right), \tag{4.9}$$

where $\mathbb{E}_{\mathbf{h},\mathbf{x}}$ refers to the expectation over the distributions of $\mathbf{h}$ and $\mathbf{x}$. This average power constraint is the same as in past work that considers transmission schemes with CSIT, e.g., [57, 63, 65, 66]. Since ρ is the total power available for transmission per channel use, to satisfy the average transmit power constraint of $P_{\text{avg}} = L\rho$, we need

$$\rho L \quad = \quad \frac{k\rho L}{N_t}\mathbb{E}_{\mathbf{h}}\left[\frac{1}{\alpha}\right]\mathbb{E}_{\mathbf{x}}[\|\mathbf{x}\|_2^2], \tag{4.10}$$

where the orthogonality property of $\tilde{\mathbf{H}}$ is used. Now, α is a $\chi^2_{N_t}$ random variable when the channel is Rayleigh fading with i.i.d. $\mathcal{N}(0,1)$ entries, and it is shown in Appendix E.1 that

$$\mathbb{E}\left[\frac{1}{\alpha}\right] = \frac{1}{N_t - 2}, \text{ for } N_t > 2. \tag{4.11}$$

Hence, assuming that each entry of $\mathbf{x}$ is normalized to have unit energy, $\mathbb{E}_{\mathbf{x}}[\|\mathbf{x}\|^2] = L$, and the average transmit power constraint can be satisfied by choosing

$$k = \frac{N_t(N_t-2)}{L}. \tag{4.12}$$

Moreover, the SNR at the receiver can be computed as $\frac{\rho(N_t-2)}{L\ \sigma_n^2}$.

4.2.2.2 Extension to Multiple Receive Antennas

First, note that the above precoding scheme converts a Rayleigh fading channel into L parallel SISO AWGN channels with a fixed gain over L channel uses. Hence, we obtain an infinite diversity order with a single receive antenna. Having additional receive antennas can improve the received SNR, but does not increase the diversity order.

One way to use multiple receive antennas with the above precoding scheme is to employ *antenna selection* at the receiver. For each channel instantiation, we select the receive antenna for which the average transmit power required is the minimum. This corresponds to choosing the antenna for which the ℓ_2 norm of the channel vector is the highest among all the receive antennas. This requires limited CSI at the receiver; i.e., the receiver would require knowledge of the antenna selected by the precoding scheme. Alternatively, the receiver could decode the data on all its receive antenna chains and pick the antenna chain for which the decoded data passes a cyclic redundancy check (CRC). Clearly, the receive antenna selection based scheme also achieves an infinite diversity order, as, on a per channel instantiation basis, the effective channel is still an AWGN channel. For example, consider the 2 receive antenna case, with $\alpha \triangleq \max\{\|\mathbf{h}_1\|_2^2, \|\mathbf{h}_2\|_2^2\}$. It is shown in Appendix E.2 that

$$\mathbb{E}\left[\frac{1}{\alpha}\right] = \frac{2^{1-N_t}}{\Gamma\left(\frac{N_t}{2}\right)} \sum_{m=0}^{\infty} \frac{\Gamma(N_t - 1 + m)}{2^m \Gamma\left(\frac{N_t}{2} + m\right)}, \tag{4.13}$$

where $\Gamma(\cdot)$ is the Gamma function [75]. Hence, k can be computed as $\frac{N_t}{\mathbb{E}[\frac{1}{\alpha}]L}$. For the case of $N_t = 4, L = 4$ with real $\mathbf{h}_i$, we get $k \approx N_t$ in the 2 receive antenna case with antenna selection, in contrast with $k = (N_t - 2)$ for the single antenna case. Thus, when $N_t = 4$, the above precoding scheme offers a nearly 3 dB improvement in the performance with 2 receive antennas and antenna selection, compared to the single receive antenna case. The extension of the antenna selection scheme to $N_r > 2$ receive antennas is straightforward. However, the expression for $\mathbb{E}\left[\frac{1}{\alpha}\right]$ is cumbersome to obtain, and hence is omitted.

Remark 5. The proposed scheme also works for the case of 2 transmit antennas, by using the *complex* Alamouti code as the underlying O-STBC. In this case, with one receive antenna, it can be shown that $\mathbb{E}\left[\frac{1}{\alpha}\right] = 1$, when the channel coefficients are i.i.d. $\mathcal{CN}(0,1)$. With $N_r = 2$ and receive antenna selection, using the derivation in Appendix E.2, one can calculate[3]

$$\mathbb{E}\left[\frac{1}{\alpha}\right] = \frac{2 \cdot 2^{1-2N_t}}{\Gamma(N_t)} \sum_{m=0}^{\infty} \frac{\Gamma(2N_t - 1 + m)}{2^m \Gamma\left(N_t + m\right)} = 0.5. \tag{4.14}$$

[3] Note that, with complex baseband signaling, we consider each entry of $\mathbf{h_i}$ is assumed to be complex circularly symmetric Gaussian distributed with a variance of 0.5 per real dimension, denoted by $\mathcal{CN}(0,1)$. Also, the entries of the noise vector are assumed to be i.i.d. and drawn from $\mathcal{CN}(0,1)$.

Hence, the average transmit power constraint can be satisfied with 2 receive antennas also; and the receive antenna selection between two antennas offers a 3 dB improvement in the average SNR compared to the single receive antenna case.

4.3 Maximum Ratio Transmission Based Precoding

Now, we present our second precoding scheme, which is based on Maximum Ratio Transmission (MRT). We start with single receive antenna and $N_t \geq 2$ transmit antennas. Let $\mathbf{h}$ denote the $N_t \times 1$ channel vector with i.i.d. $\mathcal{CN}(0,1)$ components in the complex baseband representation. In classical MRT, one uses $\frac{\mathbf{h}}{\|\mathbf{h}\|}$ as the beamforming vector at the transmitter. Here, we propose to use $\mathbf{p} \triangleq \frac{\mathbf{h}}{\|\mathbf{h}\|^2}$ to precode the unit-power data symbol x. The received signal y can be written as

$$y = \sqrt{\frac{k\rho}{N_t}}\mathbf{h}^H\mathbf{p}x + n = \sqrt{\frac{k\rho}{N_t}}x + n, \tag{4.15}$$

where $n \in \mathbb{C}$ denotes the receiver noise, distributed as $\mathcal{CN}(0,\sigma_n^2)$, and k denotes transmit power normalization. Thus, the above MRT based precoding scheme equalizes the fading channel, provided an average power constraint can be satisfied. The average transmit power can be written as

$$P_{\text{avg}} = \frac{k\rho}{N_t}\mathbb{E}[x^2]\mathbb{E}[\mathbf{p}^H\mathbf{p}] = \frac{k\rho}{N_t}\mathbb{E}\left[\frac{1}{\|\mathbf{h}\|^2}\right]. \tag{4.16}$$

This average power constraint is the same as in past work that considers transmission schemes with CSIT, e.g., [57, 63, 65, 66]. In Appendix E.1, it is shown that $\mathbb{E}\left[\frac{1}{\|\mathbf{h}\|^2}\right] = \frac{1}{N_t - 1}$ for $N_t \geq 2$. Hence, the power normalization constant $k = N_t(N_t - 1)$ satisfies the average power constraint $P_{\text{avg}} = \rho$.

$$k = \begin{cases} N_t(N_t-2), & \text{for } N_t > 2 \text{ with real } \mathbf{h} \\ N_t(N_t-1), & \text{for } N_t > 1 \text{ with complex } \mathbf{h} \end{cases}. \tag{4.17}$$

The corresponding SNR at the receiver is given by

$$\mathsf{SNR} = \begin{cases} \frac{(N_t-2)\rho}{\sigma_n^2}, & \text{for } N_t > 2 \text{ with real } \mathbf{h} \\ \frac{(N_t-1)\rho}{\sigma_n^2}, & \text{for } N_t > 1 \text{ with complex } \mathbf{h} \end{cases}, \tag{4.18}$$

where σ_n^2 is the variance of the zero mean Gaussian noise at the receiver.

4.3.1 Extension to Multiple Receive Antennas

As in the previous precoding scheme, we employ antenna selection in order to extend the scheme to multiple receive antennas. At the transmitter, we form the precoding vector $\mathbf{p}$ corresponding to the receive antenna that requires the least transmit power among the available receive antennas. Thus, the transmitter picks the receive antenna for which the resulting channel vector has the largest norm and employs the MRT based precoding scheme for the selected channel. Along the lines of (4.13) and (4.14), the normalization constant k can be computed. That is, for $N_r = 2$ and complex signalling we get $k = 2\ N_t$ which satisfies the average transmit constraint $P_{\text{avg}} = \rho$.

4.4 QR-Decomposition Based Precoding Scheme

In this section, we present another novel precoding scheme, which is based on the QR-decomposition of the channel matrix. This scheme not only applies to a wider range of antenna configurations, but also extends to multiuser scenarios with CSIT. Consider a Rayleigh fading MIMO channel with N_r receive antennas and $N_t \geq 2N_r$ transmit antennas. The complex baseband signal model for the received signal at the i^{th} receive antenna can be written as

$$y_i = \sqrt{\frac{k\rho}{N_t}}\mathbf{h}_i^H \mathbf{P}\tilde{\mathbf{x}} + n_i, \tag{4.19}$$

where $\mathbf{h}_i \in \mathbb{C}^{N_t}$ denotes the complex channel coefficients between the N_t transmit antennas and i^{th} receive antenna, $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{N_t}]^T$ denotes an *extended* data vector of dimension N_t, and is derived from a data vector $\mathbf{x} \in \mathbb{C}^{N_r}$ containing the N_r symbols to be transmitted. We assume the normalization $\mathbb{E}[\mathbf{x}^H\mathbf{x}] = 1$. Also, ρ, k and $\mathbf{P}$ denote, respectively, the average transmit power available across the N_t transmit antennas per channel use, a normalization constant, and an $N_t \times N_t$ precoding matrix. The noise is assumed to be i.i.d. across receive antennas with entries from $\mathcal{CN}(0, \sigma_n^2)$.

For ease of presentation, as in the previous section, we start with the $N_r = 1$ case. Let $\mathbf{h} \in \mathbb{C}^{N_t}$ denote the channel vector. We set the precoding matrix $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{QU}, \tag{4.20}$$

where the unitary matrix $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$ is obtained from the QR-decomposition of $\mathbf{h}$, i.e., $\mathbf{h} \triangleq \mathbf{Qr}$, with $\mathbf{r} \in \mathbb{C}^{N_t}$ and upper triangular, with first element $r_1 = \|\mathbf{h}\|_2$ and remaining elements equal to zero. Also, $\mathbf{U} \in \mathbb{C}^{N_t \times N_t}$ is chosen to be an arbitrary, non-diagonal unitary matrix.

Now, given the complex scalar data symbol x, the extended data vector $\tilde{\mathbf{x}}$ is chosen such that the following condition is satisfied:

$$\mathbf{r}^H \mathbf{U} \tilde{\mathbf{x}} = x. \tag{4.21}$$

It is easy to verify that (4.21) can be satisfied by choosing $\tilde{x}_1 = x$, $\tilde{x}_2 = \frac{x(1-r_1 u_{1,1})}{r_1 u_{1,2}}$, and $\tilde{x}_j = 0$, for $j = 3, \ldots, N_t$, where $u_{i,j}$ is the $(i,j)^{\text{th}}$ element of $\mathbf{U}$, provided $u_{1,2} \neq 0$. Substituting for $\tilde{\mathbf{x}}$ and $\mathbf{P}$, the above precoding scheme leads to the following equivalent channel:

$$y = \sqrt{\frac{k\rho}{N_t}} x + n. \tag{4.22}$$

In the above, k is a normalization constant independent of the channel instantiation, whose value is specified below. Thus, the fading channel is converted to an AWGN channel with a fixed gain. The proposed precoding scheme inherently equalizes the effect of fading and simultaneously cancels the interference caused due to the signal being transmitted from multiple antennas. Next, we show that, with k appropriately chosen, the above precoding scheme satisfies an average transmit power constraint.

Satisfying the Average Transmit Power Constraint

The average transmitted power can be written as

$$\begin{aligned} P_{\text{avg}} &= \frac{k\rho}{N_t} \mathbb{E}_{x,\mathbf{h}} \left[\tilde{\mathbf{x}}^H \mathbf{P}^H \mathbf{P} \tilde{\mathbf{x}} \right], \\ &= \frac{k\rho}{N_t} \mathbb{E}_x \left[x^2 \right] \left(1 + \frac{|u_{1,1}|^2}{|u_{1,2}|^2} + \frac{1}{|u_{1,2}|^2} \mathbb{E}_{\mathbf{h}} \left[\frac{1}{\|\mathbf{h}\|_2^2} \right] - 2 \frac{\Re\{u_{1,1}\}}{|u_{1,2}|^2} \mathbb{E}_{\mathbf{h}} \left[\frac{1}{\|\mathbf{h}\|_2} \right] \right). \end{aligned} \tag{4.23}$$

For Rayleigh fading channels, it is shown in Appendices E.1 and E.3 that

$$\begin{aligned} \mathbb{E}\left[\frac{1}{\|\mathbf{h}\|_2^2} \right] &= \frac{1}{N_t - 1}, \\ \mathbb{E}\left[\frac{1}{\|\mathbf{h}\|_2} \right] &= \frac{\Gamma\left(\frac{2N_t - 1}{2}\right)}{\Gamma\left(N_t\right)}. \end{aligned} \tag{4.24}$$

Using (4.23) and (4.24), we can satisfy the average transmit power constraint of $P_{\text{avg}} = \rho$ by choosing

$$k = N_t \left(1 + \frac{|u_{1,1}|^2}{|u_{1,2}|^2} + \frac{1}{|u_{1,2}|^2} - 2\frac{\Re\{u_{1,1}\}}{|u_{1,2}|^2}\frac{\Gamma\left(\frac{2N_t-1}{2}\right)}{\Gamma(N_t)}\right)^{-1}, \tag{4.25}$$

where $\Re\{\cdot\}$ denotes the real part of $\{\cdot\}$, and $u_{1,2} \neq 0$ so that $k > 0$ and finite. For example, when $N_t = 2$, one can choose

$$\mathbf{U} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{4.26}$$

which results in $k = 1$ and a received SNR of $\frac{\rho}{2\sigma_n^2}$, i.e., a 3 dB loss compared to the unit-gain AWGN channel. Finding the unitary matrix $\mathbf{U}$ that minimizes the SNR loss is an interesting extension for future work.

Remark 6. One way to accommodate multiple receive antennas in this scheme is to employ antenna selection, as in the O-STBC based precoding scheme. The power normalization constant k can be easily derived as follows. For example, with $N_r = 2$ and $\mathbf{U}$ chosen as in (4.26), we can write (4.23) as

$$P_{\text{avg}} \quad = \quad \frac{k\rho}{N_t}\mathbb{E}_x\left[x^2\right]\left(1 + \mathbb{E}\left[\frac{1}{\alpha}\right]\right), \tag{4.27}$$

where $\alpha = \max\{\|\mathbf{h}_1\|_2^2, \|\mathbf{h}_2\|_2^2\}$, as before. Using (4.14), we can obtain a closed-form expression for k to satisfy the average power constraint of $P_{\text{avg}} = \rho$ per channel use.

In the following, we present an alternative way to extend the above QR-based precoding scheme to the case where multiple receive antenna chains are available. The proposed scheme results in an equivalent channel that consists of N_r parallel fixed-gain AWGN channels.

Extension to Multiple Receive Antennas

When the receiver is equipped with N_r antennas, with $N_t \geq 2N_r$, our proposed extension leads to N_r parallel, non-interfering AWGN channels. The input-output relation is given by

$$\mathbf{y} = \sqrt{\frac{k\rho}{N_t}}\mathbf{H}^H\mathbf{P}\tilde{\mathbf{x}} + \mathbf{n}, \tag{4.28}$$

where the received vector $\mathbf{y} \in \mathbb{C}^{N_r}$, the channel matrix $\mathbf{H} \in \mathbb{C}^{N_t \times N_r}$, and the noise $\mathbf{n} \in \mathbb{C}^{N_r}$. Denote the QR decomposition of $\mathbf{H}$ by $\mathbf{H} = \mathbf{QR}$, where $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$ is unitary and $\mathbf{R} \in \mathbb{C}^{N_t \times N_r}$

is upper triangular. Note that, in particular, since $N_t \geq 2N_r$, the rows $N_r + 1$ through N_t of the matrix $\mathbf{R}$ are all zeros.

We consider the data vector $\mathbf{x} = [x_1,\ x_2,\ \ldots,\ x_{N_r}]^T$, and choose the extended vector $\tilde{\mathbf{x}}$ such that $\mathbf{R}^H\mathbf{U}\tilde{\mathbf{x}} = \mathbf{x}$, where $\mathbf{U} \in \mathbb{C}^{N_t \times N_t}$ is a fixed non-diagonal unitary matrix. Now, the matrix $\mathbf{R}$ can be partitioned as $\mathbf{R} = [\mathbf{R}_1^H\ \mathbf{0}^H]^H$, where the submatrices $\mathbf{R}_1$ and $\mathbf{0}$ are of dimension $N_r \times N_r$ and $(N_t - N_r) \times N_r$, respectively. We set the first N_r entries of $\tilde{\mathbf{x}}$ as $\mathbf{x}$. If we partition $\tilde{\mathbf{x}}$ as $\tilde{\mathbf{x}}^H = [\mathbf{x}^H\ \mathbf{x}'^H\ \mathbf{0}^H]$, where $\mathbf{0}$ is a vector of $(N_t - 2N_r)$ zeros, $\mathbf{x}'$ can be written as

$$\mathbf{x}' = \mathbf{R}_{u2}^{-1}\left(\mathbf{I} - \mathbf{R}_{u1}\right)\mathbf{x}, \tag{4.29}$$

where $\mathbf{R}_{u1} \triangleq \mathbf{R}_1^H\mathbf{U}_{11}$ and $\mathbf{R}_{u2} \triangleq \mathbf{R}_1^H\mathbf{U}_{12}$. The matrix $\mathbf{U}_{11}$ is the $N_r \times N_r$ principal submatrix of $\mathbf{U}$, and the matrix $\mathbf{U}_{12}$ is the $N_r \times N_r$ submatrix of $\mathbf{U}$ obtained by taking the entries from rows 1 through N_r and columns $N_r + 1$ through $2N_r$. Finally, we let $\mathbf{P} = \mathbf{QU}$, as before.

The above described precoding scheme leads to the input-output relation:

$$\mathbf{y} = \sqrt{\frac{k\rho}{N_t}}\mathbf{x} + \mathbf{n}, \tag{4.30}$$

and hence, we obtain N_r parallel, fixed-gain AWGN channels. By choosing k appropriately, we can satisfy the average power constraint on the data signal, as we show next.

Satisfying the Average Transmit Power Constraint

Noting that $\|\tilde{\mathbf{x}}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{x}'\|_2^2$, the average transmit power per channel use can be computed from

$$\begin{aligned} P_{\text{avg}} &= \frac{k\rho}{N_t}\mathbb{E}_{\mathbf{x},\mathbf{h}}\left[\mathbf{x}^H\mathbf{x} + \mathbf{x}^H(\mathbf{I} - \mathbf{R}_{u1})^H\mathbf{R}_{u2}^{-H}\mathbf{R}_{u2}^{-1}(\mathbf{I} - \mathbf{R}_{u1})\mathbf{x}\right], \\ &= \frac{k\rho}{N_t}\left[1 + \frac{1}{N_r}\text{tr}\left(\mathbb{E}_{\mathbf{h}}\left[(\mathbf{I} - \mathbf{R}_{u1})^H\mathbf{R}_{u2}^{-H}\mathbf{R}_{u2}^{-1}(\mathbf{I} - \mathbf{R}_{u1})\right]\right)\right]. \end{aligned} \tag{4.31}$$

Since the choice of the unitary matrix $\mathbf{U}$ is arbitrary, we can simply choose $\mathbf{U}_{11} = \mathbf{0}_{N_r}$ and $\mathbf{U}_{12} = \mathbf{I}_{N_r}$. Now, we get $\mathbf{R}_{u1} = \mathbf{0}$ and $\mathbf{R}_{u2} = \mathbf{R}_1^H$. Further, using Lemma 6 in [76], we have

$$\text{tr}\left(\mathbb{E}_{\mathbf{h}}\left[\mathbf{R}_{u2}^{-1}\mathbf{R}_{u2}^{-H}\right]\right) = \frac{N_r}{N_t - N_r}. \tag{4.32}$$

Hence, we can simplify the average transmit power as

$$P_{\text{avg}} = \frac{k\rho}{N_t}\left[1 + \frac{1}{N_t - N_r}\right], \tag{4.33}$$

and k can be chosen as

$$k = N_t\left[1 + \frac{1}{N_t - N_r}\right]^{-1}, \tag{4.34}$$

to satisfy the average transmit power constraint of $P_{\text{avg}} = \rho$. The SNR per receive antenna for this scheme is given by

$$\mathsf{SNR} = \frac{\rho(N_t - N_r)}{(1 + N_t - N_r)\sigma_n^2}. \tag{4.35}$$

Next, we present CSIT-based precoding schemes for the fading multiuser MAC, BC and IC.

4.5 Precoding Schemes for Multi-user Channels

In this section, we extend the above transmit precoding schemes to the multiuser MAC, BC and IC. We assume that the wireless channels between transmit and receive antenna pairs are i.i.d. and Rayleigh distributed. An interesting feature of the proposed precoding schemes is that they require each transmitter to have knowledge only of the channel between itself and the receiver(s), and not the other users' channels. We start with the multiuser MAC with CSIT.

