

SOURCE-CHANNEL CODING TECHNIQUES
IN THE PRESENCE OF INTERFERENCE AND NOISE

by

AHMAD ABOU SALEH

A thesis submitted to the
Department of Electrical and Computer Engineering
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada

September 2015

Copyright © Ahmad Abou Saleh, 2015

Abstract

As wireless systems proliferate worldwide, interference is becoming one of the main problems for system designers. Interference, which occurs when multiple transmissions take place over a common communication medium, limits system performance. Wireless devices can coordinate the use of scarce radio resources in order to manage the interference and establish successful communication. To effectively deal with the interference problem, some wireless devices must have a certain level of knowledge about the interference. In practice, this knowledge comes at the expense of using more resources (such as employing a proper channel training mechanism). With the remaining available resources, the question is how to achieve reliable communication? This thesis investigates an information theoretic approach and employs several coding techniques to improve system performance by either cancelling the interference or extracting knowledge from it about the information signal.

The first part of this thesis considers the transmission of information signals over a fading channel that is disturbed with additional interference. The system's information theoretic limit in terms of mean square error distortion is assessed. Moreover, hybrid coding schemes are proposed and analyzed to obtain an achievable performance. As an extension to this problem, source channel-state estimation is investigated; in this case, the receiver is interested in estimating the information signal

and the channel-state (interference). It is shown that the achievable performance is close to the derived limit. Motivated by multi-terminal systems, the transmission of a pair of sources over the two-user Gaussian broadcast channel in the presence of interference is then studied. Inner and outer bounds on the system's mean square error distortion are obtained.

Unlike the first part of the thesis, in the second part, two additional constraints are added to the picture: 1) low coding delay and 2) low complexity. Similar communication scenarios as the ones in the first part are investigated under low delay and complexity requirements. These two constraints are motivated by the increased popularity of wireless sensor networks; a sensor node, often conceived as having limited lifetime and processing power, communicates its sensed field information to a fusion centre over a noisy wireless channel. To meet these challenges, we propose to use analog source-channel mappings. Parametric and nonparametric mappings are investigated in this last part of the thesis; it is noted that parametric mappings based on spiral and sawtooth curves are able to outperform linear scheme. Moreover, nonparametric mappings based on joint optimization of the encoder and the decoder are shown to outperform other low delay schemes.

Acknowledgments

I would like to express my sincere appreciation and gratitude to my advisors Prof. Fady Alajaji and Prof. Wai-Yip Geoffrey Chan for their continuous support of my research, for their motivation, encouragement and immense knowledge. This dissertation is the result of their invaluable guidance and cooperation at all times during my Ph.D study. It was both a pleasure and a great experience working with them on this research.

I also acknowledge the efforts of my dissertation reviewers Prof. Steven Blostein and Prof. Tamas Linder, as well as my external examiner Prof. Fabrice Labeau for assessing my research work, and thank them for their comments and suggestions on the dissertation.

Special thanks go to people very close to me whose encouragement have made a big difference on my journey through graduate school. My eternal gratitude go to my parents and my two brothers for their unconditional love and endless support at all times.

Thank you all.

Co-Authorship

A large part of Chapter 3 has appeared in the following papers:

A. Abou Saleh, W.-Y. Chan, and F. Alajaji, “Source-channel coding for fading channels with correlated interference,” *IEEE Trans. Commun.*, vol. 62, pp. 3997-4011, Nov. 2014.

A. Abou Saleh, W.-Y. Chan, and F. Alajaji, “Low and high-delay source-channel coding with bandwidth expansion and correlated interference,” in *Proc. 13th Canadian Workshop on Inform. Theory*, Toronto, Canada, June 2013.

Part of Chapter 4 has appeared in the following papers:

A. Abou Saleh, W.-Y. Chan, and F. Alajaji, “Distortion bounds for broadcasting a Gaussian source in the presence of interference,” in *Proc. 27th Biennial Symp. on Commun.*, Kingston, June 2014.

A. Abou Saleh, F. Alajaji, and W.-Y. Chan, “Hybrid digital-analog coding for interference broadcast channels,” in *Proc. IEEE Int. Symp. on Inform. Theory*, Istanbul, Turkey, July 2013.

Part of Chapter 5 has appeared in the following papers:

A. Abou Saleh, F. Alajaji, and W.-Y. Chan, “Analog coding for Gaussian source and state interference estimation,” in *Proc. IEEE Int. Workshop on Signal Process. Advances in Wireless Commun.*, Stockholm, Sweden, June 2015.

A. Abou Saleh, F. Alajaji, and W.-Y. Chan, “Low-latency source-channel coding for fading channels with correlated interference,” *IEEE Wireless Commun. Letters*, vol. 3, pp. 137-140, Apr. 2014.

A. Abou Saleh, W.-Y. Chan, and F. Alajaji, “Low and high-delay source-channel coding with bandwidth expansion and correlated interference,” in *Proc. 13th Canadian Workshop on Inform. Theory*, Toronto, Canada, June 2013.

A. Abou Saleh, F. Alajaji, and W.-Y. Chan, “Power-constrained bandwidth-reduction source-channel mappings for fading channels,” in *Proc. 26th Biennial Symp. on Commun.*, Kingston, Canada, May 2012.

Contents

Abstract	i
Acknowledgments	iii
Contents	vi
List of Tables	ix
List of Figures	x
Acronyms	xvi
Chapter 1: Introduction	1
1.1 Source and Channel Coding	2
1.1.1 Tandem Coding System	2
1.1.2 Joint Source-Channel Coding	4
1.2 Thesis Overview	5
1.3 Summary of Contributions	7
Chapter 2: Background	9
2.1 General Problem Formulation	9
2.2 Theoretical Limit	11
2.2.1 Bounds on Source-Channel Coding	11
2.2.2 Optimal Performance Theoretically Attainable	13
2.3 Joint Source-Channel Coding Schemes	13
2.3.1 Optimal Linear System	14
2.3.2 Power Constrained Channel-Optimized Vector Quantization	16
2.3.3 Shannon-Kotel'nikov Mappings	18
2.3.4 Hybrid Digital-Analog Systems	21
2.4 $K : 1$ Shannon-Kotel'nikov Mapping	23
2.4.1 Example System using Archimedes' Spiral	25
2.4.2 Numerical Results	28

2.5	Coding Schemes with Side Information	29
2.5.1	Interference Known to the Transmitter	29
2.5.2	Side Information at the Receiver	32
2.6	Broadcast Channels	34
2.6.1	Degraded Broadcast Channels	35
2.6.2	Broadcast Scenario with Two Message Sets	36
2.6.3	Capacity Region for the Degraded Broadcast Channels	37
Chapter 3: Hybrid-Digital Analog Coding for Gaussian Source and State Interference Estimation		40
3.1	Introduction	40
3.2	Problem Formulation	43
3.3	Outer Bounds and Reference Systems	46
3.3.1	Outer Bounds	46
3.3.2	Linear Scheme	51
3.3.3	Tandem Digital Scheme	53
3.4	Distortion Region for the Layered Schemes	55
3.4.1	Scheme 1: Layering Costa and HDA Coding for Bandwidth Expansion	55
3.4.2	Scheme 2: Layering Costa and HDA Coding for Bandwidth Reduction	62
3.4.3	Numerical Results	67
3.5	JSCC for Source-Channel-State Transmission	69
3.5.1	Outer Bound	71
3.5.2	Proposed Hybrid Coding Scheme	74
3.5.3	Numerical Results	78
3.6	Summary and Conclusions	80
Chapter 4: Distortion Bounds for Broadcasting Bivariate Gaussian Sources in the Presence of Interference		82
4.1	Introduction	82
4.2	Problem Formulation	83
4.3	Outer Bounds and Reference Schemes	86
4.3.1	Outer Bound for $\rho_{V_1V_2} \neq 1$	86
4.3.2	Outer Bound for $\rho_{V_1V_2} = 1$	88
4.3.3	Linear Scheme	91
4.3.4	Tandem Digital Scheme	92
4.4	HDA Coding Schemes	93
4.4.1	HDA Scheme 1 for Matched Bandwidth	93
4.4.2	HDA Scheme 2 for Bandwidth Expansion	97

4.5	Numerical Results	99
4.5.1	Case of $\rho_{V_1V_2} \neq 1$	100
4.5.2	Case of $\rho_{V_1V_2} = 1$	103
4.6	Summary and Conclusions	105
Chapter 5: Low Delay Analog Source-Channel Coding		107
5.1	Introduction	107
5.2	Low Delay Coding for Fading Channels	109
5.2.1	Problem Formulation	109
5.2.2	Preliminaries: Theoretical Bound and Linear System	110
5.2.3	Parametric Mappings	112
5.2.4	Nonparametric Mappings	117
5.2.5	Numerical Results	122
5.3	Low Delay Coding for Fading Channels in the Presence of Correlated Interference	127
5.3.1	Problem Formulation and Main Contributions	127
5.3.2	Optimality of the Uncoded Scheme	128
5.3.3	Design Algorithm for Low Delay Coding	129
5.3.4	Numerical Results and Discussion	133
5.4	Low Delay Coding for Source and State Interference Estimation	136
5.4.1	Problem Formulation and Main Contributions	136
5.4.2	Distortion Lower Bound	137
5.4.3	Parametric Mapping	137
5.4.4	Nonparametric Mapping	142
5.4.5	Numerical Results	144
5.5	Summary and Conclusions	148
Chapter 6: Conclusions		149
6.1	Summary and Conclusions	149
6.2	Suggestions for Further Research	152
Bibliography		153

List of Tables

5.1	SDR in dB versus ρ_{VS} for $P = 0.6$, $\sigma_S^2 = 1$, $\sigma_W^2 = 0.01$ and $r = 1$. . .	134
-----	---	-----

List of Figures

1.1	Block diagram of the tandem source-channel coding system.	3
2.1	A general point-to-point communication system.	10
2.2	A point-to-point communication system based on the separation theorem for AWGN channel.	11
2.3	Performance of the optimal linear system for different dimension expansion/reduction ratios, $\sigma_V^2 = 1$ and $P = 1$. The theoretical limit (OPTA) is also plotted for comparison.	15
2.4	The reconstruction codebook structure of size 512 for a 2 : 1 PCCOVQ system. The system is designed for (a) CSNR = 0 dB and (b) CSNR = 20 dB. The graph is made for an i.i.d. Gaussian source with unit variance.	18
2.5	Shannon's example on 1 : 2 dimension expansion mapping [59].	20
2.6	A block diagram of a general hybrid digital-analog system. The digital part uses a tandem scheme and the analog part uses a linear transmission.	22
2.7	Two dimensional spiral mapping.	26
2.8	Performance of 2 : 1 bandwidth reduction using spiral mapping for a Gaussian source vector with standard deviation $\sigma_V = 1$. The optimal linear system (BPAM) and the theoretical limit (OPTA) of the system are also plotted.	28

2.9	Block diagram for the dirty paper coding problem.	29
2.10	The Wyner-Ziv problem.	32
2.11	General broadcast channel.	34
2.12	Physically degraded memoryless broadcast channel.	35
2.13	Broadcast channel with two message sets.	36
2.14	Illustration of clouds and satellite codewords used in superposition coding.	38
3.1	A $K : N$ system structure over a fading channel with interference known at the transmitter side. The interference $S^{\max(K,N)}$ is assumed to be the output of a noisy side channel with input $V^{\max(K,N)}$. V^K represents the first K samples of $V^{\max(K,N)}$ (S^N is defined similarly). The fading coefficient is assumed to be known at the receiver side; the transmitter side, however, knows the fading distribution only.	43
3.2	Scheme 1 (bandwidth expansion) encoder structure.	56
3.3	Scheme 1 (bandwidth expansion) decoder structure.	59
3.4	Scheme 2 (bandwidth reduction) encoder structure.	63
3.5	Scheme 2 (bandwidth reduction) decoder structure.	64
3.6	Performance of HDA Scheme 1 ($r = 2$) over the AWGN channel under matched noise levels for different correlation coefficient values, $P = 1$, $\sigma_S^2 = 1$ and CSNR=10 dB (as given by (3.28)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.4), respectively.	68

3.7	Performance of Scheme 1 ($r = 1$) over the fading channel under matched noise levels for different correlation coefficient, $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB and $\mathbb{E}[F^2] = 1$ (as given by (3.28)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.4), respectively. . . .	69
3.8	Performance of Scheme 1 ($r = 2$) over the fading channel under matched noise levels for different correlation coefficient ρ_{VS} , $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB and $\mathbb{E}[F^2] = 1$ (as given by (3.28)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.4), respectively.	70
3.9	Performance of Scheme 2 ($r = 1/2$) over the fading channel under matched noise levels for different ρ_{VS} , $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB and $\mathbb{E}[F^2] = 1$ as given by (3.37)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.6), respectively.	71
3.10	Performance of Scheme 1 ($r = 1$) over the fading channel under mismatched noise levels for $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB, $\rho_{VS} = 0.7$ and $\mathbb{E}[F^2] = 1$	72
3.11	Encoder structure for source-channel-state transmission.	75
3.12	Decoder structure for source-channel-state transmission.	77
3.13	Distortion region for hybrid coding scheme over the AWGN channel for $\sigma_V^2 = 1$, $P = 1$, $\sigma_S^2 = 0.5$ and $\rho_{VS} = 0.8$	80
3.14	Distortion region for hybrid coding scheme over the fading channel for $\sigma_V^2 = 1$, $P = 1$, $\sigma_S^2 = 1$, $\rho_{VS} = 0.8$ and $\mathbb{E}[F^2] = 1$	81
4.1	System model structure.	84
4.2	HDA Scheme 1 encoder structure for rate $r = \frac{N}{K} = 1$	94

4.3	HDA Scheme 2 encoder structure for rate $r = \frac{N}{K} > 1$	98
4.4	HDA Scheme encoder structure for rate $r = \frac{N}{K} \geq 1$ and $\rho_{V_1V_2} = 1$. . .	100
4.5	Distortion regions for HDA Scheme 1 for $r = 1$ as given by (4.23) and (4.26). Distortion regions for the outer bound (given by (4.4)), tandem and linear schemes (given by (4.17)) are also plotted.	101
4.6	Distortion region for HDA Scheme 1 for $\rho_{V_1V_2} = 0.5$, $\rho_{V_1S} = \rho_{V_2S} = 0.2$ and $r = 1$ as given by (4.23) and (4.26). Distortion regions for the outer bound (given by (4.4)), tandem and linear schemes (given by (4.17)) are also plotted.	102
4.7	Distortion regions for HDA Scheme 2 for $r = 2$ as given by (4.27) and (4.28). Distortion regions for the outer bound (given by (4.4)) and tandem scheme are also plotted.	103
4.8	Distortion region for the HDA scheme for $r = 1$ as given by (4.23) and (4.26). The outer bound (given by (4.8)) is also plotted.	104
4.9	Distortion region for the HDA scheme for $r = 2$ and $\rho_{V_1S} = 0.2$ as given by (4.27) and (4.28). The outer bound (given by (4.8)) is also plotted.	105
4.10	Distortion region for the HDA scheme for $r = 2$ and $\rho_{V_1S} = 0.8$ as given by (4.27) and (4.28). The outer bound (given by (4.8)) is also plotted.	106
5.1	System structure for $K : 1$ bandwidth reduction over a fading channel with AWGN. CSI is only available at the decoder side, while the encoder knows only the distribution of the fading.	109

5.2	Performance of the parametric and the nonparametric mappings for 2:1 bandwidth reduction over Rayleigh fading channel. The performance of the linear system and the theoretical limit (OPTA) with CSI at the decoder are also included. Note that for FSCI, the performance of the linear system and the theoretical limit improves over the DCSI case by at most 0.2 dB in SDR.	123
5.3	Performance of the parametric and the nonparametric mappings for 2:1 bandwidth reduction over Rayleigh fading channel. The performance of the linear system is also included. This figure is for no CSI at both encoder and decoder.	124
5.4	Performance of the parametric and the nonparametric mappings with 3 : 1 bandwidth reduction over Rayleigh fading channel. The performance of the theoretical limit and the linear system are also plotted.	125
5.5	Performance of the parametric and the nonparametric mappings with 4 : 1 bandwidth reduction over Rayleigh fading channel. The performance of the theoretical limit and the linear system are also plotted.	126
5.6	Decoded pairs for 2 : 1 bandwidth reduction, and fading gain f : (a) 0.3, (b) 0.6, and (c) 0.9. This graph is made for CSNR = 30 dB and DCSI. The lines that are drawn between asterisk points correspond to neighbours in the one dimensional mapped source (i.e., channel space).	127
5.7	SDR Performance versus CSNR for $P = 1$, $\sigma_S^2 = 1$, $\rho_{VS} = 0.3$	135
5.8	System model for joint source and state interference estimation.	136
5.9	Performance versus source-interference correlation ρ_{VS} , for $\sigma_V^2 = \sigma_S^2 = 1$ and CSNR = 25 dB.	144

5.10	Performance versus CSNR levels, for $\rho_{VS} = 0.5$ and $\sigma_V^2 = \sigma_S^2 = 1$. . .	145
5.11	Encoder (left) and its corresponding decoder (right) mappings optimized using the proposed algorithm for CSNR = 25, $\rho_{VS} = 0$, $\sigma_V = \sigma_S = 1$ and $P = 1$; parametric mapping is used for the initialization of the algorithm. In the figure to the right, the asterisks show the reconstructed (\hat{V}, \hat{S}) and the small dots are samples from the distribution of (V, S)	146
5.12	Encoder (left) and its corresponding decoder (right) mappings optimized using the proposed algorithm for CSNR = 25, $\rho_{VS} = 0.7$, $\sigma_V = \sigma_S = 1$ and $P = 1$; parametric mapping is used for the algorithm initialization. In the figure to the right, the asterisks show the reconstructed (\hat{V}, \hat{S}) and the small dots are samples from the distribution of (V, S)	147

Acronyms

AWGN	Additive White Gaussian Noise
BPAM	Block Pulse Amplitude Modulation
COVQ	Channel-Optimized Vector Quantizer
CSI	Channel State Information
CSNR	Channel Signal-to-Noise Ratio
DCSI	Decoder Channel State Information
FCSI	Full Channel State Information
JSCC	Joint Source-Channel Coding
HDA	Hybrid Digital-Analog
HSQLC	Hybrid Scalar Quantizer Linear Coder
i.i.d.	Independent and Identically Distributed
KKT	Karush-Kuhn-Tucker
LMMSE	Linear Minimum Mean Square Error
MAC	Multiple Access Channel
MC	Monte Carlo
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error

OPTA	Optimal Performance Theoretically Attainable
PAM	Pulse Amplitude Modulation
PCCOVQ	Power Constrained Channel-Optimized Vector Quantizer
PDF	Probability Density Function
PMF	Probability Mass Function
SDR	Signal-to-Distortion Ratio
VQ	Vector-Quantizer
WSNs	Wireless Sensor Networks

Chapter 1

Introduction

Telecommunication is an increasingly important part of modern societies. Not only it enables the transportation of information in a short time without being hindered by distance, but it also orients ourselves to almost every point in our globe. One of the ultimate goals in modern communication systems is to provide reliable, robust and efficient transmission of information bearing signals, such as text, images, video, speech, over a noisy medium. Such signals have a high information rate, while the medium has a limited capacity. Various theories and systems have been developed to achieve the ultimate goal of communication and to accommodate the increasing demand for high data rates and more link reliability.

Source coding (source compression) and channel coding (efficient use of the channel) were subjected to extensive research over the last decades. Traditionally, source and channel coding have been designed separately, resulting in what is called a tandem system. Despite great achieved progress for point-to-point communication systems, particularly vis-a-vis separate source and channel coding, it is not always possible to achieve the desired quality of the received signal for certain channels and for low coding delay and complexity constraints; moreover, using a tandem scheme is not

optimal for multi-terminal systems (e.g., broadcast channels). The combination of source and channel coding may achieve better performance under limited resources and for multi-terminal systems. Thus, joint source-channel coding (JSCC) is a topic that has been receiving increasing attention. Over the years, many researchers have developed JSCC techniques that show benefits in terms of performance improvement and robustness to variations in channel over separate coding for a given channel condition and fixed complexity and delay constraints.

In this thesis, we investigate two different joint source-channel coding methods for the reliable and efficient transmission of analog-valued sources over noisy channels in the presence of interference. The first one uses a high delay hybrid digital-analog coding technique, and the second one is based on low delay analog coding, referred to as Shannon-Kotel'nikov mapping.

1.1 Source and Channel Coding

1.1.1 Tandem Coding System

In a typical communication system, signals are often modelled by a discrete-time continuous-amplitude (analog) random sequences. Due to restrictions on bandwidth and storage, a source encoder is usually employed on the source sequence to compress it and remove its redundancy. This operation, which is referred to as *source coding*, leads to a loss of information. This inevitable loss of information from source coding may introduce a greater level of sensitivity to channel noise. Therefore, a channel encoder may be required to add some controlled redundancy at the output of the source encoder to enable detecting and correcting errors at the receiver side. This operation is often called *channel coding*. The channel encoder output is then modulated and

transmitted over the waveform channel. To get an estimate of the source sequence, channel and source decoders are applied on the noisy received channel output. This communication system, that consists of separately designed source and channel coders as shown in Fig. 1.1, is often called a *tandem coding system*.

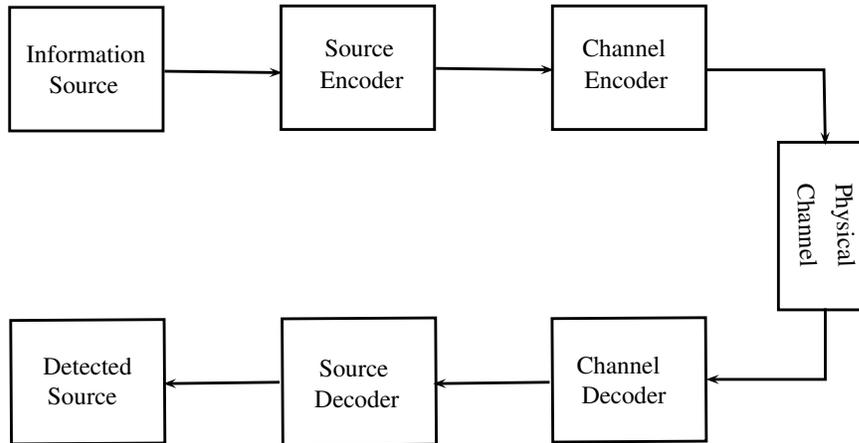


Figure 1.1: Block diagram of the tandem source-channel coding system.

According to Shannon's source coding theorem, the rate distortion function $R(D)$ is the minimum rate that is sufficient and needed to represent the source samples with an average distortion not exceeding D . According to Shannon's channel coding theorem, the capacity-cost function $C(P)$ is the maximum rate at which information can be reliably transmitted given an average power constraint P on the channel input. As a result, it is impossible to communicate at a rate above channel capacity. Combining Shannon's source and channel coding theorems, it is possible to obtain a source sequence reconstruction with fidelity D if $R(D) \leq C$. For point-to-point communication, a source signal can be transmitted (optimally) with fidelity D via a tandem coding scheme if $R(D) \leq C$. This result is known as the *source-channel separation theorem* [13, 58].

1.1.2 Joint Source-Channel Coding

Many available practical communication systems are based on Shannon's source-channel separation principle. There are, however, a few problems with tandem systems that have motivated researchers to investigate joint source-channel coding. In addition to the fact that tandem schemes are not optimal for multi-terminal systems, one main problem that is worth mentioning about the *source-channel separation theorem* is that the coders must have unlimited delay and complexity in order to achieve optimality. This means that, in practice, for fixed delay and complexity constrained applications, the tandem system may not be optimal. Another problem is that source and channel codes are designed separately. More precisely, source codes are designed assuming that the channel codes can correct all errors introduced by the channel noise; and channel codes are designed to protect all source codes equally assuming information is uniformly distributed in these codes. These assumptions are not true and unequal error protection with source codes can indeed result in better performance. Another important drawback is that the separate source-channel system is highly non-robust to mismatch in noise level between the transmitter and the receiver. More specifically, if the true channel noise is higher than the design channel noise, the performance degrades drastically. This is often called the *threshold effect* [59] and is due to the quantizer's sensitivity to the channel errors and the eventual breakdown of the error correcting code at high noise levels. Furthermore, if the actual channel noise falls below the design one, the performance remains constant beyond a certain threshold. This is often referred to as the *leveling-off effect* [59] and is due to the non-recoverable distortion introduced by the quantizer.

Examples of joint source-channel coding techniques include: (a) unequal error

protection where the idea is to use several channel codes to protect source information according to its level of importance [36, 51]; (b) optimal quantizer design for noisy channels such as channel-optimized vector quantization (COVQ) [19, 43, 44]; (c) direct source-channel mapping [11, 55, 59, 66].

Although most digital JSCC techniques perform fairly well in terms of reconstruction quality, coding delay and complexity over tandem systems when the channel condition falls below the design parameter, they usually fail to improve their performance as the channel condition enhances (levelling-off effect). Moreover, digital JSCC schemes may not be optimal over multi-terminal systems. This leads us to investigate the advantage of (1) hybrid digital-analog coding systems which combines digital and analog schemes and (2) purely analog systems (i.e., Shannon-Kotel'nikov mappings) that are based on a direct source-channel mapping approach, to achieve a graceful performance over a wide range of channel conditions and obtain a better performance over multi-terminal communication scenarios.

1.2 Thesis Overview

Chapter 2 describes the general point-to-point communication system. The theoretical limit of such system is then presented. Various joint source-channel coding techniques for point-to-point systems including optimal linear systems, power constraint channel-optimized vector quantization, Shannon-Kotel'nikov mappings, and hybrid digital-analog systems are reviewed. Furthermore, a detailed description of $K : 1$ Shannon-Kotel'nikov mappings and $2 : 1$ bandwidth reduction systems based on the double Archimedes' spiral function are presented. Several coding schemes for Gaussian channels with side information at the transmitter and the receiver are

described. For multi-terminal systems, a brief description of the broadcast channel and superposition coding are included; this coding technique is able to achieve the capacity of the degraded broadcast channel.

Chapter 3 considers the problem of sending a Gaussian source over a fading channel with Gaussian interference known non-causally to the transmitter. Joint source-channel coding schemes for the case of unequal bandwidth between the source and the channel and when the source and the interference are correlated are studied. An outer (lower) bound on the system's distortion is first derived by assuming additional information at the decoder side. A layered coding scheme is then proposed based on proper combination of power splitting, bandwidth splitting, Wyner-Ziv and hybrid coding. More precisely, a hybrid layer, that uses the source and the interference, is concatenated (superimposed) with a purely digital layer to achieve bandwidth expansion (reduction). The achievable (square error) distortion regions of these schemes (upper or inner bounds) under matched and mismatched noise levels are then analyzed. The proposed schemes are shown to perform close to the best derived bound and to be resilient to channel noise mismatch. As an application of the proposed schemes, both inner and outer bounds on the source-channel-state distortion region are derived for the fading channel with correlated interference; the receiver, in this case, aims to jointly estimate both the source signal as well as the channel-state (interference).

Chapter 4 studies the transmission of bivariate Gaussian sources over the two-user Gaussian degraded broadcast channel in the presence of interference that is correlated to the source and known non-causally at the transmitter. Each user is interested in estimating one of the sources. Hybrid digital-analog schemes are studied and their

achievable (square-error) distortion regions under matched and expansion bandwidth regimes are analyzed. These schemes, which use the idea of superposition coding for broadcast channels, require proper combinations of power splitting, bandwidth splitting, rate splitting, Wyner-Ziv and Costa coding. An outer bound on the distortion region is also derived by assuming knowledge of one of the sources at the stronger user and full/partial knowledge of the interference at both users. The proposed schemes are shown to outperform other reference schemes and behave close to the derived outer bound for certain system settings.

The fifth chapter examines low delay analog source-channel coding schemes for different noisy channels in the presence of interference. Parametric mappings based mainly on spiral and sawtooth curves are studied. Nonparametric mappings are also designed in this chapter. This is done by first deriving the necessary conditions for optimality and then proposing an iterative algorithm based on joint optimization between the encoder and the decoder. A reduced-complexity approach for the implementation of the design algorithm is also presented. The proposed (nonlinear) mappings are shown to outperform linear scheme and give a graceful performance over wide range of noise levels. Moreover, these nonlinear mappings, which are shown to fit well the channel space, overcome the inevitable saturation effect of linear scheme.

The sixth and final chapter contains a compendium of the principal results presented in the thesis.

1.3 Summary of Contributions

The main contributions of the thesis are briefly summarized as follows:

- Inner bounds on the system's distortion are found for fading channels in the

presence of interference that is correlated to the source and known non-causally to the transmitter; this is done by proposing hybrid schemes based on Costa and Wyner-Ziv coding. As an application, the proposed hybrid schemes are extended to the case of joint source state-interference estimation. Outer bounds on the system's distortion region are derived for the above mentioned communication scenarios. Such bounds are obtained by assuming additional knowledge of the interference at the decoder side.

- Derived the distortion regions for transmitting bivariate Gaussian sources over two-user Gaussian degraded broadcast channel in the presence of interference; the interference is correlated with the source and known non-causally to the transmitter. Hybrid schemes that use superposition coding technique are employed to ensure reliable transmission over the degraded broadcast channel.
- Developed a reliable sensor-communication system based on low delay analog coding; transmission over noisy medium in the presence of interference is tackled. Parametric mappings are proposed to accommodate the low delay coding constraint in sensor networks.
- To improve the performance whenever storage and offline design complexity are not an issue, low delay nonparametric mappings are designed through an iterative process based on joint optimization between the encoder and the decoder using the necessary conditions for optimality.

Chapter 2

Background

2.1 General Problem Formulation

We consider the problem of transmitting a discrete-time, continuous-amplitude source over a memoryless discrete-time, continuous-amplitude channel. We assume that the channel symbols are corrupted by additive/multiplicative noise. As shown in Fig. 2.1, the source vector $V^K = (V(1), \dots, V(K))^T \in \mathbb{R}^K$, which is composed of independent and identically distributed (i.i.d.) K samples, is transformed into an N dimensional channel input $X^N = \alpha(V^K) \in \mathbb{R}^N$ using a nonlinear mapping, in general, $\alpha(\cdot) : \mathbb{R}^K \rightarrow \mathbb{R}^N$, where $(\cdot)^T$ denotes the transpose operator. The system operates under an average transmission power constraint P

$$\frac{1}{N} \mathbb{E}[\|X^N\|^2] = \frac{1}{N} \int \|x^n\|^2 p(x^n) dx^n \leq P \quad (2.1)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and $p(x^n)$ is the probability density function (pdf) of the channel input X^N .

After transmission over the noisy channel, the received signal (channel output) Y^N is decoded to produce an estimate of the source vector $\hat{V}^K = \gamma(Y^N)$, where the

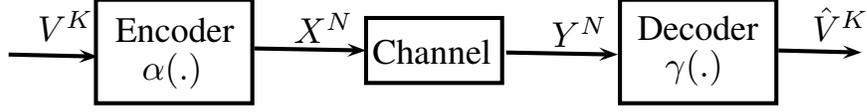


Figure 2.1: A general point-to-point communication system.

decoder is a mapping from $\mathbb{R}^N \rightarrow \mathbb{R}^K$. The aim in such communication system is to reconstruct the estimate source vector \hat{V}^K with some fidelity criterion. The distortion measure that is considered in this thesis is the mean square error (MSE) defined as follows

$$d(V^K, \hat{V}^K) = \frac{1}{K} \|V^K - \hat{V}^K\|^2. \quad (2.2)$$

Hence, the general system design can be formulated by finding the encoder and the decoder that minimize the average MSE distortion given by

$$D = \frac{1}{K} \mathbb{E}[d(V^K, \hat{V}^K)] = \frac{1}{K} \iint \|v^k - \hat{v}^k\|^2 p(v^k, \hat{v}^k) dv^k d\hat{v}^k \quad (2.3)$$

where $p(v^k, \hat{v}^k)$ is the joint pdf of the source vector and the reconstructed one. The rate of the system is given by $r = \frac{N}{K}$ channel use/source symbol. When $r = 1$, the system has a matched bandwidth between the source and the channel. For $r < 1$, the system performs bandwidth reduction; for $r > 1$, the system performs bandwidth expansion. Note that throughout the thesis, when all samples in a vector are i.i.d., we drop the indexing when referring to a sample in a vector (i.e., $X(i) = X$). Particular realizations of a random variable are written in corresponding lower case letters.