4.5.1 The Multiple Access Channel

4.5.1.1 Real O-STBC Signaling Scheme

Consider the M user MAC with N_t antennas at each transmitter (user) and a single antenna at the receiver. The received signal $\mathbf{y} \in \mathbb{R}^L$ can be written as

$$\mathbf{y} = \sum_{i=1}^{M}\sqrt{\frac{k\rho_i}{N_t}}\tilde{\mathbf{H}}^{(i)}\mathbf{P}^{(i)}\mathbf{x}_i + \mathbf{n}, \tag{4.36}$$

where $\mathbf{n} \in \mathbb{R}^L$ is the additive noise at the receiver, distributed as $\mathcal{N}(0, \sigma_n^2)$; $\mathbf{x}_i \in \mathbb{R}_i^L$ is the O-STBC data vector; and ρ_i denotes the average transmit power from the i^{th} user. Also, $\mathbf{P}^{(i)}$ denotes the precoding matrix employed by the i^{th} transmitter corresponding to its channel to the receiver, $\tilde{\mathbf{H}}^{(i)}$ is the equivalent channel matrix as defined in Sec. 4.2, and k denotes the power normalization constant. Now, we choose $\mathbf{P}^{(i)} \triangleq \frac{1}{\alpha_i}\tilde{\mathbf{H}}^{(i)\,T}$ where $\alpha_i = \|\mathbf{h}_i\|^2$, and $\mathbf{h}_i$ is the channel from the i^{th} transmitter to the receiver, with i.i.d. $\mathcal{N}(0,1)$ entries. Then, as in Sec. 4.2, the precoding scheme equalizes the channel, and we obtain L parallel Gaussian MACs with transmit powers $\rho_i, i = 1, 2, \ldots, K$. That is, the received signal can be written as

$$\mathbf{y} = \sum_{i=1}^{M} \sqrt{\frac{k\rho_i}{N_t}} \mathbf{x}_i + \mathbf{n}, \tag{4.37}$$

where $k = N_t(N_t - 2)/L$, for $N_t > 2$. Hence, the precoding scheme converts a Rayleigh flat-fading MISO MAC channel into a fixed-gain Gaussian MAC channel. Moreover, the scheme only requires each transmitter to have knowledge of its own channel to the receiver, and not the other users' channels. The capacity region of the Gaussian MAC channel is known [71], and hence, this forms an achievable rate region for the fading MAC channel with CSIT.

4.5.1.2 QR-Based Precoding Scheme

Consider the M user Rayleigh fading MAC with N_r antennas at the receiver and $N_t \geq 2N_r$ antennas at each transmitter (user). Using precoding scheme described in the previous section, the received signal $\mathbf{y} \in \mathbb{C}^{N_r}$ can be written as

$$\mathbf{y} = \sum_{i=1}^{M} \sqrt{\frac{k\rho_i}{N_t}} \mathbf{H}_i^H \mathbf{P}_i \tilde{\mathbf{x}}_i + \mathbf{n}, \tag{4.38}$$

where $\mathbf{H}_i \in \mathbb{C}^{N_t \times N_r}$ denotes the channel between the i^{th} user and the receiver, distributed as i.i.d. $\mathcal{CN}(0,1)$, $\tilde{\mathbf{x}}_i \in \mathbb{C}^{N_t}$ denotes an extended data vector, and is derived from the complex data vector $\mathbf{x}_i$ as explained earlier in the single user case. Also, ρ_i and $\mathbf{P}_i \in \mathbb{C}^{N_t \times N_t}$ denote the average transmit power available and the precoding matrix, respectively, corresponding to the i^{th} user, and k is a normalization constant. The components of the AWGN $\mathbf{n}$ are assumed to be i.i.d. $\mathcal{CN}(0, \sigma_n^2)$. At the i^{th} transmitter, we choose the matrix $\mathbf{P}_i$ as in Sec. 4.4. With this precoding scheme, the received data vector becomes

$$\mathbf{y} = \sum_{i=1}^{M} \sqrt{\frac{k\rho_i}{N_t}} \mathbf{x}_i + \mathbf{n}. \tag{4.39}$$

Thus, for the multiuser MAC channel, our proposed coding scheme converts the $N_t \times N_r$ MIMO Rayleigh fading MAC channel into N_r parallel Gaussian MAC channels with a fixed gain, when CSI is available at the transmitters.

4.5.2 The Broadcast Channel

We now present an adaptation of the proposed precoding scheme to the M user BC with N_r antennas at each user terminal and $N_t \geq 2MN_r$ antennas at the transmitter. Here, the combined channel matrix $\mathbf{H} \in \mathbb{C}^{N_t \times MN_r}$ between N_t transmit antennas and M user terminals can be considered as a virtual MIMO channel, but with MN_r individual messages. Let $\mathbf{x} = [\sqrt{\rho_1}\mathbf{s}_1, \sqrt{\rho_2}\mathbf{s}_2, \ldots, \sqrt{\rho_M}\mathbf{s}_M]^T$ denote the vector containing the messages intended to the M users, where ρ_i denotes the transmit power used by user i such that $\sum_i \rho_i = \rho$, the total available transmit power, and the transmitted symbols $\mathbf{s}_i \in \mathbb{C}^{N_r}$ are drawn from a constellation satisfying $\mathbb{E}[\mathbf{s}_i^H \mathbf{s}_i] = 1$. Let $\tilde{\mathbf{x}} \in \mathbb{C}^{N_t}$ denote an extended message vector, derived from $\mathbf{x} \in \mathbb{C}^{MN_r}$ as described in the previous section. Hence, one can write the signal model as

$$\mathbf{y} = \sqrt{\frac{k}{N_t}} \mathbf{H}^H \mathbf{P} \tilde{\mathbf{x}} + \mathbf{n}, \tag{4.40}$$

where $\mathbf{P} \in \mathbb{C}^{N_t \times N_t}$ is now a common precoding matrix for all users, k is a normalization constant and $\mathbf{n} \in \mathbb{C}^{MN_r}$ denotes the complex Gaussian noise vector at all the M receivers.

Now, the scheme proposed in Sec. 4.4 in the single user case is directly applicable to the multiuser BC. Note that, due to the possibly unequal power allocation across the users, we have $\mathbf{C}_{\mathbf{x}} = \mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \operatorname{diag}(\rho_1 \mathbf{I}_{N_r}, \rho_2 \mathbf{I}_{N_r}, \ldots, \rho_M \mathbf{I}_{N_r})$. Hence, the average power equation (4.31) is modified to:

$$P_{\text{avg}} = \frac{k\rho}{N_t} \operatorname{tr}\left(\mathbf{C}_{\mathbf{x}} \left\{ \mathbf{I}_{MN_r} + \mathbb{E}_{\mathbf{h}}\left[(\mathbf{I} - \mathbf{R}_{u1})^H \mathbf{R}_{u2}^{-H} \mathbf{R}_{u2}^{-1} (\mathbf{I} - \mathbf{R}_{u1}) \right] \right\}\right). \tag{4.41}$$

Correspondingly, the transmit power normalization constant k is given by

$$k = \frac{N_t}{\operatorname{tr}\left(\mathbf{C}_{\mathbf{x}} \left\{ \mathbf{I} + \mathbb{E}_{\mathbf{h}}\left[\mathbf{R}_1^{-H} \mathbf{R}_1^{-1} \right] \right\}\right)}, \tag{4.42}$$

where we have used $\mathbf{U}_{11} = \mathbf{0}_{MN_r}$ and $\mathbf{U}_{12} = \mathbf{I}_{MN_r}$. Thus, the average power constraint can be satisfied, and the MIMO channel $\mathbf{H}^H \in \mathbb{C}^{N_t \times MN_r}$ is simultaneously converted into MN_r parallel AWGN channels. Due to this, data received at the other users are not required for symbol detection and decoding at a given receiver.

4.5.3 The Interference Channel

In this subsection, we extend the transmit precoding proposed in the previous subsection to an M user IC. For ease of presentation, we consider the $M = 2$ user IC, with $N_t \geq 2MN_r$ antennas at each transmitter and N_r antennas at each receiver. In contrast with the BC, we now have M interfering transmitters. The received signal at i^{th} receiver can be modeled as

$$\mathbf{y}_i = \sqrt{\frac{k}{N_t}} \sum_{j=1}^{2} \mathbf{H}_{i,j}^H \mathbf{P}_j \tilde{\mathbf{x}}_j + \mathbf{n}_i, \tag{4.43}$$

where $\mathbf{H}_{i,j} \in \mathbb{C}^{N_t \times N_r}$ denotes the channel matrix between the i^{th} transmitter and j^{th} receiver, having i.i.d. $\mathcal{CN}(0,1)$ entries, and $\mathbf{n}_i$ denotes the Gaussian noise vector at the i^{th} receiver, having i.i.d. $\mathcal{CN}(0, \sigma_n^2)$ entries.

Now, we exploit the fact that a 2 user IC can be viewed as a combination of two interfering BCs. We employ the power allocation scheme described for the BC, and choose $\rho_1 = \rho$ and $\rho_2 = 0$ at transmitter 1, and $\rho_1 = 0$ and $\rho_2 = \rho$ at transmitter 2, with ρ denoting the per-user transmit power constraint, assumed to be the same for both users. We apply the precoding scheme presented for the BC in the previous subsection. Due to the zero power allocation to the signal component from each transmitter to the unintended receiver, the transmitters do not need to know the data symbols being transmitted by the other transmitter. Also, the receivers see only their intended messages, and hence do not need joint decoding or multi-user detection, and the Rayleigh fading IC is converted into MN_r parallel AWGN channels. Further, it is interesting to note that, when $M = 2$, the number of parallel AWGN channels corresponds precisely to the degrees of freedom of the two user $N_t \times N_r$ MIMO IC with perfect CSIT and CSIR [77].

Remark 7. In most of the existing precoding methods, for example, in techniques such as block diagonalization [78], the channel matrix between all possible transmit-receive pairs is needed for computing the precoding matrix. In contrast, in our proposed method, each transmitter needs to know only the channel between itself and the receivers. That is, it need not know the channels between other transmitters and receivers. This is a significant advantage in practical systems, in terms of the bandwidth and latency involved in exchanging CSI prior to data transmission. For example, in cellular systems, it leads to a reduction in the amount of CSI that needs to be shared between base stations via the backhaul link.

4.6 Simulation Results

In this section, we demonstrate the performance of the proposed precoding schemes using Monte Carlo simulations. For simplicity, we consider a Rayleigh flat-fading MIMO system with $N_t = 2$ or 4 antennas, and $N_r = 1$ or 2 antennas. We consider uncoded QPSK or 4-PAM constellations and compute the BER by averaging over 10^6 noise and 10^4 channel instantiations. We compare the BER performance of the proposed scheme with other existing schemes in the literature that assume perfect CSIR and/or perfect CSIT. We also study the impact of imperfect CSIT (obtained by uplink training using finite transmit power) as well as the impact of imperfect channel inversion (due to the finite peak transmit power available on the practical transmitters) on the BER performance of the proposed methods.

4.6.1 Single User Channels

Figure 4.1 shows the BER performance corresponding to the $N_t \times N_r = 2 \times 1$ and 2×2 MIMO systems. We compare the performance of the Alamouti encoding scheme [59] under perfect CSIR with that of the proposed O-STBC based precoding scheme under perfect CSIT. When $N_r = 2$, both schemes use antenna selection at the receiver. From the figure, we see the significant improvement in the diversity order offered by the proposed O-STBC based precoding scheme compared to the CSIR-based Alamouti scheme. Also, the performance of the O-STBC precoding scheme without antenna selection is about 3 dB worse than the unit-gain SISO AWGN channel, as predicted by the theory. Employing the antenna selection between two receive antennas fills most of this gap. Thus, the proposed scheme converts a MIMO fading channel into an equivalent SISO fixed gain AWGN channel. This plot also shows the performance of the QR-based precoding scheme in the 2×1 system. It can be observed that O-STBC based precoding needs about 0.5 dB higher transmit power to achieve the same BER. Note that the O-STBC based scheme is simpler to implement compared to the QR-based scheme.

To demonstrate the O-STBC based scheme with a higher number of transmit antennas, we show the performance of a 4×1 system employing the full-rate 4×4 real O-STBC code in (4.3) with 4-PAM constellation symbols in Fig. 4.2. Also shown is the performance of the 4×2 system with antenna selection at the receiver. In both cases, we see that the proposed precoding scheme renders the effective channel to be a fixed-gain AWGN channel at all SNRs, as expected. Also, the antenna selection between two antennas results in about 3 dB gain in the BER performance

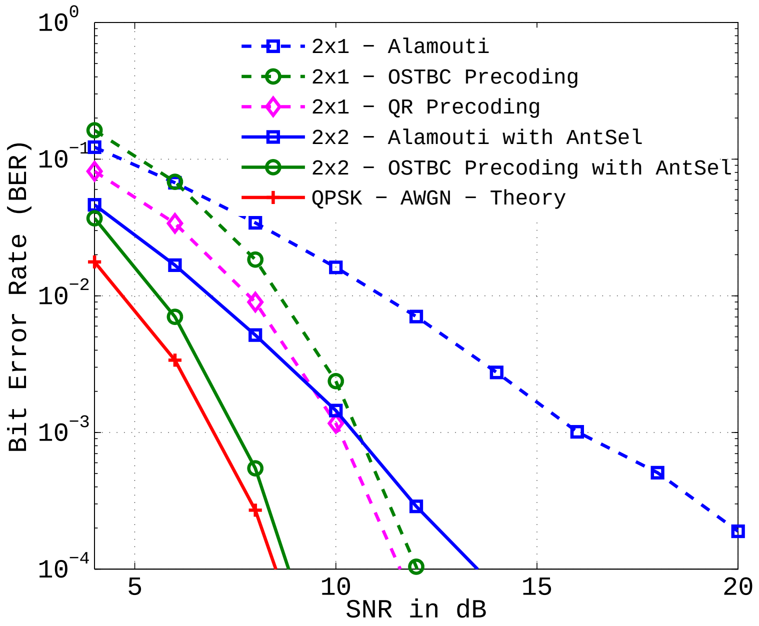

Figure 4.1: BER comparison of the Alamouti code with perfect CSIR and the two proposed schemes with perfect CSIT, for a 2×1 system and a 2×2 system with antenna selection at the receiver, using the QPSK constellation.

for the proposed precoding scheme, while it results in a diversity order improvement from 4 to 8 for the CSIR-based O-STBC transmission scheme.

Figure 4.3 shows the BER performance the QR based precoding scheme for the 2×1 and 4×2 systems. We also show the performance of the complex Alamouti code with uncoded QPSK transmission and perfect CSIR. It can be seen the BER of the proposed scheme is parallel to that of the unit-gain SISO AWGN channel. The gap between the two is about 3 dB and 1.7 dB for the 2×1 and 4×2 systems, respectively, which corroborates well with the theory in (4.35). Further, the proposed scheme far outperforms the perfect CSIR-based Alamouti coding scheme.

4.6.1.1 Precoding with CSI Estimated at the Transmitter

Now, we present simulation results when the CSIT is imperfect. The channel is estimated at the transmitter using a reverse-link training sequence consisting of 10 known symbols transmitted with 10 dB power boosting compared to the forward-link data SNR. We also evaluate the performance when the training signal is transmitted at a fixed power of 30 dB. The MMSE

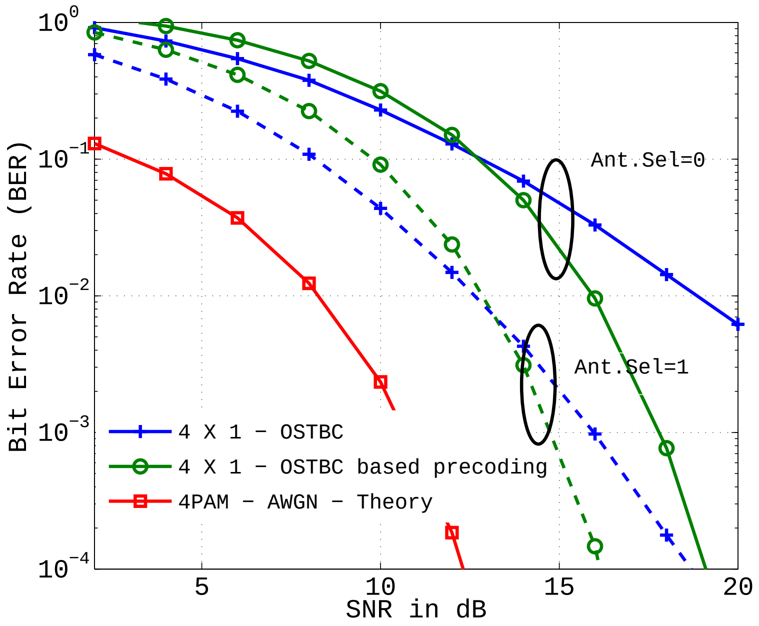

Figure 4.2: BER comparison of the the real O-STBC transmission scheme with perfect CSIR and proposed O-STBC based precoding scheme with perfect CSIT for a 4×1 system with 4-PAM constellation. The dashed curves correspond to the scheme with $N_r = 2$ and antenna selection at the receiver.

channel estimator is used for estimating the CSIT. Simulation results are provided for the O-STBC based precoder in Fig. 4.4; the behavior of the MRT and QR based precoding schemes is similar. It can be seen that the BER performance is close to that obtained with perfect CSIT, and that the waterfall-type behavior of the curves is retained.

4.6.1.2 Transmit Precoding with a Peak Power Constraint

Here, we present the simulation results when the peak power used by the transmitter is restricted to a practical limit (say, to 15 dB higher than the average average power). Limiting the peak power does not invert the channel perfectly for those channel realizations where the peak power required is more than 15 dB above the average power constraint, but the transmit power constraint is still satisfied with the normalization factor k derived earlier. The BER performance is plotted as a function of the SNR for the O-STBC scheme in Fig. 4.5; the behavior of the other two precoding schemes is similar. It can be observed that the BER performance is very close

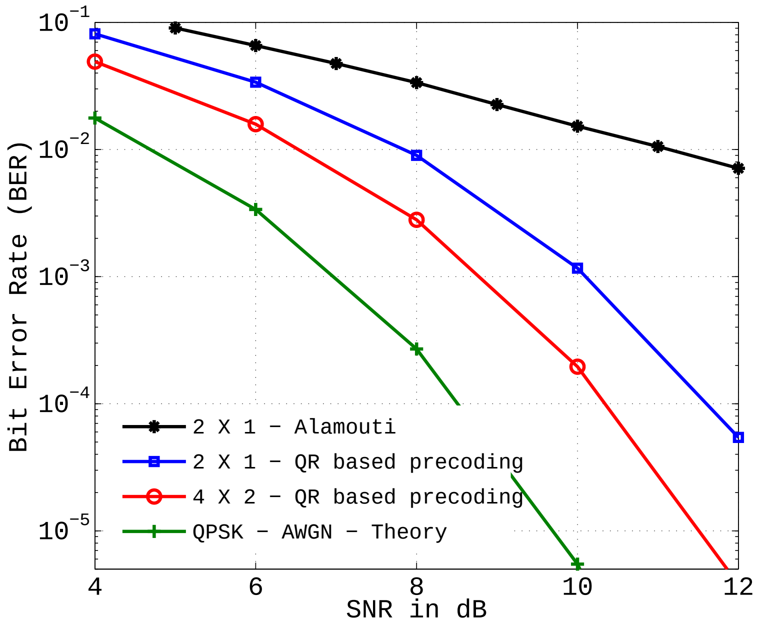

Figure 4.3: BER comparison of the Alamouti code under perfect CSIR and proposed QR-based scheme under perfect CSIT, for the 2 × 1 and 4 × 2 systems, with uncoded QPSK signaling.

to the one with no peak power limit, and the peak power constraint does not significantly alter the behavior of the curves at practical SNRs.

4.6.2 Multi-user Channels

In Fig. 4.6, we demonstrate the performance of the O-STBC precoding scheme for the MAC channel with $M = 2$ users, $N_t = 2, N_r = 1$ and $L = 2$. We compare the performance of the complex Alamouti code constructed using QPSK symbols with that of the proposed O-STBC based and QR based precoding schemes. Here, users 1 and 2 are allocated $\frac{9}{10}$ and $\frac{1}{10}$ of the total transmit power, respectively. For decoding symbols from the two users, a joint Maximum Likelihood (ML) decoder is used at the receiver. We see, again, that the proposed precoding schemes are able to convert the fading MAC into a fixed-gain Gaussian MAC, with the QR based precoding scheme marginally outperforming the O-STBC based precoding scheme.