2.2 Theoretical Limit

In this section, we present the theoretical performance limit of the point-to-point communication problem. We limit our discussion to the theoretical limit on the system's distortion for the case of a memoryless Gaussian source and an additive white Gaussian noise (AWGN) channel. The definition of the bound that is developed in this section, is based on the fundamental source and channel separation theorem introduced by Shannon [58].

2.2.1 Bounds on Source-Channel Coding

As shown from the block diagram of a tandem source-channel coding system in Fig. 2.2, a source encoder that compresses the source information and a channel encoder that protects the source information from channel noise by adding some controlled redundancy are applied to the source vector V^K . This operation produces the

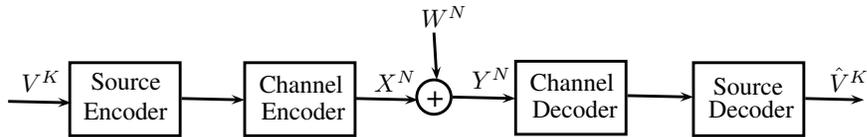


Figure 2.2: A point-to-point communication system based on the separation theorem for AWGN channel.

channel input X^N that is transmitted over an AWGN channel. The received noisy signal can be written as follows

$$Y^N = X^N + W^N \quad (2.4)$$

where addition is component-wise and each sample in the additive noise W^N is drawn from a Gaussian distribution with zero mean and variance σ_W^2 ($W \sim \mathcal{N}(0, \sigma_W^2)$) independently from the source. Using the noisy received signal Y^N , we estimate the information source using a channel and a source decoder.

Given a fidelity criterion, an absolute lower bound on the rate of the source was derived by Shannon. This is known as the rate-distortion function. For a memory-less i.i.d. Gaussian source V under the MSE distortion measure, the rate-distortion function is given by [13]

$$R(D) = \max \left[\frac{1}{2} \log \left(\frac{\sigma_V^2}{D} \right), 0 \right] \quad (\text{bits/source symbol}) \quad (2.5)$$

where σ_V^2 is the variance of the source signal and D is the distortion.

Information from the source encoder is transmitted over a noisy medium (channel) at a certain rate. This rate is governed by the capacity of the channel which represents an upper limit on the rate at which information can be transmitted reliably (with probability of error that asymptotically vanishes with respect to the coding block length) given a certain channel signal-to-noise ratio (CSNR) [58]. For the AWGN channel with an average transmission power constraint P on the channel input, the capacity-cost function can be expressed as follows [13]

$$C(P) = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_W^2} \right) \quad (\text{bits/channel use}) \quad (2.6)$$

where the ratio P/σ_W^2 is the CSNR.

2.2.2 Optimal Performance Theoretically Attainable

Finally, to determine the optimal achievable performance for the given communication system, we combine the source and the channel coding theorems; a memoryless source can be reproduced with system rate r and distortion at most D at the receiving end of a memoryless channel of capacity $C(P) > (1/r)R(D)$. Conversely, the distortion D is unattainable if $C(P) < (1/r)R(D)$. Hence, for the memoryless Gaussian source-channel pair, the optimal performance theoretically attainable (OPTA) can be found by equating the source rate distortion to the channel capacity times the system rate ($R(D) = rC(P)$). The OPTA in term of signal-to-distortion ratio (SDR) can then be expressed as follows

$$\text{SDR} \triangleq \frac{\sigma_V^2}{D} = \left(1 + \frac{P}{\sigma_W^2}\right)^r. \quad (2.7)$$

When the system rate $r > 1$, redundant dimensions are available and error correcting codes can be used; when $r < 1$, however, the source has to be compressed to lower its dimension before transmission over noisy channel. Note that this performance can be achieved using a tandem source-channel coding system.

2.3 Joint Source-Channel Coding Schemes

In this section, we present several JSCC techniques based on digital and analog signalling. Let us first note that in order to implement a mapping from a source sequence with symbol rate B_S to a channel sequence with symbol rate B_C , we use the block-based approach as shown previously in Fig. 2.2. In such system, samples of the memoryless source V are grouped into blocks of size K to form the source vector V^K which is encoded using $\alpha(\cdot)$. The encoding process maps V^K to a channel input

vector X^N of length N . By having $N/K = B_C/B_S$, bandwidth expansion/reduction is obtained by mapping the K source samples into N channel samples. Note that bandwidth expansion/reduction and dimension expansion/reduction are used interchangeably throughout this thesis. In the rest of this section, we focus more on JSCC techniques for Gaussian channels.

2.3.1 Optimal Linear System

Block pulse amplitude modulation (BPAM) is the optimal (in the mean square error sense) linear system for transmitting a vector source on a vector channel with additive noise [47]. Consider a Gaussian source vector V^K to be transmitted over a memoryless AWGN channel with average power constraint P . In [47], it was shown that for bandwidth expansion ($r > 1$), the optimal linear encoder and decoder that minimize the MSE distortion are given as follows

$$\begin{aligned} X^N &= \alpha(V^K) = \frac{\sqrt{rP}}{\sigma_V} \mathbf{I}_{N \times K} V^K \\ \hat{V}^K &= \gamma(Y^N) = \frac{\sqrt{\sigma_V^2 P/r}}{P + \sigma_W^2} \mathbf{I}_{K \times N} Y^N \end{aligned} \quad (2.8)$$

where $Y^N = X^N + W^N$ is the received signal and $\mathbf{I}_{N \times K}$ is a $N \times K$ matrix with ones on the main diagonal and zeros elsewhere.

On the other hand, for bandwidth reduction system ($r < 1$), the optimal linear encoder and decoder can be expressed as follows

$$\begin{aligned} X^N &= \alpha(V^K) = \frac{\sqrt{P}}{\sigma_V} \mathbf{I}_{N \times K} V^K \\ \hat{V}^K &= \gamma(Y^N) = \frac{\sqrt{\sigma_V^2 P}}{P + \sigma_W^2} \mathbf{I}_{K \times N} Y^N. \end{aligned} \quad (2.9)$$

In effect, the optimal linear system inserts zero samples when more channel bandwidth is available ($N > K$) and removes source samples when channel bandwidth is limited ($N < K$). Fig. 2.3 shows the performance of the optimal linear system for a 1 : 2

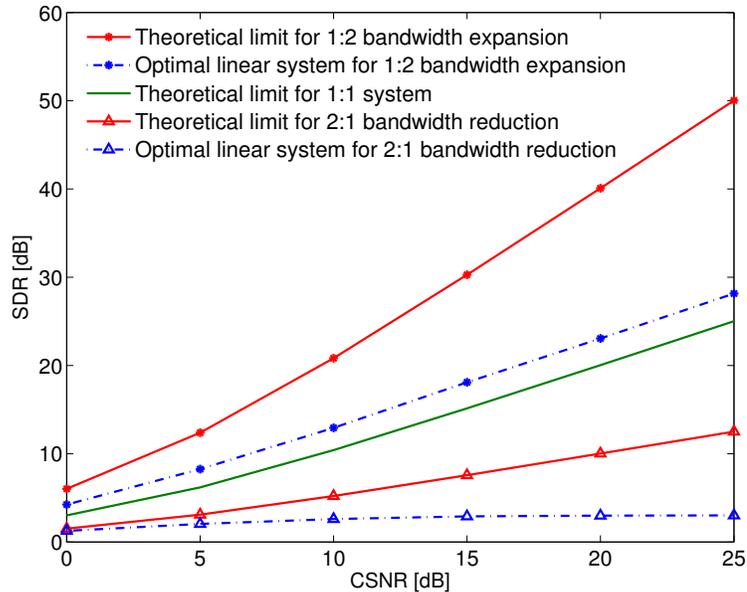


Figure 2.3: Performance of the optimal linear system for different dimension expansion/reduction ratios, $\sigma_V^2 = 1$ and $P = 1$. The theoretical limit (OPTA) is also plotted for comparison.

bandwidth expansion ratio (i.e., $K = 1$, $N = 2$) and 2 : 1 bandwidth reduction. The 1 : 1 theoretical limit is also shown for comparison. The plot is made for a memoryless Gaussian source V with unit variance and Gaussian channel with average power constraint $P = 1$. From Fig. 2.3, we can notice that under dimension reduction, the performance of the linear scheme saturates after certain CSNR level. This can be explained in a similar way as the levelling-off effect in a purely digital system. One can also notice that the linear system behaves very close to the optimal performance for low CSNRs; however as CSNR increases, the linear system is far from optimal.

Since the linear system performs well only for very poor channels which is usually not of interest, this has motivated the investigation of nonlinear systems which can achieve better performance.

2.3.2 Power Constrained Channel-Optimized Vector Quantization

In [23, 24], the authors propose to use a power constrained channel-optimized vector quantization (PCCOVQ) in order to close the gap between the linear scheme and the optimal performance. Here again, the objective is to find the encoder α and the decoder γ that minimize the average MSE distortion

$$D(\alpha, \gamma) = \mathbb{E}[\|V^K - \hat{V}^K\|^2]/K \quad (2.10)$$

subject to an average transmission power $P(\alpha) = \mathbb{E}[\|\alpha(V^K)\|^2]/N \leq P$. Using the Lagrangian method, the constrained optimization problem can be recast into an unconstrained minimization problem given by

$$\min_{\alpha, \gamma} [D(\alpha, \gamma) + \lambda P(\alpha)] \quad (2.11)$$

where λ is the Lagrange multiplier used to control the consumed average power. If for a given λ , we solve the unconstrained problem and find that the power constraint is fulfilled, the solution we have obtained is also a solution to the constrained optimization problem [30]. It is worth mentioning that this problem is related to the COVQ design problem but with an additional constraint on the transmission power.

To solve this problem, the authors in [23, 24] propose to use a vector quantizer followed by a mapping from the source space to a finite channel space set at the

encoder side. The channel signal set is composed of an N -fold cartesian product of a uniform pulse amplitude modulation (PAM) alphabet. The decoder structure is based on the nearest neighbour detector that chooses the decoded source from a reconstruction codebook. The PCCOVQ is optimized by choosing the encoder partition, the reconstruction codebook, and the distance between samples in the channel signal set that minimize the MSE distortion under a power constraint. This process is performed using a modified version of the generalized Lloyd algorithm. Bandwidth reduction and expansion are developed in [23]. It is shown that PCCOVQ performs well for dimension reduction; there is only 1 dB gap between the SDR performance of a 2 : 1 PCCOVQ and the theoretical limit OPTA. For bandwidth expansion, however, PCCOVQ gives a poor performance with respect to other reference systems. More recently, the authors in [22] study the same problem and focus on PCCOVQ under bandwidth expansion. They apply the same algorithm as the one proposed in [23] but using different initial conditions and a larger number of samples in the channel signal set. These slight modifications are shown to improve the performance of the PCCOVQ for bandwidth expansion. Numerical results indicate that the performance of 1 : 2 and 1 : 3 PCCOVQ systems are comparable to other state of the art reference systems [22]. However, this gain in performance comes at the expense of higher computational complexity for the system design due to the use of a large number of symbols in the channel signal set. In Fig. 2.4, we show the reconstruction codebook structure of size 512 for a 2 : 1 PCCOVQ. Notice that for low CSNR (0 dB), the codebook structure has a straight line shape which is similar to the linear system (BPAM). In BPAM, we disregard one component and hence the mapping is a line along one of the axes. For high CSNR (20 dB), the structure of the codebook,

however, is nonlinear and has the shape of a double Archimedes' spiral.

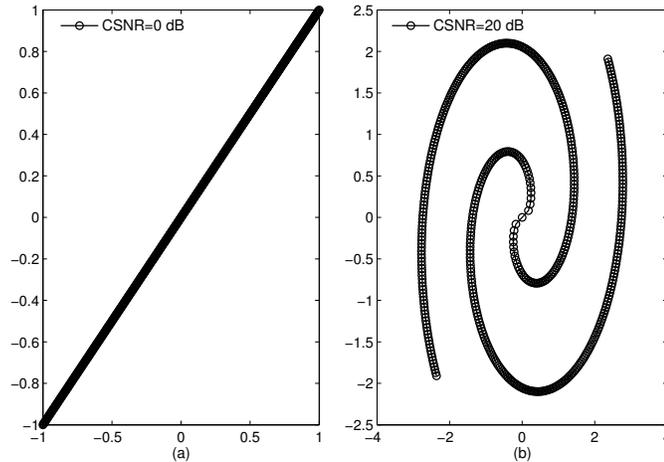


Figure 2.4: The reconstruction codebook structure of size 512 for a 2 : 1 PCCOVQ system. The system is designed for (a) CSNR = 0 dB and (b) CSNR = 20 dB. The graph is made for an i.i.d. Gaussian source with unit variance.

2.3.3 Shannon-Kotel'nikov Mappings

As opposed to quantizing the source into a discrete set of representation points which are then mapped into the channel space as in the case of PCCOVQ, Shannon-Kotel'nikov mapping [42, 59] is an approach based on direct source-channel mapping in which source and channel coders are merged into one operation. This operation maps the source space directly into the channel space. The main idea behind these mappings is based on a geometrical interpretation of the communication problem. The source, in this case, is represented using a point in the source space \mathbb{R}^K (*message space*), and the channel input is a point in the channel space \mathbb{R}^N (*signal space*). This geometrical approach is first introduced by Shannon in [59]. Furthermore, Kotel'nikov presents a theory for bandwidth expansion in his doctoral dissertation [42] by using

a similar signal mapping approach; and hence its name.

Shannon-Kotel'nikov mappings perform either a projection of the source onto a lower dimensional space or map the source into a higher dimensional space. The former represents a lossy compression of the source (dimension reduction), while the latter uses the redundant dimensions for error control (dimension expansion). For the case of matched bandwidth (dimension) between the source and the channel, it is well known that a linear, or uncoded, transmission is optimal for a memoryless source and Gaussian channel [6]. However, for mismatched bandwidth, linear transmission is suboptimal. Instead of discarding the excess of source samples to achieve bandwidth reduction, or repeating part of the source samples for bandwidth expansion (as discussed in Section 2.3.1), nonlinear mappings need to be explored in order to achieve a better performance.¹ As an example of 1 : 2 bandwidth expansion mapping, Shannon proposes the curve shown in Fig. 2.5 [59]. In this mapping, the one dimensional source is given by the line space (e.g., the length along the curve) which is mapped to a two dimensional channel input $(X(1), X(2))$. This approach will give a better performance than the one where we just send the same source symbol twice (repetition code). Shannon also suggests that the same mapping shown in Fig. 2.5 can be used for bandwidth reduction by interchanging the source and the channel space; more precisely, every source vector $(V(1), V(2))$ is projected onto the nearest point on the mapping curve which will be represented using a one dimensional channel space (e.g., the distance from some reference origin to the projected point on the curve).

The structure of the curve is very much related to the overall system performance.

¹Note that repeating part of the source or inserting zeros samples will not alter the performance of the system since both mappings are related by some orthonormal transformation.

to low power channel symbols so that the transmission power is minimized; 3) points in the channel space that are close to each other should be mapped to source symbols that are also close in the source space in order to minimize the distortion when errors occur. Moreover, one has to make sure that when choosing a mapping, all channel representations should have low correlation so that no redundant information is transmitted on different channel symbols.

Recently, the authors in [20, 31–33] have shown that for a memoryless Gaussian source, the double Archimedes' spiral represents a good mapping for 1 : 2 bandwidth expansion and 2 : 1 bandwidth reduction. Looking at the structure of the reconstruction codebook for a 2 : 1 PCCOVQ in Fig. 2.4, we can notice that this structure resembles very much the spiral mapping for high CSNR levels. In this case, the advantage of using a parameterized Shannon-Kotel'nikov mapping is the easiness in designing the system for a given source and channel characteristics. All we need is to modify the mapping parameters so that the shape can be changed accordingly.

2.3.4 Hybrid Digital-Analog Systems

The main advantage of using digital techniques in communication systems is the ability to achieve asymptotically the theoretical performance limit for a given CSNR via separate source-channel coding (tandem system). There are, however, two main drawbacks with tandem systems: 1) the *threshold effect* and 2) the *leveling-off effect*. The threshold effect means that the system performs well at the design CSNR level, while its performance degrades drastically when the true CSNR level is lower than the design one. The levelling-off effect means that the system performance remains constant beyond a certain CSNR level; this is due to the non-recoverable distortion

introduced by the quantizer.

Using a digital joint source-channel coding, we can overcome the threshold effect. Such systems, however, still suffer from the levelling-off effect. On the other hand, an analog system does not suffer from the levelling-off effect. This motivates researchers to exploit the advantage of using both digital and analog techniques in one system (hybrid technique). This is done by allowing part of the system to use digital modulation and coding and another part to use analog signalling. Mainly, schemes based on hybrid transmission are built by splitting the source into a quantized (digital) part, and a quantization error (analog) part. A general block diagram of a hybrid digital-analog (HDA) system is illustrated in Fig. 2.6 [14, 15, 50, 60, 61, 68].

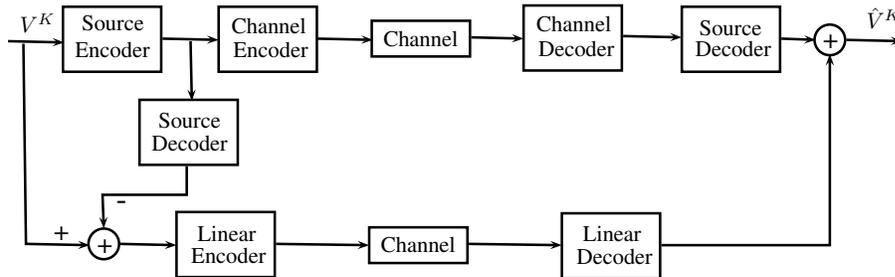


Figure 2.6: A block diagram of a general hybrid digital-analog system. The digital part uses a tandem scheme and the analog part uses a linear transmission.

In [60], the authors present an HDA system for bandwidth expansion based on vector quantization and linear (uncoded) transmission; they propose a design algorithm to optimize the system performance. In [61], an HDA scheme is proposed for bandwidth expansion/reduction system; this hybrid scheme uses tandem coding, that employs turbo code, for the digital part and a linear/nonlinear mappings for the analog part. In [14, 15], a hybrid scheme, which is referred to as hybrid scalar quantizer linear coder (HSQLC), is proposed. This scheme is able to achieve a 1 : 2 bandwidth

expansion by using a scalar quantizer for the digital part and a linear mapping for the analog part. One main difference between the HSQLC and the proposed HDA schemes in [60,61], is that the HSQLC has a low delay coding and low complexity; the HSQLC encodes a single source sample at a time (no coding delay), and uses a simple scalar quantizer and a linear coder. In contrast, the HDA schemes in [60,61] use either a vector quantization or a powerful channel codes in the digital part and hence incur large coding delay and complexity. In general, HDA systems have shown to offer a robust and graceful performance improvement/degradation for a wide range of CSNRs which make them suitable in many practical applications including broadcasting.

2.4 $K : 1$ Shannon-Kotel'nikov Mapping

In this section, we describe the theory for $K : 1$ dimension reduction mapping and provide results which will be used to analyze the $2 : 1$ reduction system in detail [32]. In such case, the source vector $V^K \in \mathbb{R}^K$ in the source space is first approximated by mapping it to a parametric curve \mathcal{S} in the source space. The dimension is subsequently changed from K to 1 by a lossless operator, for instance the radial distance from the origin out to the given point on the curve. The one dimensional parameter value is then given a representation on the channel space through an invertible mapping function. This function determines the way we measure the distance from the origin of the curve to the given approximated point. Denoting the one dimensional channel signal by X , the reconstructed signal \hat{V}^K at the receiver is estimated using the mapping $\mathcal{S}(\cdot)$ as follows $\hat{V}^K = \mathcal{S}(X + W)$, where W denotes the additive noise. There will be two distortion effects on this system, *approximation distortion* from projecting a K dimensional source vector into a one dimensional channel input, and

channel noise distortion due to the additive channel noise.

Approximation distortion: The reduction of dimensionality of a source for transmission over a power constrained channel introduces information loss. The approximation operation maps the source vectors onto a parametric curve \mathcal{S} that is a subset of \mathbb{R}^K . To reduce the approximation distortion, the curve has to densely fill the source space; this, however, is in contrast with the requirement for reducing the channel noise distortion. This trade-off is similar to the one in the traditional lossy source coding; as few representation values as possible are desired while at the same time the distance between them should be as small as possible.

Channel noise distortion: The received signal has to be passed through the non-linear mapping \mathcal{S} before it is detected. Given a transmitted one dimensional channel input X_0 , the received signal can be expressed using the following linear approximation

$$\mathcal{S}(X_0 + W) \approx \mathcal{S}(X_0) + \mathcal{S}'(X_0)W \quad (2.12)$$

where \mathcal{S}' is the derivative of \mathcal{S} at X_0 and the linear approximation is accurate for a small deviation in noise W . The last term in (2.12) contributes to the distortion of the received value. Given that X_0 was transmitted, the mean square error distortion caused by the channel noise is given by

$$\varepsilon_{ch} = \frac{1}{K} \mathbb{E}[\|\mathcal{S}'(X_0)W\|^2] = \frac{\sigma_W^2}{K} \|\mathcal{S}'(X_0)\|^2. \quad (2.13)$$

Hence, the average distortion can be expressed as follows

$$\bar{\varepsilon}_{ch} = \frac{\sigma_W^2}{K} \int \|\mathcal{S}'(x)\|^2 p(x) dx \quad (2.14)$$

where $p(x)$ is the pdf of the channel signal.

2.4.1 Example System using Archimedes' Spiral

In the following, we present a $2 : 1$ bandwidth reduction system based on spiral mapping. We consider i.i.d. Gaussian memoryless source $V^2 = [V(1) V(2)]^T$ and AWGN channel.

Encoder

We perform the bandwidth reduction by transmitting a combination of the source samples $V(1)$ and $V(2)$ as one channel input X . This is done by first approximating V^2 to the closest point on the spiral mapping. Note that for a given variable $x \in \mathbb{R}$, the two dimensional spiral mapping output can be mathematically expressed as follows [32]

$$\mathcal{S}_{1:2}(x) = \begin{bmatrix} V(1) \\ V(2) \end{bmatrix} = \frac{1}{\pi} \begin{bmatrix} \text{sgn}(x)\Delta\varphi(x) \cos \varphi(x) \\ \text{sgn}(x)\Delta\varphi(x) \sin \varphi(x) \end{bmatrix} \quad (2.15)$$

where $\text{sgn}(\cdot)$ is the signum function, Δ is the radial distance between any two neighbouring spiral arms, and $\varphi(x) = \sqrt{6.25|x|/\Delta}$. The approximated point, represented by its radial distance d_r from the origin, is then mapped to the channel via an invertible operator $\ell(\cdot)$

$$X = u\ell(d_r) = u \left(\pm 0.16 \left(\frac{\pi^2}{\Delta} \right) (d_r^2) \right) \quad (2.16)$$

where u is a gain factor related to channel power constraint P , $+$ sign represents points residing on the solid spiral arm, and the $-$ sign represents points residing on the dashed spiral arm in Fig. 2.7.

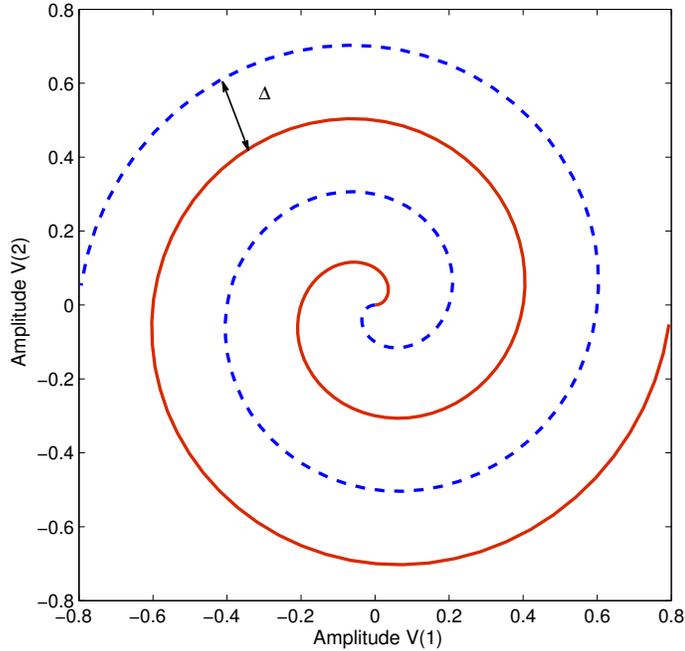


Figure 2.7: Two dimensional spiral mapping.

Decoder

Given the received noisy channel output $Y = X + W$, the maximum likelihood (ML) estimate is given by the source vector $V^2 = [V(1) \ V(2)]$ that maximizes the likelihood function

$$p(y|x) = \left(\frac{1}{2\pi\sigma_W^2} \right) e^{-\frac{\|y-x\|^2}{2\sigma_W^2}}. \quad (2.17)$$

This is achieved by the source vector $[V(1), V(2)]$ that minimizes the L_2 norm $\|y-x\|$. Although ML decoding is simple and performs well at high CSNRs, it is not optimal for the MSE distortion criterion. Next, we describe the minimum mean square error (MMSE) decoder, which is optimal in the MSE sense. The MMSE estimate can be

expressed as follows

$$\begin{aligned}\hat{V}^2 &= [\hat{V}(1) \hat{V}(2)] = \mathbb{E}[V^2|y] = \int v^2 p(v^2|y) dv^2 \\ &= \frac{\int v^2 p(y|v^2) p(v^2) dv^2}{\int p(y|v^2) p(v^2) dv^2}.\end{aligned}\quad (2.18)$$

System Optimization

For a given CSNR level, the radial distance Δ is the only parameter that needs to be optimized in order to minimize the overall MSE distortion. When the received signal is decoded, we may encounter two types of distortion: 1) the approximation distortion $\bar{\varepsilon}_a$ which is related to projecting a two dimensional signal into a one dimensional channel input, and 2) the channel distortion $\bar{\varepsilon}_{ch}$ which is due to transmitting over a noisy environment. The approximation operation mainly introduces radial errors and can be modelled by a standard scalar quantizer. Therefore, the approximation distortion is well given by $\bar{\varepsilon}_a \approx \Delta^2/12$ [32]. The channel noise distortion is expressed in a similar way as (2.14). Moreover, the received signal is rescaled using the inverse of the transmit scaling parameter (i.e., $1/u$). This implies that the channel noise is also scaled and has a power of $(1/u^2)\sigma_W^2$. After some manipulation, the distortion from channel noise $\bar{\varepsilon}_{ch}$ is well approximated by the scaled noise variance $(1/u^2)\sigma_W^2$. Hence the overall MSE distortion can be approximated as follows

$$D_{2:1 \text{ Spiral}} \approx (\bar{\varepsilon}_a + \bar{\varepsilon}_{ch}) \approx \frac{1}{2} \left(\frac{\Delta^2}{12} + \frac{\sigma_W^2}{u^2} \right).\quad (2.19)$$

The optimal radial distance Δ_{opt} is found by minimizing the MSE distortion in (2.19). Note that in the system simulations, we use both ML and MMSE decoders. When using MMSE decoder, the radial distance Δ is optimized numerically by searching

for Δ that achieves the best performance. This is done by generating a large set of $(V(1), V(2), W)$ and computing the overall distortion D empirically for each possible Δ in the search space.

2.4.2 Numerical Results

In this section, we assume an i.i.d. Gaussian source with standard deviation $\sigma_V = 1$ that is transmitted over an AWGN channel with power constraint $P = 1$. Both ML and MMSE decoders are investigated for 2 : 1 bandwidth reduction using spiral mapping. From Fig. 2.8, we can notice that MMSE decoder gives better performance

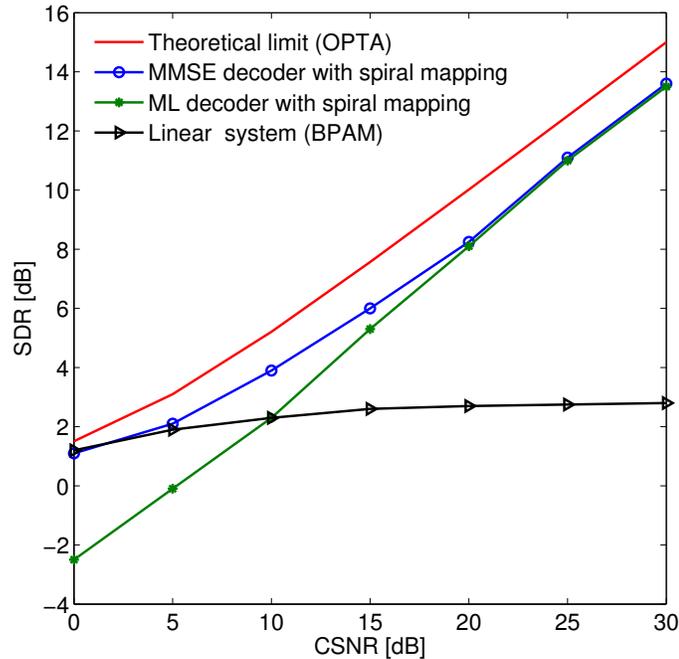


Figure 2.8: Performance of 2 : 1 bandwidth reduction using spiral mapping for a Gaussian source vector with standard deviation $\sigma_V = 1$. The optimal linear system (BPAM) and the theoretical limit (OPTA) of the system are also plotted.

than ML decoder. There is a substantial improvement at low CSNR levels and the performance from the spiral mapping is now similar to the linear system which comes close to achieve OPTA for asymptotically bad channels [47]. However as CSNR increases, the ML decoder gives a similar performance as the MMSE decoder. This is expected since the performance of the ML decoder approaches that of the optimal MMSE decoder (in the MSE sense) for high CSNR levels [57, pp. 291-292].

2.5 Coding Schemes with Side Information

2.5.1 Interference Known to the Transmitter

In this section, we focus on the AWGN channel with Gaussian interference known non-causally to the transmitter side. The system model is shown in Fig. 2.9. Assuming that the Gaussian source V^K and the interference S^N are uncorrelated, we next present two schemes that achieve the optimal distortion. Note that in this case, the optimal performance is the same as if the decoder had knowledge of the interference; this means that the interference can be completely removed and has no effect on the overall performance. This problem, which is sometimes referred to as the dirty paper coding problem, is considered in the seminal works of [12, 29].

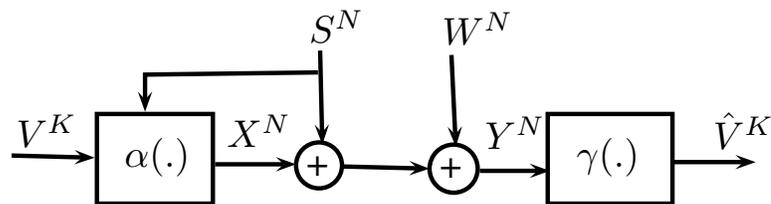


Figure 2.9: Block diagram for the dirty paper coding problem.

Tandem Costa Coding

In [29], Gel'fand and Pinsker showed that the capacity of a point-to-point communication system with side information (interference) known non-causally at the encoder side is given by

$$C = \max_{p(u,x|s)} \left(I(U; Y) - I(U; S) \right) \quad (2.20)$$

where the maximum is over all joint distributions of the form $p(s)p(u, x|s)p(y|x, s)$, U denotes an auxiliary random variable and $I(\cdot, \cdot)$ is the mutual information. In [12], Costa proposes to use the following auxiliary variable for the AWGN channel

$$U^N = X^N + \alpha S^N \quad (2.21)$$

where each sample in X^N is a zero mean i.i.d. Gaussian with variance P that is independent of S^N and α is a real parameter. By choosing $\alpha = \frac{P}{P + \sigma_w^2}$, Costa shows that the capacity is $C_{DPC} = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_w^2} \right)$ and coincides with the capacity when both encoder and decoder know the interference S^N . As a result, this choice of U^N is optimal in terms of achieving capacity.