We next illustrate the BER performance of the proposed precoding scheme for the two-user BC, in Fig. 4.7. We consider a 4 × 1 system with uncoded QPSK signaling. Equal power is allocated to both users, and, hence, the power normalization constant k with the QR-based

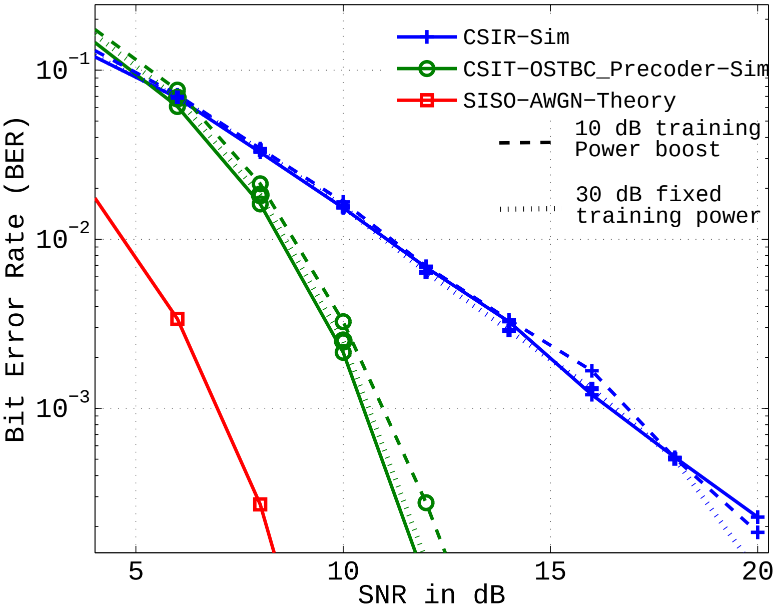

Figure 4.4: BER performance the $N_t = 2$, $N_r = 1$ system with O-STBC based precoding and estimated CSIT. QPSK constellation used for signalling. The MMSE channel estimate was computed using 10 known training symbols with 10 dB SNR boost during the training phase.

precoding scheme is given by (4.34). We see that the performance of the QR-based precoding scheme is parallel to the that of uncoded QPSK symbols in a unit-gain AWGN channel. Thus, the fading MIMO BC is converted into 2 parallel fixed-gain AWGN channels. In the plot, we also show the performance of the vector perturbation method for multi-user BC in [79] for the same antenna configuration, which also requires CSIT. The proposed scheme is not only simpler from an implementation point of view at both the transmitter and receiver, but also outperforms the vector perturbation approach by about 1 dB.

Note that, since the precoding scheme for the IC follows from that of the BC, it results in exactly the same performance as in the BC at the two receivers. Hence, we do not explicitly illustrate the performance of the proposed scheme for the 2-user IC.

4.7 Summary

In this chapter, we proposed three novel, simple-to-implement precoding schemes which utilize CSIT to convert a Rayleigh fading MIMO channel into a fixed-gain AWGN channel, thereby

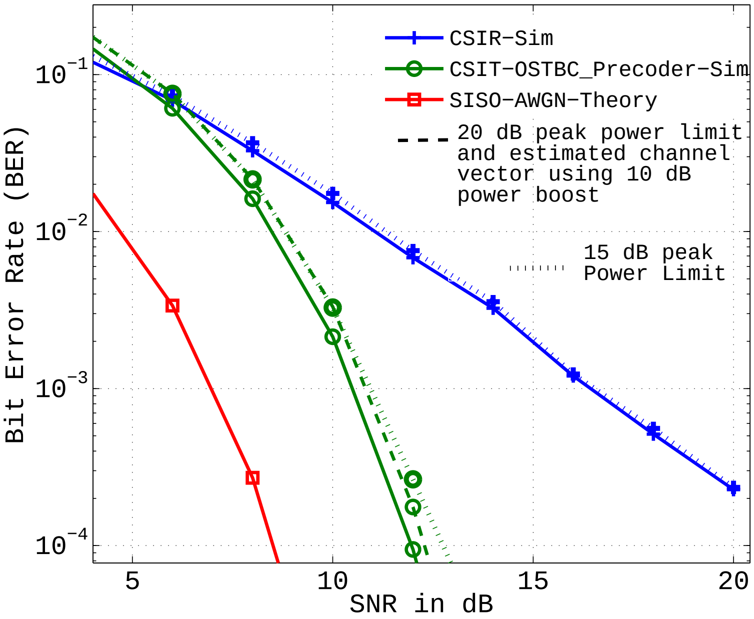

Figure 4.5: BER performance of the $N_t = 2$, $N_r = 1$ system with O-STBC based precoding and a peak power constraint. The peak power was limited to be 15 dB higher than the average transmit power. As another example, peak power limit of 20 dB is used along with estimated channel vector using training sequence with 10 dB additional power than the data transmission.

achieving an infinite diversity order, while simultaneously satisfying an average power constraint. Thus, if perfect CSI could be made available either at the transmitter, or at the receiver, but not both, the perfect CSIT option provides significantly better resilience to fading. The proposed schemes not only offer an improvement over CSIR-based techniques in terms of the diversity order, but also admit single symbol ML decoding at the receiver. We extended the precoding schemes to the fading multiuser MIMO multiple access, broadcast and interference channels. In all three cases, we showed that the fading MIMO channel is converted into parallel fixed-gain AWGN channels. Numerical simulations illustrated the significant performance advantage of the proposed scheme compared to CSIR-based diversity transmission schemes. Moreover, under the practical SNR conditions, the performance under imperfect CSIT also does not degrade significantly when compared to the performance under perfect CSIT conditions. Thus, the proposed precoding schemes are promising for use in reciprocal MIMO systems, where it is practically feasible to directly acquire CSI at the transmitter. Future work could involve optimizing the unitary matrix $\mathbf{U}$ used in the QR-based precoding scheme, and extending the proposed schemes

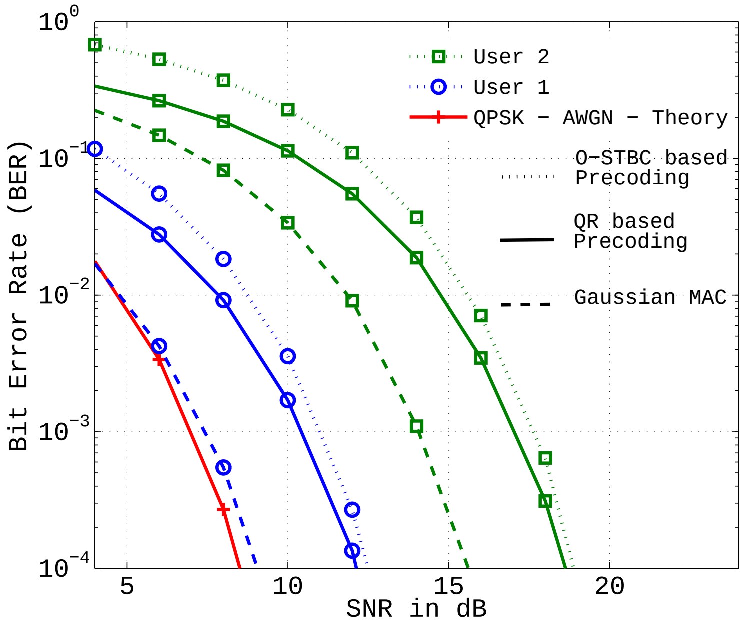

Figure 4.6: BER performance of users 1 and 2, with QPSK signaling in a 2×1 MAC. Here, the transmit powers at the users are set using $\rho_1 = \frac{9}{10}$SNR and $\rho_2 = \frac{1}{10}$SNR, and joint ML decoding is employed at the receiver.

to handle channel estimation errors at the transmitter.

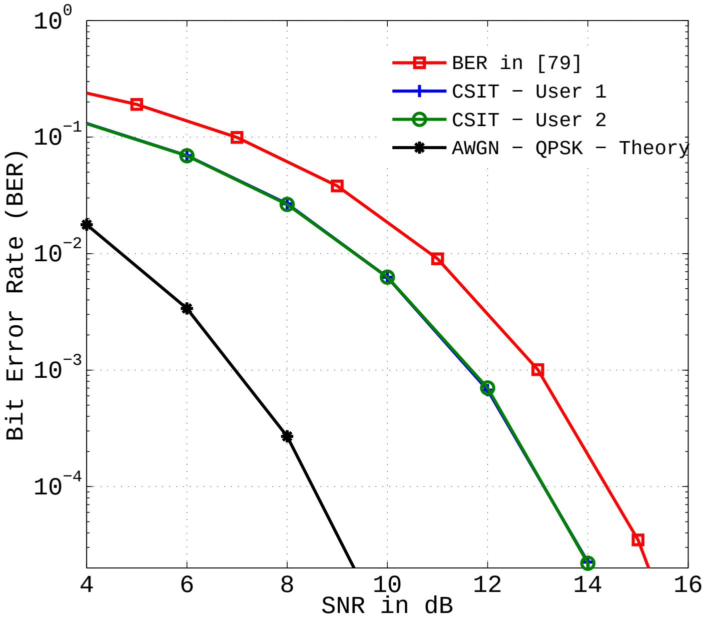

Figure 4.7: BER performance of users 1 and 2 with uncoded QPSK signaling in a 4×1 BC with $\rho_1 = \rho_2 = \frac{1}{2}\text{SNR}$.

Chapter 5

Application to CSI Feedback Link Design in MIMO Systems

"Oh God! Thou art the giver of life, the remover of pain and sorrow, the bestower of happiness. Oh! Creator of the Universe, may we receive thy supreme sin-destroying light. May Thou guide our intellect in the right direction." - **Gayatri Mantra**.

5.1 Introduction

In this chapter, the source and channel coding techniques developed in the previous chapters are applied to a MIMO wireless system, in the context of CSI feedback on the reverse-link. That is, we design the various blocks of the low-rate CSI feedback channel (See Fig. 1.2). We use the quality (MSE) of the received CSI on the reverse-link, and data-rate achieved on the forward-link, as two metrics of interest.

Our goal is to study how the techniques proposed in Chapters 2, 3, and 4 work when implemented in a communication system. To this end, we construct an end-to-end simulation platform that includes all the source coding, receive filtering, channel coding and transmit diversity techniques presented in this thesis. The simulation platform allows us to evaluate the impact of these techniques on quality of the CSI received at the base station as well as on the resulting downlink data rate. In the simulation, the CSI data is compressed using a VQ based source encoder. The index output by the source encoder is channel-coded using the Hamming TCB $(7, 3)$ code presented in 3. The output of the TCB code is mapped to symbols from the signal constellation. The symbols are sent over the fading MIMO reverse link channel using either the CSIR-based Alamouti scheme or the CSIT-based O-STBC transmit diversity scheme presented in Chapter 4. At the receiver, symbols are demodulated and decoded using the TCB decoder. This is followed by the source decoder and the receive filtering operation described in Chapter 2. This comprehensive setup allows us to evaluate the interoperability and performance of different combinations of the proposed techniques.

The rest of the chapter is organized as follows. We describe the system model in section 5.2. The performance of the system under various system configurations is given in Section 5.3. The key take-home messages from this chapter are captured in the summary remarks presented in Section 5.4.

5.2 System Model

The wireless system considered here assumes a Rayleigh fading channel, with two antennas ($N_t = 2$) at the base station (BS) and two antennas ($N_r = 2$) at the user terminal (UT). We consider a TDD system, with perfect channel reciprocity. Further, for simplicity, we consider the transmit powers from the UT and the BS to be such that the average SNR is the same at the two receivers (i.e., at the BS and the UT, respectively). We consider OFDM modulation with $N = 64$ sub-carriers. The entries of the 2×2 channel matrix for each subcarrier are therefore modeled as i.i.d. circularly symmetric complex Gaussian distributed with zero mean and unit variance. The BS sends forward-link training to the UT for estimating the channel. For simplicity, and to focus our attention on the effect of fading and noise in the reverse-link on the feedback of CSI, we make the following assumptions in the simulations:

1. A TDD system with perfect channel reciprocity.

2. CSI at the transmitter (UT) is perfect.

3. The transmit powers at the UT and BS are chosen such that the SNR at both receivers are the same.

In this chapter we consider two types of CSI feedback to the BS: (i) the channel matrix entries corresponding to each sub-carrier are source encoded using the 2-dimensional VQ method described in Chapter 2 and transmitted to the BS, and (ii) the dominant beamforming vector of the channel matrix corresponding to each subcarrier is computed by the UT and is compressed using a source encoding scheme described in the next section. The source encoder output for each of the subcarrier is sent over a noisy fading channel to the BS using the space-time codes presented in the previous chapter. From the received symbols, the BS estimates the CSI, possibly after applying a receive filter. The estimated CSI is used to compute the beamforming vector for down-link data transmission. We evaluate the performance by finding the average data rate achievable in the forward-link as well as by computing the average MSE of the estimated CSI at the BS.

We start by describing the source quantization at the UT. The descriptions of the CSI feedback transmission and reception schemes, the receive filtering at the BS, the down-link beamformed data transmission, and the achievable rate calculation are provided in later subsections.

In the following, we describe two methods of source coding the CSI at the UT. The first method involves directly compressing the entries of the channel matrices, while the second involves first computing the transmit beamforming vector and then compressing the beamforming vector at the UT.

5.2.1 CSI Compression: Source Coding of Channel Matrices

Here, the 2×2 channel matrices corresponding to each sub-carrier are compressed using a 2-dimensional Gaussian codebook. In this method, the 4 complex entries present in each channel matrix are converted into four 2-dimensional Gaussian vectors by stacking the real and imaginary parts of each element. These vectors are compressed using Lloyd-Max MSE optimal codebook designed with 2^{2B} size codebook where B denotes the number of bits per real dimension. The indices corresponding to the 256 (i.e., 64 sub-carriers $\times$ 4 complex entries per channel matrix) 2-dimensional vectors are transmitted to the BS. At the BS, the received indices are used to estimate the 64 channel matrix instantiations. These reconstructed channel matrices are used to compute the beamforming vector that will be used for the down-link data transmission.

5.2.2 CSI Compression: Source Coding of Beamforming Vectors

In this method, the UT first computes the beamforming vector on each subcarrier for forward-link data transmission as the dominant singular vector of the channel matrix on the corresponding subcarrier. The beamforming vectors so computed are then source encoded using VQ-based quantization using a $2B$-bit codebook, i.e., B bits per real dimension. The Lloyd-Max algorithm is used to design locally optimal VQ codebooks of beamforming vectors using $50,000$ beamforming vector instantiations. The VQ codebooks are designed using the projective distance as the distortion metric, which is known to minimize the downlink capacity loss [80]. The projective distance between two unit-norm vectors $\mathbf{v}$ and $\hat{\mathbf{v}}$ is defined as $d(\mathbf{v}, \hat{\mathbf{v}}) = 1 - \left|\mathbf{v}^H \hat{\mathbf{v}}\right|^2$.

Thus, on each subcarrier, the indices corresponding to the entry in the codebook that is closest to the beamforming vector on the subcarrier is found. These indices are transmitted to the BS using classical Alamouti coding [59], or using the transmit diversity scheme proposed in Chapter 4.

5.2.3 CSI Feedback Method 1: Transmission using STBC Code

The $2NB$ bits of information corresponding to the codeword indices output by the source encoder is converted into data bits, which are BPSK modulated. The data symbols are space-time encoded using the Alamouti code and transmitted to the BS. For simplicity and for ease of comparison with the proposed diversity scheme, we assume ideal knowledge of CSI at the base station for data decoding, especially for CSIR based feedback transmission in the uplink.

5.2.4 CSI Feedback Method 2: Transmission using Transmit Precoding

In a TDD system, one can exploit the reciprocity of the channel at the UT to transmit the CSI feedback information to the BS without using the uplink training. This also reduces the overall transmission time as well as provides exponential diversity order as described in Chapter 4. More specifically, the $2NB$ bits of information are transmitted using the transmit diversity scheme 1 given in Chapter 4. For simulating the effect of channel coding, the CSI data bits are channel encoded using a Hamming TCB (7,3) code given in Chapter 3, before being precoded for uplink transmission. This provides a coding gain in the CSI feedback channel compared to uncoded transmission of the CSI. In all cases, BPSK is used as the underlying modulation scheme for data transmission.

5.2.5 Receive Filtering

Since the uplink transmission is not error free, the code indices are possibly corrupted by the noise in the channel. In order to reduce the overall distortion at the source decoder output, we apply the linear receive filter derived in Chapter 2. The following subsections describe the computation of receive filter for MSE distortion metric and projective distance distortion metric.

5.2.5.1 MSE Distortion Metric

The linear receive filter that minimizes the MSE distortion is given by

$$\mathbf{R} = \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1}$$

where $\mathbf{y}$ the is received vector at the decoder when $\mathbf{x}$ is transmitted from the encoder. In the simulation setup, the quantities $\Sigma_{\mathbf{xy}}$ and $\Sigma_{\mathbf{yy}}$ are numerically computed using $50,000$ random source and channel output instantiations of the source index being transmitted through the noisy feedback channel. Thus, it is straightforward to include the receive filtering technique into the simulation setup.

As shown in Chapter 2, when the channel SNR is low, the receive filter drives the source decoder output towards the origin. However, during data transmission, the vector output by the source decoder is re-normalized to have unit-norm, to ensure that the average transmit power is maintained. In beamforming-based systems, therefore, it is the angle between the ideal and estimated beamforming vectors – that is, projection of one vector onto the other – that determines the downlink data rate. Hence, the following modification is made in the receive filter computation, in order to adapt it to account for the projective distance as the distortion metric.

5.2.5.2 Projective Distance Distortion Metric

In this scheme, we compute the covariance of the beamforming vector given the received index, and use its dominant eigenvector as the beamforming vector for data transmission. For example, in the noiseless case, this leads to choosing the dominant eigenvector of the covariance $\Sigma_{\mathbf{xx}}$ of the beamforming vector $\mathbf{x}$, conditioned on $\mathbf{x} \in \mathcal{R}_i$, where $\mathcal{R}_i \triangleq \left\{\mathbf{x} \in \mathbb{C}^{N_t} : |\mathbf{x}^H \hat{\mathbf{x}}_i| > |\mathbf{x}^H \hat{\mathbf{x}}_j|, i \neq j\right\}$

as the beamforming vector for data transmission. In the noisy case, the covariance matrix computation accounts for possible channel errors and their probabilities, in computing the beamforming vector for data transmission. We omit the details as they are straightforward.

5.3 System Performance

In the following, we compare the performance of the proposed system with two baseline (reference) systems for source compression. These baseline systems differ in terms of the way the channel instantiation is quantized, and is in-line with proposals in recent wireless standards. In both baseline systems, the feedback channel is assumed to be noiseless and delay-free. The baseline systems are described in Sec. 5.3.1.

We contrast the performance of the baseline systems with the performance obtained under various schemes:

- **Channel coding on the feedback channel:** We present results both with and without employing the trellis coded block code proposed in Chapter 3.
- **Transmit diversity scheme for the feedback channel:** we consider the Alamouti STBC and the CSIT-based Transmit precoding scheme proposed in Chapter 4.[1]
- **Receive filtering at the base station:** We present results with the receive filter designed for both the MSE distortion metric as well as the projective distance distortion metric. We use the receive filtering scheme proposed in Chapter 2. We also present results without receive filtering.

Here, two performance metrics are considered for the comparing the various systems: (i) The end-to-end average MSE in the channel matrix entries, termed as Channel Quality Metric (CQM), which measures the quality of the CSI that is decoded at the BS, and (ii) The downlink rate that can be achieved with the noisy feedback of CSI.

[1] We emphasize that by *CSIT based transmit precoding scheme* we refer to the O-STBC scheme proposed in the previous chapter. This scheme requires the channel state information at the UT in order to precode the CSI data to be sent to the BS on the uplink feedback channel. The received CSI is used at the BS for computing the beamforming vectors for downlink data transmission. Hence, “CSIT” in this context corresponds to the availability of CSI at the UT, which is the transmitter of the CSI on the feedback (uplink) channel. The CSI at the BS, which transmits data to the UT on the downlink channel, is always estimated from the signal received on the noisy feedback link.

5.3.1 Baseline Systems

5.3.1.1 REF-I System

In the first system, we convert the four complex entries of the channel matrix into four 2-dimensional real-valued vectors and quantize them using a VQ codebook with 2^{2B} entries. The codeword indices are transmitted to the BS via a noiseless feedback channel. At the BS, the beamforming vector corresponding to the largest singular value is computed from the received quantized CSI matrix. Such a direct quantization of the channel entries has been used, for example, in the IEEE 802.11n standard [3].

5.3.1.2 REF-II System

The second system differs from the REF-I system in that the beamforming vector is quantized at the UT using a VQ codebook. In particular, we use the projective distance as the metric in the Lloyd algorithm for designing the codebook. For example, similar codebooks are designed in communication standards such as 3GPP and LTE [8, 9]. The resulting codeword index is transmitted to the BS via the feedback channel.

5.3.1.3 Downlink Rate as the Performance Metric

To make the comparison fair across various number of bits of feedback and SNR conditions, the performance metric chosen for comparison of various schemes is the ratio of the average data rate with the ideal CSI and the average data rate with the quantized CSI, with or without noise in the feedback channel. This ratio is expressed as a percentage of the achievable average downlink data rate. Given the beamforming vector $\hat{\mathbf{v}}$ at the transmitter, we compute the downlink data rate as:

$$\mathcal{C}_{\mathcal{Q}} \triangleq \mathbb{E}_{\mathbf{h}} \left[\log \left(1 + \|\mathbf{H}\hat{\mathbf{v}}\|^2 P_T \right) \right],$$

where P_T is the average downlink data transmit power. Note that, strictly speaking, the above expression represents an upper bound on the ergodic capacity of the channel, since it requires knowledge of $\mathbf{H}\hat{\mathbf{v}}$ at the UT. This knowledge can be acquired without additional training symbols, for example, by measuring the average power over a large number of symbols received within the coherence interval of the channel. However, for simplicity, we assume a genie-aided receiver at the UT which has knowledge of $\|\mathbf{H}\hat{\mathbf{v}}\|$, and, hence, the above expression represents the ergodic capacity of a genie-aided, beamforming-based downlink system.

5.3.2 Simulation Results and Discussion

5.3.2.1 Effect of CSI Compression

Channel Quality Metric: Figure 5.1 considers the REF-I system described above, and compares the CQM (MSE distortion) obtained with noiseless feedback with that obtained using the Alamouti code-based and the proposed CSIT-based transmission of CSI on the uplink. Here, no receive filtering or TCBC is employed at the BS. It can be observed that the total end-to-end MSE distortion in the channel coefficients increases significantly at low SNR, for both the STBC and the CSIT based precoding methods. At low SNR, the STBC based scheme gives better performance, since the advantage of CSIT based precoding methods are exhibited only at moderately high SNRs. However, if an error correction code is employed, the CSIT-based precoding scheme outperforms the STBC based scheme. This is demonstrated in section 5.3.2.2.

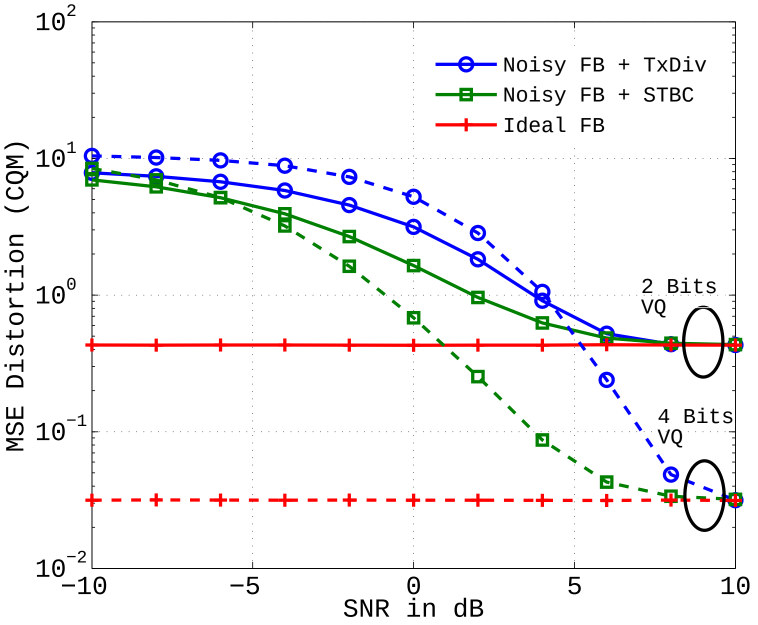

Figure 5.1: Comparison of CQM with CSI feedback for $N_t = N_r = 2$, using 2-dimensional Gaussian codebooks with 16 and 256 entries, STBC and CSIT based transmit precoding method on the feedback link, and with the BPSK constellation.

Downlink Rate Metric: Figure 5.2 shows the performance of the REF-I system that uses a 2-dimensional Gaussian codebook, for various values of B bits per real dimension. The percentage of the achievable data rate is compared for different VQ codebook cardinalities under noiseless and noisy feedback channel conditions. Under noiseless conditions, it can be seen that using as little as 3 bits per real dimension for quantizing the channel matrix achieves most of the achievable data rate with perfect CSI at the BS. Note that, the total number of feedback bits transmitted per channel instantiation is $2BN_rN_tN = 1536$ bits, for the case of $B = 3$ bits.

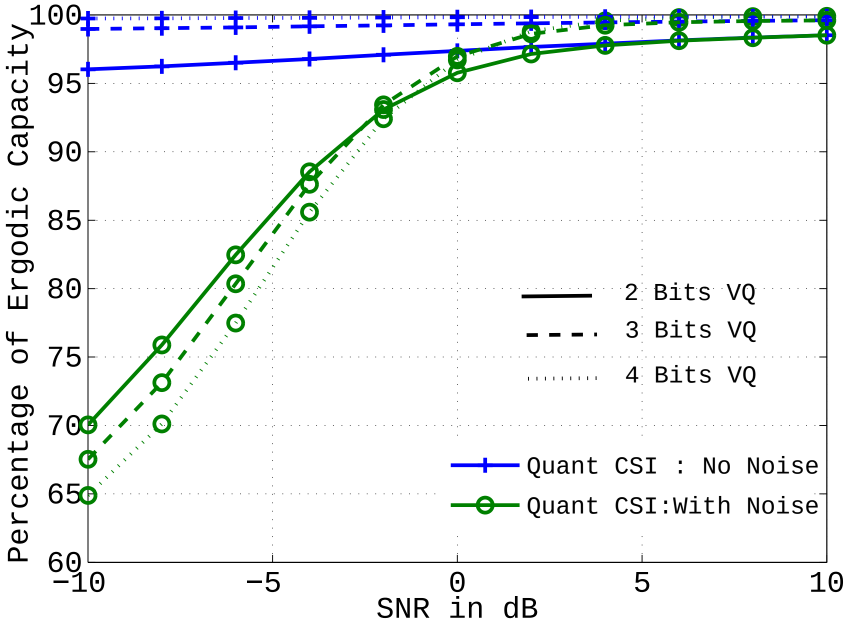

Figure 5.2: Comparison of ergodic capacity with CSI feedback for $N_t = N_r = 2$, using 2-dimensional Gaussian codebooks, STBC on the feedback link, and with the BPSK constellation.

Figure 5.3 shows the performance of the REF-II system, obtained by quantizing the beamforming vector using codebooks designed with the projective distortion metric, and with various values of B. It can be seen that as low as $2B = 6$ bits per beamforming vector achieves most of the capacity achievable with perfect CSI at the BS, in the noiseless feedback channel case. The total feedback bits transmitted per channel instantiation is $2BN = 384$, when $B = 3$. This translates to an 80% reduction in the number of feedback bits that are needed compared to the REF-I case.

Figures 5.2 and 5.3 show the dramatic effect of noise in the feedback channel. Unlike in the ideal feedback case, at low SNR, using a higher number of bits for quantization results in a higher performance loss. This is intuitively reasonable, since, the finer the quantization, the more sensitive the resulting codebook indices would be to errors in the feedback link.

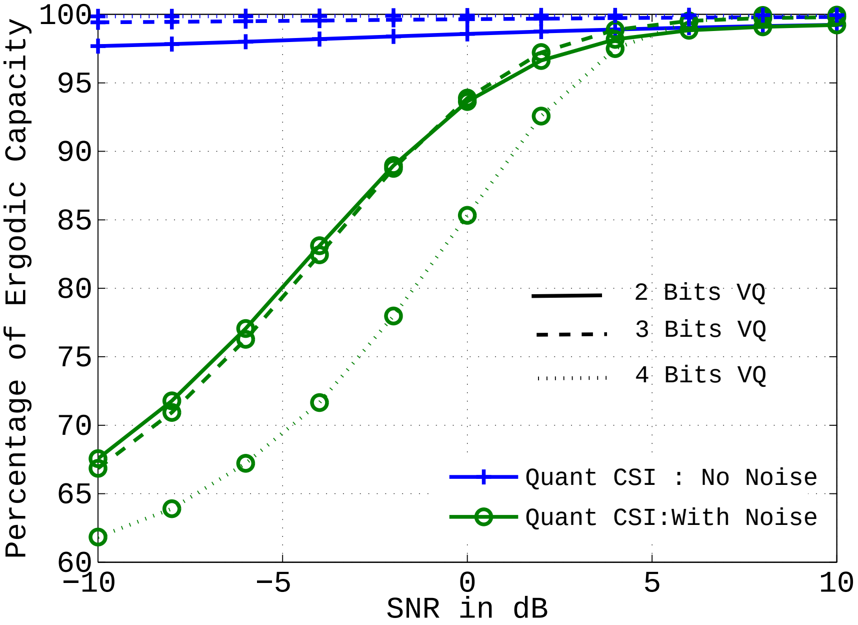

Figure 5.3: Comparison of ergodic capacity with CSI Feedback using beamforming vector compressed with the capacity-optimal, projective distance-based codebook for $N_t = N_r = 2$ and STBC using BPSK constellation.

5.3.2.2 CSI Feedback Using Proposed Techniques

Channel Quality Metric: Figures 5.4, 5.5 and 5.6 compare the CQM for STBC based and CSIT based transmit precoding during the uplink transmission of CSI. Curves are plotted both with and without the receiver filtering and TCBC based error correction codes. Specifically, Fig 5.4 shows the performance with and without receive filtering, Fig. 5.5 shows the performance with and without the TCBC, while Fig. 5.6 shows the performance with both receive filtering and TCBC. It can be observed that the CQM improves when any of the proposed techniques are applied. We see from Fig. 5.4 that the receive filter helps in improving the CQM at low SNRs for both the uplink transmission schemes. We see from Fig. 5.5 that the scheme using the Alamouti STBC and TCBC for channel error correction gives the best performance. However, with the Alamouti STBC scheme, the CQM improves gradually with SNR, unlike the waterfall type behavior in the performance of the CSIT based transmit precoding scheme around 0 dB SNR. Finally, Fig. 5.6 shows that using both the receive filter and the TCBC offers the best possible CQM performance.

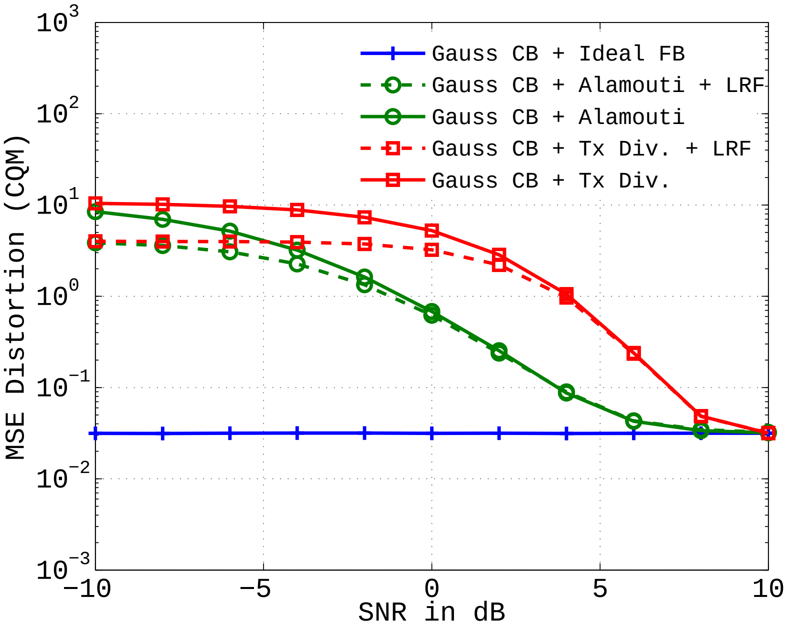

Figure 5.4: CQM performance of the CSI feedback link with $N_t = N_r = 2$, BPSK constellation, and using 2-dimensional Gaussian codebooks with 256 entries, as a function of the SNR. Compared are the CSIR based Alamouti STBC and the CSIT based transmit precoding method, with and without receive filtering. The CQM improvement due to the linear receive filtering is demonstrated here.

Downlink Rate as the Metric: We now illustrate the achievable data rate improvement from the proposed technique compared to the reference systems described in the previous section.

(a) CSI Feedback Using Receive Filter: Figure 5.7 compares the performance of the CSIT-based transmit precoding scheme and the CSIR based Alamouti STBC, with and without the receive filter, for $B = 4$. It can be seen that the receiver filter results in only a marginal improvement in the performance achieved by both the feedback transmission schemes. This plot also highlights the higher sensitivity of the Ref-II system to the noisy feedback channel.

(b) CSI Feedback with the CSIT Based Transmit Precoding Scheme and TCBC: Now, we study

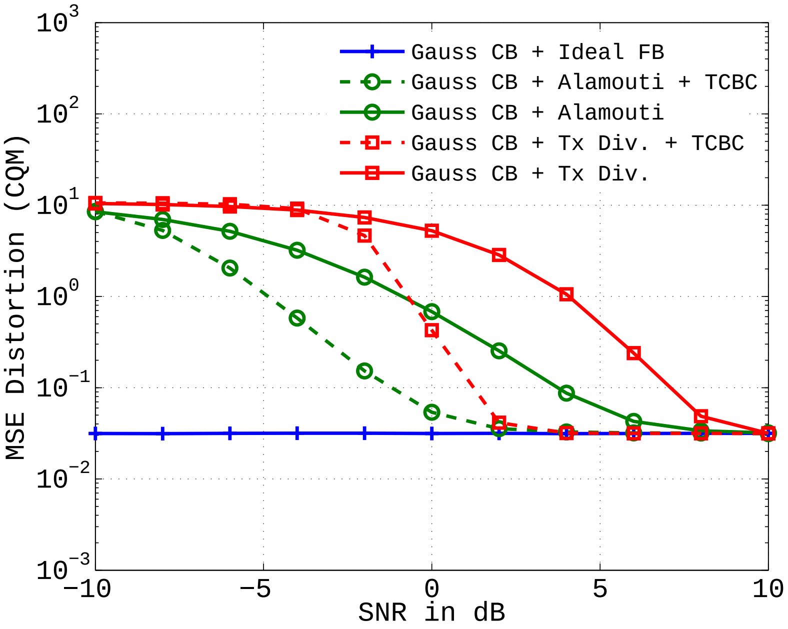

Figure 5.5: CQM performance of the CSI feedback channel for $N_t = N_r = 2$, BPSK signaling, and using 2-dimensional Gaussian codebooks with 256 entries. Compared are CSIT based transmit precoding method on the feedback link, with and without the TCBC. The CQM improvement due to TCBC is demonstrated here.

the effect of using a short latency TCBC for mitigating the effect of errors in the uplink channel. Figure 5.8 compares the performance of transmit diversity based uplink transmission with the TCB code, for both Gaussian and beamforming codebooks. Here, we used an 8 bit codebook for both the encoders. The performance improvement obtainable by employing the TCBC is clear from the graph. The performance with the TCBC is nearly as good as with a noiseless feedback channel, when the channel SNR is close to 0 dB.

(c) CSI Feedback Using Transmit Precoding, TCBC and Receive Filter: Here, we apply all the three methods for mitigating the noise in the feedback channel, namely, the CSIT-based precoding, the TCBC and the linear receive filter. In Fig. 5.9, we compare the normalized downlink data rate performance obtained due to the noise mitigation methods with that of the reference scheme, when the CSI for downlink beamforming is estimated from the feedback bits received over the noisy uplink channel . It can be observed that upto 15% improvement in the normalized downlink rate is possible for a channel SNR of 0 dB even with a simple TCBC code,

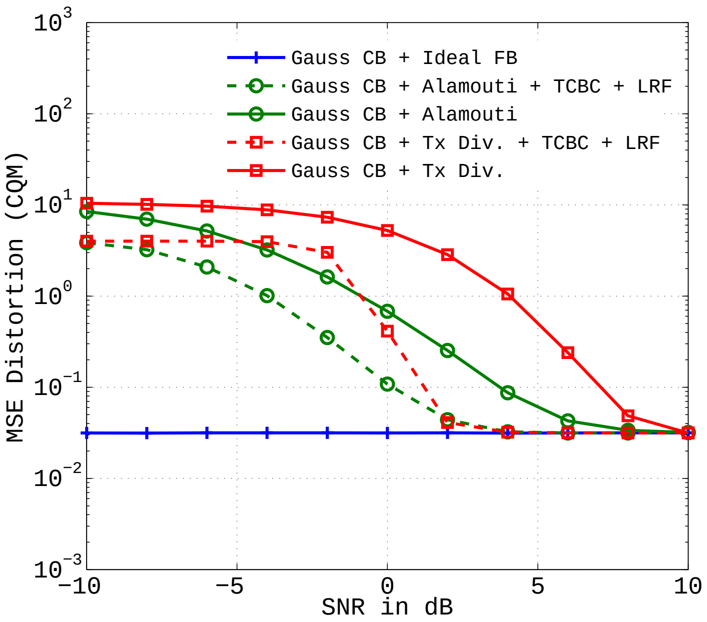

Figure 5.6: CQM performance of the CSI feedback channel for $N_t = N_r = 2$, BPSK signaling, and using 2-dimensional Gaussian codebooks with 256 entries. Compared are the CSIT based transmit precoding method on the feedback link, with and without joint receive filtering and TCBC. The CQM improvement due to jointly using both the receive filtering and the TCBC is demonstrated here.

the CSIT based transmit precoding technique and linear receive filtering. This also corresponds to about 0.25 bits/s/Hz improvement in the downlink data rate, at 0 dB SNR.

5.4 Summary

In addition to the performance plots shown in this chapter, we have also extensively simulated the system under various other system configurations. We make the following observations from our experiments:

- For the same number of bits used for encoding, quantization of the beamforming vectors based on VQ offers better performance (in terms of achievable downlink rate for the same SNR) than that obtained from compressing the entries of the channel matrix directly using either a scalar or a vector quantizer. Quantizing the beamforming vectors directly also

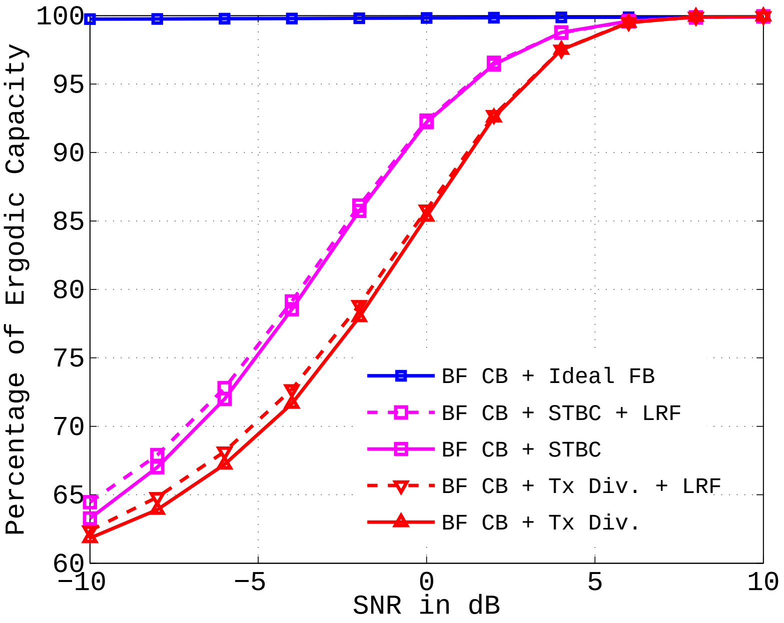

Figure 5.7: Comparison of the normalized data rate with CSI Feedback using beamforming vector codebooks for $N_t = N_r = 2$. The schemes compared are the CSIT based transmit precoding scheme and the CSIR based Alamouti STBC scheme, with BPSK constellation. Here, the receive filter is applied at the source decoder output before it is used for beamforming, as described in the previous section.

leads to fewer bits being needed, to convey the CSI to the transmitter through the feedback link. However, the beamforming VQ indices are more sensitive to channel induced errors than the channel coefficient quantization indices.

- The transmit diversity scheme converts the Rayleigh fading MIMO channel into fixed gain AWGN channel. This allows us to use existing codes designed for AWGN channels for error correction in fading channels. This was demonstrated by using the Hamming TCB $(7, 3)$ code in the simulation. This makes the channel virtually noise free for SNRs above 0 dB. However, without the channel code, the CSIT-based transmit diversity scheme performs worse than the CSIR based schemes such as the Alamouti code, at SNRs below 6 dB. When the target BER is below 10^{-4}, the CSIT based transmit diversity schemes offer significant gain in the operating SNR.

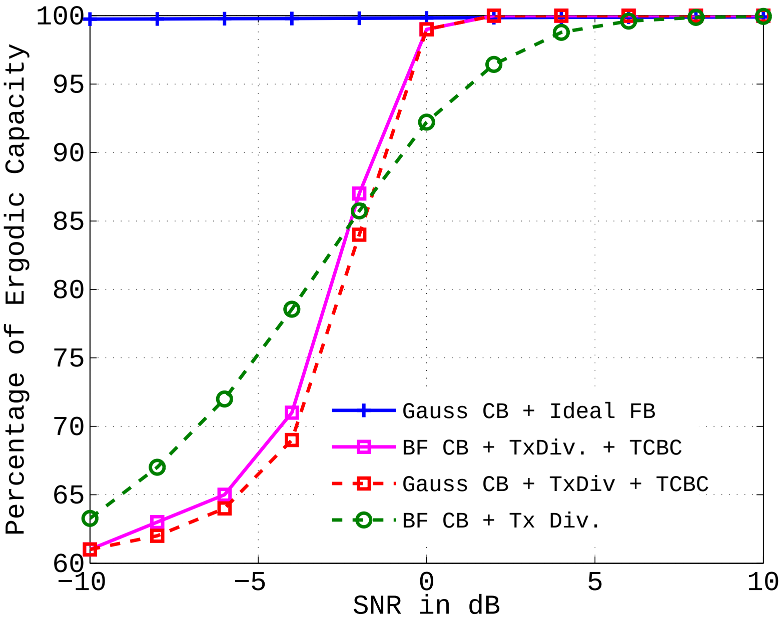

Figure 5.8: Comparison of the normalized data rate with CSI Feedback using Gaussian and beamforming vector codebooks for $N_t = N_r = 2$. The uplink transmission employed the CSIT based transmit precoding scheme with a BPSK constellation and the TCBC(7,3) channel code.

- The low-latency TCB codes offer good coding gain even with a relatively simple code such as the Hamming TCB $(7, 3)$ code. Also note that, the TCB code can easily work as an outercode, since it can be designed for both discrete as well as continuous channels.

- The receive filter offers considerable gain in the MSE distortion but only a marginal gain in the downlink rate, when the beamforming vectors are compressed.

- We have found that a similar percentage improvement can be obtained in both the average data rate and in the 10 % outage rates, when any of the noisy mitigation measures proposed in this thesis are employed. However, we have omitted the outage rate curves to avoid repetition.