The tandem Costa scheme is based on the concatenation of an optimal source code and Costa coding. Hence, the source V^K is first quantized using an optimal quantizer $\mathcal{Q} : \mathbb{R}^K \rightarrow \{1, 2, \dots, 2^{KR}\}$, where R is defined later. The quantization index $m = \mathcal{Q}(V^K)$ is then transmitted using Costa coding [12] which is briefly described as follows

- Codebook Generation: Create a codebook \mathcal{C}_u with block length N and size $2^{NI(U; Y)}$, where each codeword is generated according to the random variable U^N . The codewords are randomly assigned to 2^{KR} bins. For each U^N , let $i(U^N)$

be the index of the bin that contains U^N .

- Encoding: For a given quantization index m , the encoder looks for a codeword U^N such that $i(U^N) = m$ and (U^N, S^N) are jointly typical. If such U^N is found, then $X^N = U^N - \alpha S^N$ is transmitted.
- Decoding: Given the received signal $Y^N = X^N + S^N + W^N$, the decoder looks for a $U^N \in \mathcal{C}_u$ such that (U^N, Y^N) are jointly typical and declares $i(U^N)$ to be the decoded message if this codeword is found and is unique. From [12], the overall probability of error (encoding and decoding failures) can be made arbitrarily small by setting $R = C_{DPC}$.

HDA Costa Coding

Now, we describe a joint source channel coding where the source V^K is not explicitly quantized. Note that this scheme is only applicable for the matched source-channel bandwidth case ($K = N$). We define the auxiliary random variable U_h^N as follows [70]

$$U_h^N = X^N + \alpha_h S^N + \kappa_h V^K \quad (2.22)$$

where each sample in X^N is a zero mean i.i.d. Gaussian with variance P that is independent of S^N and V^N , $\alpha_h = \frac{P}{P + \sigma_w^2}$ and $\kappa_h^2 = \frac{P^2}{P + \sigma_w^2}$. The encoding and decoding processes of the HDA Costa scheme are described as follows

- Codebook Generation: Generate a codebook \mathcal{C}_h of block length N and size 2^{NR_h} following the random variable U_h^N , where R_h is defined later.
- Encoding: Given the source V^K ($K = N$) and the interference S^N , the encoder

looks for a $U_h^N \in \mathcal{C}_h$ such that (U_h^N, S^N, V^K) are jointly typical. If this is found, $X^N = U_h^N - \alpha_h S^N - \kappa_h V^K$ is transmitted.

- Decoding: Given the received signal Y^N , the decoder looks for U_h^N that is jointly typical with Y^N and is unique. In the absence of decoding error, the decoder forms a linear MMSE estimate of V^K based on the decoded codeword U_h^N and the received signal Y^N .

The encoding and decoding failure of the codeword U_h^N can be made arbitrarily small by using $I(U; S, V) \leq R_h \leq I(U, Y)$ [70]. Note that in [67], the authors adapted the HDA Costa coding scheme for the bandwidth reduction case; this is done by superposing two layers consisting of HDA Costa coding and tandem Costa coding.

2.5.2 Side Information at the Receiver

In this section, as shown in Fig. 2.10, we consider the transmission of a Gaussian source V^K over an AWGN channel with a Gaussian side information \tilde{V}^K at the decoder side, where $V^K = \tilde{V}^K + E^K$ with each sample in E^K follows an i.i.d. Gaussian distribution with variance \tilde{D} ($E \sim \mathcal{N}(0, \tilde{D})$). Suppose that the side information is

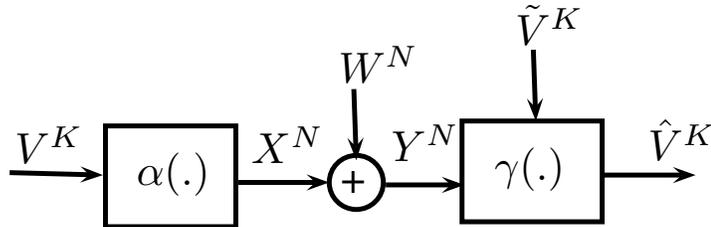


Figure 2.10: The Wyner-Ziv problem.

available at both sides (i.e., encoder and decoder), the least required rate for achieving

a desired MSE distortion D is

$$R_{wz} = \frac{1}{2} \log \frac{\tilde{D}}{D}. \quad (2.23)$$

Let us set this rate to be close to the capacity of an AWGN channel (the rate at which the channel can support with low probability of error). The best achievable distortion is then given by

$$D_{wz} = \frac{\tilde{D}}{\left(1 + \frac{P}{\sigma_w^2}\right)^r}. \quad (2.24)$$

Now we will illustrate how this distortion can be achieved using Wyner-Ziv coding [71]. Let T^K be an auxiliary random variable given by

$$T^K = \alpha_{wz} V^K + B^K \quad (2.25)$$

where $\alpha_{wz} = \sqrt{1 - \frac{D_{wz}}{\tilde{D}}}$, D_{wz} is given in (2.24) and each sample in B^K is zero mean i.i.d. Gaussian with variance D_{wz} . The encoding and decoding processes of the Wyner-Ziv coding can be summarized as follows

- Generate a length K i.i.d. Gaussian codebook \mathcal{T} of size $2^{KI(T;V)}$ and randomly assign the codewords into 2^{KR} bins with R equal to the AWGN capacity.
- For each source realization V^K , we find a codeword $T^K \in \mathcal{T}$ such that (V^K, T^K) is jointly typical. The encoder then transmits the index bin of this codeword using a channel code with rate R .
- The decoder first decodes the bin index and then looks for a codeword T^K in this bin such that (T^K, \tilde{V}^K) are jointly typical. In case of succeed (which happens with high probability of error with the chosen rate), we form a linear MMSE

estimate of V^K based on the decoded codeword T^K and the side information \tilde{V}^K .

It can be verified that using the above approach, we can achieve the distortion in (2.24).

2.6 Broadcast Channels

The broadcast channel is a communication channel in which there is one transmitter and two or more receivers. The basic problem in this multi-terminal channel is to find the set of simultaneously achievable rates for reliable communication in this channel. The general broadcast, which is illustrated in Fig. 2.11, can be mathematically defined as follows [13, 17].

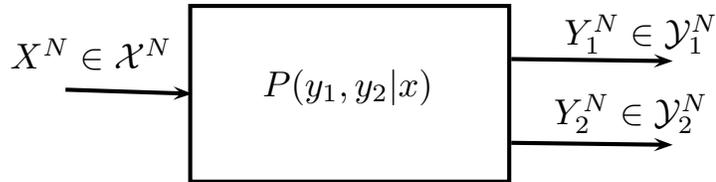


Figure 2.11: General broadcast channel.

Definition 2.1. A broadcast channel consists of an input alphabet \mathcal{X} , two output alphabets $\mathcal{Y}_1, \mathcal{Y}_2$ and a transition probability mass function (pmf) $P(y_1, y_2|x)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$. The broadcast channel is said to be memoryless if $P(y_1^n, y_2^n|x^n) = \prod_{i=1}^n P(y_1(i), y_2(i)|x(i))$.

We note that an error for the first receiver (Y_1^N) depends only on the distribution $P(x^n, y_1^n)$ and not on the joint distribution $P(x^n, y_1^n, y_2^n)$. As a result, the capacity region of a broadcast channel depends only on the conditional marginal distributions

$P(y_1^n|x^n)$ and $P(y_2^n|x^n)$; this means that for two broadcast channels with the same marginal distributions, their capacity regions are the same.

2.6.1 Degraded Broadcast Channels

Definition 2.2. *A broadcast channel is said to be physically degraded if*

$$P(y_1, y_2|x) = P(y_2|x)P(y_1|y_2). \quad (2.26)$$

This means we have the following Markov chain $X \rightarrow Y_2 \rightarrow Y_1$.

The degraded broadcast channel can be illustrated as in Fig. 2.12. As an example

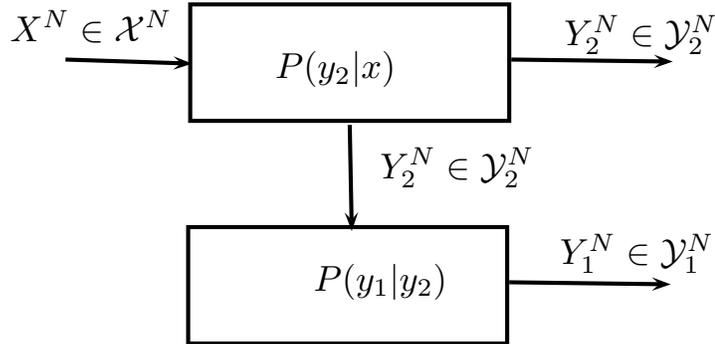


Figure 2.12: Physically degraded memoryless broadcast channel.

of degraded broadcast channel, we present the Gaussian degraded broadcast channel.

In this case, one output is a degraded version of the other output. The received signals at user 2 and 1 can be written as follows, respectively,

$$\begin{aligned} Y_2^N &= X^N + W_2^N \\ Y_1^N &= X^N + W_1^N = Y_2^N + \tilde{W}^N \end{aligned} \quad (2.27)$$

where each sample in W_2^N follows an i.i.d. Gaussian distribution with variance $\sigma_{W_2}^2$ ($W_2 \sim \mathcal{N}(0, \sigma_{W_2}^2)$) and each sample in \tilde{W}^N is drawn from an i.i.d. Gaussian distribution ($\tilde{W} \sim \mathcal{N}(0, \sigma_{W_1}^2 - \sigma_{W_2}^2)$). Note that in this case, user 1 is the weak user and user 2 is the strong user.

2.6.2 Broadcast Scenario with Two Message Sets

Fig. 2.13 shows the structure of a broadcast channel with two message sets. A

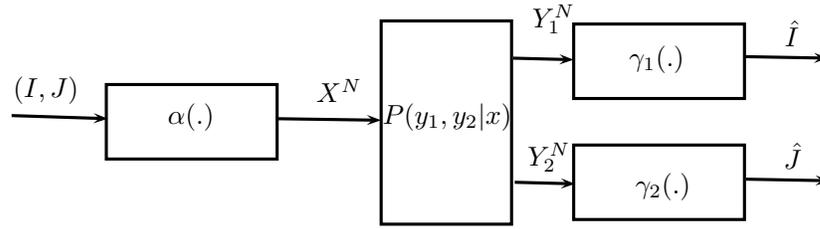


Figure 2.13: Broadcast channel with two message sets.

$(2^{NR_1}, 2^{NR_2}, N)$ code for a broadcast channel with two message sets consists of an encoder,

$$\alpha : \{1, 2, \dots, 2^{NR_1}\} \times \{1, 2, \dots, 2^{NR_2}\} \rightarrow \mathcal{X}^N \quad (2.28)$$

and two decoders

$$\gamma_1 : \mathcal{Y}_1^N \rightarrow \{1, 2, \dots, 2^{NR_1}\} \quad \text{and} \quad \gamma_2 : \mathcal{Y}_2^N \rightarrow \{1, 2, \dots, 2^{NR_2}\}. \quad (2.29)$$

The average probability of error is defined as the probability that the decoded message is not equal to the transmitted message; that is

$$P_e = Pr(\hat{I} = I \text{ or } \hat{J} = J) \quad (2.30)$$

where the messages (I, J) are assumed to be uniformly distributed and $Pr(\cdot)$ denotes the probability of an event. A rate of pair (R_1, R_2) is said to be achievable if there exists a sequence of $(2^{NR_1}, 2^{NR_2}, N)$ codes with probability of error $P_e \rightarrow 0$ as $N \rightarrow \infty$. The capacity in this case, is given by the closure of the set of achievable rates.

2.6.3 Capacity Region for the Degraded Broadcast Channels

We now consider sending independent information over a degraded broadcast channel at rate R_1 to Y_1 and rate R_2 to Y_2 .

Theorem 2.1. *The capacity region for sending independent information over the degraded broadcast channel $X \rightarrow Y_2 \rightarrow Y_1$ is the convex hull of the closure of all (R_1, R_2) satisfying*

$$\begin{aligned} R_1 &\leq I(U_b; Y_1) \\ R_2 &\leq I(X; Y_2 | U_b) \end{aligned} \tag{2.31}$$

for some joint distribution $P(u_b)P(x|u_b)P(y_1, y_2|x)$ and U_b is an auxiliary random variable. Note that $I(\cdot, \cdot)$ denotes the mutual information.

The capacity of the degraded broadcast channel is achieved using superposition coding [17]. The auxiliary random variable U_b^N will serve as a cloud centre that can be distinguished by both receivers (i.e., Y_1^N and Y_2^N). Each cloud consists of 2^{NR_2} codewords X^N distinguishable by the strong receiver Y_2^N . The worst receiver can only see the clouds, while the better receiver can see the individual codewords (the satellite codewords) within the clouds. A representation of the satellite and cloud codewords are given in Fig. 2.14.

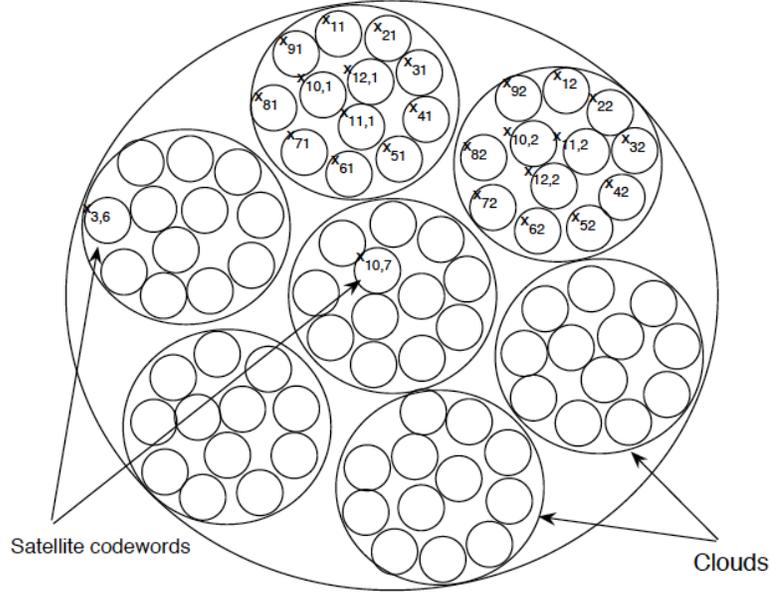


Figure 2.14: Illustration of clouds and satellite codewords used in superposition coding.

Capacity Region of Gaussian Degraded Broadcast Channel

The capacity region of the Gaussian degraded broadcast channel is a function of only the CSNR and a power allocation parameter.

Theorem 2.2. *The capacity region of the Gaussian degraded broadcast channel is the set of rate pairs (R_1, R_2) such that*

$$\begin{aligned} R_1 &\leq \frac{1}{2} \log \left(1 + \frac{\eta_t P}{(1 - \eta_t) P + \sigma_{W_1}^2} \right) \\ R_2 &\leq \frac{1}{2} \log \left(1 + \frac{(1 - \eta_t) P}{\sigma_{W_2}^2} \right) \end{aligned} \quad (2.32)$$

where $0 \leq \eta_t \leq 1$.

Achievability follows by using the superposition coding technique, setting each sample in U_b^N to be an i.i.d. Gaussian ($U_b \sim \mathcal{N}(0, \eta_t P)$) and choosing $X^N = U_b^N + \tilde{U}^N$, where each sample in \tilde{U}^N is an i.i.d. Gaussian ($\tilde{U} \sim \mathcal{N}(0, (1 - \eta_t)P)$) independent of U_b . With this choice of (U_b^N, X^N) , it can be readily shown that the region in (2.31) reduces to (2.32).

Chapter 3

Hybrid-Digital Analog Coding for Gaussian Source and State Interference Estimation

3.1 Introduction

The traditional approach for analog source transmission in point-to-point communication systems is to employ separate source and channel coders. This separation is (asymptotically) optimal given unlimited delay and complexity in the coders [58]. As mentioned in the previous chapter, there are, however, two disadvantages associated with digital transmission. One is the threshold effect: the system typically performs well at its designed noise level, while its performance degrades drastically when the true noise level is higher than the design level. The other trait is the levelling-off effect: as the noise level decreases, the performance remains constant beyond a certain threshold. JSCC schemes are more robust to noise level mismatch than tandem systems. Analog JSCC schemes, which are based on the so-called direct source-channel mappings (Shannon-Kotel'nikov mappings) [2, 3, 9, 21, 32, 34, 38, 39, 55], are used to increase the robustness of communication systems. Moreover, a family of HDA schemes are introduced in [50, 61, 69] to overcome the threshold and the levelling-off effects.

In [27, 41, 52], HDA schemes are also proposed for broadcast channels.

As described in Sec. 2.5.1, it is well known that for the problem of transmitting a Gaussian source over an AWGN channel with interference that is known non-causally to the transmitter, a tandem Costa coding scheme, which comprises an optimal source encoder followed by Costa's dirty paper channel code (Costa coding) [12], and an HDA Costa coding [70] are optimal in the absence of correlation between the source and the interference. In [35], the authors study a joint source channel coding scheme for transmitting analog Gaussian source over AWGN channel with interference known to the transmitter and correlated with the source. In that work, they propose two schemes for the matched source-channel bandwidth case; the first one is the superposition of the uncoded signal and a digital signal resulting from the concatenation of a Wyner-Ziv coder [71] and a Costa coder, while in the second scheme the digital part is replaced by the HDA Costa coding. The limiting case of this problem, where the source and the interference are fully correlated is studied in [63]; the authors show that a purely analog scheme (uncoded) is optimal. Moreover, they also consider the problem of sending a digital (finite alphabet) source in the presence of interference where the interference is independent from the source. More precisely, the optimal tradeoff between the achievable rate for transmitting the digital source and the distortion in estimating the interference is studied; they show that the optimal rate-state-distortion tradeoff is achieved by a coding scheme that uses a portion of the power to amplify the interference and the remaining power to transmit the digital source via Costa coding. In [4], the authors consider the same problem as the one in [63] but with imperfect knowledge of the interference at the transmitter side.

In this chapter, we study the reliable transmission of a memoryless Gaussian

source over a Rayleigh fading channel with interference known non-causally to the transmitter and correlated to the source. More precisely, we consider equal and unequal source-channel bandwidths and analyze the achievable distortion region under matched and mismatched noise levels. We propose a layered scheme based on hybrid coding. One application of JSCC with correlated interference can be found in sensor networks and cognitive radio channels where two nodes interfere with each other. One node directly transmits its signal; the other, however, is able to detect its neighbouring node's transmission and treat it as a correlated interference. One interesting extension of this problem, which is also considered in this chapter, is to study the source-channel-state distortion region for the fading channel with correlated interference; in that case, the receiver side is interested in estimating both the source and the channel-state (interference). Inner and outer bounds on the source-interference distortion region are established. Our setting contains several interesting limiting cases. In the absence of fading and for the matched source-channel bandwidth, our system reverts to that of [35]; for the uncorrelated source-interference scenario without fading, our problem reduces to the one in [67] for the bandwidth reduction case. Moreover, the source-channel-state transmission scenario generalizes the setting in [63] to include fading and correlation between source and interference. The rest of the chapter is organized as follows. In Section 3.2, we present the problem formulation. In Section 3.3, we derive an outer bound and introduce linear and tandem digital schemes. In Section 3.4, we derive inner bounds (achievable distortion region) under both matched and mismatched noise levels by proposing layered hybrid coding schemes. We extend these inner and outer bounds to the source-channel-state communication scenario in Section 3.5. Finally, conclusions are drawn in Section 3.6.

Note that throughout the thesis, we use the following notation. Vectors are denoted by characters superscripted by their dimensions. For a given vector $X^N = (X(1), \dots, X(N))^T$, we let $[X^N]_1^K$ and $[X^N]_{K+1}^N$ denote the sub-vectors $[X^N]_1^K \triangleq (X(1), \dots, X(K))^T$ and $[X^N]_{K+1}^N \triangleq (X(K+1), \dots, X(N))^T$, respectively. When there is no confusion, we also write $[X^N]_1^K$ as X^K . When all samples in a vector are i.i.d., we drop the indexing when referring to a sample in a vector (i.e., $X(i) = X$).

3.2 Problem Formulation

We consider the transmission of a Gaussian source $V^K = (V(1), \dots, V(K))^T \in \mathbb{R}^K$ over a Rayleigh fading channel in the presence of Gaussian interference $S^N \in \mathbb{R}^N$ known at the transmitter (see Fig. 3.1). The source vector V^K represents the first K samples

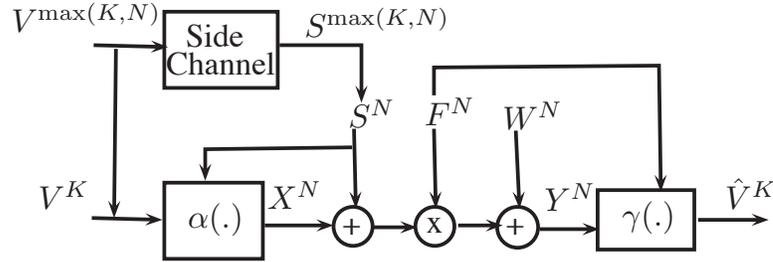


Figure 3.1: A $K : N$ system structure over a fading channel with interference known at the transmitter side. The interference $S^{\max(K,N)}$ is assumed to be the output of a noisy side channel with input $V^{\max(K,N)}$. V^K represents the first K samples of $V^{\max(K,N)}$ (S^N is defined similarly). The fading coefficient is assumed to be known at the receiver side; the transmitter side, however, knows the fading distribution only.

of $V^{\max(K,N)}$; S^N is similarly defined. The source vector V^K , which is composed of i.i.d. samples, is transformed into an N dimensional channel input $X^N \in \mathbb{R}^N$ using a nonlinear mapping function, in general, $\alpha(\cdot) : \mathbb{R}^K \times \mathbb{R}^N \rightarrow \mathbb{R}^N$. The received symbol

is given by

$$Y^N = F^N(X^N + S^N) + W^N \quad (3.1)$$

where addition and multiplication are component-wise, F^N represents an N -block Rayleigh fading that is independent of $(V^K; S^N; W^N)$ and known to the receiver side only, $X^N = \alpha(V^K, S^N)$, S^N is an i.i.d. Gaussian interference vector (with each sample $S \sim \mathcal{N}(0, \sigma_S^2)$) that is considered to be the output of a side channel with input $V^{\max(K, N)}$ as shown in Fig. 3.1, and each sample in the additive noise W^N is drawn from a Gaussian distribution ($W \sim \mathcal{N}(0, \sigma_W^2)$) independently from both the source and the interference. Unlike the typical dirty paper problem which assumes an AWGN channel with interference (that is uncorrelated to the source) [12], we consider a fading channel and assume that V^K and S^N are jointly Gaussian. Since the fading realization is known only at the receiver, we have partial knowledge of the actual interference $F^N S^N$ at the transmitter. In this scenario, we assume that only $V(i)$ and $S(i)$, $i = 1, \dots, \min(K, N)$, are correlated according to the following covariance matrix

$$\Sigma_{VS} = \begin{bmatrix} \sigma_V^2 & \rho_{VS}\sigma_V\sigma_S \\ \rho_{VS}\sigma_V\sigma_S & \sigma_S^2 \end{bmatrix} \quad (3.2)$$

where ρ_{VS} is the source-interference correlation coefficient. The system operates under an average power constraint P

$$\mathbb{E}[|\alpha(V^K, S^N)|^2]/N \leq P. \quad (3.3)$$

The reconstructed signal is given by $\hat{V}^K = \gamma(Y^N, F^N)$, where the decoder is a mapping from $\mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}^K$. According to the correlation model described above,

note that for system's rate $r < 1$ (bandwidth reduction), the first N source samples $[V^K]_1^N$ and S^N are correlated via the covariance matrix in (3.2), while the remaining $K - N$ samples $[V^K]_{N+1}^K$ and S^N are independent. For $r > 1$ (bandwidth expansion), however, V^K and $[S^N]_1^K$ are correlated via the covariance matrix in (3.2), while V^K and $[S^N]_{K+1}^N$ are uncorrelated.

We aim to find a source-channel encoder α and decoder γ that minimize the MSE distortion $D = \mathbb{E}[\|V^K - \hat{V}^K\|^2]/K$ under the average power constraint in (3.3). For a particular coding scheme (α, γ) , the performance is determined by the channel power constraint P , the fading distribution, the system rate r , and the incurred distortion D at the receiver. For a given power constraint P , fading distribution and rate r , the distortion region is defined as the closure of all distortions D_o for which (P, D_o) is achievable. A power-distortion pair is achievable if for any $\delta > 0$, there exist sufficiently large integers K and N with $N/K = r$, a pair of encoding and decoding functions (α, γ) satisfying (3.3), such that $D < D_o + \delta$. In this work, we analyze the distortion for equal and unequal bandwidths between the source and the channel with no constraint on the delay (i.e., both N and K tend to infinity with $\frac{N}{K} = r$ fixed). Our main contributions in this chapter can be summarized as follows

- We derive inner and outer bounds for the system's distortion region for a Gaussian source over fading channel with correlated interference under equal and unequal source-channel bandwidths. The outer bounds are found by assuming full/partial knowledge of the interference at the decoder side. The inner bounds are derived by proposing hybrid coding schemes and analyzing their achievable distortion region. These schemes are based on proper combination of power splitting, bandwidth splitting, Wyner-Ziv and hybrid coding; a hybrid

layer that uses the source and the interference is concatenated (superimposed) with a purely digital layer to achieve bandwidth expansion (reduction). Different from the problem considered in [35], we consider the case of fading and mismatch in the source-channel bandwidth. Our scheme offers better performance than the one in [35] under matched bandwidth (when accommodating the Costa coder in their scheme for fading channels). Moreover, our scheme is optimal when there is no fading and when the source-interference are either uncorrelated or fully correlated.

- As an application of the proposed schemes, we consider source-channel-state transmission over fading channels with correlated interference. In such case, the receiver aims to jointly estimate both the source signal as well as the channel-state. Inner and outer bounds are derived for this scenario. For the special case of uncorrelated source-interference over AWGN channels, we obtain the optimal source-channel-state distortion tradeoff; this result is analogous to the optimal rate-state distortion for the transmission of a finite discrete source over a Gaussian state interference derived in [63]. For correlated source-interference and fading channels, our inner bound performs close to the derived outer bound and outperforms the adapted scheme of [63].

3.3 Outer Bounds and Reference Systems

3.3.1 Outer Bounds

In [35] and [10], outer (i.e., lower) bounds on the achievable distortion are derived for point-to-point communication over Gaussian channel with correlated interference under matched bandwidth between the source and the channel. This is done by

assuming full/partial knowledge of the interference at the decoder side. In this section, for the correlation model considered above, we derive outer bounds for the fading interference channel under unequal source-channel bandwidth. Since $S(i)$ and $V(i)$ are correlated for $i = 1, \dots, \min(K, N)$, we have $S(i) = S_I(i) + S_D(i)$, with $S_D(i) = \frac{\rho_{VS}\sigma_S}{\sigma_V}V(i)$ and $S_I \sim \mathcal{N}(0, (1 - \rho_{VS}^2)\sigma_S^2)$ are independent of each other. To derive an outer bound, we assume knowledge of both $(\tilde{S}^K, [S^N]_{K+1}^N)$ and F^N at the decoder side for the case of bandwidth expansion, where $\tilde{S}^K = \eta_1 S_I^K + \eta_2 S_D^K$ (the linear combination \tilde{S} is motivated by [10]), and (η_1, η_2) is a pair of real parameters. For the bandwidth reduction case, we assume knowledge of \tilde{S}^N and F^N at the decoder to derive a bound on the average distortion for the first N samples; the derivation of a bound on the average distortion for the remaining $K - N$ samples assumes knowledge of $[V^K]_1^N$ in addition to \tilde{S}^N .

Definition 3.1. Let $MSE(Y; \tilde{S})$ be the distortion incurred from estimating Y based on \tilde{S} using a linear minimum MSE estimator (LMMSE) denoted by $\gamma_{lmmse}(\tilde{S}^K, f^K)$. This distortion, which is a function of $\eta_1, \eta_2, \mathbb{E}[XS_I]$ and $\mathbb{E}[XS_D]$, is given by $MSE(Y; \tilde{S}) = \mathbb{E}[(Y - \gamma_{lmmse}(\tilde{S}^K, f^K))^2] = \left(\mathbb{E}[Y^2] - \frac{(\mathbb{E}[Y\tilde{S}])^2}{\mathbb{E}[\tilde{S}^2]} \right)$, where $\mathbb{E}[Y^2] = f^2(P + \sigma_S^2 + 2(\mathbb{E}[XS_I + XS_D])) + \sigma_W^2$, $\mathbb{E}[Y\tilde{S}] = f(\mathbb{E}[X(\eta_1 S_I + \eta_2 S_D)] + \mathbb{E}[\eta_1 S_I^2 + \eta_2 S_D^2])$ and $\mathbb{E}[\tilde{S}^2] = \mathbb{E}[\eta_1^2 S_I^2 + \eta_2^2 S_D^2]$. These terms will be used in Lemmas 3.1 and 3.2.

Lemma 3.1. For a $K : N$ bandwidth expansion system with $N \geq K$ (the matched case is treated as a special case), the outer bound on the system's distortion D can be

expressed as follows:

$$D \geq D_{ob} \triangleq \sup_{\eta_1, \eta_2} \inf_{\substack{X: \\ |\mathbb{E}[XS_I]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_I^2]} \\ |\mathbb{E}[XS_D]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_D^2]}} \left\{ \frac{\text{Var}(V|\tilde{S})}{\exp \left\{ \mathbb{E}_F \left[\log \left(\left(\frac{\text{MSE}(Y; \tilde{S})}{\sigma_W^2} \right) \left(\frac{f^2 P + \sigma_W^2}{\sigma_W^2} \right)^{r-1} \right) \right] \right\}} \right\} \quad (3.4)$$

where $\text{Var}(V|\tilde{S}) = \sigma_V^2 \left(1 - \frac{\eta_2^2 \rho_{VS}^2}{\eta_1^2 (1 - \rho_{VS}^2) + \eta_2^2 \rho_{VS}^2} \right)$ is the variance of V given \tilde{S} .

Proof. For a $K : N$ system with $N \geq K$, we have the following

$$\begin{aligned} \frac{K}{2} \log \frac{\text{Var}(V|\tilde{S})}{D} &\leq I(V^K; \hat{V}^K | \tilde{S}^K, [S^N]_{K+1}^N, F^N) \leq I(V^K; Y^N | \tilde{S}^K, [S^N]_{K+1}^N, F^N) \\ &= h(Y^N | \tilde{S}^K, [S^N]_{K+1}^N, F^N) - h(Y^N | V^K, S^N, F^N) \\ &\leq h(Y^K | \tilde{S}^K, F^K) + h([Y^N]_{K+1}^N | [S^N]_{K+1}^N, [F^N]_{K+1}^N) - h(Y^N | V^K, S^N, F^N) \\ &= \mathbb{E}_F \left[h(Y^K | \tilde{S}^K, f^K) + h([Y^N]_{K+1}^N | [S^N]_{K+1}^N, [f^n]_{k+1}^n) \right] - h(W^N) \\ &\leq \mathbb{E}_F \left[\frac{K}{2} \log 2\pi e (\text{MSE}(Y; \tilde{S})) + \frac{N-K}{2} \log 2\pi e (f^2 P + \sigma_W^2) \right] - \frac{N}{2} \log 2\pi e \sigma_W^2 \\ &= \mathbb{E}_F \left[\frac{K}{2} \log \left(\frac{\text{MSE}(Y; \tilde{S})}{\sigma_W^2} \right) + \frac{N-K}{2} \log \left(\frac{f^2 P + \sigma_W^2}{\sigma_W^2} \right) \right] \end{aligned} \quad (3.5)$$

where we used $h(Y^K | \tilde{S}^K, f^K) \leq h(Y^K - \gamma_{\text{mse}}(\tilde{S}^K, f^K)) \leq \frac{K}{2} \log 2\pi e \left(\text{MSE}(Y; \tilde{S}) \right)$. By the Cauchy-Schwarz inequality, we have $|\mathbb{E}[XS_I]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_I^2]}$ and $|\mathbb{E}[XS_D]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_D^2]}$. For a given η_1 and η_2 , we have to choose the highest value of $\text{MSE}(Y; \tilde{S})$ over $\mathbb{E}[XS_D]$ and $\mathbb{E}[XS_I]$; then we need to maximize the right-hand side of (3.4) over η_1 and η_2 . Note that most inequalities follow from rate-distortion theory, the data processing inequality, the facts that conditioning reduces differential entropy and that the Gaussian distribution maximizes differential entropy. \square

Lemma 3.2. For $K : N$ bandwidth reduction ($K > N$), the outer bound on D is

given by

$$\begin{aligned}
 D \geq D_{ob}(\xi^*) \triangleq & \sup_{\eta_1, \eta_2} \inf_{\xi} \left\{ \inf_{\substack{X: \\ |\mathbb{E}[XS_I]| \leq \sqrt{(1-\xi)P\mathbb{E}[S_I^2]} \\ |\mathbb{E}[XS_D]| \leq \sqrt{(1-\xi)P\mathbb{E}[S_D^2]}}} \left\{ r \frac{\text{Var}(V|\tilde{S})}{\exp \left\{ \mathbb{E}_F \left[\log \left(\frac{\text{MSE}(Y; \tilde{S})}{\xi P f^2 + \sigma_W^2} \right) \right] \right\}} \right. \right. \\
 & \left. \left. + (1-r) \frac{\sigma_V^2}{\exp \left\{ \mathbb{E}_F \left[\frac{N}{K-N} \log \left(\frac{\xi P f^2 + \sigma_W^2}{\sigma_W^2} \right) \right] \right\}} \right\} \right. \quad (3.6)
 \end{aligned}$$

where $\xi \in [0, 1]$.

Proof. We start by decomposing the average MSE distortion as follows

$$\begin{aligned}
 D &= \frac{1}{K} \mathbb{E}[\|V^K - \hat{V}^K\|^2] = \frac{1}{K} \left(\mathbb{E}[\|V^N - \hat{V}^N\|^2] + \mathbb{E}[\|[V^K]_{N+1}^K - [\hat{V}^K]_{N+1}^K\|^2] \right) \\
 &= \frac{N}{K} \left(\frac{1}{N} \mathbb{E}[\|V^N - \hat{V}^N\|^2] \right) + \frac{K-N}{K} \left(\frac{1}{K-N} \mathbb{E}[\|[V^K]_{N+1}^K - [\hat{V}^K]_{N+1}^K\|^2] \right) \\
 &= rD_1 + (1-r)D_2 \quad (3.7)
 \end{aligned}$$

where D_1 and D_2 are the average distortion in reconstructing V^N and $[V^K]_{N+1}^K$, respectively. To find an outer bound on D , we derive bounds on both D_1 and D_2 . To bound D_1 , We can write the following expression

$$\begin{aligned}
 \frac{N}{2} \log \frac{\text{Var}(V|\tilde{S})}{D_1} &\leq I(V^N; \hat{V}^N | \tilde{S}^N, F^N) \leq I(V^N; Y^N | \tilde{S}^N, F^N) \\
 &= h(Y^N | \tilde{S}^N, F^N) - h(Y^N | \tilde{S}^N, V^N, F^N) \\
 &= h(Y^N | \tilde{S}^N, F^N) - h(Y^N | S^N, V^N, F^N) \\
 &\stackrel{(a)}{\leq} \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e (\text{MSE}(Y; \tilde{S})) - \frac{N}{2} \log 2\pi e (\xi P f^2 + \sigma_W^2) \right] \\
 &\leq \sup_{Y \in \mathcal{A}} \mathbb{E}_F \left[\frac{N}{2} \log \left(\frac{\text{MSE}(Y; \tilde{S})}{\xi P f^2 + \sigma_W^2} \right) \right] \quad (3.8)
 \end{aligned}$$

where the set $\mathcal{A} = \{Y : h(Y^N|S^N, V^N, F^N) = \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e(\xi P f^2 + \sigma_W^2) \right]\}$. Note that in (3.8)-(a) we use the fact that $h(Y^N|S^N, V^N, F^N) = \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e(\xi P f^2 + \sigma_W^2) \right]$, for some $\xi \in [0, 1]$. This can be shown by noting that the following inequality holds $\frac{N}{2} \log 2\pi e(\sigma_W^2) = h(W^N) \leq h(Y^N|S^N, V^N, F^N) \leq h(F^N X^N + W^N|F^N) = \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e(P f^2 + \sigma_W^2) \right]$; as a result, there is a $\xi \in [0, 1]$ such that $h(Y^N|S^N, V^N, F^N) = \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e(\xi P f^2 + \sigma_W^2) \right]$. Moreover in (3.8)-(a), we used the fact that

$$\begin{aligned} h(Y^N|\tilde{S}^N, F^N) &= \mathbb{E}_F[h(Y^N|\tilde{S}^N, f^n)] = \mathbb{E}_F[h(Y^N - \gamma_{\text{tmse}}(\tilde{S}^N, f^n)|\tilde{S}^N, f^n)] \\ &\leq \mathbb{E}_F[h(Y^N - \gamma_{\text{tmse}}(\tilde{S}^N, f^n))] \leq \frac{N}{2} \mathbb{E}_F[\log 2\pi e(\text{MSE}(Y; \tilde{S}))]. \end{aligned} \quad (3.9)$$

Similarly, to derive a bound on D_2 , we have the following

$$\begin{aligned} \frac{K-N}{2} \log \frac{\sigma_V^2}{D_2} &\leq I([V^K]_{N+1}^K; [\hat{V}^K]_{N+1}^K|S^N, V^N, F^N) \leq I([V^K]_{N+1}^K; Y^N|S^N, V^N, F^N) \\ &= h(Y^N|S^N, V^N, F^N) - h(Y^N|S^N, V^N, [V^K]_{N+1}^K, F^N) \\ &= h(Y^N|S^N, V^N, F^N) - h(Y^N|S^N, V^K, F^N) \\ &= \mathbb{E} \left[\frac{N}{2} \log \left(\frac{\xi P f^2 + \sigma_W^2}{\sigma_W^2} \right) \right] \end{aligned} \quad (3.10)$$

where in the last equality, we used $h(Y^N|S^N, V^N, F^N) = \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e(\xi P f^2 + \sigma_W^2) \right]$ as shown earlier. Note that since we do not know the value of ξ , the overall distortion has to be minimized over the parameter ξ . Now using (3.8) and (3.10) in (3.7), we

have the following bound

$$D \geq \inf_{\xi} \inf_{Y \in \mathcal{A}} \left\{ r \frac{\text{Var}(V|\tilde{S})}{\exp \left\{ \mathbb{E}_F \left[\log \left(\frac{\text{MSE}(Y; \tilde{S})}{\xi P f^2 + \sigma_W^2} \right) \right] \right\}} + (1-r) \frac{\sigma_V^2}{\exp \left\{ \mathbb{E}_F \left[\frac{N}{K-N} \log \left(\frac{\xi P f^2 + \sigma_W^2}{\sigma_W^2} \right) \right] \right\}} \right\} \quad (3.11)$$

where the sup in (3.8) is manifested as inf on the distortion. Note that the above sequence of inequalities in (3.9) becomes equalities when Y is conditionally Gaussian given F and when $Y - \gamma_{lmse}(\tilde{S}, f)$ and \tilde{S} are jointly Gaussian and orthogonal to each other given F ; this happens when X^* is jointly Gaussian with S , V and W given F . Hence, the sup in (3.8) happens when X^* is Gaussian. Now we write $X^* = N_{\xi}^* + X_{\xi}^*$, where $N_{\xi}^* \sim \mathcal{N}(0, \xi P)$ is independent of (V, S) and $X_{\xi}^* \sim \mathcal{N}(0, (1 - \xi)P)$ is a function of (V, S) . Note that X_{ξ}^* is independent of N_{ξ}^* . As a result, the equality $h(Y^N | S^N, V^N, F^N) = \mathbb{E}_F \left[\frac{N}{2} \log 2\pi e (\xi P f^2 + \sigma_W^2) \right]$ still holds and hence $Y^* \in \mathcal{A}$, $\mathbb{E}[Y^2] = f^2(P + \sigma_S^2 + 2(\mathbb{E}[X_{\xi}^* S_I] + X_{\xi}^* S_D)) + \sigma_W^2$ and $\mathbb{E}[Y \tilde{S}] = f(\mathbb{E}[X_{\xi}^*(\eta_1 S_I + \eta_2 S_D)] + \mathbb{E}[\eta_1 S_I^2 + \eta_2 S_D^2])$. By the Cauchy-Schwarz inequality, $|\mathbb{E}[X^* S_I]| = |\mathbb{E}[X_{\xi}^* S_I]| \leq \sqrt{\mathbb{E}[(X_{\xi}^*)^2] \mathbb{E}[S_I^2]}$ and $|\mathbb{E}[X^* S_D]| = |\mathbb{E}[X_{\xi}^* S_D]| \leq \sqrt{\mathbb{E}[(X_{\xi}^*)^2] \mathbb{E}[S_D^2]}$. Hence we maximize the value of $\text{MSE}(Y; \tilde{S})$ over X or equivalently over $\mathbb{E}[X_{\xi}^* S_I]$ and $\mathbb{E}[X_{\xi}^* S_D]$ satisfying the above constraints. Finally, the parameters η_1 and η_2 are chosen so that the right hand side of (3.6) is maximized. \square

3.3.2 Linear Scheme

In this section, we assume that the encoder transforms the K dimensional signal V^K into an N dimensional channel input X^N using a linear transformation according to

$$X^N = \alpha(V^K, S^N) = \mathbf{T}V^K + \mathbf{M}S^N \quad (3.12)$$

where \mathbf{T} and \mathbf{M} are $\mathbb{R}^{N \times K}$ and $\mathbb{R}^{N \times N}$ matrices, respectively. In such case, Y^N is conditionally Gaussian given F^N and the MMSE decoder is a linear estimator, with, $\hat{V}^K = \Sigma_{VY} \Sigma_Y^{-1} Y^N$, where $\Sigma_{VY} = \mathbb{E}[(V^K)(Y^N)^T]$ and $\Sigma_Y = \mathbb{E}[(Y^N)(Y^N)^T]$. The matrices \mathbf{T} and \mathbf{M} can be found (numerically) by minimizing the MSE distortion $D_{linear} = \mathbb{E}_F \left[\frac{1}{K} \text{tr} \left\{ \sigma_V^2 \mathbf{I}_{K \times K} - \Sigma_{VY} \Sigma_Y^{-1} \Sigma_{VY}^T \right\} \right]$ under the power constraint in (3.3), where $\text{tr}(\cdot)$ is the trace operator. Note that by setting \mathbf{M} to be the zero matrix and $\mathbf{T} = \sqrt{P/\sigma_V^2} \mathbf{I}_{N \times K}$, the system reduces to the uncoded scheme. Focusing on the matched case ($K = N$), we have the following lemma for finite block length K .

Lemma 3.3. *For the matched-bandwidth source-channel coding of a Gaussian source transmitted over an AWGN fading channel with correlated interference, the distortion lower bound for any linear scheme is achieved with single-letter linear codes.*

Proof. Recall that since V^K and S^K are correlated, we have $S^K = \frac{\rho_{VS}\sigma_S}{\sigma_V} V^K + S_I^K$, where the samples in S_I^K are i.i.d. Gaussian with common variance $\sigma_S^2(1 - \rho_{VS}^2)$. As a result and using (3.12)

$$\begin{aligned} Y^K &= \mathbf{F} \left(\mathbf{T} + \frac{\rho_{VS}\sigma_S}{\sigma_V} \mathbf{M} + \frac{\rho_{VS}\sigma_S}{\sigma_V} \mathbf{I}_{K \times K} \right) V^K + \mathbf{F} (\mathbf{M} + \mathbf{I}_{K \times K}) S_I^K + W^K \\ &= \mathbf{F} \tilde{\mathbf{T}} V^K + \mathbf{F} \tilde{\mathbf{M}} S_I^K + W^K \end{aligned} \quad (3.13)$$

where $\mathbf{F} = \text{diag}(F^K)$ is a diagonal matrix that represents the fading channel, $\tilde{\mathbf{M}} = (\mathbf{M} + \mathbf{I}_{K \times K})$ and $\tilde{\mathbf{T}} = \left(\mathbf{T} + \frac{\rho_{VS}\sigma_S}{\sigma_V} \mathbf{M} + \frac{\rho_{VS}\sigma_S}{\sigma_V} \mathbf{I}_{K \times K} \right)$. After some manipulation, the distortion D_{linear} is given by

$$\begin{aligned} D_{linear} &= \frac{1}{K} \mathbb{E}_F \left[\text{tr} \left\{ \left(\tilde{\mathbf{T}}^T \mathbf{F}^T [\sigma_S^2(1 - \rho_{VS}^2) \mathbf{F} \tilde{\mathbf{M}} \tilde{\mathbf{M}}^T \mathbf{F}^T + \sigma_W^2 \mathbf{I}_{K \times K}]^{-1} \mathbf{F} \tilde{\mathbf{T}} + \sigma_V^{-2} \mathbf{I}_{K \times K} \right)^{-1} \right\} \right] \\ &= \frac{1}{K} \mathbb{E}_F \left[\text{tr} \left\{ (\mathbf{Q} \mathbf{F}^T \mathbf{R} \mathbf{F} + \sigma_V^{-2} \mathbf{I}_{K \times K})^{-1} \right\} \right] \end{aligned} \quad (3.14)$$

where we define $\mathbf{Q} = \tilde{\mathbf{T}}\tilde{\mathbf{T}}^T$, $\mathbf{R} = [\sigma_S^2(1-\rho_{VS}^2)\mathbf{F}\tilde{\mathbf{M}}\tilde{\mathbf{M}}^T\mathbf{F}^T + \sigma_W^2\mathbf{I}_{K\times K}]^{-1}$ and use the fact that for any square matrices \mathbf{A} and \mathbf{B} , $\text{tr}(\mathbf{I} + \mathbf{AB})^{-1} = \text{tr}(\mathbf{I} + \mathbf{BA})^{-1}$ [1]. Now by noting that for any positive-definite $K \times K$ square matrix \mathbf{D} , $\text{tr}(\mathbf{D}^{-1}) \geq \sum_{i=1}^K D_{ii}^{-1}$ [1], where D_{ii} denotes the diagonal elements in \mathbf{D} and equality holds iff \mathbf{D} is diagonal, we can write the following

$$D_{linear} \geq \frac{1}{K} \mathbb{E}_F \left[\sum_{i=1}^K \frac{1}{Q_{ii}|F_{ii}|^2 R_{ii} + \sigma_V^{-2}} \right]. \quad (3.15)$$

Equality in (3.15) holds iff \mathbf{Q} and \mathbf{R} are diagonal; hence the optimal solution gives a diagonal \mathbf{T} and \mathbf{M} . Thus, the linear coding can be achieved in a scalar form without performance loss. \square

3.3.3 Tandem Digital Scheme

Recall that in [29] and as described in Sec. 2.5.1, Gel'fand and Pinsker showed that the capacity of a point-to-point communication with side information (interference) known non-causally at the encoder side is given by

$$C = \max_{p(u,x|s)} I(U; Y) - I(U; S) \quad (3.16)$$

where the maximum is over all joint distributions of the form $p(s)p(u, x|s)p(y|x, s)$ and U denotes an auxiliary random variable. In [12], Costa showed that using $U = X + \alpha S$, with $\alpha = \frac{P}{P + \sigma_W^2}$ over AWGN channel with interference known at the transmitter, the achievable capacity is $C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_W^2} \right)$ which coincides with the capacity of the AWGN channel (no interference). As a result, this choice of U is optimal in terms of maximizing capacity. Next, we adapt the Costa scheme for the fading channel; we

choose $U = X + \alpha S$ as above, where α is redesigned to fit our problem. Using (3.16) and by interpreting the fading F as a second channel output, an achievable rate R is given by

$$R = I(U; Y, F) - I(U; S) = I(U; Y|F) - I(U; S) \quad (3.17)$$

where we use the fact that $I(U; F) = 0$. After some manipulations, the rate R can be expressed as follows

$$R = \mathbb{E}_F \left[\frac{1}{2} \log \left(\frac{P[f^2(P + \sigma_S^2) + \sigma_W^2]}{P\sigma_S^2 f^2(1 - \alpha)^2 + \sigma_W^2(P + \alpha^2\sigma_S^2)} \right) \right]. \quad (3.18)$$

To optimize the value of α , we minimize the expected value of the denominator in (3.18) (i.e., $\mathbb{E}_F[P\sigma_S^2 f^2(1 - \alpha)^2 + \sigma_W^2(P + \alpha^2\sigma_S^2)]$). As a result, we choose $\alpha = \frac{P\mathbb{E}[f^2]}{P\mathbb{E}[f^2] + \sigma_W^2}$ for finite noise levels. Note that this choice of α is independent of S and depends on the second order statistics of the fading. In [74], the authors show that by choosing $\alpha = \frac{P}{P + \sigma_W^2}$, Costa coding maximizes the achievable rate for fading channels in the limits of both high and low noise levels.

The tandem scheme is based on the concatenation of an optimal source code and the adapted Costa coding (described above). The optimal source code quantizes the analog source with a rate close to that in (3.18), and the adapted Costa coder achieves a rate equal to (3.18). Hence, from the lossy JSCC theorem [17], the MSE distortion for a $K : N$ system can be expressed as follows

$$D_{tandem} = \frac{\sigma_V^2}{\exp \left\{ \mathbb{E}_F \left[r \log \left(\frac{P[f^2(P + \sigma_S^2) + \sigma_W^2]}{P\sigma_S^2 f^2(1 - \alpha)^2 + \sigma_W^2(P + \alpha^2\sigma_S^2)} \right) \right] \right\}}. \quad (3.19)$$

Note that due to the purely digital nature of this scheme, its performance does not

improve when the noise level decreases (levelling-off effect) or in the presence of correlation between the source and the interference.

Remark 3.1. *Assuming no fading, the tandem scheme is optimal for the uncorrelated case ($\rho_{VS} = 0$). The system's distortion in (3.19) is then simplified as follows*

$$D_{tandem} = \frac{\sigma_V^2}{\left(1 + \frac{P}{\sigma_W^2}\right)^r}. \quad (3.20)$$

3.4 Distortion Region for the Layered Schemes

In this section, we propose layered schemes based on Wyner-Ziv and HDA coding for transmitting a Gaussian source over a fading channel with correlated interference. These schemes require proper combination of power splitting, bandwidth splitting, rate splitting, Wyner-Ziv and HDA coding. A performance analysis in the presence of noise mismatch is also conducted.

3.4.1 Scheme 1: Layering Costa and HDA Coding for Bandwidth Expansion

This scheme comprises two layers that output X_1^K and X_2^{N-K} . The channel input is obtained by multiplexing (concatenating) the output codeword of both layers $X^N = [X_1^K \ X_2^{N-K}]$ as shown in Fig. 3.2. The first layer is composed of two sublayers that are superimposed to produce the first K samples of the channel input $X_1^K = X_a^K + X_d^K$. The first sublayer is purely analog and consumes an average power of P_a ; the output of this sublayer is given by $X_a^K = \sqrt{a}(\beta_1 V^K + \beta_2 S^K)$, where $\beta_1, \beta_2 \in [-1 \ 1]$ and $a = \frac{P_a}{\beta_1^2 \sigma_V^2 + \beta_2^2 \sigma_S^2 + 2\beta_1 \beta_2 \rho_{VS} \sigma_V \sigma_S}$ is a gain factor related to power constraint P_a , with $0 \leq P_a \leq P$. The second sublayer, that outputs X_d^K and consumes the remaining

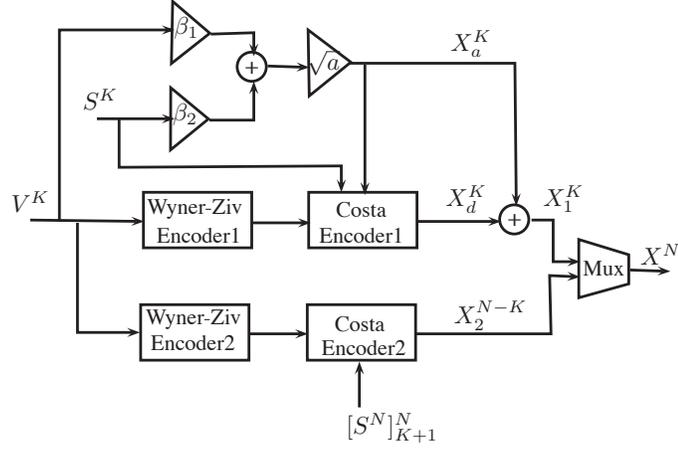


Figure 3.2: Scheme 1 (bandwidth expansion) encoder structure.

power $P_d = P - P_a$, encodes the source V^K using a Wyner-Ziv coder followed by a (generalized) Costa coder. The Wyner-Ziv encoder, which uses the fact that an estimate of V^K can be obtained at the decoder side, forms a random variable T_1^K as follows

$$T_1^K = \alpha_{wz_1} V^K + B_1^K \quad (3.21)$$

where each sample in B_1^K is a zero mean i.i.d. Gaussian, α_{wz_1} and the variance of B_1 are defined later. The encoding process starts by generating a K -length i.i.d. Gaussian codebook \mathcal{T}_1 of size $2^{KI(T_1;V)}$ and randomly assigning the codewords into 2^{KR_1} bins with R_1 defined later. For each source realization V^K , the encoder searches for a codeword $T_1^K \in \mathcal{T}_1$ such that (V^K, T_1^K) are jointly typical. In the case of success, the Wyner-Ziv encoder transmits the bin index of this codeword using Costa coding. The Costa coder, which treats the analog sublayer X_a^K in addition to S^K as interference, forms the following auxiliary random variable

$$U_{c_1}^K = X_d^K + \alpha_{c_1} \check{S}^K \quad (3.22)$$

where $\check{S}^K = (X_a^K + S^K)$, the samples in X_d^K are i.i.d. zero mean Gaussian with variance $P_d = P - P_a$ and $0 \leq \alpha_{c_1} \leq 1$ is a real parameter. Note that X_d^K is independent of V^K and S^K . The encoding process of the Costa coding can be summarized as follows

- *Codebook Generation:* Generate a K -length i.i.d. Gaussian codebook \mathcal{U}_{c_1} with $2^{KI(U_{c_1}; Y_1, F)}$ codewords, where Y_1^K is the first K samples of the received signal Y^N . Every codeword is generated following the random variable $U_{c_1}^K$ and uniformly distributed over 2^{KR_1} bins. The codebook is revealed to both encoder and decoder.
- *Encoding:* For a given bin index (the output of the Wyner-Ziv encoder), the Costa encoder searches for a codeword $U_{c_1}^K$ such that the bin index of $U_{c_1}^K$ is equal to the Wyner-Ziv output and $(U_{c_1}^K, \check{S}^K)$ are jointly typical. In the case of success, the Costa encoder outputs $X_d^K = U_{c_1}^K - \alpha_{c_1} \check{S}^K$. Otherwise, an encoding failure is declared.

The second layer, which outputs X_2^{N-K} , encodes V^K using a Wyner-Ziv with rate R_2 and a Costa coder that treats $[S^N]_{K+1}^N$ as interference. The Wyner-Ziv encoder, which uses the fact that an estimate of V^K is obtained from the first layer, forms the random variable T_2^K as follows

$$T_2^K = \alpha_{wz2} V^K + B_2^K \quad (3.23)$$

where the samples in B_2^K are i.i.d. and follow a zero mean Gaussian distribution, α_{wz2} and the variance of B_2 are defined later. The Costa coder forms the auxiliary random variable $U_{c_2}^{N-K} = X_2^{N-K} + \alpha_{c_2} [S^N]_{K+1}^N$, where the samples in X_2^{N-K} are i.i.d.

zero mean Gaussian with variance P , and the real parameter α_{c_2} is defined later. The encoding process of the Wyner-Ziv and the Costa coder for the second layer is very similar to the one described for the first layer.

At the receiver side, as shown in Fig. 3.3, from the first K components of the received signal $Y^N = [Y_1^K, Y_2^{N-K}] = F^N(X^N + S^N) + W^N$, where $Y_1^K = [Y^N]_1^K$ and $Y_2^{N-K} = [Y^N]_{K+1}^N$, the Costa decoder estimates the codeword $U_{c_1}^K$ by searching for a codeword $U_{c_1}^K$ such that $(U_{c_1}^K, Y_1^K, F^K)$ are jointly typical. By the result of Gel'fand-Pinsker [29] (or Costa [12]) and by treating the fading coefficient F^K as a second channel output, the error probability of encoding and decoding the codeword $U_{c_1}^K$ vanishes as $K \rightarrow \infty$ if

$$\begin{aligned}
R_1 &= I(U_{c_1}; Y_1, F) - I(U_{c_1}; \check{S}) = I(U_{c_1}; Y_1|F) - (h(U_{c_1}) - h(U_{c_1}|\check{S})) \\
&= h(U_{c_1}) + h(Y_1|F) - h(U_{c_1}, Y_1|F) - h(U_{c_1}) + h(U_{c_1}|\check{S}) \\
&= \mathbb{E}_F \left[\frac{1}{2} \log \left(\frac{P_d[f^2(P_d + \sigma_S^2) + \sigma_W^2]}{P_d\sigma_S^2 f^2(1 - \alpha_{c_1})^2 + \sigma_W^2(P_d + \alpha_{c_1}^2\sigma_S^2)} \right) \right] \tag{3.24}
\end{aligned}$$

where $\sigma_S^2 = \mathbb{E}[(X_a + S)^2]$. We then obtain a linear MMSE estimate of V^K (based on Y_1^K and $U_{c_1}^K$), denoted by V_a^K . The distortion from estimating the source using V_a^K is given by

$$D_a = \mathbb{E}_F [\sigma_V^2 - \Gamma\Lambda^{-1}\Gamma^T] \tag{3.25}$$

where $\Lambda = \mathbb{E}[[U_{c_1} \ Y_1]^T[U_{c_1} \ Y_1]]$ is the covariance of $[U_{c_1} \ Y_1]$ and $\Gamma = \mathbb{E}[V[U_{c_1} \ Y_1]]$ is the correlation vector between V and $[U_{c_1} \ Y_1]$. By using rate R_1 on the Wyner-Ziv encoder, the bin index of the Wyner-Ziv can be decoded correctly (with high probability). The Wyner-Ziv decoder then looks for a codeword T_1^K in this bin such that (T_1^K, V_a^K) are jointly typical (as $K \rightarrow \infty$, the probability of error in decoding T_1^K

vanishes). A better estimate of V^K is then obtained based on V_a^K and the decoded codeword T_1^K . The distortion in the estimated source \tilde{V}^K is then

$$\tilde{D} = \frac{D_a}{\exp \left\{ \mathbb{E}_F \left[\log \left(\frac{P_d [f^2 (P_d + \sigma_S^2) + \sigma_W^2]}{P_d \sigma_S^2 f^2 (1 - \alpha_{c_1})^2 + \sigma_W^2 (P_d + \alpha_{c_1}^2 \sigma_S^2)} \right) \right] \right\}}. \quad (3.26)$$

Note that this distortion is equal to the distortion incurred when assuming that the side information V_a^K is also known at the transmitter side; this can be achieved by choosing $\alpha_{wz_1} = \sqrt{1 - \frac{\tilde{D}}{D_a}}$ and $B_1 \sim \mathcal{N}(0, \tilde{D})$ in (3.21) and using a linear MMSE estimator based on V_a^K and T_1^K . In contrast to the AWGN channel with correlated interference [35], a purely analog layer is not sufficient to accommodate for the correlation over AWGN fading channel with correlated interference; indeed using the knowledge of $U_{c_1}^K$ as a side information to obtain a better description of the Wyner-Ziv codewords T_1^K will achieve a better performance. From the last $N - K$ received

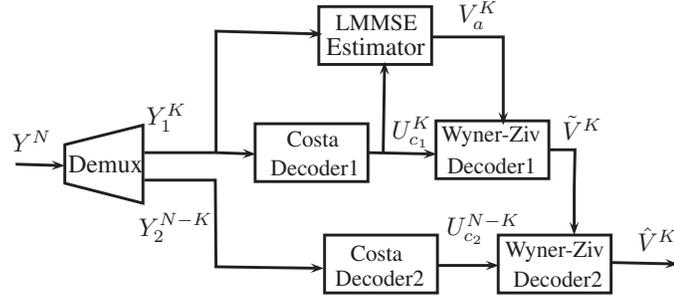


Figure 3.3: Scheme 1 (bandwidth expansion) decoder structure.

symbols Y_2^{N-K} , the Costa decoder estimates the codeword $U_{c_2}^{N-K}$ by searching for a

codeword $U_{c_2}^{N-K}$ such that $(U_{c_2}^{N-K}, Y_2^{N-K}, [F^N]_{K+1}^N)$ are jointly typical. The probability of error in encoding and decoding the codeword $U_{c_2}^{N-K}$ goes to zero by choosing

$$R_2 = I(U_{c_2}; Y_2, F) - I(U_{c_2}; S) = \mathbb{E}_F \left[\frac{1}{2} \log \left(\frac{P[f^2(P + \sigma_S^2) + \sigma_W^2]}{P\sigma_S^2 f^2(1 - \alpha_{c_2})^2 + \sigma_W^2(P + \alpha_{c_2}^2 \sigma_S^2)} \right) \right] \quad (3.27)$$

where $\alpha_{c_2} = P\mathbb{E}[f^2]/(P\mathbb{E}[f^2] + \sigma_W^2)$ is found in a similar way as done in Sec. 3.3.3. By using this rate, the Wyner-Ziv bin index can be decoded correctly (with high probability). The Wyner-Ziv decoder then looks for a codeword T_2^K in the decoded bin such that T_2^K and the side information from the first layer \tilde{V}^K are jointly typical. A refined estimate of the source can be found using the side information \tilde{V}^K and the decoded codeword T_2^K . The resulting distortion is then

$$D_{\text{Scheme 1}} = \inf_{\beta_1, \beta_2, P_a, \alpha_{c_1}} \left\{ \frac{\tilde{D}}{\exp \left\{ \mathbb{E}_F \left[\log \left(\frac{P[f^2(P + \sigma_S^2) + \sigma_W^2]}{P\sigma_S^2 f^2(1 - \alpha_{c_2})^2 + \sigma_W^2(P + \alpha_{c_2}^2 \sigma_S^2)} \right)^{r-1} \right] \right\}} \right\}. \quad (3.28)$$

Note that this distortion is equal to the distortion realized when assuming \tilde{V}^K is also known at the transmitter side; this can be achieved using a linear MMSE estimator based on $[T_1 \ T_2 \ Y_1]$, and by setting $\alpha_{wz_2} = \sqrt{1 - \frac{D_{\text{Scheme 1}}}{D}}$ and $B_2 \sim \mathcal{N}(0, D_{\text{Scheme 1}})$ in (3.23).