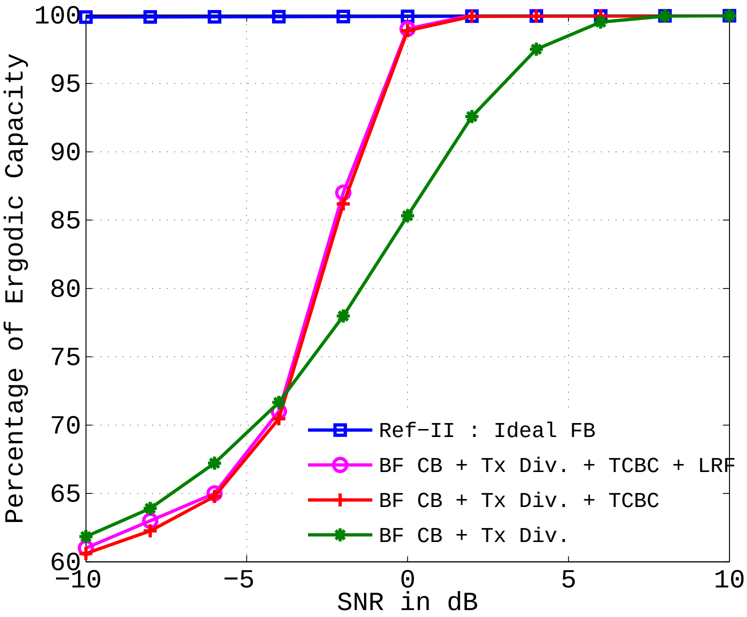

Figure 5.9: Comparison of the normalized downlink data rate with CSI Feedback using beamforming vector codebooks designed for projection distance metric for $N_t = N_r = 2$. The transmission of CSI on the uplink feedback channel employed the CSIT based transmit precoding scheme with BPSK constellation and the TCBC(7,3) channel code.

Chapter 6

Conclusions

"The lotus of the heart, where Brahman exists in all His glory that and not the body, is the true city of Brahman. Brahman, dwelling therein, is untouched by any deed, ageless, deathless, free from grief, free from hunger and from thirst. His desires are right desires, and His desires are fulfilled." - **Chandogya Upanishad**

6.1 Contributions

This thesis addressed three key problems in the MIMO reverse-link channel, with particular emphasis on its use for sending the CSI feedback to the base station. The main contributions of this thesis are as follows:

- *Channel noise tolerant source coding:* We proposed channel noise mitigation techniques for source compression that can be implemented at the receiver. That is, the proposed techniques reduce the overall end-to-end distortion when the source compressed data is sent over noisy channel. The key advantage in this scheme compared to the existing methods is that the source encoder at the transmitter can remain channel agnostic. The receiver filter used at the decoder is computed as a function of channel SNR and applied to the source decoder output, in order to reduce the total distortion.

- *Low latency error correction coding:* The proposed low latency error correction codes perform as good as known Turbo codes for short block lengths, but with lower complexity and decoding delay. Our proposed codes were based on a uniform distance sub-code partitioning that is possible in many of the existing linear block codes. Due to this, the proposed code construction procedure can leverage the vast literature on LBCs with good minimum distance properties. The TCBCs can be used to encode the source compressed channel state information (CSI) data.

- *Reliable communication in a fading environment:* We proposed diversity methods for reducing the SNR required to send the feedback data reliably by converting the Rayleigh fading MIMO or MISO channel into SISO AWGN channels with fixed gain. These methods require CSI at the transmitter, which can be acquired by sending a known training signal in reverse-link direction in time division duplex systems. These schemes were extended to fading multi-user channels as well. This study showed that, in reciprocal Rayleigh fading MIMO systems, acquiring CSI at the transmitter is fundamentally better than acquiring CSI at the receiver. Moreover, with perfect CSIT, one can obtain an infinite diversity order, which is in contrast with the finite diversity order obtainable in perfect CSIR based diversity techniques.

Finally, all the above methods were applied to the MIMO reverse-link CSI feedback channel. We constructed an end-to-end simulation platform that includes all the source coding, receive filtering, channel coding and transmit diversity techniques presented in this thesis. This comprehensive setup allowed us to evaluate different combinations of the proposed techniques in a single platform. Using the platform, we demonstrated the improvement in channel quality at the transmitter and the consequent improvement in the achievable downlink data rate, offered by the proposed techniques.

6.2 Future Work

The techniques presented in this thesis to address the various issues in MIMO reverse-link channel can be extended to other applications. Some specific examples are as follows:

- Receive filtering can be extended to handle non-linear filtering methods and nonstandard channels such as finite state channels, channels with synchronization errors, insertion/deletion errors, and so on.

- The uniform partitioning theorem for binary linear block codes can be extended to non-binary fields, which can result in efficient coding and decoding of TCBC built on non-binary fields also.

- The transmit diversity methods can be extended for multiple receive antenna systems by optimally combining them with or without the knowledge of CSI at the receiver. The antenna selection method addressed in this thesis is one of the ways of addressing the issue. There could be other ways in which additional improvements can be obtained. Also, the real O-STBC based precoding scheme can be extended to handle complex O-STBC also.

- Finally, the CSI feedback channel design can be extended to the multiuser scenario, where the inter-user interference also needs to be taken into account.

Appendix A

High-Rate Distortion Analysis

A.1 High-Rate Distortion Analysis

For a noiseless channel, the MSE distortion of VQ (without the receive filtering) can be written as

$$E_d = \sum_{i=1}^{N} \int_{\mathbf{x}\in\mathcal{R}_i} d(\mathbf{x}, \hat{\mathbf{x}}_i) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \tag{A.1}$$

For the result to follow, the standard high-rate approximations in [24, 35], and the quantization cell approximation in [27] are employed. [1] Now, computing the Taylor series expansion of the distortion measure $d(\mathbf{x}, \hat{\mathbf{x}})$ about $\mathbf{x} = \hat{\mathbf{x}}$ results in

$$d(\mathbf{x}, \hat{\mathbf{x}}) = d(\hat{\mathbf{x}}, \hat{\mathbf{x}}) + (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{d}(\hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{D}(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + \mathcal{O}(\|\mathbf{x} - \hat{\mathbf{x}}\|^3),$$

where $\mathbf{d}(\mathbf{x})$ denotes the derivative defined by $d_i(\hat{\mathbf{x}}) \triangleq \frac{d\ d(\mathbf{x},\hat{\mathbf{x}})}{dx_j}\Big|_{\mathbf{x}=\hat{\mathbf{x}}}$ and $\mathbf{D}(\mathbf{x})$ denotes the Hessian matrix given by $D(i,j) \triangleq \frac{d^2\ d(\mathbf{x},\hat{\mathbf{x}})}{dx_j\ dx_k}\Big|_{\mathbf{x}=\hat{\mathbf{x}}}$. Since the distortion measure is a proper metric and its derivative goes to zero at the local minima of twice continuously differentiable function, the distortion can be approximated as

$$d(\mathbf{x}, \hat{\mathbf{x}}_i) \approx (\mathbf{x} - \hat{\mathbf{x}}_i)^T \mathbf{D}(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}_i) \ \ \forall \mathbf{x} \in \mathcal{R}_i,$$

[1] The approximations used in this paper are well established in the classical source coding literature, and are known to be accurate for high-rate quantization. A good rule-of-thumb is that about 3 bits per dimension (i.e., N of the order 2^{3n}) are required for the high-rate results to apply.

where the constant $\frac{1}{2}$ is absorbed in the definition of Hessian matrix. Note that, the Hessian matrix is interchangeably referred as "sensitivity matrix" in the literature. Thus, any distortion measure can be approximated by a weighted MSE (W-MSE) within the vicinity of $\mathcal{R}_i$ for high-rate quantizer. Under high-rate condition, the specific point density, which is a piece-wise constant function defined as $g_N(\mathbf{x}) \triangleq 1/(NV(\mathcal{R}_i))$ where $\mathbf{x} \in \mathcal{R}_i$ and $V(\mathcal{R}_i)$ is the volume of the Voronoi region $\mathcal{R}_i$, approaches a continuous *point density function* $\lambda(\mathbf{x})$ as N increases. Specifically, for high-rate quantization, the regions $\mathcal{R}_i$ are small, and $f_{\mathbf{x}}(\mathbf{x}) \approx f_{\mathbf{x}}(\hat{\mathbf{x}}_i)$ for $\mathbf{x} \in \mathcal{R}_i$ Hence, (A.1) can be reduced to [24, 25, 37]

$$E_d^{\mathrm{SO}} \doteq \frac{n}{n+2} N^{\frac{-2}{n}} \kappa_n^{\frac{-2}{n}} |\mathbf{W}|^{\frac{1}{n}} \int_{\mathbf{x}} \lambda^{\frac{-2}{n}}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}, \tag{A.2}$$

where $\doteq$ denotes asymptotic equality (i.e., equality when number of quantization cells is very large), κ_n is the volume of an $n-$dimensional unit sphere and the superscript E_d^{SO} denotes the distortion for source optimized VQ. The point density function that minimizes (A.1) is [25, 37] given by

$$\lambda_{\mathrm{conv}}(\mathbf{x}) = \frac{f_{\mathbf{x}}^{\frac{n}{n+2}}(\mathbf{x})}{\int_{\mathbf{y} \in \mathcal{D}_{\mathbf{x}}} f_{\mathbf{x}}^{\frac{n}{n+2}}(\mathbf{y}) d\mathbf{y}}. \tag{A.3}$$

A code book with the above point density can be designed using, for example, the Lloyd-Max algorithm [35], which involves using a large set of training vectors and starting with a random code book, and iteratively updating the quantization regions $\mathcal{R}_i$ and the code points $\hat{\mathbf{x}}_i$ using the Nearest-Neighbor Criterion (NNC) and Centroid Criterion (CC) respectively, till the sample-averaged distortion converges. The resulting high-rate expected distortion for $n-$dimensional Gaussian vector with zero mean and unit variance per dimension is given by

$$E_d^{\mathrm{SO}} \doteq 2\pi N^{\frac{-2}{n}} k_n^{\frac{-2}{n}} \left(\frac{n+2}{n}\right)^{\frac{n}{2}} |\mathbf{W}|^{\frac{1}{n}}. \tag{A.4}$$

A.2 Some Key Approximations

We present some key high-rate approximations that are used in the derivation of the receive filter. Let $\mathbf{x} \in \mathcal{R}_i$ be the source instantiation and let $\hat{\mathbf{x}}_j$ be the received codeword when the index i is transmitted over a noisy channel. Let $\mathbf{e} \triangleq (\mathbf{x} - \hat{\mathbf{x}}_j)$ denote the error vector. Then, the mean and covariance of $\mathbf{e}$ can be written as

$$\begin{aligned}\mathbb{E}[\mathbf{e}] &= \sum_{i=1}^{N}\int_{\mathcal{R}_i}(\mathbf{x}-\hat{\mathbf{x}}_i)f_{\mathbf{x}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \overset{(a)}{\approx} 0 && \text{(A.5)}\\ \mathbb{E}[\mathbf{e}\mathbf{e}^T] &= \sum_{i=1}^{N}\int_{\mathcal{R}_i}(\mathbf{x}-\hat{\mathbf{x}}_i)(\mathbf{x}-\hat{\mathbf{x}}_i)^T f_{\mathbf{x}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \overset{(b)}{\approx} \sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i)\int_{\mathcal{E}_i}\mathbf{e}\mathbf{e}^T\,\mathrm{d}\mathbf{e}, && \text{(A.6)}\end{aligned}$$

where $\mathcal{E}_i$ denotes the Voronoi region $\mathcal{R}_i$ shifted to the origin. The equality (a) is obtained by assuming that the codewords are at the centroids of the Voronoi regions [25], and (b) is due to the approximating $f_{\mathbf{x}}(\mathbf{x})$ with $f_{\mathbf{x}}(\hat{\mathbf{x}}_i)$ inside the quantization cell $\mathcal{R}_i$ [25]. Using an ellipsoid approximation for $\mathcal{R}_i$, it can be shown that [27, 37]

$$\int_{\mathcal{T}(0,\mathbf{W},V_i)} \mathbf{e}^T\mathbf{Q}\mathbf{e}\,\mathrm{d}\mathbf{e} = \frac{V_i}{n+2}\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}} \mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{Q}\right). \tag{A.7}$$

where $\mathbf{Q}$ is any positive semi-definite matrix and $\mathcal{T}(\mathbf{y},\mathbf{W},V_i)$ is the hyper-ellipsoid defined as

$$\mathcal{T}(\mathbf{y},\mathbf{M},V) \triangleq \left\{\mathbf{x}\left|\left(\frac{\kappa_n^2}{V^2|\mathbf{M}|}\right)^{\frac{1}{n}}(\mathbf{x}-\mathbf{y})^T\mathbf{M}(\mathbf{x}-\mathbf{y}) \le 1\right.\right\}. \tag{A.8}$$

Now, to compute the (i, j)-th element in (A.6), one can simply set $\mathbf{Q} = \mathbf{E}_{ij}$ in (A.7), with $\mathbf{E}_{ij}$ being the all zero matrix except for a 1 as the (i, j)-th element, as follows:

$$\sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i)\int_{\mathcal{E}_i}\mathbf{e}\mathbf{e}^T\,\mathrm{d}\mathbf{e} \;\approx\; \sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i)\frac{V_i}{n+2}\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}\mathbf{W}^{-1} \overset{(c)}{\approx} \mathbf{\Phi}_n\Gamma_n N^{\frac{-2}{n}}, \tag{A.9}$$

where $\mathbf{\Phi}_n \triangleq \frac{\kappa_n^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{n+2}\mathbf{W}^{-1}$, $\Gamma_n \triangleq \left[\int_{\mathcal{D}_{\mathbf{x}}} f_{\mathbf{x}}^{\frac{n}{n+2}}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right]^{\frac{n+2}{n}}$, and $\mathbf{\Theta} \triangleq \mathbf{\Phi}_n\Gamma_n N^{\frac{-2}{n}}$. In the above, (c) is obtained by substituting for the source-optimized point density in (A.3) and converting the summation into the corresponding integral. Note that, the trace of the expression above results in the high-rate characterization of the WMSE of VQ-based source coding [27, 37].

A.3 Derivation for Ideal IA

The following relationships are used in the simplification of the average distortion expression. By adding and subtracting $\mathbf{R}\hat{\mathbf{x}}_i$ in $d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j)$ one can simplify the following sum

$$\begin{aligned} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j) &= d(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i + \mathbf{R}\hat{\mathbf{x}}_i - \mathbf{R}\hat{\mathbf{x}}_j) \\ \sum_{j\in S(i)} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j) &= \sum_{j\in S(i)} \big[(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i) + (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) \\ &+ (\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) + (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)\big]. \qquad \text{(A.10)} \end{aligned}$$

$$\begin{aligned} \sum_{j\in S(i)} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j) &= B(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i) + \sum_{j\in S(i)} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) \\ &+ (\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}\mathbf{R}\left[\sum_{j\in S(i)} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)\right] + \left[\sum_{j\in S(i)} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\right]\mathbf{R}^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i), \end{aligned}$$

$$d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_i) = d(\hat{\mathbf{x}}_i, \mathbf{R}\hat{\mathbf{x}}_i) + \mathbf{e}^T\mathbf{W}\mathbf{e} + \hat{\mathbf{x}}_i^T(\mathbf{I} - \mathbf{R})^T\mathbf{W}\mathbf{e} + \mathbf{e}^T\mathbf{W}(\mathbf{I} - \mathbf{R})\hat{\mathbf{x}}_i, \qquad \text{(A.11)}$$

where we have used the approximation that $\sum_{j\in S(i)}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) \approx 0$. Using the above, the following integral can be simplified as,

$$\sum_{j\in S(i)} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j) = B(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i) + \sum_{j\in S(i)} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)$$

$$\begin{aligned} &\int_{\mathbf{x}\in\mathcal{R}_i} (\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)d\mathbf{x} = \int_{\mathbf{x}\in\mathcal{R}_i} (\hat{\mathbf{x}}_i + \mathbf{e} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\hat{\mathbf{x}}_i + \mathbf{e} - \mathbf{R}\hat{\mathbf{x}}_i)d\mathbf{x} \\ &= \int_{\mathbf{x}\in\mathcal{R}_i} \hat{\mathbf{x}}_i^T(\mathbf{I} - \mathbf{R})^T\mathbf{W}(\mathbf{I} - \mathbf{R})\hat{\mathbf{x}}_i d\mathbf{x} + \int_{\mathbf{x}\in\mathcal{E}_i} \mathbf{e}^T\mathbf{W}(\mathbf{I} - \mathbf{R})\hat{\mathbf{x}}_i d\mathbf{e} \\ + &\int_{\mathbf{x}\in\mathcal{E}_i} \hat{\mathbf{x}}_i^T(\mathbf{I} - \mathbf{R})^T\mathbf{W}\mathbf{e}d\mathbf{e} + \int_{\mathbf{x}\in\mathcal{E}_i} \mathbf{e}^T\mathbf{W}\mathbf{e}d\mathbf{e}. \qquad \text{(A.12)} \end{aligned}$$

$$\begin{aligned} \int_{\mathbf{x}\in\mathcal{R}_i} (\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x} - \mathbf{R}\hat{\mathbf{x}}_i)d\mathbf{x} &= \left[\hat{\mathbf{x}}_i^T(\mathbf{I} - \mathbf{R})^T\mathbf{W}(\mathbf{I} - \mathbf{R})\hat{\mathbf{x}}_i\right] V_i \\ &+ \left[\frac{n}{n+2}\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}\right] V_i. \qquad \text{(A.13)} \end{aligned}$$

Substituting the above simplifications in (B.9), we can write the total distortion as

$$\begin{aligned} E_d \doteq & \sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i) \left(\left[\hat{\mathbf{x}}_i^T (\mathbf{I} - \mathbf{R})^T \mathbf{W} (\mathbf{I} - \mathbf{R}) \hat{\mathbf{x}}_i \right] V_i + \left[\frac{n}{n+2} \left(\frac{V_i^2 |\mathbf{W}|}{\kappa_n^2} \right)^{\frac{1}{n}} \right] V_i \right. \\ & \left. + \frac{1-Q}{B} \left[\sum_{j \in S(i)} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T \mathbf{R}^T \mathbf{W} \mathbf{R} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) \right] V_i \right). \end{aligned} \tag{A.14}$$

Let $\mathbf{W} = \mathbf{G}^T\mathbf{G}$ denotes the Cholesky decomposition of $\mathbf{W}$. Using the Rayleigh quotient relation,

$$\Lambda = \frac{\mathbf{x}^T \mathbf{R}^T \mathbf{W} \mathbf{R} \mathbf{x}}{\mathbf{x}^T \mathbf{W} \mathbf{x}}$$

$$\mathbf{x}^T \mathbf{R}^T \mathbf{W} \mathbf{R} \mathbf{x} \approx \frac{1}{n} \mathrm{tr} \left(\mathbf{W}^{-1} \mathbf{R}^T \mathbf{W} \mathbf{R} \right) \mathbf{x}^T \mathbf{W} \mathbf{x} \tag{A.15}$$

Note that, the above approximation matches well when all the eigenvalues are equal and $\mathbf{y} = \mathbf{G}\mathbf{x}$ is one of the eigenvectors. For high rate quantization, $\int_{\mathbf{e}} \mathbf{e} d\mathbf{e} = 0$ and $\int_{\mathbf{x} \in \mathcal{R}_i} d\mathbf{x} = V_i$. In order to evaluate $\sum_{j \in S_i} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T \mathbf{R}^T \mathbf{W} \mathbf{R} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)$, consider the region $\mathcal{G}_i$ whose volume is V_i' which is greater than V_i such that $V_i' = \frac{B}{N\lambda(\hat{\mathbf{x}}_i)}$. The above summation can be approximated as shown below.

$$\begin{aligned} \sum_{j \in S_i} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T \mathbf{R}^T \mathbf{W} \mathbf{R} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) & \approx \frac{1}{n} \mathrm{tr} \left(\mathbf{W}^{-1} \mathbf{R}^T \mathbf{W} \mathbf{R} \right) \sum_{j \in S_i} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T \mathbf{W} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) \\ & = \frac{4B}{n} \mathrm{tr} \left(\mathbf{W}^{-1} \mathbf{R}^T \mathbf{W} \mathbf{R} \right) \left(\frac{V_i^2 |\mathbf{W}|}{\kappa_n^2} \right)^{\frac{1}{n}}, \end{aligned}$$

where we have approximated the sum of W-MSE between the B codewords ($\hat{\mathbf{x}}_j$'s) surrounding the i^{th} codeword ($\hat{\mathbf{x}}_i$) as $4B$ times the square of the average radius of the region $\mathcal{R}_i$. Hence, the average distortion for ideal IA can be written as follows:

$$\begin{aligned} E_d^{\text{ideal}} \doteq & \sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i) \left(\left[\hat{\mathbf{x}}_i^T (\mathbf{I} - \mathbf{R})^T \mathbf{W} (\mathbf{I} - \mathbf{R}) \hat{\mathbf{x}}_i \right] V_i + \left[\frac{n}{n+2} \left(\frac{V_i^2 |\mathbf{W}|}{\kappa_n^2} \right)^{\frac{1}{n}} \right] V_i \right. \\ & \left. + (1-Q) \left[\frac{4}{n} \mathrm{tr} \left(\mathbf{W}^{-1} \mathbf{R}^T \mathbf{W} \mathbf{R} \right) \left(\frac{V_i^2 |\mathbf{W}|}{\kappa_n^2} \right)^{\frac{1}{n}} \right] V_i \right), \end{aligned} \tag{A.16}$$

where $V_i = \frac{1}{N\lambda(\mathbf{x}_i)}$ and $V_i' = B$. Now, substituting for V_i in terms of the point density function $\lambda(\mathbf{x})$ and converting the summation into integral using the Monte Carlo integration formula, we get

$$\begin{aligned} E_d^{\text{ideal}} &\doteq \int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x}) \left[\mathbf{x}^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\mathbf{x}\right] d\mathbf{x} + \frac{nN^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{(n+2)\kappa_n^{\frac{2}{n}}} \left[\int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})d\mathbf{x}\right] \\ &+ \frac{4(1-Q)N^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{n\ \kappa_n^{\frac{2}{n}}} \operatorname{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right) \left[\int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})d\mathbf{x}\right]. \end{aligned} \tag{A.17}$$

A.4 Optimum Minimum-WMSE Filter

It should be noted that the optimum Rx filter $\mathbf{R}_{\text{opt}}$ for the weighted-MSE can be computed by minimizing the WMSE with respect to the Rx filter $\mathbf{R}$. That is,

$$\begin{aligned} \frac{\partial J}{\partial \mathbf{R}} &= \mathbb{E}[\mathbf{x}^T\mathbf{W}\mathbf{x} - \mathbf{y}^T\mathbf{R}^T\mathbf{W}\mathbf{x} - \mathbf{x}^T\mathbf{W}\mathbf{R}\mathbf{y} + \mathbf{y}^T\mathbf{R}^T\mathbf{W}\mathbf{R}\mathbf{y}] = 0 \\ &\Rightarrow\ \mathbb{E}[-\mathbf{W}\mathbf{x}\mathbf{y}^T - \mathbf{W}^T\mathbf{x}\mathbf{y}^T + \mathbf{W}^T\mathbf{R}\mathbf{y}\mathbf{y}^T + \mathbf{W}\mathbf{R}\mathbf{y}\mathbf{y}^T] = 0, \\ &\Rightarrow\ -2\mathbf{W}\left(\mathbb{E}[\mathbf{x}\mathbf{y}^T] - \mathbf{R}\mathbb{E}[\mathbf{y}\mathbf{y}^T]\right) = 0, \end{aligned} \tag{A.18}$$

$$\frac{\partial J}{\partial \mathbf{R}} = \mathbb{E}[\mathbf{x}^T\mathbf{W}\mathbf{x} - \mathbf{y}^T\mathbf{R}^T\mathbf{W}\mathbf{x} - \mathbf{x}^T\mathbf{W}\mathbf{R}\mathbf{y} + \mathbf{y}^T\mathbf{R}^T\mathbf{W}\mathbf{R}\mathbf{y}] = 0, \tag{A.19}$$

where we can assumed that $\mathbf{W}$ is symmetric and used the standard derivative formulae.

$$\begin{aligned} \frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} &= \mathbf{a}\mathbf{b}^T \\ \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} &= \mathbf{b}\mathbf{a}^T \\ \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} &= \mathbf{a}\mathbf{a}^T \\ \frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} &= \mathbf{D}^T\mathbf{X}\mathbf{b}\mathbf{c}^T + \mathbf{D}\mathbf{X}\mathbf{c}\mathbf{b}^T, \end{aligned} \tag{A.20}$$

Solving (A.18) we get the expression of $\mathbf{R}_{\text{opt}}$ as

$$\mathbf{R}_{\text{opt}} = \Sigma_{\mathbf{x}\mathbf{y}}\Sigma_{\mathbf{y}\mathbf{y}}^{-1}. \tag{A.21}$$

A.5 Average Distortion of Ideal IA-Semi Hard Decision

Assuming *ideal IA*, [2] the closest centroids differ in their indices by 1 bit. For high rate coding, when the shape of the source Voronoi regions are similar, the distortion between the new centroids $\hat{\mathbf{x}}_j$ and the source vectors in region $\mathcal{R}_i$ can be upper bounded by sum of distortion between the centroid of $\mathcal{R}_i$ and an offset vector $\mathcal{E}$. That is,

$$d(\mathbf{x}, \hat{\mathbf{x}}_j) \leq d(\mathbf{x}, \mathbf{x}_i) + \mathcal{E}_{i,j}, \tag{A.22}$$

where $\mathcal{E}_{i,j} = d(\mathbf{x}_i, \hat{\mathbf{x}}_j)$. Now, $\frac{1}{B}\sum_{j \in S(i)} \mathcal{E}_{i,j}$ can be approximated by the average distortion between the centroid $\mathbf{x}_i$ and the boundary of the hyper-ellipsoid. [3] With high rate quantization, it is known that the Voronoi regions $\mathcal{R}_i$ can be well approximated by using hyper-ellipsoids [27], with the values of $\mathbf{x}$ satisfying

$$(\mathbf{x} - \mathbf{x}_i)^T \mathbf{W} (\mathbf{x} - \mathbf{x}_i) \leq \left(\frac{V_i^2}{|\mathbf{W}| \kappa_n^2} \right)^{\frac{1}{n}}, \tag{A.23}$$

In [81], it is shown to be

$$\overline{\mathcal{E}}_i \triangleq \frac{1}{B} \sum_{j \in S(i)} \mathcal{E}_{i,j} \quad \approx \left(\frac{V_i^2}{|\mathbf{W}| \kappa_n^2} \right)^{\frac{1}{n}}. \tag{A.24}$$

Substituting (A.24) and (A.22) in (2.58), it can be shown that

$$E_{d,I}^{1E} \approx E_d^{\text{SO}} + (1 - \phi_E)\, \kappa_n^{\frac{-2}{n}} |\mathbf{W}|^{\frac{-1}{n}} \sum_{i=1}^{2^B} f_{\mathbf{X}}(\mathbf{x}_i) V_i^{\frac{2}{n}} V_i, \tag{A.25}$$

where V_i is the volume of the Voronoi region, which can be approximated as $V_i \approx 1/(2^B \lambda(\mathbf{x}_i))$. Hence,

$$E_{d,I}^{1E} \approx E_d^{\text{SO}} + \frac{(1 - \phi_E)\, \kappa_n^{\frac{-2}{n}}}{2^{\frac{2B}{n}} |\mathbf{W}|^{\frac{1}{n}}} \sum_{i=1}^{2^B} \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{\lambda^{\frac{2}{n}}(\mathbf{x}_i)} V_i,$$

$$E_{d,I}^{1E} \approx E_d^{\text{SO}} + \frac{|\mathbf{W}|^{\frac{-1}{n}} (1 - \phi_E)\, \kappa_n^{\frac{-2}{n}}}{2^{\frac{2B}{n}}} \int_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) \lambda^{\frac{-2}{n}}(\mathbf{x})\, d\mathbf{x}. \tag{A.26}$$

[2] It is not guaranteed that such an index assignment is possible for all values of B and dimension n. However, for any given values of B and n, this can be satisfied for a fraction of the centroids.

[3] since the new centroids for the erasure case can be expected to lie on the boundary of the hyper-ellipsoidal region.

Also, recognizing that the integral term in (A.26) is proportional to the E_d^{SO} in (2.55),

$$E_{d,I}^{1E} \approx E_d^{\mathrm{SO}} \left[1 + (1 - \phi_E) \left(\frac{n+2}{n} \right) \right]. \tag{A.27}$$

Appendix B

Alternate Receive Filter Derivation

In this section, we derive the optimum Rx filter which minimizes the expected distortion due to noisy reception of the transmitted index in a conventional approach rather than the MMSE approach presented in the Chapter 2. Here, we show that conventional approach also results in similar optimum receive filter. [1]

B.1 Rx Filter for Random IA

The expected distortion is obtained by taking a double expectation over the source density and the channel transition probabilities, as follows:

$$E_d = \sum_{i,j=1}^{N} P_{j|i} \int_{\mathbf{x} \in \mathcal{R}_i} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}. \tag{B.1}$$

Using high-rate analysis, $\mathbf{x} \in \mathcal{R}_i$ can be expressed as $\mathbf{x} = \hat{\mathbf{x}}_i + \mathbf{e}$, where $\mathbf{e}$ is a "small" vector. Then, $d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j)$ can be expanded as $d(\hat{\mathbf{x}}_i, \mathbf{R}\hat{\mathbf{x}}_j) + \mathbf{e}^T\mathbf{W}\mathbf{e} + 2\mathbf{e}^T\mathbf{W}(\hat{\mathbf{x}}_i - R\hat{\mathbf{x}}_j)$. Next, the quantization cell $\mathcal{R}_i$ is approximated by an n-dimensional hyper-ellipsoid with the same volume as $\mathcal{R}_i$ [27, 28]:

$$\mathcal{R}_i \approx \hat{\mathcal{R}}_i \triangleq \{\mathbf{x} : d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_i) \leq \tau\}. \tag{B.2}$$

[1] The expressions given by MMSE analysis is marginally more accurate than the expression obtained by the more conventional approach given here. The primary difference is that the influence of the source compression error is missing in the conventional approach. However, when B is reasonably large, the difference between the two solutions is negligible.

Hence, the average distortion can be written as

$$
\begin{aligned}
E_d \doteq\; & \epsilon_N \sum_{i,j=1}^{N} d(\hat{\mathbf{x}}_i, \mathbf{R}\hat{\mathbf{x}}_j) f_{\mathbf{x}}(\hat{\mathbf{x}}_i) V_i \\
& +(1-N\epsilon_N)\sum_{i=1}^{N} \hat{\mathbf{x}}_i^T (\mathbf{I}-\mathbf{R})^T \mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i f_{\mathbf{x}}(\hat{\mathbf{x}}_i) V_i \\
& +\sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i)\left(\int_{\mathbf{e}\in\bar{\mathcal{E}}_i} \mathbf{e}^T\mathbf{W}\mathbf{e}\; d\mathbf{e}\right)\sum_{j=1}^{N} P_{j|i},
\end{aligned} \tag{B.3}
$$

where $V_i \triangleq \text{Vol}(\mathcal{R}_i) = (N\lambda(\hat{\mathbf{x}}_i))^{-1}$, $\bar{\mathcal{E}}_i \triangleq \{\mathbf{e} : \hat{\mathbf{x}}_i + \mathbf{e} \in \mathcal{R}_i\}$ is the Voronoi region translated to the origin. Using the quantization cell approximation (B.2), it can be shown that [27],

$$
\int_{\mathbf{e}\in\bar{\mathcal{E}}_i} \mathbf{e}^T\mathbf{W}\mathbf{e}d\mathbf{e} \approx \frac{n}{n+2}|\mathbf{W}|^{\frac{1}{n}} V_i (V_i^2 \kappa_n^{-2})^{\frac{1}{n}}, \tag{B.4}
$$

where $|\mathbf{W}|$ denotes the determinant of the matrix $\mathbf{W}$. Also, using the definition of the *point density* and converting summations into integrals using the Monte Carlo integration formula [29], (B.3) can be reduced to

$$
\begin{aligned}
E_d \doteq\; & N\epsilon_N \int_{\mathbf{x},\mathbf{y}} (\mathbf{x}-\mathbf{R}\mathbf{y})^T\,\mathbf{W}\,(\mathbf{x}-\mathbf{R}\mathbf{y})\,\lambda(\mathbf{y}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} d\mathbf{y} \\
& + (1-N\epsilon_N)\int_{\mathbf{x}} \mathbf{x}^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\mathbf{x} f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \\
& + \frac{n}{n+2} N^{\frac{-2}{n}} \kappa_n^{\frac{-2}{n}} |\mathbf{W}|^{\frac{1}{n}} \int_{\mathbf{x}} \lambda^{\frac{-2}{n}}(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}.
\end{aligned} \tag{B.5}
$$

The above expression reduces to known expressions (e.g., [29]) when $\mathbf{R} = \mathbf{W} = \mathbf{I}$.

Returning to (B.5), with a little manipulation, one obtains

$$
E_d \doteq N\epsilon_N \left[\text{tr}(\mathbf{W}\Sigma_{\mathbf{x}}) + \text{tr}(\mathbf{W}\mathbf{R}\Sigma_\lambda \mathbf{R}^T)\right] + (1-N\epsilon_N)\,\text{tr}\left(\mathbf{W}\,(\mathbf{I}-\mathbf{R})\,\Sigma_{\mathbf{x}}\,(\mathbf{I}-\mathbf{R})^T\right) + E_d^{\text{SO}}, \tag{B.6}
$$

where $\Sigma_{\mathbf{x}}$ and Σ_λ are the covariance matrices of a random vectors $\mathbf{x}$ and $\mathbf{y}$ respectively, with their corresponding probability density given by $f_{\mathbf{x}}(\mathbf{x})$ and $\lambda(\mathbf{x})$. Clearly, when $\epsilon_N = 0$, (B.6) is minimized by $\mathbf{R} = \mathbf{I}$ as expected. Using straightforward differentiation, the $\mathbf{R}$ that minimizes E_d above can be shown to be the following Minimum Mean Squared Error (MMSE)-type matrix

$$
\mathbf{R}_{\text{opt}} = \Sigma_{\mathbf{x}} \left(\frac{N\epsilon_N}{1-N\epsilon_N}\Sigma_\lambda + \Sigma_{\mathbf{x}}\right)^{-1}, \tag{B.7}
$$

and the corresponding expected distortion can be found by substituting (B.7) into (B.6) as

$$E_d \doteq E_d^{\mathrm{SO}} + \mathrm{tr}(\mathbf{W}\Sigma_{\mathbf{x}}) + (1 - N\epsilon_N)\mathrm{tr}\left(\mathbf{W}\Sigma_{\mathbf{x}}\left(\frac{N\epsilon_N}{1 - N\epsilon_N}\Sigma_\lambda + \Sigma_{\mathbf{x}}\right)^{-1}\Sigma_{\mathbf{x}}\right).$$

Interestingly, the optimum receive filter depends only on the second-order properties of the source and the codebook. For the n-dimensional independent and identically distributed (IID) Gaussian source, the optimum linear filter becomes

$$\mathbf{R}_{\mathrm{opt}} = \frac{n(1 - N\epsilon_N)}{n + 2N\epsilon_N} I,$$

and the corresponding distortion is given by

$$E_d^{\mathrm{Ropt}} \doteq \mathrm{tr}(\mathbf{W})N\epsilon_N + \frac{N\epsilon_N\,(1 - N\epsilon_N)\,n\,(n+2)}{n + 2N\epsilon_N} + 2\pi N^{\frac{-2}{n}}\kappa_n^{\frac{-2}{n}}\left(\frac{n+2}{n}\right)^{\frac{n}{2}}. \tag{B.8}$$

B.2 Rx Filter for Ideal IA

For simplicity of presentation, we restrict the analysis to 1 bit error only, which dominates the performance when the channel SNR is reasonably high, although it can be extended to multi-bit errors. The expected distortion can be written as

$$E_d = \sum_{i=1}^{N} f_X(x_i) \sum_{j=1}^{N} P_{j|i} \int_{\mathbf{x}\in\mathcal{R}_i} d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j)d\mathbf{x}, \tag{B.9}$$

where $\mathbf{x} \in \mathcal{R}_i$ can be written as $\hat{\mathbf{x}}_i + \mathbf{e}$. If atmost one bit error occurs, channel transition probabilities for the transmission of B, can be written as follows.

$$\begin{aligned} P_{i|i} &= Q \triangleq (1-q)^B \\ P_{j|i} &= \frac{1-Q}{B} = \frac{1-(1-q)^B}{B}\ \forall j \in S_i, \end{aligned} \tag{B.10}$$

where S_i denotes the set of neighbours for the i^{th} code vector considering all 1 bit errors in the index corresponding to the i^{th} code vector. The following relationships are used in the simplification of the average distortion expression. By adding and subtracting $\mathbf{R}\hat{\mathbf{x}}_i$ in $d(\mathbf{x}, \mathbf{R}\hat{\mathbf{x}}_j)$

one can simplify the following sum

$$\begin{aligned}\sum_{j\in S(i)} d(\mathbf{x},\mathbf{R}\hat{\mathbf{x}}_j) &= B(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)+\sum_{j\in S(i)}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)\\ &+ (\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}\mathbf{R}\left[\sum_{j\in S(i)}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)\right]+\left[\sum_{j\in S(i)}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)^T\right]\mathbf{R}^T\mathbf{W}(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i),\end{aligned}$$

$$d(\mathbf{x},\mathbf{R}\hat{\mathbf{x}}_i) = d(\hat{\mathbf{x}}_i,\mathbf{R}\hat{\mathbf{x}}_i)+\mathbf{e}^T\mathbf{W}\mathbf{e}+\hat{\mathbf{x}}_i^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}\mathbf{e}+\mathbf{e}^T\mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i, \qquad \text{(B.11)}$$

where we have used the approximation that $\sum_{j\in S(i)}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)\approx 0$. Using the above, the following integral can be simplified as,

$$\begin{aligned}\sum_{j\in S(i)} d(\mathbf{x},\mathbf{R}\hat{\mathbf{x}}_j) &= B(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)\\ &+ \sum_{j\in S(i)}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)\int_{\mathbf{x}\in\mathcal{R}_i}(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\mathbf{x}-\mathbf{R}\hat{\mathbf{x}}_i)d\mathbf{x}\\ &\int_{\mathbf{x}\in\mathcal{R}_i}(\hat{\mathbf{x}}_i+\mathbf{e}-\mathbf{R}\hat{\mathbf{x}}_i)^T\mathbf{W}(\hat{\mathbf{x}}_i+\mathbf{e}-\mathbf{R}\hat{\mathbf{x}}_i)d\mathbf{x}\\ &= \int_{\mathbf{x}\in\mathcal{R}_i}\hat{\mathbf{x}}_i^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i d\mathbf{x}+\int_{\mathbf{x}\in\mathcal{E}_i}\mathbf{e}^T\mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i d\mathbf{e}\\ &+ \int_{\mathbf{x}\in\mathcal{E}_i}\hat{\mathbf{x}}_i^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}\mathbf{e}d\mathbf{e}+\int_{\mathbf{x}\in\mathcal{E}_i}\mathbf{e}^T\mathbf{W}\mathbf{e}d\mathbf{e}\\ &= \left[\hat{\mathbf{x}}_i^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i\right]V_i+\left[\frac{n}{n+2}\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}\right]V_i,\end{aligned}$$

Substituting the above simplifications in (B.9), we can write the total distortion as

$$\begin{aligned}E_d &\doteq \sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i)\Bigg(\left[\hat{\mathbf{x}}_i^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i\right]V_i+\left[\frac{n}{n+2}\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}\right]V_i\\ &+ \frac{1-Q}{B}\left[\sum_{j\in S(i)}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i-\hat{\mathbf{x}}_j)\right]V_i\Bigg).\end{aligned} \qquad \text{(B.12)}$$

Let $\mathbf{W}=\mathbf{G}^T\mathbf{G}$ denotes the Cholesky decomposition of $\mathbf{W}$. Using the Rayleigh quotient relation,

$$\Lambda=\frac{\mathbf{x}^T\mathbf{R}^T\mathbf{W}\mathbf{R}\mathbf{x}}{\mathbf{x}^T\mathbf{W}\mathbf{x}}$$

It can be shown that

$$\mathbf{x}^T\mathbf{R}^T\mathbf{W}\mathbf{R}\mathbf{x} \approx \frac{1}{n}\mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right)\mathbf{x}^T\mathbf{W}\mathbf{x}. \tag{B.13}$$

Note that, the above approximation matches well when all the eigenvalues are equal and $\mathbf{y} = \mathbf{G}\mathbf{x}$ is one of the eigenvectors.

For high rate quantization, $\int_{\mathbf{e}} \mathbf{e} d\mathbf{e} = 0$ and $\int_{\mathbf{x}\in\mathcal{R}_i} d\mathbf{x} = V_i$. In order to evaluate $\sum_{j\in S_i}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)$, consider the region $\mathcal{G}_i$ whose volume is V_i' such that $V_i' = \frac{B}{N\lambda(\hat{\mathbf{x}}_i)}$. The above summation can be approximated as shown below.

$$\begin{aligned}\sum_{j\in S_i}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{R}^T\mathbf{W}\mathbf{R}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) &\approx \frac{1}{n}\mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right)\sum_{j\in S_i}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T\mathbf{W}(\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) \\ &= \frac{4B}{n}\mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right)\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}.\end{aligned}$$

where we have approximated the sum of W-MSE between the B centroids ($\hat{\mathbf{x}}_j$'s) surrounding the i^{th} centroid ($\hat{\mathbf{x}}_i$) as $4B$ times the square of the average radius of the region $\mathcal{R}_i$. Hence, the average distortion for *ideal IA* can be written as follows:

$$\begin{aligned}E_d^{\text{ideal}} &\doteq \sum_{i=1}^{N} f_{\mathbf{x}}(\hat{\mathbf{x}}_i)\Bigg(\left[\hat{\mathbf{x}}_i^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\hat{\mathbf{x}}_i\right]V_i + \left[\frac{n}{n+2}\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}\right]V_i \\ &+ (1-Q)\left[\frac{4}{n}\mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right)\left(\frac{V_i^2|\mathbf{W}|}{\kappa_n^2}\right)^{\frac{1}{n}}\right]V_i\Bigg).\end{aligned} \tag{B.14}$$

Now, substituting for V_i in terms of the point density function $\lambda(\mathbf{x})$ and converting the summation into integral using the Monte Carlo integration formula, we get

$$\begin{aligned}E_d^{\text{ideal}} &\doteq \int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\left[\mathbf{x}^T(\mathbf{I}-\mathbf{R})^T\mathbf{W}(\mathbf{I}-\mathbf{R})\mathbf{x}\right]d\mathbf{x} \\ &+ \frac{nN^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{(n+2)\kappa_n^{\frac{2}{n}}}\left[\int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})d\mathbf{x}\right] \\ &+ \frac{4(1-Q)N^{\frac{-2}{n}}|\mathbf{W}|^{\frac{1}{n}}}{n\ \kappa_n^{\frac{2}{n}}}\mathrm{tr}\left(\mathbf{W}^{-1}\mathbf{R}^T\mathbf{W}\mathbf{R}\right)\left[\int_{\mathbf{x}} f_{\mathbf{x}}(\mathbf{x})\lambda^{\frac{-2}{n}}(\mathbf{x})d\mathbf{x}\right].\end{aligned} \tag{B.15}$$

After some manipulations, the average distortion can be written as

$$E_d^{\text{ideal}} = \operatorname{tr}\left(\mathbf{W}(\mathbf{I}-\mathbf{R})\Sigma_{\mathbf{x}}(\mathbf{I}-\mathbf{R})^T\right) + E_d^{\text{SO}}\left[1+\frac{4(1-Q)(n+2)}{n^2}\operatorname{tr}\left(\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\mathbf{R}^T\right)\right]. \quad \text{(B.16)}$$

Now, straight forward differentiation of the above with respect to $\mathbf{R}$ results in the optimum Rx filter matrix $\mathbf{R}_{opt}$.

$$\left[\mathbf{W}^T\mathbf{R}\Sigma_{\mathbf{x}}^T + \mathbf{W}\mathbf{R}\Sigma_{\mathbf{x}} - \mathbf{W}^T\Sigma_{\mathbf{x}}^T - \mathbf{W}\Sigma_{\mathbf{x}}\right] + \frac{4(1-Q)(n+2)E_d^{\text{SO}}}{n^2}\left[\mathbf{W}^T\mathbf{R}\mathbf{W}^{-T} + \mathbf{W}\mathbf{R}\mathbf{W}^{-1}\right] = 0. \quad \text{(B.17)}$$

For symmetric $\mathbf{W}$ and $\Sigma_{\mathbf{x}}$, a closed form expression for $\mathbf{R}_{opt}$ can be obtained as

$$\mathbf{R}_{\text{opt}} = \left(\Sigma_{\mathbf{x}} + \frac{4(1-Q)(n+2)E_d^{\text{SO}}}{n^2}\mathbf{W}^{-1}\right)^{-1}\Sigma_{\mathbf{x}}. \quad \text{(B.18)}$$

Clearly, when the channel is error free, i.e., $Q = 1$, the optimum filter matrix turns out to be an identity matrix $\mathbf{R} = \mathbf{I}$.