Remark 3.2. For AWGN channels with no fading, the same scheme can be used. In this case, the distortion from reconstructing the source can be expressed as follows:

$$D_{\text{Scheme 1}} = \inf_{\beta_1, \beta_2, P_a} \left\{ \frac{D_a}{[(1 + P/\sigma_W^2)^{r-1}(1 + P_d/\sigma_W^2)]} \right\}. \quad (3.29)$$

This distortion can be found by setting the fading coefficient $F = 1$, $\alpha_{c_1} = P_d/(P_d +$

σ_W^2) and $\alpha_{c_2} = P/(P + \sigma_W^2)$ in (3.28). The distortion in (3.29) can be also achieved by replacing the sublayer that outputs X_d^K by an HDA Costa layer. Note that using only Y_1^K as input to the LMMSE estimator in Fig. 3.3 is enough for the AWGN case. In such case, D_a in (3.29) can be simplified as follows

$$D_a = \left(\sigma_V^2 - \frac{(\sqrt{a}\beta_1^2\sigma_V^2 + (\sqrt{a}\beta_2 + 1)\rho_{VS}\sigma_V\sigma_S)^2}{P + (2\sqrt{a}\beta_2 + 1)\sigma_S^2 + 2\sqrt{a}\beta_1\rho_{VS}\sigma_V\sigma_S + \sigma_W^2} \right). \quad (3.30)$$

Moreover, one can check that this scheme is optimal (for the AWGN channel) for $\rho_{VS} = 0$ and $\rho_{VS} = 1$. For $\rho_{VS} = 0$, this happens by shutting down the analog sublayer (i.e., $P_a = 0$) in the scheme and using $(\eta_1 = 1, \eta_2 = 1)$ on the outer bound in (3.4). For the case of $\rho_{VS} = 1$, the optimal power allocation for the scheme is $(P_a = P, P_d = 0)$. The resulting system's distortion can be shown to be equal to the outer bound in (3.4) for $(\eta_1 = 1, \eta_2 = 0)$.

Scheme 1 under mismatch in noise levels: Next, we study the distortion of the proposed scheme in the presence of noise mismatch between the transmitter and the receiver. The actual channel noise power $\sigma_{W_a}^2$ is assumed to be lower than the design one σ_W^2 (i.e., $\sigma_{W_a}^2 < \sigma_W^2$). Under such assumption, the Costa and Wyner-Ziv decoders are still able to decode correctly all codewords with low probability of error. After decoding T_1^K and T_2^K , a symbol-by-symbol linear MMSE estimator of V^K based on Y_1^K , T_1^K and T_2^K is calculated. Hence Scheme 1's distortion under noise mismatch is given by

$$D_{(\text{Scheme 1})-\text{mis}} = \mathbb{E}_F \left[\sigma_V^2 - \Gamma_{1-\text{mis}}^T \Lambda_{1-\text{mis}}^{-1} \Gamma_{1-\text{mis}} \right] \quad (3.31)$$

where $\Lambda_{1-\text{mis}}$ is the covariance matrix of $[T_1 \ T_2 \ Y_1]$, and $\Gamma_{1-\text{mis}}$ is the correlation vector between V and $[T_1 \ T_2 \ Y_1]$. Note that $\sigma_{W_a}^2$ is used in the covariance matrix

Λ_{1-mis} instead of σ_W^2 .

Remark 3.3. *When $\sigma_{W_a}^2 > \sigma_W^2$, all codewords cannot be decoded correctly at the receiver side; as a result we can only estimate the source vector V^K by applying a linear MMSE estimator based on the noisy received signal Y_1^K . The system's distortion in this case is given by*

$$D_{(Scheme\ 1)-mis} = \mathbb{E}_F \left[\sigma_V^2 - \frac{f^2(\sqrt{a}\beta_1^2\sigma_V^2 + (\sqrt{a}\beta_2 + 1)\rho_{VS}\sigma_V\sigma_S)^2}{f^2(P + (2\sqrt{a}\beta_2 + 1)\sigma_S^2 + 2\sqrt{a}\beta_1\rho_{VS}\sigma_V\sigma_S) + \sigma_{W_a}^2} \right]. \quad (3.32)$$

3.4.2 Scheme 2: Layering Costa and HDA Coding for Bandwidth Reduction

In this section, we present a layered scheme for bandwidth reduction. This scheme comprises three layers that are superposed to produce the channel input $X^N = X_a^N + X_1^N + X_2^N$, where X_a^N , X_1^N and X_2^N denote the outputs of the first, second and third layers, respectively. The scheme's encoder structure is depicted in Fig. 3.4. Recall that we denote the first N samples of V^K by V^N and the last $K - N$ samples by $[V^K]_{N+1}^K$. The first layer is an analog layer that outputs $X_a^N = \sqrt{a}(\beta_1 V^N + \beta_2 S^N)$, a linear combination of V^N and S^N , and consumes $P_a \leq P$ as average power, where $\beta_1, \beta_2 \in [-1, 1]$, and $a = \frac{P_a}{\beta_1^2\sigma_V^2 + \beta_2^2\sigma_S^2 + 2\beta_1\beta_2\rho_{VS}\sigma_V\sigma_S}$ is a gain factor related to the power constraint P_a . The second layer, which operates on the first N samples of the source, encodes V^N using a Wyner-Ziv with rate R_1 followed by a Costa coder. The Wyner-Ziv encoder forms a random variable

$$T_1^N = \alpha_{wz1} V^N + B_1^N \quad (3.33)$$

where the samples in B_1^N are i.i.d and follow a zero mean Gaussian distribution, the parameter α_{wz_1} and the variance of B_1 are related to the side information from the first layer and hence defined later. The Costa coder that treats both X_a^N and S^N as interference forms the following auxiliary random variable $U_{c_1}^N = X_1^N + \alpha_{c_1} \check{S}^N$, where the samples in X_1^N are i.i.d. zero mean Gaussian with variance $P_1 \leq P - P_a$ and independent of the source and the interference, $\check{S}^N = X_a^N + S^N$ and $0 \leq \alpha_{c_1} \leq 1$ is a real parameter. The last layer encodes $[V^K]_{N+1}^K$ using an optimal source encoder with rate R_2 followed by a Costa coder. The Costa encoder, which treats the outputs of the first two layers (X_a^N, X_1^N) as well as S^N as known interference, forms the following auxiliary random variable $U_{c_2}^N = X_2^N + \alpha_{c_2} \tilde{S}^N$, where $\tilde{S}^N = (X_a^N + X_1^N + S^N)$, the samples in X_2^N are zero mean i.i.d. Gaussian with variance $P_2 = P - P_1 - P_a$ and $\alpha_{c_2} = P_2 \mathbb{E}[f^2] / (P_2 \mathbb{E}[f^2] + \sigma_W^2)$.

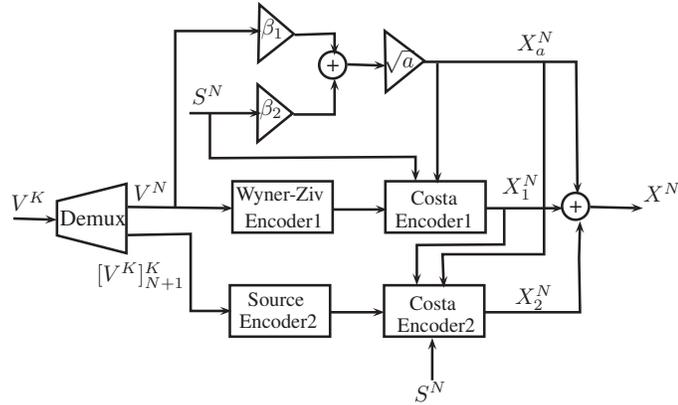


Figure 3.4: Scheme 2 (bandwidth reduction) encoder structure.

At the receiver side, as illustrated in Fig. 3.5, from the received signal Y^N the Costa decoder estimates $U_{c_1}^N$. By using a rate $R_1 = I(U_{c_1}; Y, F) - I(U_{c_1}; \check{S}) = \mathbb{E}_F \left[\frac{1}{2} \log \left(\frac{P_1 [f^2 (P_1 + \sigma_S^2 + P_2) + \sigma_W^2]}{P_1 (\sigma_S^2) f^2 (1 - \alpha_{c_1})^2 + (\sigma_W^2 + f^2 P_2) (P_1 + \alpha_{c_1}^2 \sigma_S^2)} \right) \right]$, where $\sigma_S^2 = \mathbb{E}[(X_a + S)^2]$, the Costa decoder (of the second layer) is able to estimate the codewords $U_{c_1}^N$ with vanishing

error probability. We then obtain an estimate of V^N , denoted by V_a^N , using a linear MMSE estimator based on Y^N and $U_{c_1}^N$. The distortion from estimating V^N using

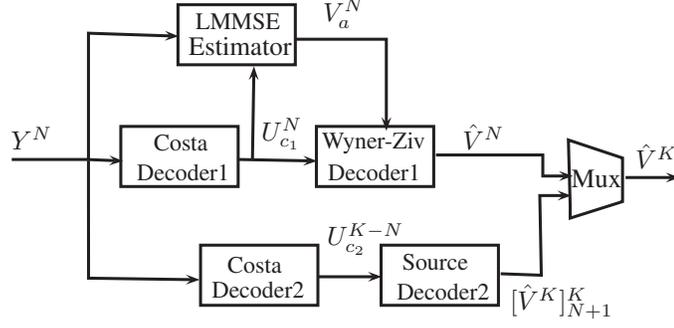


Figure 3.5: Scheme 2 (bandwidth reduction) decoder structure.

V_a^N is then given by

$$D_a = \mathbb{E}_F [\sigma_V^2 - \Gamma \Lambda^{-1} \Gamma^T] \quad (3.34)$$

where Λ is the covariance of $[U_{c_1} \ Y]$ and Γ is the correlation vector between V and $[U_{c_1} \ Y]$. The Wyner-Ziv decoder (of the second layer) then looks for a codeword T_1^N such that (T_1^N, V_a^N) are jointly typical (as $N \rightarrow \infty$, the probability of error in decoding T_1^N vanishes). A better estimate of V^N is then obtained based on the side information V_a^N and the decoded codeword T_1^N . The distortion from reconstructing V^N is then given by

$$D_1^{Scheme\ 2} = \frac{D_a}{\exp \left(\mathbb{E}_F \left[\log \left(\frac{P_1 [f^2 (P_1 + \sigma_S^2 + P_2) + \sigma_W^2]}{P_1 (\sigma_S^2) f^2 (1 - \alpha_{c_1})^2 + (\sigma_W^2 + f^2 P_2) (P_1 + \alpha_{c_1}^2 \sigma_S^2)} \right) \right] \right)}. \quad (3.35)$$

Note that the distortion in (3.35) can be found by choosing $\alpha_{wz_1} = \sqrt{1 - \frac{D_1^{Scheme\ 2}}{D_a}}$ and $B_1 \sim \mathcal{N}(0, D_1^{Scheme\ 2})$ in (3.33) and using a linear MMSE estimator based on V_a^N and T_1^N . To get an estimate of $[V^K]_{N+1}^K$, we use a Costa decoder followed by a source

decoder. Codewords of this layer can be decoded correctly (with high probability) by choosing the rate $R_2 = I(U_{c_2}; Y, F) - I(U_{c_2}; \tilde{S}) = \mathbb{E}_F \left[\frac{1}{2} \log \left(\frac{P_2 [f^2(P_2 + \sigma_S^2) + \sigma_W^2]}{P_2 \sigma_S^2 f^2(1 - \alpha_2)^2 + \sigma_W^2 (P_2 + \alpha_2^2 \sigma_S^2)} \right) \right]$, where $\sigma_S^2 = \mathbb{E}[(X_a + X_1 + S)^2]$. The distortion in reconstructing $[V^K]_{N+1}^K$ can be found by equating the rate-distortion function to the transmission rate R_2 ; this means that $\frac{K-N}{2} \log \frac{\sigma_V^2}{D_2^{Scheme\ 2}} = (N)R_2$. As a result, the distortion in reconstructing $[V^K]_{N+1}^K$, denoted by $D_2^{Scheme\ 2}$, is given by

$$D_2^{Scheme\ 2} = \frac{\sigma_V^2}{\exp \left\{ \mathbb{E}_F \left[\frac{r}{1-r} \log \left(\frac{P_2 [f^2(P_2 + \sigma_S^2) + \sigma_W^2]}{P_2 \sigma_S^2 f^2(1 - \alpha_2)^2 + \sigma_W^2 (P_2 + \alpha_2^2 \sigma_S^2)} \right) \right] \right\}}. \quad (3.36)$$

Hence, the system's distortion is given by

$$D_{Scheme\ 2} = \inf_{\beta_1, \beta_2, P_a, P_1, \alpha_{c_1}} \{r D_1^{Scheme\ 2} + (1-r) D_2^{Scheme\ 2}\}. \quad (3.37)$$

Remark 3.4. For the AWGN channel, the distortion $D_1^{Scheme\ 2}$ and $D_2^{Scheme\ 2}$ for the reduction case are simplified as follows

$$D_1^{Scheme\ 2} = \frac{D_a}{1 + \frac{P_1}{P_2 + \sigma_W^2}} \quad \text{and} \quad D_2^{Scheme\ 2} = \frac{\sigma_V^2}{\left(1 + \frac{P_2}{\sigma_W^2}\right)^{\frac{r}{1-r}}}. \quad (3.38)$$

Since for AWGN channel, the use of $U_{c_1}^N$ as input to the LMMSE estimator in Fig. 3.5 does not improve the performance, the distortion D_a admits a simplified expression as given in (3.30). The distortions in (3.38) can be derived by choosing $\alpha_{c_1} = \frac{P_1}{P_1 + P_2 + \sigma_W^2}$ and $\alpha_{c_2} = \frac{P_2}{P_2 + \sigma_W^2}$. Note that this scheme is optimal for uncorrelated source-interference and for full correlation between the source and the interference. For the uncorrelated case, the analog layer is not needed ($P_a = 0$, $D_a = \sigma_V^2$) and

the optimal power allocation between the two other layers can be derived by minimizing the resulting distortion with respect to P_1 ; the optimal power P_1 is $P_1^* = \sigma_W^2 \left[1 - \left(1 + \frac{P}{\sigma_W^2} \right)^{1-r} \right] + P$. For the case of full correlation between the (first N samples of the) source and the interference ($\rho_{VS} = 1$), the second layer can be shut down ($P_1 = 0$) and the optimal P_a^* satisfies the following equation

$$\begin{aligned} & \sigma_W^2 \left(1 + \frac{\sigma_W}{\sqrt{P_a}} \right) \left(1 + \frac{P - P_a}{\sigma_W^2} \right)^{\frac{1}{1-r}} \left(P + \sigma_W^2 + \sqrt{P_a \sigma_V^2} \right) \\ & - \left(P + \sigma_W^2 + \sigma_V^2 + 2\sqrt{P_a \sigma_V^2} \right)^2 = 0. \end{aligned} \quad (3.39)$$

Scheme 2 under mismatch in noise levels: We next examine the distortion of the proposed scheme in the presence of noise mismatched between the transmitter and the receiver. The actual channel noise power $\sigma_{W_a}^2$ is assumed to be lower than the design one σ_W^2 (i.e., $\sigma_{W_a}^2 < \sigma_W^2$). Under such assumption, the Costa and Wyner-Ziv decoders can decode all codewords with vanishing probability of error. The distortion in reconstructing $[V^K]_{N+1}^K$, $D_{2-mis}^{Scheme\ 2}$, is hence the same as in the matched noise level case; and the distortion from reconstructing V^N is $D_{1-mis}^{Scheme\ 2} = \mathbb{E}_F [\sigma_V^2 - \Gamma^T \Lambda^{-1} \Gamma]$, where Λ is the covariance matrix of $[T_1\ Y]$, and Γ is the correlation vector between V and $[T_1\ Y]$. As a result, the system's distortion is $D_{(Scheme\ 2)-mis} = r D_{1-mis}^{Scheme\ 2} + (1 - r) D_{2-mis}^{Scheme\ 2}$. Note that $\sigma_{W_a}^2$ is used in Λ instead of σ_W^2 when computing $D_{1-mis}^{Scheme\ 2}$.

Remark 3.5. When $\sigma_{W_a}^2 > \sigma_W^2$, all codewords cannot be decoded correctly at the receiver side; as a result we can only estimate the source vector V^N by applying a linear MMSE estimator based on the noisy received signal Y^N . The system's distortion is

then given by

$$D_{(\text{Scheme 2})\text{-mis}} = r\mathbb{E}_F \left[\left(\sigma_V^2 - \frac{f^2(\sqrt{a}\beta_1^2\sigma_V^2 + (\sqrt{a}\beta_2 + 1)\rho_{VS}\sigma_V\sigma_S)^2}{f^2(P + (2\sqrt{a}\beta_2 + 1)\sigma_S^2 + 2\sqrt{a}\beta_1\rho_{VS}\sigma_V\sigma_S) + \sigma_{W_a}^2} \right) \right] + (1-r)\sigma_V^2. \quad (3.40)$$

3.4.3 Numerical Results

In this section, we assume an i.i.d. zero-mean Gaussian source with unitary variance that is transmitted over an AWGN Rayleigh fading channel with Gaussian interference. The interference power is $\sigma_S^2 = 1$, the power constraint is set to $P = 1$ and the Rayleigh fading has $\mathbb{E}[F^2] = 1$. To evaluate the performance, we consider the SDR performance; the design CSNR $\triangleq \frac{P\mathbb{E}[F^2]}{\sigma_W^2}$ is set to 10 dB for all numerical results. Fig. 3.6, which considers the AWGN channel, shows the SDR performance versus the correlation coefficient ρ_{VS} for bandwidth expansion ($r = 2$) and matched noise levels between the transmitter and receiver. We note that the proposed scheme outperforms the tandem Costa reference scheme (described in Sec. 3.3.3) and performs very close to the “best” derived outer bound for a wide range of correlation coefficients. Although not shown, the proposed scheme also outperforms significantly the linear scheme of Sec. 3.3.2. For the limiting cases of $\rho_{VS} = 0$ and 1, we can notice that the SDR performance of the proposed scheme coincides with the outer bound and hence is optimal.

Figs. 3.7, 3.8 and 3.9 show the SDR performance versus ρ_{VS} for the fading channel with interference under matched noise levels and for $r = 1, 2$ and $1/2$, respectively. As in the case of the AWGN channel, we remark that the proposed HDA schemes outperform the tandem Costa and the linear schemes and perform close to the best

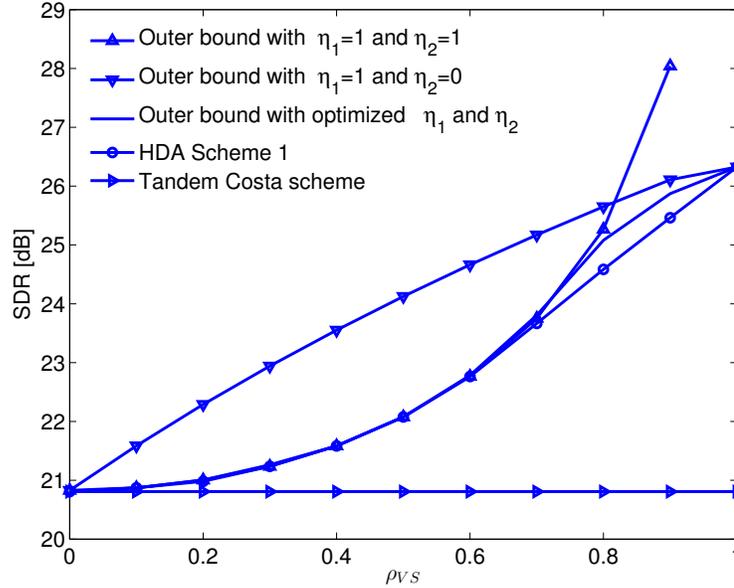


Figure 3.6: Performance of HDA Scheme 1 ($r = 2$) over the AWGN channel under matched noise levels for different correlation coefficient values, $P = 1$, $\sigma_S^2 = 1$ and CSNR=10 dB (as given by (3.28)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.4), respectively.

outer bound. In contrast to the AWGN case, the proposed scheme never coincides with the outer bound for finite noise levels. Using the result in [74], one can easily show that our schemes are optimal for $\rho_{VS} = 0$ in the limits of high and low noise levels. As a result, the auxiliary random variable used for the Costa coder is optimal in the noise level limits.

Fig. 3.10 shows the SDR performance versus CSNR levels under mismatched noise levels. All schemes in Fig. 3.10 are designed for CSNR=10 dB, $r = 1$ and $\rho_{VS} = 0.7$. The true CSNR varies between 0 and 35 dB. We observe that the proposed scheme is resilient to noise mismatch due to its hybrid digital-analog nature. As the correlation coefficient values decreases, the power allocated to the analog layer

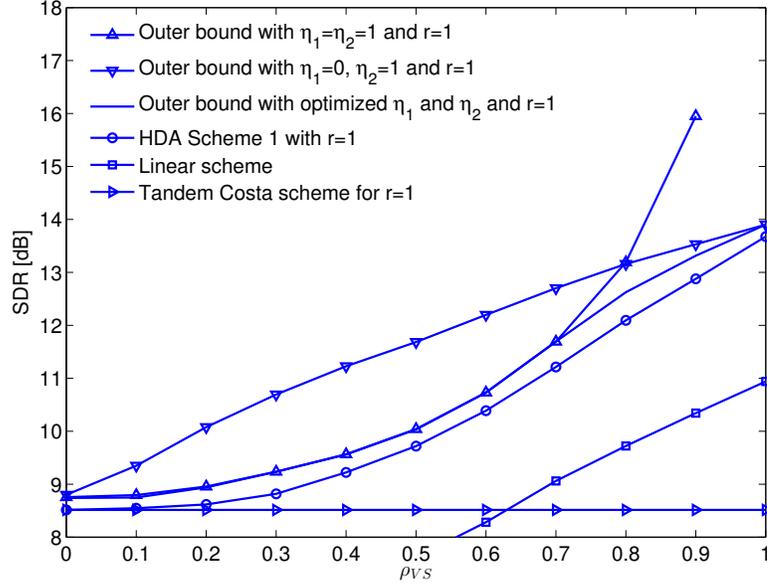


Figure 3.7: Performance of Scheme 1 ($r = 1$) over the fading channel under matched noise levels for different correlation coefficient, $P = 1$, $\sigma_{\xi}^2 = 1$, CSNR=10 dB and $\mathbb{E}[F^2] = 1$ (as given by (3.28)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.4), respectively.

decreases. Hence, the SDR gap between the proposed and the tandem Costa scheme under mismatched noise levels decreases and the robustness (which is the trait of analog schemes) reduces.

3.5 JSCC for Source-Channel-State Transmission

As an application of the joint source-channel coding problem examined in this chapter we consider the transmission of analog source-channel-state pairs over a fading channel with Gaussian state interference. We establish inner and outer bounds on the source-interference distortion for the fading channel. The only difference between this problem and that examined in the previous sections is that the receiver is also

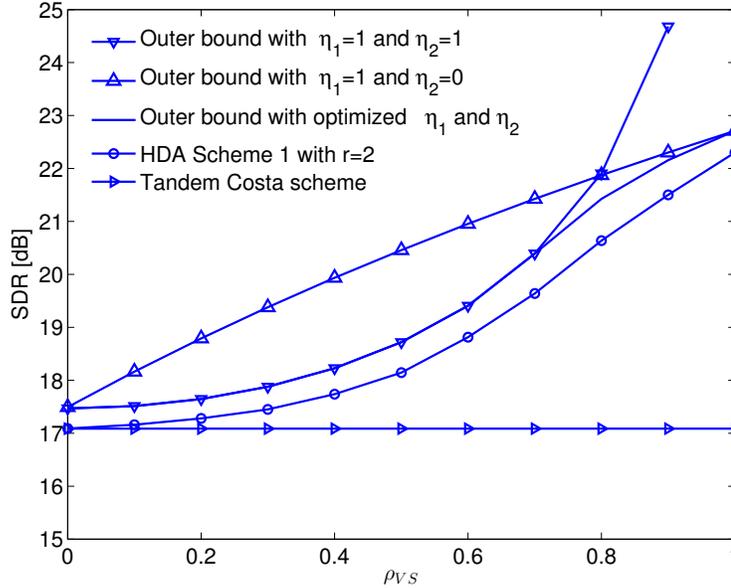


Figure 3.8: Performance of Scheme 1 ($r = 2$) over the fading channel under matched noise levels for different correlation coefficient ρ_{VS} , $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB and $\mathbb{E}[F^2] = 1$ (as given by (3.28)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.4), respectively.

interested in estimating the interference S^N . For simplicity, we focus on the matched bandwidth case (i.e., $K = N$); the unequal source-channel bandwidth case can be treated in a similar way as in Section 3.4. We also assume that the decoder has knowledge of the fading. We denote the distortion from reconstructing the source and the interference by $D_v = \frac{1}{K}\mathbb{E}[|V^K - \hat{V}^K|^2]$ and $D_s = \frac{1}{N}\mathbb{E}[|S^N - \hat{S}^N|^2]$, respectively. For a given power constraint P , a rate r and a Rayleigh fading channel, the distortion region is defined as the closure of all distortion pair (D_v^o, D_s^o) for which (P, D_v^o, D_s^o) is achievable, where a power-distortion triple is achievable if for any $\delta_v, \delta_s > 0$, there exist sufficiently large integers K and N with $N/K = r$, encoding and decoding functions satisfying (3.3), such that $D_v < D_v^o + \delta_v$ and $D_s < D_s^o + \delta_s$.

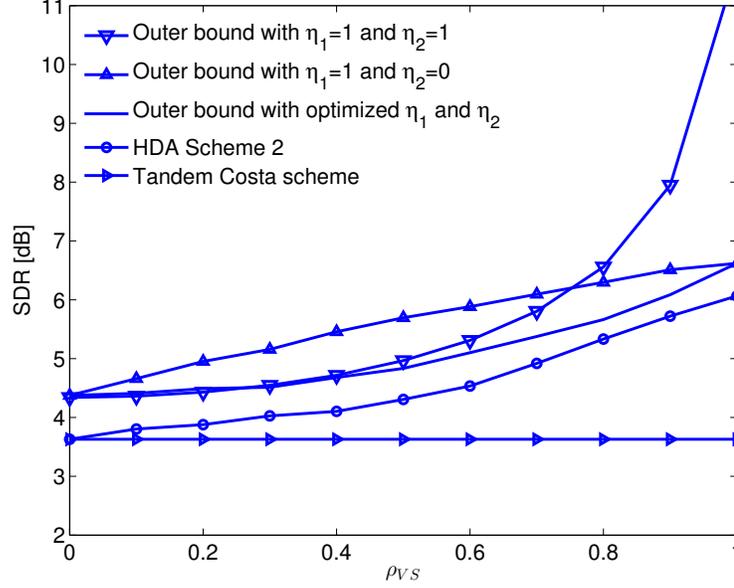


Figure 3.9: Performance of Scheme 2 ($r = 1/2$) over the fading channel under matched noise levels for different ρ_{VS} , $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB and $\mathbb{E}[F^2] = 1$ as given by (3.37)). Tandem scheme and outer bounds on SDR are plotted using (3.19) and (3.6), respectively.

3.5.1 Outer Bound

Lemma 3.4. *For the matched bandwidth case, the outer bound on the distortion region (D_v, D_s) can be expressed as follows*

$$D_v \geq \frac{\text{Var}(V|S)}{\exp \left\{ \mathbb{E}_F \left[\log \frac{\zeta P |f|^2 + \sigma_W^2}{\sigma_W^2} \right] \right\}}, \quad D_s \geq \frac{\sigma_S^2}{\exp \left\{ \mathbb{E}_F \left[\log \frac{|f|^2 (P + \sigma_S^2 + 2\sqrt{(1-\zeta)P\sigma_S^2}) + \sigma_W^2}{\zeta P |f|^2 + \sigma_W^2} \right] \right\}} \quad (3.41)$$

where $\text{Var}(V|S) = \sigma_V^2(1 - \rho_{VS}^2)$ is the variance of V given S and $0 \leq \zeta \leq 1$.

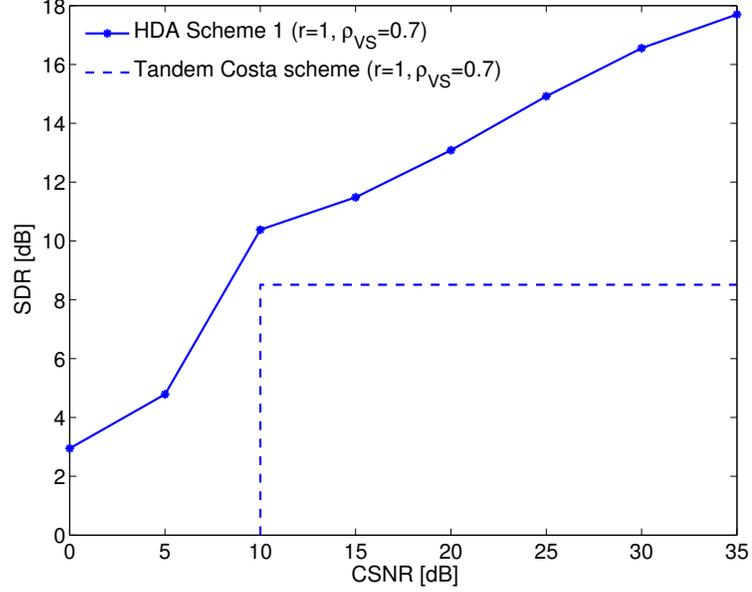


Figure 3.10: Performance of Scheme 1 ($r = 1$) over the fading channel under mismatched noise levels for $P = 1$, $\sigma_S^2 = 1$, CSNR=10 dB, $\rho_{VS} = 0.7$ and $\mathbb{E}[F^2] = 1$.

Proof. For the source distortion, we can write the following

$$\begin{aligned}
\frac{K}{2} \log \frac{\sigma_V^2}{D_v} &\stackrel{(a)}{\leq} I(V^K; \hat{V}^K | F^K) \stackrel{(b)}{\leq} I(V^K; \hat{V}^K | F^K) + I(V^K; S^K | \hat{V}^K, F^K) \\
&= I(V^K; \hat{V}^K, S^K | F^K) \stackrel{(c)}{=} I(V^K; S^K) + I(V^K; \hat{V}^K | S^K, F^K) \\
&\stackrel{(d)}{\leq} \frac{K}{2} \log \frac{\sigma_V^2}{\text{Var}(V|S)} + I(V^K; Y^K | S^K, F^K) \\
&= \frac{K}{2} \log \frac{\sigma_V^2}{\text{Var}(V|S)} + h(Y^K | S^K, F^K) - h(W^K) \\
&\stackrel{(e)}{=} \frac{K}{2} \log \frac{\sigma_V^2}{\text{Var}(V|S)} + \frac{K}{2} \mathbb{E}_F \left[\log \frac{\zeta P |f|^2 + \sigma_W^2}{\sigma_W^2} \right] \tag{3.42}
\end{aligned}$$

where (a) follows from the rate-distortion theorem, (b) follows from the non-negativity of mutual information, (c) follows from the chain rule of mutual information and the

fact that F^K is independent of (V^K, S^K) , (d) holds by the data processing inequality and in (e) we used $h(Y^K|S^K, F^K) = \frac{K}{2}\mathbb{E}_F[\log(\zeta P|f|^2 + \sigma_W^2)]$ for some $\zeta \in [0, 1]$; this can be proved from the fact that $\frac{K}{2}\log\sigma_W^2 = h(W^K) \leq h(Y^K|S^K, F^K) \leq h(F^K X^K + W^K|F^K) = \frac{K}{2}\mathbb{E}_F[\log(P|f|^2 + \sigma_W^2)]$. Hence, there exists a $\zeta \in [0, 1]$ such that $h(Y^K|S^K, F^K) = \frac{K}{2}\mathbb{E}_F[\log(\zeta P|f|^2 + \sigma_W^2)]$.