B.3 Rx Filter for Specific IA

In this section, we model the expected distortion for a given IA as a convex combination of the expected distortion for the *ideal IA* and expected distortion for the *Random IA*. The weighing constant used in the combination is determined using simulation later.

$$\begin{aligned}
E_d^{IA} &\doteq \eta E_d^{\text{ideal}} + (1-\eta)E_d^{\text{random}} \\
&= \eta\left\{\operatorname{tr}\left(\mathbf{W}(\mathbf{I}-\mathbf{R})\Sigma_{\mathbf{x}}(\mathbf{I}-\mathbf{R})^T\right) + E_d^{\text{SO}}\left[1+\frac{4(1-Q)(n+2)}{n^2}\operatorname{tr}\left(\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\mathbf{R}^T\right)\right]\right\} \\
&+ (1-\eta)\left\{N\epsilon_N\left[\operatorname{tr}(\mathbf{W}\Sigma_{\mathbf{x}}) + \operatorname{tr}(\mathbf{W}\mathbf{R}\Sigma_{\lambda}\mathbf{R}^T)\right]\right. \\
&+ \left.(1-N\epsilon_N)\operatorname{tr}\left(\mathbf{W}\left(\mathbf{I}-\mathbf{R}\right)\Sigma_{\mathbf{x}}\left(\mathbf{I}-\mathbf{R}\right)^T\right)\right\} + (1-\eta)E_d^{\text{SO}} \\
&= \left[1-(1-\eta)N\epsilon_N\right]\operatorname{tr}\left(\mathbf{W}(\mathbf{I}-\mathbf{R})\Sigma_{\mathbf{x}}(\mathbf{I}-\mathbf{R})^T\right) \\
&+ E_d^{\text{SO}}\left(1+\eta\left[\frac{4(1-Q)(n+2)}{n^2}\operatorname{tr}\left(\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\mathbf{R}^T\right)\right]\right) \\
&+ (1-\eta)N\epsilon_N\left[\operatorname{tr}(\mathbf{W}\Sigma_{\mathbf{x}}) + \operatorname{tr}(\mathbf{W}\mathbf{R}\Sigma_{\lambda}\mathbf{R}^T)\right]. \qquad \text{(B.19)}
\end{aligned}$$

Using straightforward differentiation with respect to $\mathbf{R}$ and equating to zero, we get

$$(1-(1-\eta)N\epsilon_N)\left[\mathbf{W}\mathbf{R}\Sigma_{\mathbf{x}}-\mathbf{W}\Sigma_{\mathbf{x}}\right]+\left(\frac{4\eta(1-Q)(n+2)E_d^{\mathrm{SO}}}{n^2}\right)\left[\mathbf{W}\mathbf{R}\mathbf{W}^{-1}\right] \\ +(1-\eta)N\epsilon_N\left[\mathbf{W}\mathbf{R}\Sigma_\lambda\right]=0. \quad \text{(B.20)}$$

$$\mathbf{R}_{opt}=\Sigma_{\mathbf{x}}\left[\Sigma_{\mathbf{x}}+\frac{(1-\eta)N\epsilon_N}{1-(1-\eta)N\epsilon_N}\Sigma_\lambda+\frac{4\eta(1-Q)(n+2)E_d^{\mathrm{SO}}}{\left[1-(1-\eta)N\epsilon_N\right]n^2}\mathbf{W}^{-1}\right]^{-1}. \quad \text{(B.21)}$$

The proportionality constant η can be obtained by simulations for the given IA. For a Gaussian i.i.d. source with variance per dimension $\sigma^2=1$, the optimum Rx filter is given by

$$\mathbf{R}_{opt}=\left[\frac{n+2(1-\eta)N\epsilon_N}{n(1-(1-\eta)N\epsilon_N)}\mathbf{I}+\frac{4\eta(1-Q)(n+2)E_d^{\mathrm{SO}}}{\left[1-(1-\eta)N\epsilon_N\right]n^2}\mathbf{W}^{-1}\right]^{-1}.$$

It is interesting to note that even if the weight matrix is identity $\mathbf{W}=\mathbf{I}$, there appears a correction term corresponding to contribution from the *ideal IA*. That is, the Rx filter output is a scaled version of the received codevector. The scale value reduces as E_d^{SO} increases (i.e., less bits used in the quantization) or η increases. Evaluating the average distortion for this optimal filter matrix is mathematically intractable. Hence, we evaluate (B.19) numerically for the computed $\mathbf{R}_{opt}$ in (B.21). Next, we will describe another Rx filtering technique for mitigating the channel noise.

Appendix C

TCBC: Proofs of Lemmas and Theorem 2

C.1 Proof of Lemma 2

Let $\mathbf{C}_0$ be a linear uniform code with distance d_u. Since $\mathbf{C}_0$ is non-trivial and linear, there exist $\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2 \in \mathbf{C}_0$ such that $\mathbf{c}_2 = \mathbf{c}_0 + \mathbf{c}_1$ and $\mathbf{c}_i \neq \mathbf{0}$ for $i = 0, 1, 2$. Now, the Hamming weight of $\mathbf{c}_2$ is can be expanded as follows:

$$\mathcal{W}_H(\mathbf{c}_2) = \mathcal{W}_H(\mathbf{c}_0) + \mathcal{W}_H(\mathbf{c}_1) - 2\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1) \tag{C.1}$$

Since the Hamming weight of $\mathbf{c}_0, \mathbf{c}_1$ and $\mathbf{c}_2$ are all equal to d_u, the above equation simplifies to

$$d_u = 2[d_u - \mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1)], \tag{C.2}$$

and hence d_u is even. Moreover, $\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1) = d_u/2$. To show the converse, let $\mathbf{C}_0$ be a uniform code with $d_u = 2\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1)$. Then,

$$\mathcal{W}_H(\mathbf{c}_0 + \mathbf{c}_1) = 2[d_u - \mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1)] = d_u, \tag{C.3}$$

and hence $\mathbf{c}_2 \in \mathbf{C}_0$. This proves that $\mathbf{C}_0$ is linear with respect to the code-words $\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2$.

C.2 Proof of Lemma 3

Consider the sum $\mathbf{c_0} + \mathbf{c}_1 + \mathbf{c}_2$. Using Lemma 2, one can write

$$\begin{aligned}\mathcal{W}_H(\mathbf{c}_0 + \mathbf{c}_1 + \mathbf{c}_2) &= d_u + d_u - 2\mathcal{W}_H(\mathbf{c}_0 * (\mathbf{c}_1 + \mathbf{c}_2)) \\ &= 2d_u - 2\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 + \mathbf{c}_0 * \mathbf{c}_2) \\ &= 2[d_u - d_u + 2\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \mathbf{c}_2)] \\ &= 4\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \mathbf{c}_2), \qquad \text{(C.4)}\end{aligned}$$

which shows that if $\mathcal{W}_H(\mathbf{c}_1 * \mathbf{c}_2 * \mathbf{c}_3) = 0$, then $\mathbf{c_0} = \mathbf{c}_1 + \mathbf{c}_2$. For the converse, let $\mathbf{c}_2 = \mathbf{c}_0 + \mathbf{c}_1$. Substituting for $\mathbf{c}_2$, we get

$$\mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 * \mathbf{c}_2) = \mathcal{W}_H(\mathbf{c}_0 * \mathbf{c}_1 + \mathbf{c}_0 * \mathbf{c}_1) = \mathcal{W}_H(\mathbf{0}),$$

which establishes the result.

C.3 Proof of Lemma 4

First, we show that a non-trivial uniform sub-code $\mathbf{C}_u \subset \mathbb{F}_2^n$ exists with distance d_u given in (3.2) and then show that a linear subset $\mathbf{C}_0^F$ can be obtained from this sub-code. Let $M_{n=4k+i}$ denote the cardinality of the uniform set with code-words of length $n = 4k+i$ for various integer values of $k \geq 1$, and $i = 0, 1, 2, 3$. Since Hadamard matrices exist for $n = 1, 2$ and $4k$ [82], there exist uniform codes with distance $d_u = n/2$ with atleast n code-words for these values of n. That is, $M_{4k} \geq n$ and $d_u = n/2$. Moreover, the Hadamard code has the all zero vector as one of its columns. Hence, the Hadamard code can be shortened by 1 bit corresponding to the all zero column, without loss of the distance properties[1] of the code. This implies that $M_{4k+3} \geq n$ and $d_u = \frac{n+1}{2}$. When $d_u = \frac{n+1}{2}$, the Plotkin bound

$$M_{\text{Plotkin}} \leq 2\left\lfloor \frac{d_u}{2d_u - n} \right\rfloor,$$

[1] The Hadamard code of length $n = 4k + 4$ has $4k + 4$ codewords. Even after it is shortened by 1 bit, we have $n = 4k + 4$ codes and the Hamming distance between the codes is not altered, since we shorten only an all-zero column in the code matrix. Hence, $M_{4k+3} \geq n$.

is known to achieve the equality for uniform codes [83], and therefore $M_{4k+3} = n + 1$. To compute a bound on the cardinality for $n = 4k + 1$, consider appending any non-zero column of the Hadamard code for $n = 4k$ to the same code. The appended code has n code-words with code length $(n + 1)$. Since each column of the Hadamard code has $n/2$ non-zero values, $n/2$ code-words of the extended code with $n = 4k + 1$ have $d_u = \frac{n-1}{2}$. Thus, $M_{4k+1} \geq \frac{n}{2}$. To compute the cardinality for $n = 4k+2$, consider the 1 bit shortened code from $n = 4k+3$. From Theorem. 2 in [84], it follows that $M_{4k+2} \geq \left\lceil \frac{d_u M_{4k+3}}{n} \right\rceil = \left\lceil \frac{n+1}{2} \right\rceil$ and $d_u = \frac{n+2}{2}$. For $n \geq 4$, this lower bound is ≥ 2. Thus, we have shown that a uniform sub-code of $\mathbf{C}_u$ exists with even-valued d_u given by (3.2) and that the cardinality of the sub-code is at least 2 for $k \geq 1$.

Now, consider any two non-zero code-words and their sum. This creates a non-trivial uniform code. Moreover, $\mathbf{c}_0$, $\mathbf{c}_1$, $\mathbf{c}_0 + \mathbf{c}_1$ and $\mathbf{0}$ can be used to form $\mathbf{C}_0^F$, which is now a non-trivial linear uniform sub-code of $\mathbb{F}_2^n$ with uniform distance d_u given by (3.2). Therefore, $\mathbf{C}_0^F$ has a cardinality of atleast 4 including the all zero code-word $\mathbf{0}$. Hence, the dimension of the vector space spanned by $\mathbf{C}_0^F$ is at least 2.

C.4 Proof of Theorem 2

We first prove (i). Note that $\mathbf{C}_0$ is linear and its cosets tile $\mathcal{C}$, as $\mathbf{C}_0$ is an intersection of $\mathcal{C}$ and a linear set. Let $\mathbf{C}_0^{\max}$ represent a subset of $\mathcal{C}$ that is both linear and a maximal uniform set, and has the same uniform distance as $\mathbf{C}_0$. Then, there exists a unitary transform between the basis vectors of $\mathbf{C}_0^{\max}$ and $\mathbf{C}_0$. Therefore, without loss of generality, we can transform the code words in $\mathbf{C}_0^{\max}$ such that it forms a superset of $\mathbf{C}_0$ and preserves the uniform distance property. That is, $\mathbf{C}_0 \subseteq \mathbf{C}_0^{\max}$. Then, if $|\mathbf{C}_0| < |\mathbf{C}_0^{\max}|$, there exists atleast one coset of $\mathbf{C}_0$ such that its union with $\mathbf{C}_0$ preserves the uniform distance property. This can be achieved by picking any code word $\mathbf{c} \in \mathbf{C}_0^{\max}, \mathbf{c} \notin \mathbf{C}_0$. Now, $\mathbf{c} + \mathbf{C}_0$ forms a coset of $\mathbf{C}_0$ and the coset belongs to $\mathbf{C}_0^{\max}$ since it is linear. Thus, $\mathbf{C}_0 \cup (\mathbf{C}_0 + \mathbf{c})$ is still a linear uniform set. One can now repeat this procedure of combining the cosets of $\mathbf{C}_0$ to obtain $\mathbf{C}_0^{\max}$.

The bounds on the cardinality of $\mathbf{C}_0$ in (ii) follow from the arguments presented in the proof of Lemma 4. The lower bound follows since $\mathbf{C}_0$ is a non-trivial uniform set. The upper bound follows from the Plotkin bound and using the fact that the number of elements is a power of 2.

To show (iii), it can be seen from (C.1) in the proof of Lemma 2 that

$$\mathcal{W}_H \left(\mathbf{c}_0 + \mathbf{c}_1 + \ldots + \mathbf{c}_{j^*} \right) = \sum_{k=1}^{j^*} 2^k (-1)^{k+1} \binom{j^*}{k} \frac{d_u}{2^k}, \tag{C.5}$$

where the first 2^k arises because of the number of levels of the recursive expansion of the Hamming weight using (C.1), and the 2^k in the denominator arises because of the conditions in (3.3) and (3.4). Since $\sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} = 1$, the above summation equals d_u. This shows that the Hamming weight of the sum of $\mathbf{c}_0$, $\mathbf{c}_1, \ldots, \mathbf{c}_{j^*}$ is d_u. Using a similar procedure, we can show the uniform distance property of any linear combination of the code-words $\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{j^*}$. The cardinality of the set comprising all linear combinations of these $j^* + 1$ vectors is 2^{j^*+1}.

C.5 Proof of Lemma 5

Let $\mathbf{X}_N \in \mathcal{C}^L$ be code sequence of length $N = nL$ generated by a TCB encoder. Let $\mathbf{Y}_N$ be the received code sequence after possible corruption by noise. The signal model for $\mathbf{Y}_N$ is given by $\mathbf{Y}_N = \mathbf{X}_N \oplus \mathbf{Z}_N$, where $\mathbf{Z}_N$ represents the noise sequence and the operator $\oplus$ represents the XOR operation in the BSC and real/complex addition in the AWGN channel.

The MLSD decoder chooses a sequence $\hat{\mathbf{X}}_N$ such that $\hat{\mathbf{X}}_N = \arg\max_{\mathbf{X}_N} \Pr(\mathbf{Y}_N|\mathbf{X}_N)$. For a given channel error probability, the MLSD decoder will pick $\mathbf{X}_N$ which differs from $\mathbf{Y}_N$ in the least number of bit positions (or Euclidean distance for the AWGN channel). Since $\mathbf{X}_N$ is a concatenation of code-words from the parent code, and the noise samples are i.i.d., one can write $\Pr(\mathbf{Y}_N|\mathbf{X}_N) = \prod_{i=1}^{L} \Pr(\mathbf{y}_n^i|\mathbf{x}_n^i)$, where $\mathbf{y}_n^i$ and $\mathbf{x}_n^i$, $1 \leq i \leq L$ are the individual code-words used to construct $\mathbf{Y}_N$ and $\mathbf{X}_N$, respectively.

The TCB decoder employs minimum distance decoding to decide on the individual code-words $\mathbf{x}_n$ once the sequence of sub-code indices are known from the Viterbi decoder. Hence, $\Pr(\mathbf{y}_n^i|\mathbf{x}_n^i)$ are maximized by the TCB decoder for every $n-$tuple. Moreover, note that the Viterbi decoder finds the maximum likelihood sequence of sub-code indices. Hence, we need to only show that the branch metric used for the Viterbi decoder is optimum. Recall that the branch metric used for each transition of the trellis is the smallest distance of the received $n-$tuple $\mathbf{y}_n$ from each of the uniform sub-codes. Let $\mathbf{x}_n \in \mathcal{C}$ be the codeword that is at the least distance from $\mathbf{y}_n$. The minimum distance between $\mathbf{x}_n$ and $\mathbf{y}_n$ is the same as the minimum distance between the closest codeword to $\mathbf{y}_n$ in $\mathbf{C}_i$, the sub-code containing the code-word $\mathbf{x}_n$.

Hence, the chosen branch metric is optimum in the sense of maximum posteriori probability, and the decoder is an MLSD decoder.

Appendix D

Applications of TCBC

In this section, three different applications of TCBC are presented to demonstrate the usefulness of the proposed codes.

D.1 FEC Codes Based on TCBC

The construction of codes for applications such as deep-space missions, storage in magnetic/optical medium, etc require very large minimum distance codes, so that they can operate at very low SNRs and/or the target P_b is very small, of the order 10^{-10}. TCB codes provide a systematic way for constructing codes with such large minimum distance by choosing the right sub-set partitioning from the underlying parent code and a trellis code whose minimum distance is the same or larger than that of the uniform sub-set obtained from the partitioning.

As an example construction, a $\frac{9}{23} = 0.39$ rate TCB $(23q, 9q, 12)$ code is built using the binary Golay $(23, 12, 7)$ code as parent code along with a rate $\frac{6}{9}$ trellis code[1] with $d_{free}^{CC} = 3$ and constraint length $K = 7$ (See Figure D.1). Here, q is an integer > 1 which denotes the number of code-word sequences used to build the long-length code. The asymptotic coding gain from this new Forward Error Correction (FEC) code is $10\log_{10}(\frac{12\times 9}{23}) = 6.7$ dB which is 1.5 dB higher than the coding gain C_L of the parent code. Similarly, one can construct a TCB $(31q, 10q, 16)_2$ code of rate 0.32, using the BCH $(31, 11, 11)$ code as parent code and a rate $\frac{7}{8}$ CC, offering an asymptotic coding gain of $10\log_{10}(\frac{16\times 10}{31}) = 7.12$ dB. Thus, one can construct FECs with a

[1] CC$(9, 6)$ is a special type of CC with feedback shift register where delay of 3 is guaranteed for every input bit. That is, for the input with 1 non-zero element followed by all zeros does not bring the state back to all-zero state, but a specific input in non-zero state can bring the shift register to all-zero state.

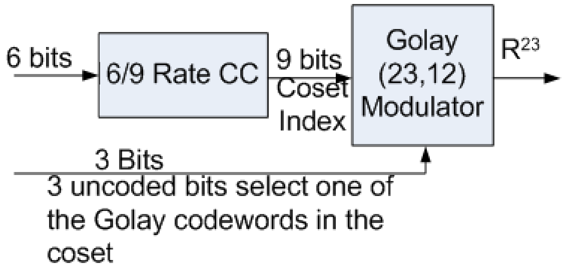

Figure D.1: Encoder structure for TCBC-based FEC Using binary Golay code(23,12,7).

desired minimum distance and rate by uniform sub-set partitioning of the LBC and using a CC with the required constraint length and rate to build the TCBC.

D.2 Low Rate Quasi-Orthogonal TCB Codes

Low rate orthogonal (LRO) codes are used in the uplink of a CDMA system, where they provide tolerance against co-channel interferers. LRO codes have rate of 2^{-K}, where K is the constraint length of the code. They can be considered as Hadamard codes indexed by the contents of a shift register. Recognizing that Hadamard codes are *maximal uniform codes*, LRO codes can be viewed as TCBC with trivial uniform sub-set partitioning (i.e, with 1 element in each sub-code). That is, the contents of a K-bit shift register selects a sub-set index, and the corresponding row of Hadamard matrix is transmitted.