For the interference distortion, we have the following

$$\begin{aligned}
 \frac{K}{2}\log\frac{\sigma_S^2}{D_s} &\stackrel{(a)}{\leq} I(S^K; \hat{S}^K|F^K) \stackrel{(b)}{\leq} I(S^K; Y^K|F^K) = h(Y^K|F^K) - h(Y^K|S^K, F^K) \\
 &\stackrel{(c)}{\leq} \sup_{X \in \mathcal{B}} \mathbb{E}_F \left[\frac{K}{2}\log 2\pi e(|f|^2(P + \sigma_S^2 + 2\mathbb{E}[SX]) + \sigma_W^2) \right] \\
 &\quad - \mathbb{E}_F \left[\frac{K}{2}\log 2\pi e(\zeta P|f|^2 + \sigma_W^2) \right] \\
 &\stackrel{(d)}{=} \mathbb{E}_F \left[\frac{K}{2}\log \frac{|f|^2(P + \sigma_S^2 + 2\sqrt{(1-\zeta)P\sigma_S^2}) + \sigma_W^2}{\zeta P|f|^2 + \sigma_W^2} \right] \tag{3.43}
 \end{aligned}$$

where (a) follows from the rate-distortion theorem, (b) follows from data processing inequality for the mutual information, in (c) the set $\mathcal{B} = \{X : h(Y^K|S^K, F^K) = \mathbb{E}_F[\frac{K}{2}\log 2\pi e(\zeta P|f|^2 + \sigma_W^2)]\}$ and the inequality in (c) holds from the fact that Gaussian maximizes differential entropy and $h(Y^K|S^K, F^K) = \frac{K}{2}\mathbb{E}_F[\log(\zeta P|f|^2 + \sigma_W^2)]$ (as used in (3.42)). Note that the supremum over X in (c) happens when Y, S and W are jointly Gaussian given F (i.e., X^* is Gaussian). Now, we represent $X^* = N_\zeta^* + X_\zeta^*$, where $N_\zeta^* \sim \mathcal{N}(0, \zeta P)$ is independent of $X_\zeta^* \sim \mathcal{N}(0, (1-\zeta)P)$. Note that X_ζ^* is a function of S . Using this form of X^* , $h(Y^K|S^K, F^K) = \frac{K}{2}\mathbb{E}_F[\log(\zeta P|f|^2 + \sigma_W^2)]$ still holds (i.e., $X^* \in \mathcal{B}$) and $\mathbb{E}[X^*S] = \mathbb{E}[X_\zeta^*S]$. Maximizing over X is equivalent to maximizing over $\mathbb{E}[X_\zeta^*S]$; using Cauchy-Schwarz $(\mathbb{E}[X^*S] = \mathbb{E}[X_\zeta^*S] \leq \sqrt{\mathbb{E}[(X_\zeta^*)^2]\mathbb{E}[S^2]})$ we get (d). □

3.5.2 Proposed Hybrid Coding Scheme

The proposed scheme is composed of three layers as shown in Fig. 3.11. The first layer, which is purely analog, consumes an average power P_a and outputs a linear combination of the source and the interference $X_a^K = \sqrt{a_1}(\beta_{11}V^K + \beta_{12}S^K)$, where $\beta_{11}, \beta_{12} \in [-1, 1]$ and $a_1 = \frac{P_a}{(\beta_{11}^2\sigma_V^2 + 2\beta_{11}\beta_{12}\rho_{VS}\sigma_V\sigma_S + \beta_{12}^2\sigma_S^2)}$ is a gain factor related to power constraint P_a . The second layer employs a source-channel vector-quantizer (VQ) on the interference; the output of this layer is $X_q^K = \mu(S^K + U_q^K)$, where $\mu > 0$ is a gain related to the power constraint and samples in U_q^K follow a zero mean i.i.d. Gaussian that is independent of V and S and has a variance Q . A similar VQ encoder was used in [64] for the broadcast of bivariate sources and for the multiple access channel [46]. In what follows, we outline the encoding process of the VQ

- *Codebook Generation:* Generate a K -length i.i.d. Gaussian codebook \mathcal{X}_q with 2^{KR_q} codewords with R_q defined later. Every codeword is generated following the random variable X_q^K ; this codebook is revealed to both the encoder and decoders.
- *Encoding:* The encoder searches for a codeword X_q^K in the codebook that is jointly typical with S^K . In case of success, the transmitter sends X_q^K .

The last layer encodes a linear combination of V^K and S^K , denoted by \tilde{X}_{wz}^K , using a Wyner-Ziv with rate R followed by a Costa coder. The Costa coder uses an average power of P_d and treats X_a^K , S^K and X_q^K as known interference. The linear combination is given by $\tilde{X}_{wz}^K = \tilde{\beta}_{21}V^K + \tilde{\beta}_{22}S^K = \sqrt{a_2}(\beta_{21}V^K + \beta_{22}S^K)$, where $\beta_{21}, \beta_{22} \in [-1, 1]$ and $a_2 = \frac{P_d}{(\beta_{21}^2\sigma_V^2 + 2\beta_{21}\beta_{22}\rho_{VS}\sigma_V\sigma_S + \beta_{22}^2\sigma_S^2)}$. The Wyner-Ziv encoder forms

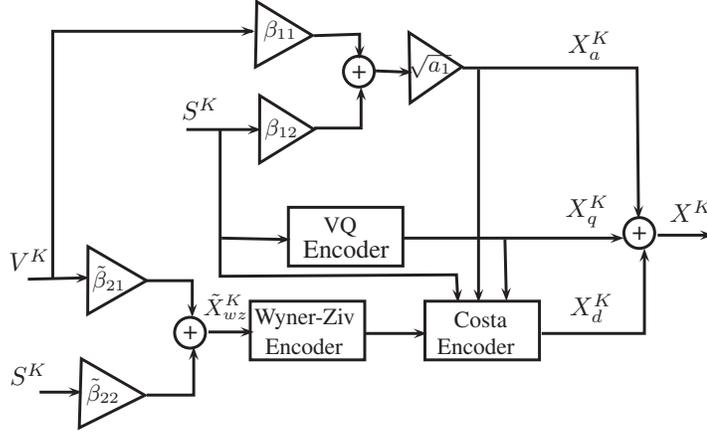


Figure 3.11: Encoder structure for source-channel-state transmission.

a random variable T^K as follows

$$T^K = \alpha_{wz} \tilde{X}_{wz}^K + B^K \quad (3.44)$$

where the samples in B^K are zero mean i.i.d. Gaussian, the parameter α_{wz} and the variance of B are defined later. The encoding process of the Wyner-Ziv starts by generating a K -length i.i.d. Gaussian codebook \mathcal{T} of size $2^{KI(T; \tilde{X}_{wz})}$ and randomly assigning the codewords into 2^{KR} bins with R defined later. For each realization \tilde{X}_{wz}^K , the Wyner-Ziv encoder searches for a codeword $T^K \in \mathcal{T}$ such that (\tilde{X}_{wz}^K, T^K) are jointly typical. In the case of success, the Wyner-Ziv encoder transmits the bin index of this codeword using Costa coding. The Costa coder, that treats $\tilde{S}^K = X_a^K + X_q^K + S^K$ as known interference, forms the following auxiliary random variable $U_c^K = X_d^K + \alpha_c \tilde{S}^K$, where each sample in X_d^K is $\mathcal{N}(0, P_d)$ that is independent of the source and the interference and $0 \leq \alpha_c \leq 1$. The encoding process for the Costa coder can be described in a similar way as done before.

At the receiver side, as shown in Fig. 3.12, from the noisy received signal Y^K , the VQ decoder estimates X_q^K by searching for a codeword $X_q^K \in \mathcal{X}_q$ that is jointly typical with the received signal Y^K and F^K . Following the result in [64] and the error analysis of [48], the error probability of decoding X_q^K goes to zero by choosing the rate R_q to satisfy $I(S; X_q) \leq R_q \leq I(X_q; Y, F)$, where

$$\begin{aligned}
 I(S; X_q) &= h(X_q) - h(X_q|S) = \frac{1}{2} \log \frac{\sigma_S^2 + Q}{Q} \\
 I(X_q; Y, F) &= I(X_q; F) + I(X_q; Y|F) = h(Y|F) - h(Y|X_q, F) \\
 &= \mathbb{E}_F \left\{ \frac{1}{2} \log 2\pi e (\mathbb{E}[Y^2]) - \frac{1}{2} \log 2\pi e \left(\mathbb{E}[Y^2] - \frac{\mathbb{E}[X_q Y]^2}{\mathbb{E}[X_q^2]} \right) \right\}.
 \end{aligned} \tag{3.45}$$

The variance Q has to be chosen to satisfy the above rate constraint. Furthermore, to ensure the power constraint is satisfied, we need μ to satisfy the following equation

$$P_a + \mu^2(\sigma_S^2 + Q) + 2\mu\mathbb{E}[SX_a] + P_d \leq P. \tag{3.46}$$

The Costa decoder then searches for a codeword U_c^K that is jointly typical with (Y^K, F^K) . Since the received signal Y^K and the codewords X_q^K and U_c^K are correlated with \tilde{X}_{wz}^K , an LMMSE estimate of \tilde{X}_{wz}^K , denoted by $\hat{\tilde{X}}_{wz}^K$, can be obtained based on Y^K and the decoded codewords X_q^K and U_c^K . Mathematically, the estimate is given by $\hat{\tilde{X}}_{wz} = \Gamma_a \Lambda_a^{-1} [X_q \ U_c \ Y]^T$, where Λ_a is the covariance of $[X_q \ U_c \ Y]$ and Γ_a is the correlation vector between \tilde{X}_{wz} and $[X_q \ U_c \ Y]$. The distortion in reconstructing \tilde{X}_{wz}^K is then given by

$$D_a = \mathbb{E}_F [P_d - \Gamma_a \Lambda_a^{-1} \Gamma_a^T]. \tag{3.47}$$

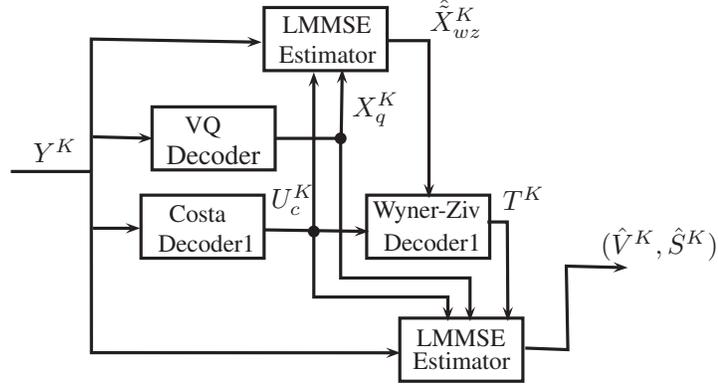


Figure 3.12: Decoder structure for source-channel-state transmission.

Moreover, the Wyner-Ziv decoder estimates the codeword T^K by searching for a $T^K \in \mathcal{T}$ that is jointly typical with \hat{X}_{wz}^K . The error probability of decoding both codewords T^K and U_c^K vanishes as $K \rightarrow \infty$ if the coding rate of the Wyner-Ziv and the Costa coder is set to

$$R = \mathbb{E}_F \left[\frac{1}{2} \log \left(\frac{P_d[f^2(P_d + \sigma_S^2) + \sigma_W^2]}{P_d(\sigma_S^2)f^2(1 - \alpha_c)^2 + \sigma_W^2(P_d + \alpha_c^2\sigma_S^2)} \right) \right] \quad (3.48)$$

where $\sigma_S^2 = \mathbb{E}[(X_a + X_q + S)^2]$. A better estimate of \tilde{X}_{wz}^K can be obtained by using the codeword T^K and \hat{X}_{wz}^K . The distortion in reconstructing \tilde{X}_{wz}^K can be expressed as follows

$$\tilde{D} = \frac{D_a}{\exp \left\{ \mathbb{E}_F \left[\log \left(\frac{P_d[f^2(P_d + \sigma_S^2) + \sigma_W^2]}{P_d(\sigma_S^2)f^2(1 - \alpha_c)^2 + \sigma_W^2(P_d + \alpha_c^2\sigma_S^2)} \right) \right] \right\}}. \quad (3.49)$$

This distortion can be achieved using a linear MMSE estimate based on T^K , X_q^K and Y^K by choosing $\alpha_{wz} = \sqrt{1 - \frac{\tilde{D}}{D_a}}$ and $B \sim \mathcal{N}(0, \tilde{D})$ in (3.44).

After decoding T^K , X_q^K , a linear MMSE estimator is used to reconstruct the source and the interference signals. As a result, the distortion in decoding V^K and

S^K are given as follows

$$D_v^{Hybrid} = \mathbb{E}_F [\sigma_V^2 - \Gamma_v \Lambda^{-1} \Gamma_v^T] \quad D_s^{Hybrid} = \mathbb{E}_F [\sigma_S^2 - \Gamma_s \Lambda^{-1} \Gamma_s^T] \quad (3.50)$$

where Λ is the covariance of $[X_q \ T \ Y]$, Γ_v is the correlation vector between V and $[X_q \ T \ Y]$ and Γ_s is the correlation vector between S and $[X_q \ T \ Y]$.

Remark 3.6. *For the AWGN channel with $\rho_{VS} = 0$, using the source itself instead of \tilde{X}_{wz}^K as input to the Wyner-Ziv encoder, shutting down the second layer and setting $\beta_{11} = 0$ in X_a^K give the best possible performance; the inner bound in such case coincides with the outer bound, hence the scheme is optimal. This result is analogous to the optimality result of the rate-state-distortion for the transmission of a finite discrete source over a Gaussian state interference derived in [63].*

3.5.3 Numerical Results

We consider source-interference pairs that are transmitted over a Rayleigh fading channel ($\mathbb{E}[F^2] = 1$) with Gaussian interference and power constraint $P = 1$; the CSNR level is set to 10 dB. For reference, we adapt the scheme of [63] to our scenario. Recall that the source and the interference are jointly Gaussian, hence $V^K = \rho_{VS} \frac{\sigma_V}{\sigma_S} S^K + S_I^K$, where samples in S_I^K are i.i.d. Gaussian with variance $\sigma_{S_I}^2 = (1 - \rho_{VS}^2) \sigma_V^2$ and independent of S^K . Now if we quantize S_I^K into digital data, the setup becomes similar to the one considered in [63]; hence the encoding is done by allocating a portion of the power, denoted by P_s , to transmit S^K and the remaining power $P_d = (P - P_s)$ is used to communicate the digitized S_I^K using the

(generalized) Costa coder. The received signal of such scheme is given by

$$Y^K = F^K \left(\sqrt{\frac{P_s}{\sigma_S^2}} S^K + X_d^K + S^K \right) + W^K \quad (3.51)$$

where X_d^K denotes the output of the digital part that communicates S_I^K . An estimate of S^K is obtained by applying an LMMSE estimator on the received signal; the distortion from reconstructing V^K is equal to the sum of the distortions from estimating $\rho_{VS} \frac{\sigma_V}{\sigma_S} S^K$ and S_I^K . Mathematically, the distortion region of such reference scheme can be expressed as follows

$$\begin{aligned} D_s^{Reference} &= \mathbb{E}_F \left[\sigma_S^2 - \frac{\mathbb{E}[SY]^2}{\mathbb{E}[Y^2]} \right] = \mathbb{E}_F \left[\sigma_S^2 - \frac{f^2(\sqrt{P_s}\sigma_S + \sigma_S^2)^2}{f^2(P + \sigma_S^2 + 2\sqrt{P_s}\sigma_S) + \sigma_W^2} \right], \\ D_v^{Reference} &= \rho_{VS}^2 \frac{\sigma_V^2}{\sigma_S^2} D_s + \frac{\sigma_{S_I}^2}{\exp \left\{ \mathbb{E}_F \left[\log \left(\frac{P_d[f^2(P_d + \sigma_S^2) + \sigma_W^2]}{P_d(\sigma_S^2)f^2(1-\alpha_c)^2 + \sigma_W^2(P_d + \alpha_c^2\sigma_S^2)} \right) \right] \right\}} \end{aligned} \quad (3.52)$$

where $\sigma_S^2 = \mathbb{E}[(\sqrt{P_s}/\sigma_S^2 S + S)^2]$ and the parameter α_c is related to the Costa coder. To evaluate the performance, we plot the outer bound (given by (3.41)) and the inner bounds (the achievable distortion region) of the proposed hybrid coding (given by (3.50)) and the reference scheme (adapted scheme of [63]). Fig. 3.13, which considers the AWGN channel, shows the distortion regions of the source-interference pair for $\rho_{VS} = 0.8$ and $\sigma_S^2 = 0.5$. We can notice that the hybrid coding scheme is very close to the outer bound and outperforms the scheme of [63]. Fig. 3.14, which considers the fading channel, shows the distortion regions of the source-interference pair for $\rho_{VS} = 0.8$ and $\sigma_S^2 = 1$. The hybrid coding scheme performs relatively close to the outer bound.

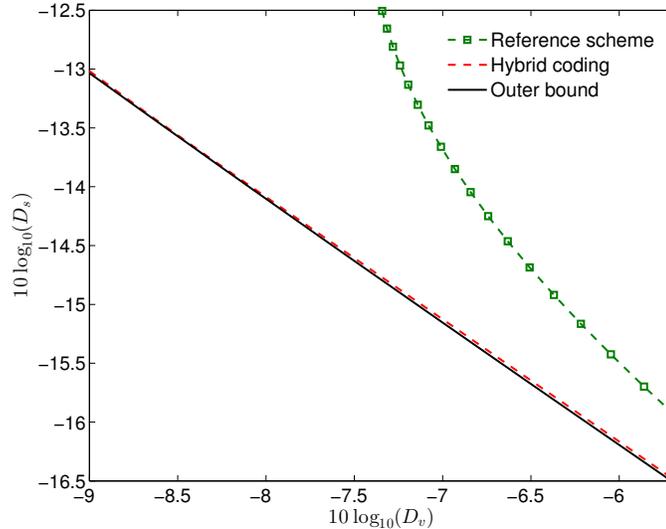


Figure 3.13: Distortion region for hybrid coding scheme over the AWGN channel for $\sigma_V^2 = 1$, $P = 1$, $\sigma_S^2 = 0.5$ and $\rho_{VS} = 0.8$.

3.6 Summary and Conclusions

In this chapter, we consider the problem of reliable transmission of Gaussian sources over Rayleigh fading channels with correlated interference under unequal source-channel bandwidth. Inner and outer bounds on the system's distortion are derived. The outer bound is derived by assuming additional knowledge at the decoder side; while the inner bound is found by analyzing the achievable distortion region of the proposed hybrid coding scheme. Numerical results show that the proposed schemes perform close to the derived outer bound and to be robust to channel noise mismatch. As an application of the proposed schemes, we derive inner and outer bounds on the source-channel-state distortion region for the fading interference channel; in this case, the receiver is interested in estimating both source and interference. Our

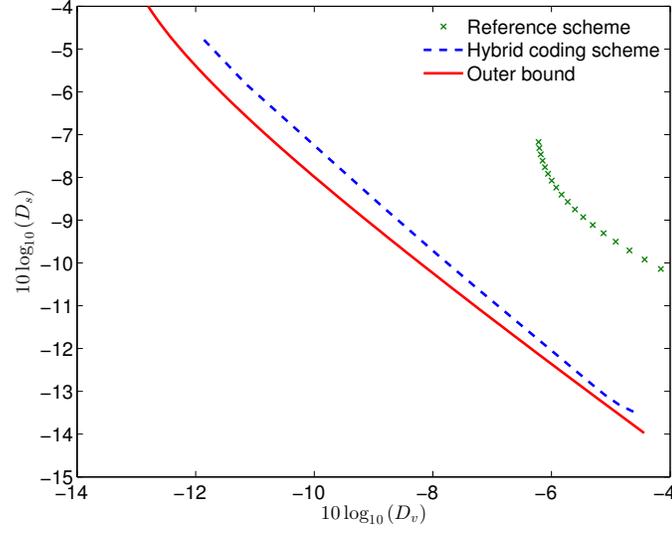


Figure 3.14: Distortion region for hybrid coding scheme over the fading channel for $\sigma_V^2 = 1$, $P = 1$, $\sigma_S^2 = 1$, $\rho_{VS} = 0.8$ and $\mathbb{E}[F^2] = 1$.

setting contains several interesting limiting cases. In the absence of fading and/or correlation and for some source-channel bandwidths, our setting resorts to the scenarios considered in [35, 63, 67].

Chapter 4

Distortion Bounds for Broadcasting Bivariate Gaussian Sources in the Presence of Interference

4.1 Introduction

For multi-terminal systems, tandem coding is no longer optimal; a JSCC scheme may be required to achieve optimality. One simple scenario where the tandem scheme is suboptimal concerns the broadcast of Gaussian sources over Gaussian channels [28]. For a single Gaussian source sent over a Gaussian broadcast channel with matched source-channel bandwidth, the distortion region is known and can be realized by a linear scheme [28]. For mismatched source-channel bandwidth, the best known coding schemes are based on JSCC with hybrid signalling [50, 54, 56, 61]. One extension to this problem is the broadcasting of two correlated Gaussian sources to two users, each of which is interested in recovering one of the two sources; in [45], it is proven that the linear scheme is optimal when the system's signal-to-noise ratio is below a certain threshold under matched bandwidth. In [64], a hybrid digital-analog scheme is proposed for the same matched bandwidth system and is shown to be optimal whenever the linear scheme of [45] is not; hence providing a complete characterization

of the distortion region. Under mismatched bandwidth, various HDA schemes are proposed in [5], consisting of different combinations of several known schemes using either superposition or dirty paper (Costa) coding. Recently, in [26], a tandem scheme based on successive coding is studied and shown to outperform the HDA schemes of [5].

In this chapter, we consider the transmission of two correlated sources over Gaussian broadcast channel in the presence of interference, where the interference is assumed to be correlated with the sources. We propose and analyze HDA schemes for this system based on Wyner-Ziv [71], Costa [12] and HDA Costa coding [70]. The rest of the chapter is organized as follows: Section 4.2 presents the problem formulation. Section 4.3 introduces an outer bound on the system's distortion region and some reference schemes. In Section 4.4, inner bounds on the distortion region under matched and expansion bandwidth are studied by proposing HDA schemes. Numerical results are included in Section 4.5. Finally, conclusions are drawn in Section 4.6.

4.2 Problem Formulation

We consider the transmission (Fig. 4.1) of a pair of correlated Gaussian sources (V_1^K, V_2^K) over a two-user Gaussian broadcast channel in the presence of Gaussian interference S^N known non-causally to the transmitter. User i ($i = 1, 2$) receives the transmitted signal corrupted by additive white Gaussian noise W_i^N and interference S^N , where each sample in S^N is drawn from an i.i.d. Gaussian distribution with variance σ_S^2 . Each user i aims to estimate $V_i^K = (V_i(1), V_i(2), \dots, V_i(K))$, where each sample $V_i(j), j = 1, \dots, K$, is drawn from an i.i.d. Gaussian with variance $\sigma_{V_i}^2$. Herein,

we assume that $(V_1(i), V_2(i), S(i))$, $i = 1, \dots, K$, are correlated via the following covariance matrix

$$\Sigma_{V_1 V_2 S} = \begin{bmatrix} \sigma_{V_1}^2 & \rho_{V_1 V_2} \sigma_{V_1} \sigma_{V_2} & \rho_{V_1 S} \sigma_{V_1} \sigma_S \\ \rho_{V_1 V_2} \sigma_{V_1} \sigma_{V_2} & \sigma_{V_2}^2 & \rho_{V_2 S} \sigma_{V_2} \sigma_S \\ \rho_{V_1 S} \sigma_{V_1} \sigma_S & \rho_{V_2 S} \sigma_{V_2} \sigma_S & \sigma_S^2 \end{bmatrix} \quad (4.1)$$

where $\rho_{V_1 V_2}$, $\rho_{V_1 S}$ and $\rho_{V_2 S}$ are the correlation coefficients between V_1 and V_2 , S and V_1 and S and V_2 , respectively. The covariance matrix in (4.1) being positive definite restricts the possible values of $\rho_{V_1 V_2}$, $\rho_{V_1 S}$ and $\rho_{V_2 S}$.

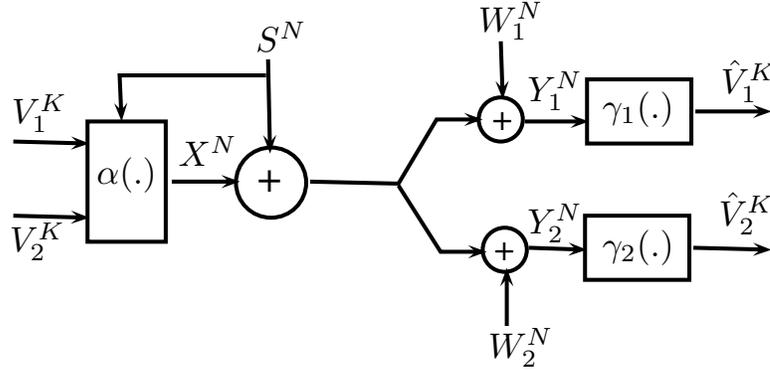


Figure 4.1: System model structure.

As shown in Fig. 4.1, the source pair vector (V_1^K, V_2^K) is transformed into an N dimensional channel input $X^N \in \mathbb{R}^N$ via $\alpha(\cdot)$, a mapping from $(\mathbb{R}^K \times \mathbb{R}^K \times \mathbb{R}^N) \rightarrow \mathbb{R}^N$. The received vector at user i is given by

$$Y_i^N = X^N + S^N + W_i^N \quad (4.2)$$

where addition is component-wise, $X^N = \alpha(V_1^K, V_2^K, S^N)$, S^N is the Gaussian interference ($S \sim \mathcal{N}(0, \sigma_S^2)$) known to the transmitter, and each sample in the additive noise W_i^N is drawn from an i.i.d. Gaussian distribution with variance $\sigma_{W_i}^2$ ($W_i \sim \mathcal{N}(0, \sigma_{W_i}^2)$) independently from both sources and interference. The system operates under an average power constraint P given by

$$\frac{\mathbb{E}[|\alpha(V_1^K, V_2^K, S^N)|^2]}{N} \leq P \quad (4.3)$$

The reconstructed signal is given by $\hat{V}_i^K = \gamma_i(Y_i^N)$, where the decoder functions $\gamma_i(\cdot)$ are mappings from $\mathbb{R}^N \rightarrow \mathbb{R}^K$; again, the reconstruction quality at each user is the MSE distortion $D_{v_i} = \frac{\mathbb{E}[\|V_i^K - \hat{V}_i^K\|^2]}{K}$ for $i = 1, 2$. We assume that $\sigma_{W_1}^2 > \sigma_{W_2}^2$ and the broadcast channel is physically degraded; hence user 1 is the weak user and user 2 is the strong one. This means that we can write the received signal at the weak user as $Y_1^N = Y_2^N + \tilde{W}^N$, where each sample in \tilde{W}^N follows a Gaussian distribution that is independent of everything else and has a variance $\sigma_{\tilde{W}}^2 = \sigma_{W_1}^2 - \sigma_{W_2}^2$. For a given power constraint P and system's rate r , the distortion region is defined as the closure of all distortion pairs $(D_{v_1}^o, D_{v_2}^o)$ for which $(P, D_{v_1}^o, D_{v_2}^o)$ is achievable, where a power-distortion triple is achievable if for any $\delta_{v_i} > 0$, there exist sufficiently large integers K and N with $N/K = r$, encoding and decoding functions $(\alpha, \gamma_1, \gamma_2)$ satisfying (4.3), such that $D_{v_i} < D_{v_i}^o + \delta_{v_i}, i = 1, 2$. In this chapter, we are interested in analyzing the distortion region of this system under matched ($r = 1$) and expansion bandwidth modes ($r > 1$). Note that for $r > 1$, V_i^K and the first K interference samples S^K in S^N are correlated via the covariance matrix in (4.1), while V_i^K and $[S^N]_{K+1}^N$ are independent.

4.3 Outer Bounds and Reference Schemes

4.3.1 Outer Bound for $\rho_{V_1 V_2} \neq 1$

In [62] and [5], an outer bound on the distortion region for sending correlated sources over the broadcast channel without interference is obtained for $r = 1$ and $r \neq 1$, respectively. This is done by assuming additional knowledge of the source V_1^K at the strong user (user 2). In this section, we derive an outer bound on the distortion region for the interference broadcast channel for $r \geq 1$. Since $S(i)$ and $V_1(i)$ are correlated for $i = 1, \dots, K$, we have $S(i) = \tilde{S}_I(i) + \tilde{S}_D(i)$, with $\tilde{S}_D(i) = \frac{\rho_{V_1 S} \sigma_S}{\sigma_{V_1}} V_1(i)$ and $\tilde{S}_I \sim \mathcal{N}(0, (1 - \rho_{V_1 S}^2) \sigma_S^2)$. To derive an outer bound, we assume knowledge of V_1^K at the strong user (this is a reasonable assumption for small correlation coefficients; this bound, however, might not be tight for high correlation values) and $(\tilde{S}^K, [S^N]_{K+1}^N)$ at both users, where $\tilde{S}^K = \eta_1 \tilde{S}_I^K + \eta_2 \tilde{S}_D^K$.

Definition 4.1. Let $MSE(Y_i; \tilde{S})$ be the distortion incurred from estimating Y_i based on \tilde{S} using an LMMSE estimator, for $i = 1, 2$. This distortion, which is a function of $\eta_1, \eta_2, \mathbb{E}[X \tilde{S}_I]$ and $\mathbb{E}[X \tilde{S}_D]$, will be used in Lemma 4.1 and Lemma 4.2.

Lemma 4.1. The outer bound on the distortion region can be expressed as follows

$$\begin{aligned}
 D_{v_1} &\geq \sup_{\eta_1, \eta_2} \inf_{\substack{X: \\ |\mathbb{E}[X \tilde{S}_I]| \leq \sqrt{\mathbb{E}[X^2] \mathbb{E}[\tilde{S}_I^2]} \\ |\mathbb{E}[X \tilde{S}_D]| \leq \sqrt{\mathbb{E}[X^2] \mathbb{E}[\tilde{S}_D^2]}} \left\{ \frac{\text{Var}(V_1 | \tilde{S}) (\xi P + \sigma_{W_1}^2)^r}{(MSE(Y_1; \tilde{S})) (P + \sigma_{W_1}^2)^{r-1}} \right\} \\
 D_{v_2} &\geq \frac{\text{Var}(V_2 | V_1, S)}{\left(1 + \frac{\xi P}{\sigma_{W_2}^2}\right)^r} \tag{4.4}
 \end{aligned}$$

where $\xi \in [0, 1]$, $\text{Var}(V_2 | V_1, S) = \sigma_{V_2}^2 \left(1 - \frac{\rho_{V_1 V_2}^2 - 2\rho_{V_1 V_2} \rho_{V_1 S} \rho_{V_2 S} + \rho_{V_2 S}^2}{1 - \rho_{V_1 S}^2}\right)$ is the variance of

V_2 given V_1 and S and $\text{Var}(V_1|\tilde{S}) = \sigma_{V_1}^2 \left(1 - \frac{\eta_2^2 \rho_{V_1 S}^2}{\eta_1^2 (1 - \rho_{V_1 S}^2) + \eta_2^2 \rho_{V_1 S}^2}\right)$ is the variance of V_1 given \tilde{S} .

Proof. For a $K : N$ system with $N \geq K$, we have

$$\begin{aligned}
\frac{K}{2} \log \frac{\sigma_{V_2}^2}{D_{v_2}} &\leq I(V_2^K; \hat{V}_2^K) \leq I(V_2^K; Y_2^N, V_1^K, \tilde{S}^K, [S^N]_{K+1}^N) \\
&= I(V_2^K; V_1^K, \tilde{S}^K, [S^N]_{K+1}^N) + I(V_2^K; Y_2^N | V_1^K, \tilde{S}^K, [S^N]_{K+1}^N) \\
&= h(V_2^K) - h(V_2^K | V_1^K, S^N) + h(Y_2^N | V_1^K, S^N) - h(Y_2^N | V_1^K, V_2^K, S^N) \\
&\stackrel{(a)}{=} \frac{K}{2} \log \frac{\sigma_{V_2}^2}{\text{Var}(V_2 | V_1, S)} + \frac{N}{2} \log 2\pi e(\xi P + \sigma_{W_2}^2) - \frac{N}{2} \log 2\pi e(\sigma_{W_2}^2) \\
&= \frac{K}{2} \log \frac{\sigma_{V_2}^2}{\text{Var}(V_2 | V_1, S)} + \frac{N}{2} \log \left(1 + \frac{\xi P}{\sigma_{W_2}^2}\right) \tag{4.5}
\end{aligned}$$

where (a) uses the fact that we can bound $h(Y_2^N | V_1^K, S^N)$ as follows: $\frac{N}{2} \log 2\pi e(\sigma_{W_2}^2) = h(Y_2^N | V_1^K, V_2^K, S^N) \leq h(Y_2^N | V_1^K, S^N) \leq h(Y_2^N | S^N) \leq \frac{N}{2} \log 2\pi e(P + \sigma_{W_2}^2)$; hence there exists an $\xi \in [0, 1]$ such that $h(Y_2^N | V_1^K, S^N) = \frac{N}{2} \log 2\pi e(\xi P + \sigma_{W_2}^2)$. To get a bound on estimating V_1^K , we can write the following

$$\begin{aligned}
\frac{K}{2} \log \frac{\sigma_{V_1}^2}{D_{v_1}} &\leq I(V_1^K; \hat{V}_1^K) \leq I(V_1^K; Y_1^N, \tilde{S}^K, [S^N]_{K+1}^N) \\
&= I(V_1^K; \tilde{S}^K, [S^N]_{K+1}^N) + I(V_1^K; Y_1^N | \tilde{S}^K, [S^N]_{K+1}^N) \\
&= h(V_1^K) - h(V_1^K | \tilde{S}^K, [S^N]_{K+1}^N) + h(Y_1^N | \tilde{S}^K, [S^N]_{K+1}^N) - h(Y_1^N | V_1^K, S^N) \\
&\leq \frac{K}{2} \log \frac{\sigma_{V_1}^2}{\text{Var}(V_1 | \tilde{S})} + h(Y_1^K | \tilde{S}^K) + h([Y_1^N]_{K+1}^N | [S^N]_{K+1}^N) - h(Y_1^N | V_1^K, S^N) \\
&\leq \frac{K}{2} \log \frac{\sigma_{V_1}^2}{\text{Var}(V_1 | \tilde{S})} + \frac{K}{2} \log 2\pi e(\text{MSE}(Y_1; \tilde{S})) + \frac{N-K}{2} \log 2\pi e(P + \sigma_{W_1}^2) \\
&\quad - \frac{N}{2} \log 2\pi e(\xi P + \sigma_{W_1}^2) \tag{4.6}
\end{aligned}$$

in the last inequality, we use the fact that $h(Y_1^N | V_1^K, S^N) \geq \frac{N}{2} \log 2\pi e(\xi P + \sigma_{W_1}^2)$ due

to the entropy power inequality and since $Y_1^N = Y_2^N + \tilde{W}^N$ with $\tilde{W} \sim \mathcal{N}(0, \sigma_{W_1}^2 - \sigma_{W_2}^2)$. Moreover, we use $h(Y_1^K | \tilde{S}^K) \leq h(Y_1^K - \gamma_{\text{lmmse}}(\tilde{S}^K))$, where $\gamma_{\text{lmmse}}(\tilde{S}^K)$ is the LMMSE estimator of Y_1 based on \tilde{S} . By the Cauchy-Schwartz inequality, we have $|\mathbb{E}[X\tilde{S}_I]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[\tilde{S}_I^2]}$ and $|\mathbb{E}[X\tilde{S}_D]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[\tilde{S}_D^2]}$. For a given η_1 and η_2 , the maximum value of $\text{MSE}(Y_1; \tilde{S})$ over $\mathbb{E}[X\tilde{S}_I]$ and $\mathbb{E}[X\tilde{S}_D]$ has to be used; then, we need to maximize the distortion expression over the parameters η_1 and η_2 . Note that most inequalities follow from rate-distortion theory, the data processing inequality, the non-negativity of mutual information, conditioning reduces differential entropy and the fact that the Gaussian distribution maximizes differential entropy. \square

Remark 4.1. *The bound in (4.4) reduces to the one in [62] when there is no interference. This can be seen by setting $\rho_{V_1S} = \rho_{V_2S} = 0$ and $\eta_1 = \eta_2 = 0$. Neglecting the strong user (i.e., reducing the broadcast problem to point-to-point communications), the bound on reconstructing V_1 in (4.4) reduces to the bound derived in Lemma 3.1 for the case of AWGN channel (no fading). This can be seen by setting $\xi = 0$ in the derived bound on D_{v_1} .*

4.3.2 Outer Bound for $\rho_{V_1V_2} = 1$

For this special case, we consider broadcasting one source V^K (i.e., $V_1^K = V_2^K = V^K$). A “trivial” outer bound can be derived by treating each user separately (as in the case of point-to-point communications). As a result, by assuming partial knowledge about the interference $(\tilde{S}^K, [S^N]_{K+1}^N)$ at the receiver side, we can obtain an outer bound on the source reconstruction distortion D_{v_i} at user i in a similar way as given in Lemma 3.1. The outer bound on the source reconstruction distortion at user i ,

D_{v_i} , under bandwidth expansion ($r \geq 1$) can be then expressed as follows

$$D_{v_i} \geq \sup_{\eta_1, \eta_2} \inf_{\substack{X: \\ |\mathbb{E}[XS_I]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_I^2]} \\ |\mathbb{E}[XS_D]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_D^2]}} \left(\frac{\text{Var}(V|\tilde{S})(\sigma_{W_i}^2)^r}{(\text{MSE}(Y_i; \tilde{S}))(P + \sigma_{W_i}^2)^{r-1}} \right). \quad (4.7)$$

Note that this outer bound gives a rectangular region for D_{v_1} and D_{v_2} . To derive a better bound, we need to introduce an auxiliary random variable G^K similar to the one in [53]. This approach is also used to find a bound for the Gaussian broadcast channel in [56]. More precisely, an outer bound can be obtained by assuming knowledge of $(\tilde{S}^K, [S^N]_{K+1}^N)$ at both users and choosing $G^K = V^K + Z^K$, where Z^K is independent of everything else and each sample of Z^K follows a zero mean i.i.d. Gaussian distribution with variance σ_Z^2 .

Lemma 4.2. *For a $K : N$ bandwidth expansion system, the outer bound can be expressed as follows*

$$\begin{aligned} D_{v_1} &= \eta \inf_{\substack{X: \\ |\mathbb{E}[XS_I]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_I^2]} \\ |\mathbb{E}[XS_D]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[S_D^2]}} \frac{\text{Var}(V|\tilde{S})(\sigma_{W_1}^2)^r}{(\text{MSE}(Y_1; \tilde{S}))(P + \sigma_{W_1}^2)^{r-1}} \\ D_{v_2} &\geq \sup_{\sigma_Z^2} \frac{\text{Var}(V|\tilde{S})\sigma_Z^2(\sigma_{W_2}^2)^r}{\text{Var}(G|\tilde{S}) \left[A \left(\frac{D_{v_1} + \sigma_Z^2}{\text{Var}(G|\tilde{S})} \right)^{\frac{1}{r}} - \sigma_{\tilde{W}}^2 \right]^r} - H \end{aligned} \quad (4.8)$$

where $\eta \geq 1$, $A = (\text{MSE}(Y_1; \tilde{S}))^{\frac{1}{r}} (P + \sigma_{W_1}^2)^{\frac{r-1}{r}}$, $H = (\sigma_{W_2}^2)^r \text{Var}(V|\tilde{S})$ and $\text{Var}(G|\tilde{S}) = \text{Var}(V|\tilde{S}) + \sigma_Z^2$ is the variance of G given \tilde{S} . Note that the outer bound has to be maximized over η_1 and η_2 and that the bound on D_{v_2} is a function of D_{v_1} .

Proof. D_{v_1} has the same form as the one in (4.7); for a given $\eta \geq 1$, we get a bound

on D_{v_2} . This is done as follows

$$\begin{aligned}
\frac{K}{2} \log \frac{\text{Var}(V|\tilde{S})}{D_{v_2}} &\leq I(V^K; \hat{V}_2^K | \tilde{S}^K, [S^N]_{K+1}^N) \leq I(V^K; Y_2^N | \tilde{S}^K, [S^N]_{K+1}^N) \\
&= I(V^K; G^K | \tilde{S}^K, [S^N]_{K+1}^N) + I(V^K; Y_2^N | G^K, \tilde{S}^K, [S^N]_{K+1}^N) \\
&\quad - I(V^K; G^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N) \\
&= h(G^K | \tilde{S}^K, [S^N]_{K+1}^N) - h(G^K | V^K, S^N) + h(Y_2^N | G^K, \tilde{S}^K, [S^N]_{K+1}^N) \\
&\quad - h(W_2^N) - h(G^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N) + h(G^K | Y_2^N, V^K, S^N) \\
&= h(G^K | \tilde{S}^K, [S^N]_{K+1}^N) + h(Y_2^N | G^K, \tilde{S}^K, [S^N]_{K+1}^N) - h(W_2^N) \\
&\quad - h(G^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N). \tag{4.9}
\end{aligned}$$

Note that in (4.9), $h(G^K | \tilde{S}^K, [S^N]_{K+1}^N) = \frac{K}{2} \log 2\pi e \text{Var}(G|\tilde{S})$ and $h(W_2^N) = \frac{N}{2} \log 2\pi e \sigma_{W_2}^2$.

Next, we bound the remaining two terms in (4.9). Using the entropy power inequality, we can write the following

$$2^{\frac{2}{K}} h(G^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N) \geq 2^{\frac{2}{K}} h(V^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N) + 2^{\frac{2}{K}} h(Z^K). \tag{4.10}$$

Now using (4.10) and the fact that $h(V^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N) = h(V^K | \tilde{S}^K, [S^N]_{K+1}^N) - I(V^K; Y_2^N | \tilde{S}^K, [S^N]_{K+1}^N)$, we can lower bound $h(G^K | Y_2^N, \tilde{S}^K, [S^N]_{K+1}^N)$ by

$$\frac{K}{2} \log 2\pi e \left(\text{Var}(V|\tilde{S}) 2^{-\frac{2}{K} I(V^K; Y_2^N | \tilde{S}^K, [S^N]_{K+1}^N)} + \sigma_Z^2 \right) \tag{4.11}$$

Since the broadcast channel is degraded, Y_1^N can be written as the sum of Y_2^N and a noise \tilde{W} that is independent of everything else and has a variance $\sigma_{\tilde{W}}^2 = \sigma_{W_1}^2 - \sigma_{W_2}^2$.

Hence using the entropy power inequality, we have

$$2^{\frac{2}{N}h(Y_1^N|G^K, \tilde{S}^K, [S^N]_{K+1}^N)} \geq 2^{\frac{2}{N}h(Y_2^N|G^K, \tilde{S}^K, [S^N]_{K+1}^N)} + 2\pi e \sigma_{\tilde{W}}^2 \quad (4.12)$$

where $h(Y_1^N|G^K, \tilde{S}^K, [S^N]_{K+1}^N) = h(Y_1^N|\tilde{S}^K, [S^N]_{K+1}^N) - I(Y_1^N; G^K|\tilde{S}^K, [S^N]_{K+1}^N)$; moreover, $h(Y_1^N|\tilde{S}^K, [S^N]_{K+1}^N)$ can be upper bounded in a similar way as in (4.6) and the mutual information $I(Y_1^N; G^K|\tilde{S}^K, [S^N]_{K+1}^N)$ is lower bounded as follows

$$I(Y_1^N; G^K|\tilde{S}^K, [S^N]_{K+1}^N) \geq I(\hat{V}_1^K; G^K|\tilde{S}^K, [S^N]_{K+1}^N) \geq \frac{K}{2} \log \frac{\text{Var}(G|\tilde{S})}{D_{v_1} + \sigma_Z^2} \quad (4.13)$$

where we use the data processing theorem, the rate-distortion theory and the fact that

$$\frac{1}{K} \mathbb{E}[\|\hat{V}_1^K - G^K\|^2] = \frac{1}{K} \mathbb{E}[\|\hat{V}_1^K - V^K + V^K - G^K\|^2] = (D_{v_1} + \sigma_Z^2) \quad (4.14)$$

where we use the fact that $V^K - G^K = Z^K$ is independent of $(\hat{V}_1^K - V^K)$. Now combining all the above inequalities and after some manipulations we can get the bound on D_{v_2} given in (4.8). \square

4.3.3 Linear Scheme

In this section, we assume that the encoder transforms the K dimensional sources (V_1^K, V_2^K) into an N dimensional channel input X^N using a linear transformation according to

$$X^N = \mathbf{T}_1 V_1^K + \mathbf{T}_2 V_2^K + \mathbf{M}_1 S^N \quad (4.15)$$

where $\mathbf{T}_1, \mathbf{T}_2$ are $N \times K$ matrices and \mathbf{M}_1 is an $N \times N$ matrix. In such case, the received signal at user i , Y_i^N is conditionally Gaussian and the MMSE decoder is a linear estimator. The estimated source at user i is then given by

$$\hat{V}_i^K = \Sigma_{V_i Y_i} \Sigma_{Y_i}^{-1} Y_i^N \quad (4.16)$$

where $\Sigma_{V_i Y_i} = \mathbb{E}[(V_i^K)(Y_i^N)^T]$ is the correlation matrix between V_i^K and Y_i^N and $\Sigma_{Y_i} = \mathbb{E}[(Y_i^N)(Y_i^N)^T]$ is the covariance matrix of Y_i^N , for $i = 1, 2$. The distortion region of the linear scheme can then be expressed as follows

$$D_{v_i}^{linear} = \frac{1}{K} \text{tr} \{ \sigma_{V_i}^2 \mathbf{I}_{K \times K} - \Sigma_{V_i Y_i} \Sigma_{Y_i}^{-1} \Sigma_{V_i Y_i}^T \}. \quad (4.17)$$

4.3.4 Tandem Digital Scheme

This strategy is based on successive coding where the sources are encoded jointly at both the common and the refinement layers. Using [26], the achievable source coding rate (R_1, R_2) for any distortion (D'_{v_1}, D'_{v_2}) is given by

$$\begin{aligned} R_1(\nu) &= \frac{1}{2} \log \frac{1 - \rho^2}{D'_{v_1} (1 - \nu^2 (1 - D'_{v_1})) - (\rho - \nu (1 - D'_{v_1}))^2} \\ R_2(\nu) &= \left[\frac{1}{2} \log \frac{1 - \nu^2 (1 - D'_{v_1})}{D'_{v_2}} \right]^+ \end{aligned} \quad (4.18)$$

where $\nu \in \left[\rho, \min \left(\frac{1}{\rho}, \frac{\rho}{1 - D'_{v_1}}, \sqrt{(1 - D'_{v_2}) / (1 - D'_{v_1})} \right) \right]$ and $[x]^+ = \max(x, 0)$. For a Gaussian interference broadcast channel, the rate (R_1, R_2) can be achieved if and only

if there exists $0 \leq \eta_t \leq 1$ such that

$$\begin{aligned} R_1 &\leq \frac{r}{2} \log \left(1 + \frac{\eta_t P}{(1 - \eta_t)P + \sigma_{W_1}^2} \right), \\ R_2 &\leq \frac{r}{2} \log \left(1 + \frac{(1 - \eta_t)P}{\sigma_{W_2}^2} \right). \end{aligned} \quad (4.19)$$

By plugging these rates into (4.18), we get the achievable distortions for $D_{v_1}^{tandem}$ and $D_{v_2}^{tandem}$. The above rates can be achieved via Costa coding. Note that this is the ‘best’ tandem scheme for uncorrelated interference in terms of achievable distortion region.

4.4 HDA Coding Schemes

4.4.1 HDA Scheme 1 for Matched Bandwidth

As shown from the encoder structure in Fig. 4.2, this scheme has five layers that are merged to output X^K ($K = N$). The first layer, which uses an average power of P_a , outputs $X_a^K = \sqrt{a}(\beta_1 V_1^K + \beta_2 V_2^K + \beta_3 S^K)$, a linear combination of the sources and the interference, where $\beta_1, \beta_2, \beta_3 \in [-1, 1]$, and $a = P_a / (\beta_1^2 \sigma_{V_1}^2 + \beta_2^2 \sigma_{V_2}^2 + \beta_3^2 \sigma_S^2 + 2\beta_1 \beta_2 \rho_{V_1 V_2} \sigma_{V_1} \sigma_{V_2} + 2\beta_1 \beta_3 \rho_{V_1 S} \sigma_{V_1} \sigma_S + 2\beta_2 \beta_3 \rho_{V_2 S} \sigma_{V_2} \sigma_S)$ is a gain factor related to the power constraint P_a . The second layer, which is meant for both users, employs a source-channel VQ on the source vector V_1^K with rate R_q ; the output of this layer is $X_q^K = \mu(V_1^K + U_q^K)$, where μ is a gain factor related to the power constraint and samples in U_q^K follow a zero mean i.i.d. Gaussian that is independent of V_1^K, V_2^K, S^N and has a variance Q . The encoding process of the VQ is previously described in Chapter 3 and the VQ codebook is denoted by \mathcal{X}_q . The third layer, which outputs X_h^K with power P_h , uses HDA Costa coding on the linear combination $X_h'^K = \sqrt{u} X_a^K$,

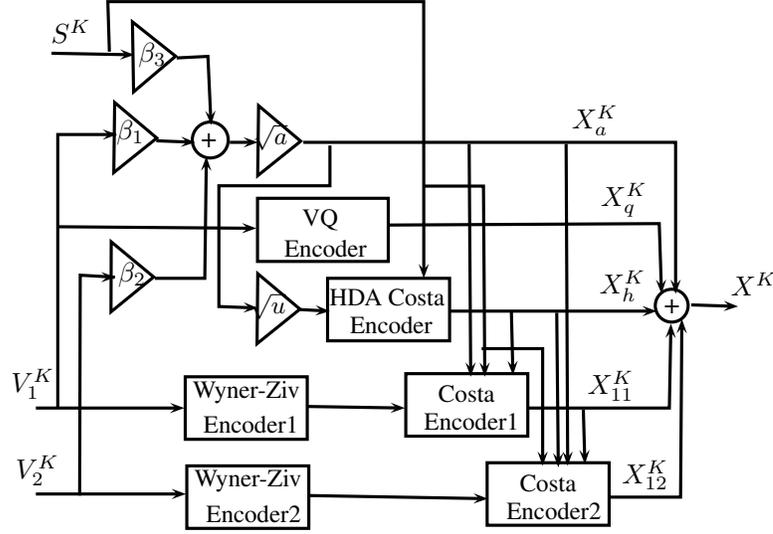


Figure 4.2: HDA Scheme 1 encoder structure for rate $r = \frac{N}{K} = 1$.

where $u = P_h/P_a$ is a gain factor related to power constraint P_h . This layer is meant for both users and treats X_a^K and S^K as known interference. The auxiliary random variable of the HDA Costa encoder is given by

$$U_h^K = X_h^K + \alpha_h(S^K + X_a^K) + \kappa_h X_h'^K \quad (4.20)$$

where each sample in X_h^K follows an i.i.d. Gaussian with variance P_h ($X_h \sim \mathcal{N}(0, P_h)$), α_h and κ_h are defined later. The HDA Costa encoder forms a codebook \mathcal{U}_h with codeword length K and 2^{KR_h} codewords (the rate R_h is defined later). Every codeword is generated following the random variable U_h^K . The codebook is revealed to both the encoder and decoder. The encoder searches for a $U_h^K \in \mathcal{U}_h$ such that $(X_h'^K, S'^K, U_h^K)$ are jointly typical, where $S'^K = (S^K + X_a^K)$. The fourth layer, which uses P_{11} as average power, encodes the source V_1^K using Wyner-Ziv coding at a rate

$R_1 = \frac{1}{2} \log \left(1 + \frac{P_{11}}{P - \mathbb{E}[(X_a + X_q + X_h)^2] + \sigma_{W_1}^2} \right)$. The Wyner-Ziv index m_1 is then encoded using Costa coding that treats S^K , X_a^K and X_h^K as interference; the output of this layer is denoted by X_{11}^K . In the fifth layer, which is meant for the strong user, the source V_2^K is first encoded using a Wyner-Ziv at rate $R_2 = \frac{1}{2} \log \left(1 + \frac{P_{12}}{\sigma_{W_2}^2} \right)$ followed by a Costa coder that treats S^K , X_a^K , X_h^K and X_{11}^K as interference and outputs X_{12}^K , where $P_{12} = P - \mathbb{E}[(X_a + X_q + X_h + X_{11})^2]$. The channel input is the superposition of all five layers $X^K = X_a^K + X_q^K + X_h^K + X_{11}^K + X_{12}^K$. Note that the proposed HDA scheme uses the superposition coding principle for broadcast channels introduced in Chapter 2. In our scheme, this concept (of cloud and satellite codewords) is manifested by using layers (i.e., codewords) that are able to be decoded by all users while other layers are only meant to the strong user.

At the weak user, from the noisy received signal Y_1^K , the VQ decoder estimates X_q^K by searching for a codeword $X_q^K \in \mathcal{X}_q$ that is jointly typical with the received signal Y_1^K . Following the result of [64]² and the error analysis of [48], the error probability of decoding X_q^K goes to zero by choosing the rate $I(V_1; X_q) \leq R_q \leq I(X_q; Y_1)$. The variance Q of U_q has to be chosen to satisfy the rate constraint on R_q . Based on Y_1^K and the decoded VQ codeword X_q^K , an LMMSE estimator is used to get an estimate of V_1^K denoted by \tilde{V}_1^K . The reconstruction distortion $\tilde{D}_{v_1} = \mathbb{E}[|V_1^K - \tilde{V}_1^K|^2]/K$ is given by

$$\tilde{D}_{v_1} = \sigma_{V_1}^2 - \Gamma_q \Lambda_q^{-1} \Gamma_q^T \quad (4.21)$$

where Γ_q is the correlation vector between V_1 and $[X_q \ Y_1]$ and Λ_q is the covariance of $[X_q \ Y_1]$. The HDA Costa decoder then estimates the codeword U_h^K by searching for a codeword in \mathcal{U}_h that is jointly typical with $(\tilde{Y}_1^K, \tilde{V}_1^K)$, where $\tilde{Y}_1^K = Y_1^K - X_q^K$. By choosing $\alpha_h = \frac{P_h}{P_h + P_{11} + P_{12} + \sigma_{W_1}^2}$ and κ_h in (4.20) so that the HDA Costa rate R_h satisfies

the following constraint $I(U_h; S', X'_h) \leq R_h \leq I(U_h; \tilde{Y}_1, \tilde{V}_1)$, the error probability of decoding U_h^K goes to zero. The HDA Costa decoder then forms an LMMSE estimate of V_1^K based on \tilde{Y}_1^K , the VQ decoded codeword X_q^K and the decoded codeword U_h^K . The resulting distortion is given by

$$D'_{v_1} = \sigma_{V_1}^2 - \Gamma_1^T \Lambda_1^{-1} \Gamma_1 \quad (4.22)$$

where Λ_1 is the covariance matrix of $[X_q \ U_h \ \tilde{Y}_1]$, and Γ_1 is the correlation vector between V_1 and $[X_q \ U_h \ \tilde{Y}_1]$. A better estimate of V_1^K , denoted by \hat{V}_1^K is obtained from the fourth layer by using the decoded Wyner-Ziv 1 codeword T_1^K . As a result the average distortion from reconstructing V_1^K is given by

$$D_{v_1}^{HDA \text{ Scheme } 1} = \frac{D'_{v_1}}{1 + \frac{P_{11}}{P_{12} + \sigma_{W_1}^2}}. \quad (4.23)$$

Note that the Wyner-Ziv 1 codewords follows the following random variable

$$T_1^K = \alpha_{wz} V_1^K + B_1^K \quad (4.24)$$

where $\alpha_{wz} = \sqrt{1 - \frac{P_{12} + \sigma_{W_1}^2}{P_{11} + P_{12} + \sigma_{W_1}^2}}$ and $B_1 \sim \mathcal{N}(0, D_{v_1}^{HDA \text{ Scheme } 1})$.

The strong user, that is able to decode all codewords used by the weak user, estimates the source V_2^K by first finding a linear MMSE estimate of V_2^K , denoted by \tilde{V}_2^K , based on the VQ codeword X_q^K , the HDA Costa codeword U_h^K , the Wyner-Ziv codeword T_1^K and $\tilde{Y}_2^K = Y_2^K - X_q^K$. The distortion in reproducing V_2^K is

$$\tilde{D}_{v_2} = \sigma_{V_2}^2 - \Gamma_2^T \Lambda_2^{-1} \Gamma_2 \quad (4.25)$$

where Λ_2 is the covariance matrix of $[X_q \ U_h \ T_1 \ \tilde{Y}_2]$, and Γ_2 is the correlation vector between V_2 and $[X_q \ U_h \ T_1^K \ \tilde{Y}_2]$. A better estimate \hat{V}_2^K is then found using the decoded Wyner-Ziv 2 codeword T_2^K and \tilde{V}_2^K . The resulting overall distortion in estimating V_2^K can be expressed as follows

$$D_{v_2}^{HDA \text{ Scheme 1}} = \frac{\tilde{D}_{v_2}}{1 + \frac{P_{12}}{\sigma_{W_2}^2}}. \quad (4.26)$$

Note that this distortion can be achieved using an LMMSE estimator based on $X_q^K, U_h^K, T_1^K, \tilde{Y}_2^K$ and the Wyner-Ziv codeword T_2^K . The inner bound for HDA Scheme 1 is given by (4.23) and (4.26).

4.4.2 HDA Scheme 2 for Bandwidth Expansion

This scheme comprises two layers that are concatenated to output the transmitted signal as shown in Fig. 4.3. The first layer, which outputs \tilde{X}_1^K , consists of the HDA Scheme 1 encoder for $r = 1$ (composed of five sublayers) as described in the previous section. The second layer is composed of two sublayers. The first sublayer encodes V_1^K using a Wyner-Ziv at a rate $R_3 = \frac{1}{2} \log \left(1 + \frac{P_{21}}{P - P_{21} + \sigma_{W_1}^2} \right)$ followed by a Costa coder with an average power P_{21} . Note that the Costa coder treats $[S^N]_{K+1}^N$ as interference and outputs X_{21}^{N-K} that is decodable by both users. The second sublayer encodes V_2^K using a Wyner-Ziv at a rate $R_4 = \frac{1}{2} \log \left(1 + \frac{P_{22}}{\sigma_{W_2}^2} \right)$ followed by a Costa coder with an average power $P_{22} = P - P_{21}$ that treats X_{21}^{N-K} and $[S^N]_{K+1}^N$ as interference and outputs X_{22}^{N-K} . The output of the second layer is then given by $\tilde{X}_2^{N-K} = X_{21}^{N-K} + X_{22}^{N-K}$. Note that X^N is the concatenation of \tilde{X}_1^K and \tilde{X}_2^{N-K} .

At the weak user, an LMMSE decoder based on the decoded VQ codeword X_q^K , the HDA Costa codeword U_h^K , the Wyner-Ziv codeword T_1^K and the first K received

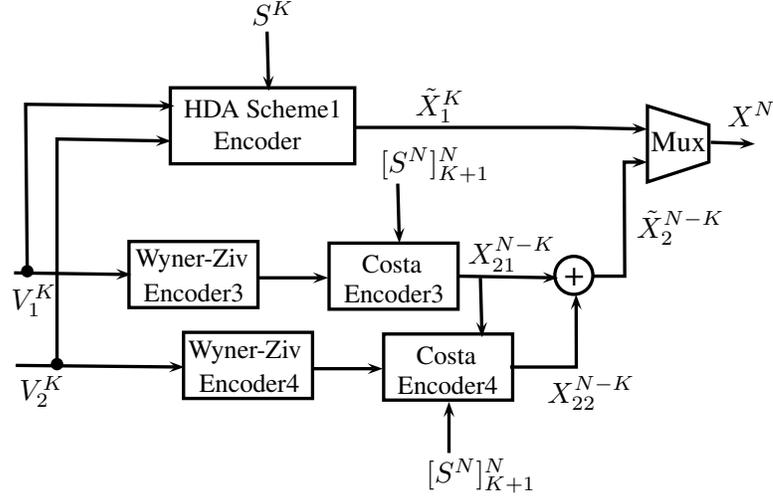


Figure 4.3: HDA Scheme 2 encoder structure for rate $r = \frac{N}{K} > 1$.

samples Y_1^K is used to get an estimate of V_1^K denoted by $V_1'^K$. The distortion in estimating V_1^K using $V_1'^K$, denoted by D'_{v_1} , can be expressed in a similar way as given in (4.23). Using the last $N - K$ samples of the received signal $[Y_1^N]_{K+1}^N$, a better refinement of V_1^K can be obtained using the Wyner-Ziv decoder 3. The overall distortion in reconstructing V_1^K is then given by

$$D_{v_1}^{HDA \text{ Scheme } 2} = \frac{D'_{v_1}}{\left(1 + \frac{P_{21}}{P_{22} + \sigma_{W_1}^2}\right)^{r-1}}. \quad (4.27)$$

At the strong user, which is able to decode all codewords used by the weak user, we can obtain an estimate of V_2^K using an LMMSE estimator based on the received signal Y_2^K , the VQ codeword X_q^K , the HDA Costa codeword U_h^K , the Wyner-Ziv 1 codeword T_1^K and the decoded codeword T_3^K of the Wyner-Ziv encoder 3. The resulting distortion is $D_{v_2}^* = \sigma_{V_2}^2 - \Gamma_{22} \Lambda_{22}^{-1} \Gamma_{22}^T$, where Λ_{22} is the covariance matrix of $[X_q \ U_h \ T_1 \ T_3 \ Y_2]$, and Γ_{22} is the correlation vector between V_2 and

$[X_q \ U_h \ T_1 \ T_3 \ Y_2]$. Note that, in a similar way as explained previously, $T_3 = \alpha_{wz3}V_1 + B_3$, where $\alpha_{wz2} = \sqrt{1 - \left(\frac{P_{22} + \sigma_{W_1}^2}{P_{21} + P_{22} + \sigma_{W_1}^2}\right)^{r-1}}$ and $B_3 \sim \mathcal{N}(0, D_{v_1}^{HDA \text{ Scheme } 2})$. A refinement of this estimate can be obtained using the Wyner-Ziv decoder 2 (of HDA scheme 1) and 4. The resulting distortion in estimating V_2^K can be expressed as follows

$$D_{v_2}^{HDA \text{ Scheme } 2} = \frac{D_{v_2}^*}{\left(1 + \frac{P_{12}}{\sigma_{W_2}^2}\right) \left(1 + \frac{P_{22}}{\sigma_{W_2}^2}\right)^{r-1}}. \quad (4.28)$$

The inner bound for HDA Scheme 2 is given by (4.27) and (4.28).

Remark 4.2. *Note that for the special case of $\rho_{V_1 V_2} = 1$ (i.e., $V_1^K = V_2^K = V^K$), the VQ layer and the digital layer (composed of Wyner-Ziv and Costa coder) that outputs X_{11}^K are unnecessary; the power of the two layers that produce X_q^K and X_{11}^K can be set to zero. In such case, the encoder structure of the proposed scheme can be simplified as given in Fig. 4.4.*

Moreover, for full correlation between the source and the interference ($\rho_{V_1 S} = \rho_{V_2 S} = 1$), an uncoded scheme, which is a scaled version of the source, is optimal for the matched bandwidth case (i.e., $r = 1$). This can be proved by comparing the resulting distortion to the outer bound. Note that the best outer bound in this case is obtained by choosing $\eta_1 = 1$ and $\eta_2 = 0$ in (4.7). This result is analogous to the one in [63] for point-to-point communications.