However, one can build a low rate quasi-orthogonal code using the principles of TCBC. We describe such a code construction for illustration, which has a coding gain advantage as well as a rate advantage, when compared to an LRO code of the same length. Consider a $(16, 3.5)_2$ code constructed with quasi-Hadamard code-words by concatenating Hadamard-8 (LRO-8 denoted as $\mathbf{H}_8$) and Hadamard-4 (LRO-4 denoted as $\mathbf{H}_4$) code words as shown below.

$$\mathbf{H}_{16,12} = \left[\begin{array}{c:c} \mathbf{H}_8 & \begin{array}{c}\mathbf{H}_4^T \\ \hline \mathbf{H}_4^T\end{array} \\ \hdashline \begin{array}{c}\mathbf{H}_4^T \\ \hline \mathbf{H}_4^T\end{array} & \mathbf{H}_8 \end{array}\right]$$

The columns of this matrix can be treated as 12 code-words of length 16 and can be partitioned into 4 orthogonal sub-sets with 3 elements in each. This code and the uniform sub-set partitioning can be used to build a $(16q, 2.5q)$ TCBC, which provides a coding gain over the LRO-16 code. For the above calculation of coding gain of TCBC, a rate $\frac{1}{2}$ CC with $d_{free}^{CC} = 3$ is assumed and 1 trit is assumed to be equivalent to 1.5 bits. In the encoder, the output of the CC selects the sub-set index and the 1 trit input selects one of the 3 code-words in the selected sub-set. The coding gain of a LRO-16 code is $10 \log_{10}(8 \times 4/16) = 3$ dB. However, with the proposed quasi-orthogonal code, one can get $10 \log_{10}(16 \times 2.5/16) = 3.97$ dB, i.e, an additional coding gain of 0.97 dB as well as a higher coding rate ($\frac{2.5}{16}$ instead of $\frac{1}{16}$) compared to the LRO-16 code.

D.3 TCBC Based Lattice Codes

Lattice codes are capacity achieving coset codes for the AWGN channel [85], obtained by set partitioning of a lattice into cosets and using a trellis code to select the cosets. Using the Lemma 3 in [43], one can associate a lattice with every linear block code. Here, we illustrate such an 8-dimensional lattice code construction with a TCB $(8q, 3q)$ code that uses an extended Hamming $(8, 4)$ code as parent LBC and the QPSK constellation. The TCBC $(8q, 3q)$ is constructed using a rate $\frac{2}{3}$ CC whose output selects one of the code-word (uniform sub-code) pairs,[2] and one uncoded bit selects one of the code-words in the pair. To construct the lattice code, we need to partition both the lattice as well as the code-words of the extended Hamming code. First, we partition the extended Hamming code into pairs at the maximum distance ($\tilde{d} = 8$) from each other as described above. Second, the lattice is partitioned into 32 sub-lattices with 8 elements each. Since the QPSK constellation is employed (4 consecutive QPSK symbols make 1 lattice point in the $\mathbb{R}^8$ lattice), the total number of lattice points available is $4^4 = 256$. As an illustration, the coset-0 (lattice Λ) in an 8-dimensional real vector space ($\mathbb{R}^8$) is given by $\{(s_0,s_0,s_0,s_0)$, (s_0,s_0,s_3,s_3), (s_0,s_3,s_1,s_1), (s_0,s_3,s_2,s_2), (s_1,s_1,s_1,s_2), (s_1,s_1,s_2,s_1), (s_1,s_2,s_0,s_3), (s_1,s_2,s_3,s_0) $\}$, where the modulation symbol mapping is given by $s_0 = (-1, 1)$, $s_1 = (1, 1)$, $s_2 = (-1, -1)$ and $s_3 = (1, -1)$, corresponding to the QPSK constellation points. It can be verified that the minimum distance of the lattice (coset-0) is 8. Although there are 32 sub-lattices, only 16 of them are employed here for ease of implementation. In the next section, a performance comparison between the conventional (CC-based) lattice code and the proposed TCBC-based

[2] The uniform cosets for the extended Hamming code are given by $\{0, 15\}$, $\{1, 14\}$, $\{2, 13\}$, $\{3, 12\}$, $\{4, 11\}$, $\{5, 10\}$, $\{6, 9\}$ and $\{7, 8\}$.

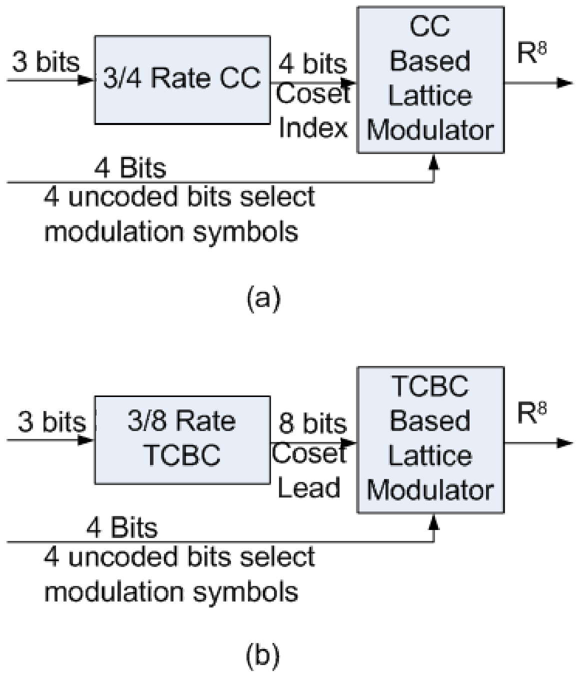

Figure D.2: Encoder structure for (a) CC-based and (b) TCBC-based lattice encoders.

lattice code is shown. The two codes use the same set of sub-lattices and have identical data rates, but it will be seen that the TCBC-based lattice code offers a better *BER* performance.

The conventional lattice code is CC-based, and consists of a rate $\frac{3}{4}$ CC followed by TCM which generates a code with overall rate $\frac{6}{8}$ (see Figure D.2a). Thus, the 8 bits after the CC are mapped to 4 QPSK symbols represented by a point in 8-dimensional lattice. In the above, an alternate construction via a TCBC $(8, 3)$ code was described, which gives out the coset leads and 3 additional bits select the constellation symbols as in the conventional case (see Figure D.2b) (i.e., 8 input bits select the code-word to be transmitted). The reason for the better performance exhibited by the TCBC is that it is driven by a rate $\frac{2}{3}$ CC which offers a higher d_{free}^{CC} than the rate $\frac{3}{4}$ CC in the conventional lattice code.

Appendix E

Statistics of χ_K and χ_K^2 Random variables

E.1 Mean of the Inverse of a χ_K^2 Distributed Random Variable

Consider a central χ^2-distributed random variable with K degrees of freedom, with PDF given by

$$f_X(x) = \frac{x^{\frac{K}{2}-1} e^{\frac{-x}{2}}}{2^{\frac{K}{2}} \Gamma(\frac{K}{2})}, x \geq 0. \tag{E.1}$$

It is straightforward to show that the mean of the inverse of X can be written as

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{K-2}, \text{ for } K > 2. \tag{E.2}$$

When the channel coefficients are circularly symmetric complex Gaussian with zero mean and unit variance, the square of the ℓ_2 norm of the channel is a scaled χ_{2K}^2 random variable. If X denotes the square of the ℓ_2 norm of the channel, the mean of $1/X$ can be shown to be

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{K-1}, \text{ for } K > 1. \tag{E.3}$$

E.2 Mean of the Inverse of the Maximum of Two χ_K^2 Distributed Random Variables

The CDF of the random variable $X \triangleq \max(X_1, X_2)$, where X_i's are χ^2-distributed with K degrees of freedom, can be written as

$$F_X(x) = \left(\frac{1}{\Gamma\left(\frac{K}{2}\right)}\right)^2 \gamma^2\left(\frac{K}{2}, \frac{x}{2}\right), \tag{E.4}$$

where $\gamma(s,x)$ is the lower-incomplete Gamma function [75]. The PDF of X can be obtained by differentiating the above with respect to x, as

$$f_X(x) = \frac{2}{\Gamma\left(\frac{K}{2}\right)}\gamma\left(\frac{K}{2}, \frac{x}{2}\right)\frac{2^{-\frac{K}{2}}x^{\frac{K}{2}-1}e^{-\frac{x}{2}}}{\Gamma\left(\frac{K}{2}\right)}, x \geq 0. \tag{E.5}$$

We expand $\gamma(s,x)$ into an infinite series as

$$\gamma(s,x) = x^s\Gamma(s)e^{-x}\sum_{i=0}^{\infty}\frac{x^i}{\Gamma(s+i-1)}. \tag{E.6}$$

Substituting in (E.5) and taking expectation of $\frac{1}{X}$, it is easy to show that

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{2^{1-K}}{\Gamma\left(\frac{K}{2}\right)}\sum_{i=0}^{\infty}\frac{\Gamma(K-1+i)}{2^i\Gamma\left(\frac{K}{2}+i\right)}, \text{ for } K > 2. \tag{E.7}$$

E.3 Mean of the Inverse of a χ_K Distributed Random Variable

Consider a central χ distributed random variable with K degrees of freedom, with PDF

$$f_X(x) = \frac{1}{\Gamma\left(\frac{K}{2}\right)}2^{1-\frac{K}{2}}x^{K-1}e^{-\frac{x^2}{2}}, x \geq 0. \tag{E.8}$$

The mean of $1/X$ can be written as

$$\begin{aligned}\mathbb{E}\left[\frac{1}{X}\right] &= \frac{2^{1-\frac{K}{2}}}{\Gamma\left(\frac{K}{2}\right)}\int_0^{\infty}x^{K-2}e^{\frac{-x^2}{2}}dx \\ &= \frac{2^{1-\frac{K}{2}}}{\Gamma\left(\frac{K}{2}\right)}\frac{\Gamma\left(\frac{K-1}{2}\right)}{2^{1-\frac{K-1}{2}}} = \frac{\Gamma\left(\frac{K-1}{2}\right)}{\Gamma\left(\frac{K}{2}\right)\sqrt{2}}.\end{aligned} \tag{E.9}$$

When the channel coefficients are circularly symmetric complex Gaussian with zero mean and unit variance, the square of the ℓ_2 norm of the vector is a scaled χ_{2K}^2 random variable. For this scaled random variable, if X denotes the ℓ_2 norm of the channel, we have

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{\Gamma\left(\frac{2K-1}{2}\right)}{\Gamma(K)}, \text{ for } K > 1. \tag{E.10}$$

Bibliography

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Laboratories Internal Tech. Memo*, Jun. 1995.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] "IEEE standard for information technology — telecommunications and information exchange between systems — local and metropolitan area networks — specific requirements — part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," LAN/MAN Standards Committee, New York, NY, USA, Oct. 2009. [Online]. Available: http://standards.ieee.org/about/get/802/802.11.html

[4] D. J. Love, R. W. Heath, Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, pp. 1341–1365, Oct. 2008.

[5] D. J. Love, R. W. Heath, Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[6] C. R. Murthy and B. D. Rao, "Quantization methods for equal gain transmission with finite rate feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 233 –245, Jan. 2007.

[7] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.

[8] "3GPP TS 25.214 Physical Layer Procedures (Release 7)," 3rd Generation Partnership Project Organization, Valbonne, France, 2007. [Online]. Available: http://www.3gpp.org

[9] "3GPP TS 36.213 Physical Layer Procedures (Release 8)," 3rd Generation Partnership Project Organization, Valbonne, France, 2007. [Online]. Available: http://www.3gpp.org

[10] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-Advanced," *IEEE Commun. Mag.*, pp. 127 – 135, Mar. 2013.

[11] K. Zeger and A. Gersho, "Vector quantizer design for memoryless noisy channels," in *Proc. ICC*, Jun. 1988, pp. 1593–1597.

[12] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 799 – 809, Jul. 1990.

[13] K. Zeger and A. Gersho, "Zero redundancy channel coding in vector quantization," *Electronic Letters*, vol. 23, pp. 654–655, Jun. 1987.

[14] J. D. Marca and N. S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multi-dimensional quantizer," in *Proc. ICC*, Jun. 1987, pp. 1128–1132.

[15] N. T. Cheng and N. K. Kingsbury, "Robust zero-error redundancy vector quantization for noisy channels," in *Proc. ICC*, Jun. 1989, pp. 1338–1342.

[16] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Eng. Japan*, vol. 67-B, no. 4, pp. 39–47, 1984.

[17] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 155 – 160, Jan. 1991.

[18] K. Zeger, E. Paksoy, and A. Gersho, "Source/channel coding for vector quantizers by index assignment permutations," in *Proc. IEEE Int. Symp. Inf. Theory*, Jan. 1990, pp. 78–79.

[19] K. Zeger and A. Gersho, "Pseudo gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2156, Dec. 1990.

[20] P. Knagenhjelm and E. Agrell, "The Hadamard transform - a tool for index assignment," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1139 – 1151, Jul. 1996.

[21] V. Cuperman, F. H. Liu, and P. Ho, “Soft decision vector quantization for noisy channels,” in *Proc. of IEEE Workshop on Speech Coding for Telecommunications*, Mar. 1993, pp. 99–100.

[22] M. Skoglund and P. Hedelin, “Hadamard-based soft decoding for vector quantization over noisy channels,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 515–532, Mar. 1999.

[23] J. Zheng and B. D. Rao, “Capacity analysis of MIMO systems using limited feedback transmit precoding schemes,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2886–2901, Jul. 2008.

[24] W. R. Bennett, “Spectra of quantized signals,” *Bell Systems Technical Journal*, vol. 27, pp. 446 – 472, Jul. 1948.

[25] A. Gersho, “Asymptotically optimal block quantization,” *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 373–380, Jul. 1979.

[26] M. Phamdo and F. Alajaji, “Performance of COVQ over AWGN Rayleigh fading channels with soft-decision BPSK modulation,” in *in Proc. of Conf. on Information Science and Systems, Princeton, NJ*, Mar. 1996, pp. 137–142.

[27] W. R. Gardner and B. D. Rao, “Theoretical analysis of the high rate vector quantizer of LPC parameters,” *IEEE Trans. Speech Audio Process.*, vol. 3, no. 5, pp. 367–381, Sep. 1995.

[28] J. Li, N. Chaddha, and R. M. Gray, “Asymptotic performance of vector quantizers with a perceptual distortion measure,” *IEEE Trans. Inf. Theory*, vol. 45, pp. 1082–1091, May 1999.

[29] C. R. Murthy and B. D. Rao, “High rate analysis of source coding for symmetric error channels,” in *Proc. of Data Compression Conf. (DCC)*, Mar. 2006, pp. 163–172.

[30] X. Yu, H. Wang, and E.-H. Yang, “Design and analysis of optimal noisy channel quantization with random index assignment,” *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5796–5804, Nov. 2010.

[31] C. R. Murthy, E. R. Duni, and B. D. Rao, “High-rate vector quantization for noisy channels with applications to wideband speech spectrum compression,” *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5390–5403, Nov. 2011.

[32] S. Rasipuram and C. R. Murthy, “Receive filtering for fading channels,” in *Proc. Nat. Conf. Commun. (NCC)*, Jan. 2010.

[33] G. BenDavid and D. Malah, “Simple adaptation of vector-quantizers to combat channel errors,” in *IEEE DSP Workshop*, Oct. 1994, pp. 41–44.

[34] C. R. Murthy, “Receiver only optimized vector quantization for noisy channels,” in *Proc. of IEEE Intl. Symposium on Personal, Indoor and Mobile Radio Commn. (PIMRC)*, Sep. 2008, pp. 1–5.

[35] S. P. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[36] H. W. Sorensen, *Parameter Estimation: Principles and Problems*, control and systems theory: vol. 9 ed. Mercel Dekker Inc., 1980.

[37] S. Na and D. L. Neuhoff, “Bennett’s integral for vector quantizers,” *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 886–900, Jul. 1995.

[38] J. G. Proakis, *Digital Communications*, 3rd ed. McGraw Hill Inc., 1995.

[39] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, 6th ed. Elsevier Science Publishing Company, 1988.

[40] G. Ungerboeck and I. Csajka, “On improving data-link performance by increasing the channel alphabet and introducing sequence coding,” in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 1976, p. 237.

[41] G. Ungerboeck, “Channel coding with multi-level/phase signals,” *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–77, Jan. 1982.

[42] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Application*, 2nd ed. Pearson-Prentice Hall, 2004.

[43] D. G. Forney, Jr., "Coset codes-Part I: Introduction and geometrical classification," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1123–1151, Sep. 1988.

[44] D. Slepian, "Group codes for the Gaussian channel," *Bell Systems Technical Journal*, pp. 575–602, Apr. 1968.

[45] F. R. Kschishang, P. G. D. Buda, and S. Pasupathy, "Block coset codes for M-ary phase shift keying," *IEEE J. Sel. Areas Commun.*, vol. 7, no. 6, pp. 900–913, Aug. 1989.

[46] P. Delsarte and V. I. Levenshtein, "Association schemes and coding theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2477–2504, Oct. 1998.

[47] A. G. i. Amat, G. Montorsi, and S. Benedetto, "A new approach to the construction of high-rate convolutional codes," *IEEE Commun. Lett.*, vol. 5, no. 11, pp. 453–455, Nov. 2001.

[48] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *Proc. Globecom*, Nov. 1989, pp. 47.1.1–47.1.17.

[49] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[50] R. Venkataramani and T. L. Marzetta, "Reciprocal training and scheduling protocol for MIMO systems," in *Proc. Allerton Conf. on Commun., Control and Comput.*, Monticello, IL, Oct. 2003, pp. 304–304.

[51] T. Dahl, N. Christophersen, and D. Gesbert, "Blind MIMO eigenmode transmission based on the algebraic power method," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2424–2431, 2004.

[52] M. Guillaud, D. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," *Signal Processing and Its Applications*, 2005.

[53] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, 2006.

[54] B. N. Bharath and C. R. Murthy, "Channel training signal design for reciprocal multiple antenna systems with beamforming," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 140 –151, Jan. 2013.

[55] T. Haustein, C. V. Helmolt, E. Jorswieck, V. Jungnickel, and V. Pohl, “Performance of MIMO systems with channel inversion,” in *Proc. VTC*, 2002, pp. 36–40.

[56] S. N. Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, “Great expectations: The value of spatial diversity in wireless networks,” *Proc. IEEE*, vol. 92, no. 2, pp. 219–270, Feb. 2004.

[57] C. Steger, A. Khoshnevis, A. Sabharwal, and B. Aazhang, “The case for transmitter training,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2006, pp. 35–39.

[58] B. N. Bharath and C. R. Murthy, “On the DMT of TDD-SIMO systems with channel-dependent reverse channel training,” *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3332–3341, Nov. 2012.

[59] S. M. Alamouti, “A simple diversity technique for wireless communications,” *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[60] V. Tarokh, H. Jafarkhani, and A. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[61] X. J. Zhang and K. B. Letaief, “Power control and channel training for MIMO channels: A DMT perspective,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 2080–2088, Jul. 2011.

[62] L. Zheng, “Diversity-Multiplexing Tradeoff: A Comprehensive View of Multiple Antenna Systems,” Ph.D. dissertation, Univ. of California, Berkeley, 2002.

[63] V. Aggarwal and A. Sabharwal, “Power controlled feedback and training for two-way MIMO channels,” *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3310–3331, Jul. 2010.

[64] R. T. Derryberry, S. D. Gray, D. M. Ionescu, G. Mandyam, and B. Raghothaman, “Transmit diversity in 3G CDMA systems,” *IEEE Commun. Mag.*, pp. 68–75, Apr. 2002.

[65] V. Sharma, K. Premkumar, and R. N. Swamy, “Exponential diversity achieving spatio-temporal power allocation scheme for fading channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 188–208, Jan. 2008.

[66] X. J. Zhang and Y. Gong, "Impact of CSIT on the tradeoff of diversity and spatial multiplexing in MIMO channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2009, pp. 769–773.

[67] C. Peel, B. Hochwald, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195 – 202, Jan. 2005.

[68] M. Airy, S. Bhadra, R. W. Heath Jr., and S. Shakkottai, "Transmit precoding for the multiple antenna broadcast channel," in *Proc. VTC*, 2006, pp. 1396–1400.

[69] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537 – 544, Mar. 2005.

[70] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.

[71] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[72] A. Boariu and D. M. Ionescu, "A class of nonorthogonal rate-one space-time block codes with controlled interference," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 270–276, Mar. 2003.

[73] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*, 1st ed. Cambridge University Press, 2003.

[74] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, 1st ed. Cambridge University Press, 2005.

[75] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed. Academic Press, Inc, 1994.

[76] A. Lozano, A. M. Tulino, and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[77] S. Jafar and M. Fakhereddin, "Degrees of freedom for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.

[78] J. Park, B. Lee, and B. Shim, “A MMSE vector precoding with block diagonalization for multiuser MIMO downlink,” *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 569–577, 2012.

[79] D. J. Ryan, I. B. Collings, I. V. L. Clarkson, and R. W. Heath Jr., “Performance of vector perturbation multiuser MIMO systems with limited feedback,” *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2633–2644, Sep. 2009.

[80] J. Roh and B. Rao, “Transmit beamforming in multiple-antenna systems with finite rate feedback: A VQ-based approach,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1101 –1112, Mar. 2006.

[81] T. Ganesan and C. R. Murthy, “Receiver only optimized semi-hard decision VQ for noisy channels,” in *Proc. Globecom*, Dec. 2009, pp. 1–6.

[82] J. H. Van Lint and R. M. Wilson, *A Course in Combinatorics*, 2nd ed. Cambridge University Press, 2001.

[83] J. H. Van Lint, *Introduction to Coding Theory*, 3rd ed. Springer, 1998.

[84] A. E. Brouwer, J. B. Shearer, N. J. A. Sloane, and W. D. Smith, “A new table of constant-weight codes,” *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1334–1380, Nov. 1990.

[85] R. Urbanke and B. Rimoldi, “Lattice codes can achieve capacity on AWGN channels,” *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 273–278, Jan. 1998.