4.5 Numerical Results

In this section, we assume that the source pairs (V_1^K, V_2^K) , with variance $\sigma_{V_1}^2 = \sigma_{V_2}^2 = 1$, are broadcasted to two users with interference variance $\sigma_S^2 = 0$ dB, and observation noise variance $\sigma_{W_1}^2 = 0$ dB and $\sigma_{W_2}^2 = -5$ dB, respectively. The system's average

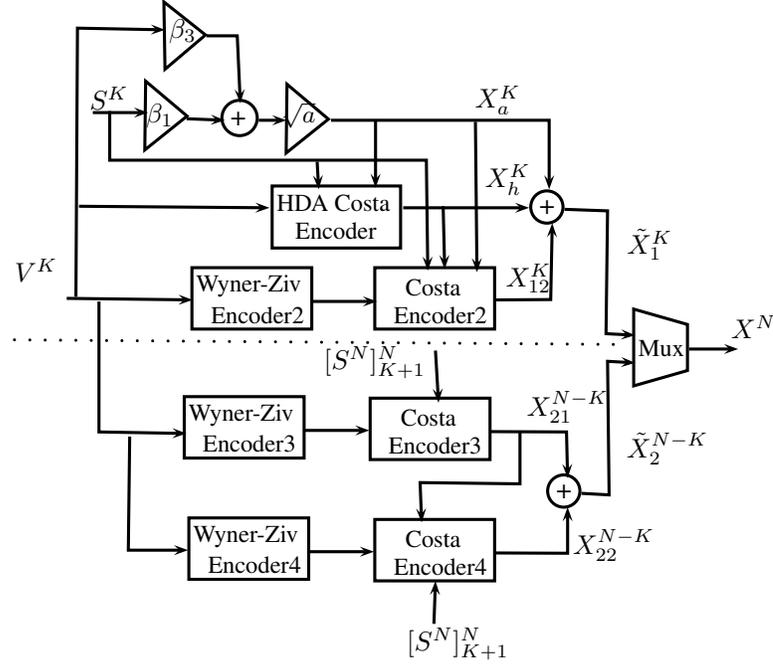


Figure 4.4: HDA Scheme encoder structure for rate $r = \frac{N}{K} \geq 1$ and $\rho_{V_1 V_2} = 1$.

power is set to $P = 1$. To evaluate the performance, we plot the inner and outer bounds derived in the previous sections for $r = 1$ and 2.

4.5.1 Case of $\rho_{V_1 V_2} \neq 1$

Fig. 4.5 focuses on the uncorrelated source-interference case ($\rho_{V_1 S} = \rho_{V_2 S} = 0$) under matched bandwidth ($r = 1$). For low correlation between the source pairs ($\rho_{V_1 V_2} = 0.2$), the proposed scheme gives some improvement over the tandem scheme and performs very close to the derived outer bound; for high source correlation levels ($\rho_{V_1 V_2} = 0.8$), however, the HDA scheme outperforms the tandem system but has a larger gap with respect to the outer bound. Note that for uncorrelated source-interference, the linear scheme gives a poor performance.

Fig. 4.6 shows the performance of the HDA scheme for $\rho_{V_1 V_2} = 0.5$, $\rho_{V_1 S} = \rho_{V_2 S} =$

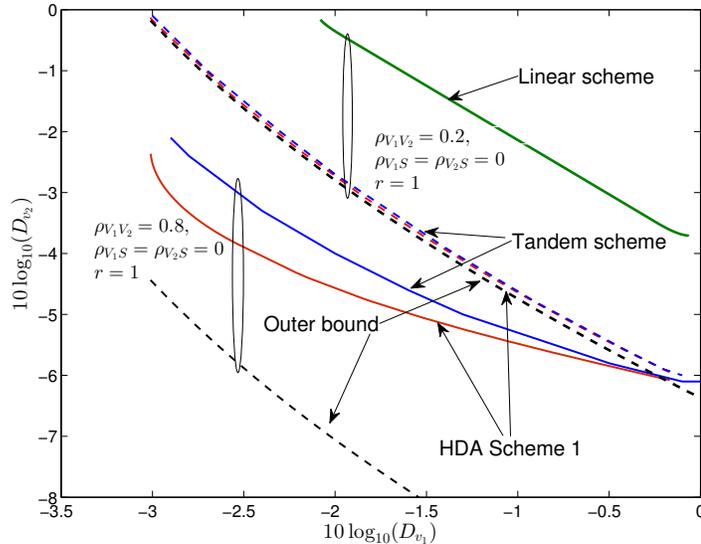


Figure 4.5: Distortion regions for HDA Scheme 1 for $r = 1$ as given by (4.23) and (4.26). Distortion regions for the outer bound (given by (4.4)), tandem and linear schemes (given by (4.17)) are also plotted.

0.2 and $r = 1$. We can notice that the purely analog scheme outperforms slightly the tandem scheme without being able to approach the HDA scheme. This can be explained from the fact that we operate at high noise levels, and since the linear scheme can benefit from the source-interference correlation. For the tandem scheme, which uses Costa coding, the transmitted signal is designed to be orthogonal to the interference; hence it cannot exploit the source-interference correlation and no performance improvement can be detected. Note that for moderate to low noise levels, the linear scheme does not outperform the tandem scheme for low source-interference correlations. Moreover, from other simulations, we notice that for $r = 1$, $\rho_{V_1V_2} = 0.8$, and $\rho_{V_1S} = \rho_{V_2S} = 0.5$, the linear scheme gives the best performance (our HDA scheme reduces to a linear scheme in this case). This can be explained by noting that in [45], the authors prove that under some conditions on the noise power and source

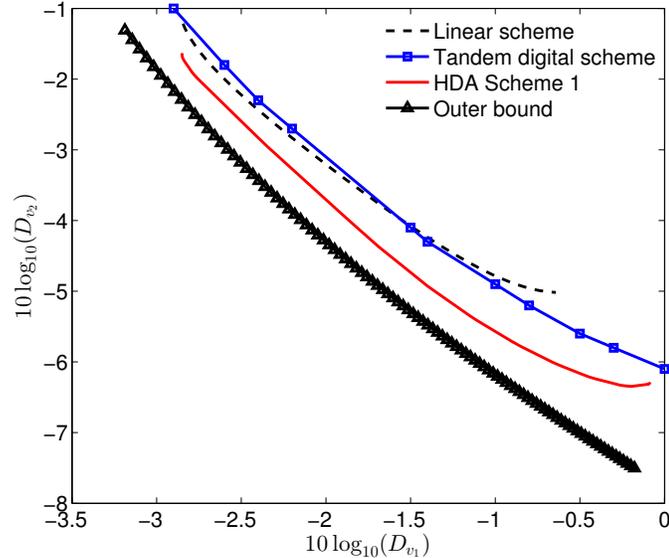


Figure 4.6: Distortion region for HDA Scheme 1 for $\rho_{V_1V_2} = 0.5$, $\rho_{V_1S} = \rho_{V_2S} = 0.2$ and $r = 1$ as given by (4.23) and (4.26). Distortion regions for the outer bound (given by (4.4)), tandem and linear schemes (given by (4.17)) are also plotted.

correlation (which are in accordance with the conditions for the last simulation), the linear scheme is optimal for broadcasting bivariate Gaussians under no interference. As a result, under similar conditions, the linear scheme is expected to give good performance for our problem when the source-interference correlation gets high.

Fig. 4.7 shows that the HDA scheme outperforms the tandem scheme under bandwidth expansion ($r = 2$). Note that for $\rho_{V_1V_2} = 0.2$ and $\rho_{V_1S} = \rho_{V_2S} = 0$, it is hard to notice (from Fig. 4.7) the gain of the HDA scheme over the tandem system on the plotted scale; the outer bound for this case is not shown since both schemes perform very closely to it. Moreover, the tandem scheme cannot benefit from the source-interference correlation and its performance depends solely on $\rho_{V_1V_2}$ in Fig. 4.7.

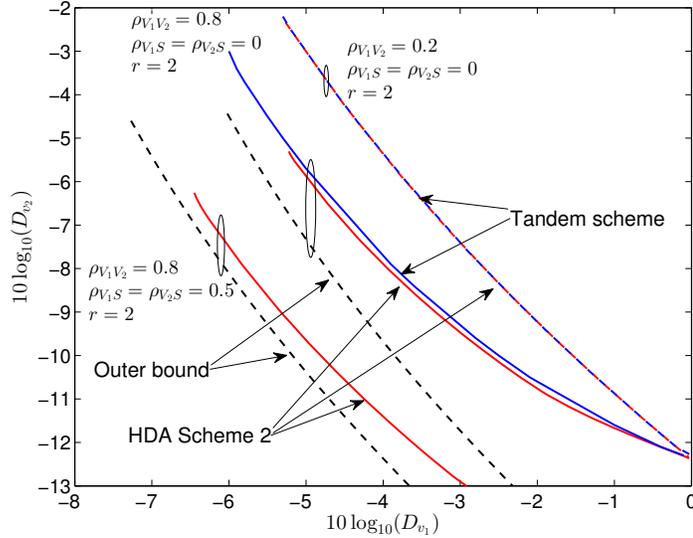


Figure 4.7: Distortion regions for HDA Scheme 2 for $r = 2$ as given by (4.27) and (4.28). Distortion regions for the outer bound (given by (4.4)) and tandem scheme are also plotted.

4.5.2 Case of $\rho_{V_1 V_2} = 1$

Recall that in this scenario, in which we have a single source vector to broadcast, the source-interference correlation $\rho_{V_1 S}$ is the same as $\rho_{V_2 S}$. Fig. 4.8 focuses on the matched source-channel bandwidth case ($r = 1$). As we can notice, the gap between the inner and the outer bounds decreases with the increase of the source-interference correlation $\rho_{V_1 S}$. We also notice that the power allocated to the analog part of our scheme increases with $\rho_{V_1 S}$. For the extreme case of full correlation between the source and the interference ($\rho_{V_1 S} = 1$), our scheme reduces to a purely analog scheme which is optimal as mentioned in Remark 4.2. For the matched source-channel bandwidth case ($r = 1$), the best outer bound is found to be the trivial bound given in (4.7); using an auxiliary random variable as done in Lemma 4.2 is not useful in this case.

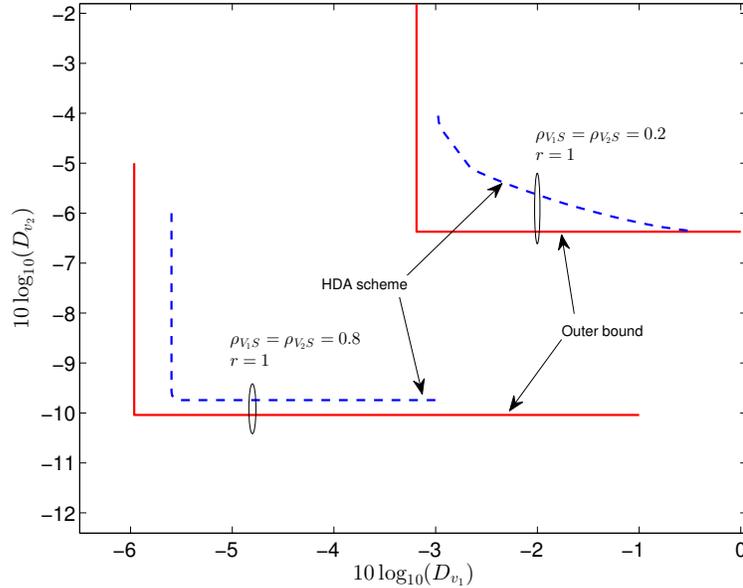


Figure 4.8: Distortion region for the HDA scheme for $r = 1$ as given by (4.23) and (4.26). The outer bound (given by (4.8)) is also plotted.

Figs. 4.9 and 4.10 show the inner and outer bounds for $r = 2$. We can notice that the gap between the achievable region and the outer bound is bigger than the ones in the matched bandwidth case. Moreover, the outer bound derived in Lemma 4.2 is shown to be beneficial for the expansion case. We can also notice that the proposed scheme performs close to the outer bound when D_{v_1} is high; this is from the fact that for high distortion D_{v_1} , the system behaves similar to a point-to-point communication. Note that for the special case of $r = 2$ (our scheme is designed for any $r \geq 1$), using HDA Costa coder at the expansion layer leads to an increase in the achievable distortion region.

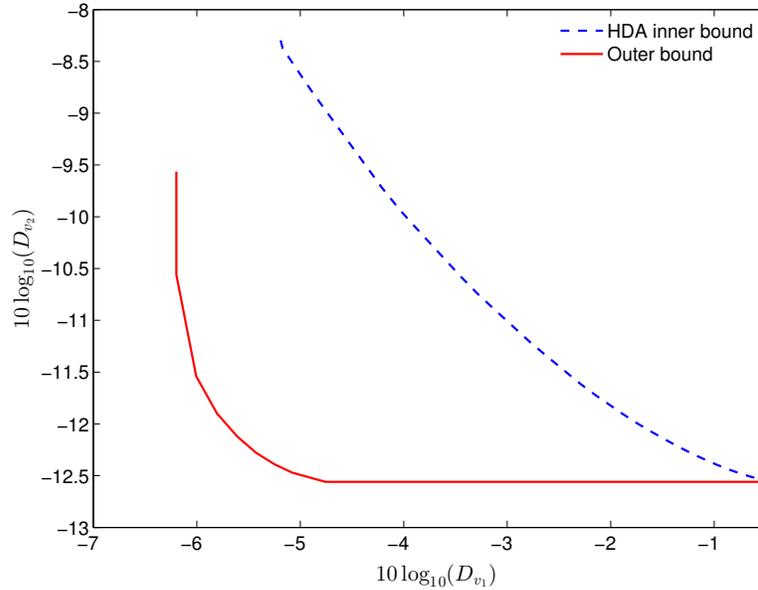


Figure 4.9: Distortion region for the HDA scheme for $r = 2$ and $\rho_{V_1S} = 0.2$ as given by (4.27) and (4.28). The outer bound (given by (4.8)) is also plotted.

4.6 Summary and Conclusions

In this chapter, we consider the transmission of a pair of correlated Gaussian sources over the two-user Gaussian degraded broadcast channel in the presence of interference that is correlated to the source. We propose layered HDA schemes under matched and expansion source-channel bandwidth scenarios based on Wyner-Ziv and HDA Costa coding; we also analyze their inner bounds. Outer bounds on the system's distortion region are also established by assuming additional knowledge at the receiver side. Numerical results indicate that the HDA schemes outperform the 'best' tandem scheme and perform close to the derived outer bounds under certain system settings.

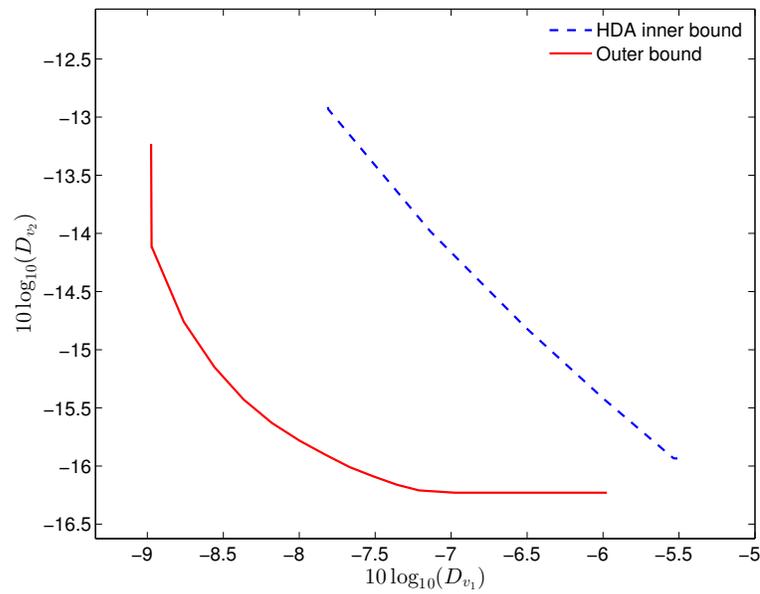


Figure 4.10: Distortion region for the HDA scheme for $r = 2$ and $\rho_{V_1 S} = 0.8$ as given by (4.27) and (4.28). The outer bound (given by (4.8)) is also plotted.

Chapter 5

Low Delay Analog Source-Channel Coding

5.1 Introduction

Shannon proves that the use of separate source and channel coders is optimal for point-to-point communications given unlimited delay and complexity in the coders [58]. In practice, JSCC can lead to a better performance when delay and complexity are constrained. For the case of Gaussian source over a Gaussian channel with matched source-channel bandwidth, linear transmission is optimal [28]. This result, however, does not hold for the case of mismatch between source and channel bandwidth, and in the presence of fading or interference. A common approach for JSCC design is to jointly optimize the components of a tandem system with respect to the channel and source characteristics. Another approach based on nonlinear analog mapping is treated in [32, 42, 55, 59].

With the increasing popularity of wireless sensor networks (WSNs), reliable transmission with delay and complexity constraints is more relevant than ever. A sensor node, often conceived as having limited lifetime and processing power, communicates its sensed field information to a fusion centre over a noisy wireless channel. To meet

these challenges, in this chapter, we investigate using a low delay and low complexity lossy source-channel mappings in WSNs.

We study the reliable transmission of a memoryless Gaussian source over different noisy channels subject to low delay/complexity constraints. More specifically, we use parametric and nonparametric mappings under different source-channel bandwidth ratios. The nonparametric mappings are based on joint optimization of the encoder and the decoder under an average power constraint. The case of bandwidth reduction/expansion over AWGN channels is studied in [2, 22, 24]. In [22, 24], the approach used is based on mapping the output of a vector quantizer to a specific point in a channel signal set. A direct source-channel mapping approach, however, is considered in [2, 32]. Source-channel mappings for the relay and the multiple access channels (MAC) are studied in [37, 38]. Our system, that uses nonlinear direct source-channel mappings, is shown to overcome the saturation effect, which is inevitable with linear systems, and achieves a graceful performance over a wide range of noise levels.

The rest of the chapter is organized as follows. Section 5.2 considers low delay coding over fading channels; both parametric and nonparametric mappings are proposed for $K : 1$ bandwidth reduction. In Section 5.3, we consider the problem of transmitting analog source over fading channels in the presence of interference that is known to the transmitter and correlated with the source; this problem is also considered in Chapter 3. Unlike Chapter 3, we aim to optimize the encoder and decoder under low delay coding constraint. Section 5.4 studies the problem of source and state interference estimation. Finally, conclusions are drawn in Section 5.5.

5.2 Low Delay Coding for Fading Channels

5.2.1 Problem Formulation

In this section, we consider the transmission of a Gaussian source vector V^K over a memoryless fading channel with AWGN. Fig. 5.1 shows the system structure where the channel state information (CSI) is available at the decoder only (DCSI), and the transmitter is assumed to know the fading distribution. We also consider two other cases: 1) when CSI is available at both the encoder and the decoder (full CSI or FCSI); 2) when CSI is not available at the transmitter and the receiver. The source vector V^K is transformed into a one dimensional channel input $X \in \mathbb{R}$ using a nonlinear mapping function $\alpha(\cdot)$ (i.e., $X = \alpha(V^K)$). The received signal can be expressed as

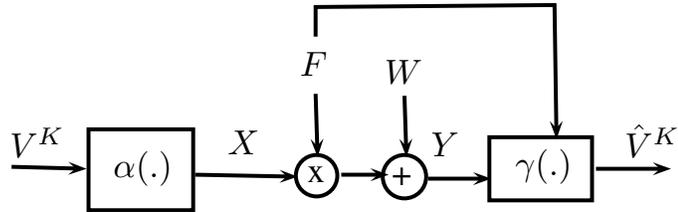


Figure 5.1: System structure for $K : 1$ bandwidth reduction over a fading channel with AWGN. CSI is only available at the decoder side, while the encoder knows only the distribution of the fading.

$$Y = FX + W \quad (5.1)$$

where F is the fading gain drawn from an i.i.d. Rayleigh distributed process. As in the previous chapters, the system operates under an average power constraint P given by

$$\mathbb{E}[\alpha^2(V^K)] \leq P. \quad (5.2)$$

At the receiver, the transmitted source vector is estimated using MMSE decoding (assuming DCSI/FCSI)

$$\hat{V}^K \triangleq \gamma(Y, F) = \mathbb{E}[V^K | y, f]. \quad (5.3)$$

We aim to find the optimal source mapping α and receiver γ that minimize the system's MSE distortion $D = \frac{\mathbb{E}[\|V^K - \hat{V}^K\|^2]}{K}$ under the average power constraint in (5.2).

5.2.2 Preliminaries: Theoretical Bound and Linear System

Theoretical Limit

Shannon's separation theorem states that in a point-to-point communication system, optimal performance can be achieved by optimizing separately (in tandem) the source and the channel coders given unlimited complexity and delay in the coders [58]. As a result, the optimal performance OPTA can be derived by equating the source rate-distortion function to the channel capacity times the system's rate.

For a memoryless Gaussian source V with variance σ_V^2 and MSE distortion measure, the rate-distortion function is [13]

$$R(D) = \max \left\{ 0, \frac{1}{2} \log \left(\frac{\sigma_V^2}{D} \right) \right\}. \quad (5.4)$$

The capacity of a fading channel when considering DCSI/FSCI and with power input P is given by [65]

$$C(P) = \mathbb{E}_F \left[\frac{1}{2} \log \left(1 + \frac{P(f)f^2}{\sigma_W^2} \right) \right] \quad (5.5)$$

where $P(f) = P$ for DCSI. When CSI is available at both the encoder and the decoder (i.e., FCSI), the transmitted power $P(f)$ in (5.5) waterfills over the fading states as

follows

$$P(f) = \max \left\{ 0, \left(\frac{1}{b} - \frac{\sigma_W^2}{f^2} \right) \right\} \quad (5.6)$$

where b satisfies $\mathbb{E}[P(f)] = P$. To find OPTA we set $R(D) = rC(P)$. Solving this will lead to the optimal distortion

$$D_{\text{OPTA}} = \sigma_V^2 \exp \left(r \mathbb{E}_F \left[\log \left(\frac{\sigma_W^2}{P(f)f^2 + \sigma_W^2} \right) \right] \right). \quad (5.7)$$

Note that this bound is achievable asymptotically using long block codes.

Linear Transmission

For a $K : 1$ bandwidth reduction, we use a simple coder that removes samples to perform bandwidth reduction. In this case the encoder $\alpha(\cdot)$ is a multiplication by a constant matrix and the transmitted symbol can be expressed as follows

$$X = \sqrt{\frac{P(f)}{\sigma_V^2}} \mathbf{I}_{1 \times K} V^K \quad (5.8)$$

where $P(f)$ is given as follows [72]

$$P(f) = \begin{cases} P, & \text{for DCSI} \\ \max \left\{ 0, \frac{1}{f} \left(\frac{1}{b'} - \frac{\sigma_W^2}{f} \right) \right\}, & \text{for FCSI} \end{cases} \quad (5.9)$$

where b' is calculated by solving $\mathbb{E}[P(f)] = P$. The optimal decoder (assuming DCSI/FCSI) $\gamma(Y, F)$ is given by

$$\hat{V}^K = \mathbb{E}[V^K | y, f] = \frac{f \sqrt{P(f) \sigma_V^2}}{P(f) f^2 + \sigma_W^2} \mathbf{I}_{K \times 1} y. \quad (5.10)$$

When no CSI is available at the receiver, the MMSE decoder can be expressed as follows

$$\hat{V}^K = \mathbb{E}[V^K|y] = \frac{\iint v^k p(y|\alpha(v^k), f) p(f) p(v^k) df dv^k}{\iint p(y|\alpha(v^k), f) p(f) p(v^k) df dv^k}. \quad (5.11)$$

5.2.3 Parametric Mappings

2 : 1 Spiral Mapping

System Structure

The Archimedes' spiral is shown to perform well for AWGN channels [32]. In this section, we extend the work of [32], which is introduced in Chapter 2, to optimize the spiral mapping over fading channels. At the time of working on this problem, we discovered that spiral mapping is also considered over fading channels in [8]. However, our system differs from the one in [8] in parts by the structure, optimization process and in applying power allocation when we have full CSI.

Bandwidth reduction is achieved by first approximating the two dimensional source vector V^2 to the closest point on the spiral mapping. The approximated point, represented by its radial distance d_r from the origin, is then mapped to the channel via an invertible operator $\ell(\cdot)$

$$X = u\tilde{X} = u\ell(d_r) = u \left(\pm 0.16 \left(\frac{\pi^2}{\Delta} \right) (d_r^2) \right). \quad (5.12)$$

Recall from Chapter 2 that u is a gain factor related to the channel power constraint P , Δ is the radial distance between any two neighbouring spiral arms, and $+$, $-$ represent positive and negative channel values, respectively. At the receiver (assuming DCSI), we scale the received samples using an optimal scaling factor $f\sqrt{P\mathbb{E}[\tilde{X}^2]}/(Pf^2 + \sigma_W^2)$

instead of a suboptimal factor $1/(uf)$ as used in [8]. The reconstructed source vector is then estimated using ML decoding. This is done by mapping the scaled received sample to a point on the spiral curve. In the simulation results, we also use the MMSE decoder.

System Optimization

The radial distance Δ is the only parameter that needs to be optimized in order to minimize the average MSE distortion. The source signal is affected by two types of distortion: 1) the approximation distortion $\bar{\varepsilon}_a$ which is related to the approximation operation, and 2) the channel distortion $\bar{\varepsilon}_{ch}$ which is due to the transmission over a noisy environment. Following the analysis introduced in Chapter 2, the system's distortion can be approximated as follows

$$D_{2:1 \text{ Spiral}} \approx \mathbb{E}_F[(\bar{\varepsilon}_a + \bar{\varepsilon}_{ch})|f] \approx \frac{1}{2} \left\{ \frac{\Delta^2}{12} + \sigma_W^2 \mathbb{E}[\tilde{X}^2] \mathbb{E}_F \left[\frac{Pf^2}{(Pf^2 + \sigma_W^2)^2} \right] \right\}. \quad (5.13)$$

The optimal radial distance Δ_{opt} is found by minimizing the MSE distortion in (5.13).

Power Allocation

In this section, we assume that the CSI is also available at the transmitter side (i.e., FCSI). Instead of updating the radial distance Δ at each time index [8], we allocate the power (along the time index) according to the CSI knowledge in a similar approach to [72].

Assuming the channel state is f and the corresponding power allocated is $P(f)$, the MSE distortion is well approximated in a similar way to (5.13). The optimal power allocation $P^*(f)$ that minimizes the system's distortion can be found by solving the

following optimization problem

$$\min \mathbb{E}_F \left[\frac{P(f)f^2}{(P(f)f^2 + \sigma_W^2)^2} \right] \quad \text{s.t.} \quad \mathbb{E}_F[P(f)] = P, \quad P(f) \geq 0. \quad (5.14)$$

Lemma 5.1. *Using the Karush-Kuhn-Tucker (KKT) conditions [7], for low noise level, the optimal power allocation is well approximated as follows*

$$P^*(f) = \begin{cases} \frac{1}{f} \left(\frac{1}{\sqrt{\tilde{b}}} - \frac{2\sigma_W^2}{f} \right), & \text{for } f \geq 2\sigma_W^2 \sqrt{\tilde{b}} \\ 0, & \text{otherwise} \end{cases} \quad (5.15)$$

where \tilde{b} is a threshold for all channel states and can be found from the average power constraint P and the statistics of the fading F .

Proof. For low noise levels and by neglecting the term σ_W^4 , the optimization problem in (5.14) can be approximated as follows

$$\min \mathbb{E}_F \left[\frac{1}{(P(f)f^2 + 2\sigma_W^2)} \right] \quad \text{s.t.} \quad \mathbb{E}_F[P(f)] = P, \quad P(f) \geq 0. \quad (5.16)$$

If f has a finite states, assuming $Pr(f = f_i) = h_i$ and $P(f_i) = P_i$, $i = 1, \dots, \tilde{L}$, the optimization problem can be expressed as follows

$$\min \sum_{i=1}^{\tilde{L}} \left[\frac{1}{P_i f_i^2 + 2\sigma_W^2} \right] h_i \quad \text{s.t.} \quad \sum_{i=1}^{\tilde{L}} P_i h_i = P, \quad P_i \geq 0. \quad (5.17)$$

The Lagrangian is given by

$$G = \sum_{i=1}^{\tilde{L}} \left[\frac{1}{P_i f_i^2 + 2\sigma_W^2} \right] h_i + \tilde{b} \left(\sum_{i=1}^{\tilde{L}} P_i h_i - P \right) - \sum_{i=1}^{\tilde{L}} \varsigma_i P_i \quad (5.18)$$

where \tilde{b} and ς_i are Lagrange multipliers. We obtain the KKT conditions as follows

$$\begin{aligned} \frac{\partial G}{\partial P_i} &= \frac{-f_i^2}{(2\sigma_W^2 + P_i f_i^2)^2} h_i + \tilde{b} h_i - \varsigma_i = 0, \\ P_i \varsigma_i &= 0, i = 1, \dots, \tilde{L}, \\ \sum_{i=1}^{\tilde{L}} P_i h_i &= P. \end{aligned} \quad (5.19)$$

For all those $P_i \neq 0$, we have $\varsigma_i = 0$ and $\tilde{b} = \frac{f_i^2}{(2\sigma_W^2 + P_i f_i^2)^2}$. Therefore

$$P^*(f_i) = \begin{cases} \frac{1}{f_i} \left(\frac{1}{\sqrt{\tilde{b}}} - \frac{2\sigma_W^2}{f_i} \right), & \text{for } f_i \geq 2\sigma_W^2 \sqrt{\tilde{b}} \\ 0, & \text{otherwise} \end{cases} \quad (5.20)$$

where \tilde{b} is a common threshold for all states and can be solved from $\sum_{i=1}^{\tilde{L}} P_i^* h_i = P$. If f has a non-discrete pdf, then by discretizing the pdf and taking the limit, we obtain the solution as given in (5.15). \square

$K : 1$ Parametric Mappings

In this section, we consider both 3 : 1 and 4 : 1 bandwidth reductions. In [20], parametric mappings over AWGN channel for these reduction ratios are presented. In what follows, we extend these mappings to accommodate the fading channel. In a similar way to the 2 : 1 spiral mapping, the 3 : 1 and the 4 : 1 bandwidth reductions are done by first approximating V^K to the closest point on the mapping curve, and then performing a one dimensional representation using (5.12). For a 3 : 1 bandwidth reduction, a ball of yarn mapping is used; for a given variable $x \in \mathbb{R}^+$, the three

dimensional mapping output can be mathematically expressed as follows

$$\mathcal{S}_{1:3}(x) = \frac{\Delta\varphi(x)}{\pi} \begin{bmatrix} \cos \varphi(x) \\ \cos \frac{1}{2\pi} \varphi(x) \sin \varphi(x) \\ \sin \frac{1}{2\pi} \varphi(x) \sin \varphi(x) \end{bmatrix} \quad (5.21)$$

and when $x \in \mathbb{R}^-$, the mapping output is given by

$$\mathcal{S}_{1:3}(x) = \frac{-\Delta\varphi(x)}{\pi} \begin{bmatrix} \sin \frac{1}{2\pi} \varphi(x) \cos \varphi(x) \\ \sin \varphi(x) \\ \cos \frac{1}{2\pi} \varphi(x) \cos \varphi(x) \end{bmatrix}. \quad (5.22)$$

For a 4 : 1 bandwidth reduction, for a given variable $x \in \mathbb{R}^+$, the four dimensional output can be expressed as follows

$$\mathcal{S}_{1:4}(x) = \frac{\Delta\varphi(x)}{\pi} \begin{bmatrix} \sin \frac{1}{3\pi} \varphi(x) \cos \varphi(x) \\ \cos \frac{1}{2\pi} \varphi(x) \sin \varphi(x) \\ \sin \frac{1}{2\pi} \varphi(x) \sin \varphi(x) \\ \cos \frac{1}{3\pi} \varphi(x) \cos \varphi(x) \end{bmatrix} \quad (5.23)$$

and when $x \in \mathbb{R}^-$, the mapping output is given by

$$\mathcal{S}_{1:4}(x) = \frac{-\Delta\varphi(x)}{\pi} \begin{bmatrix} \sin \frac{1}{3\pi} \varphi(x) \cos \varphi(x) \\ \sin \frac{1}{2\pi} \varphi(x) \sin \varphi(x) \\ \cos \frac{1}{3\pi} \varphi(x) \cos \varphi(x) \\ \cos \frac{1}{2\pi} \varphi(x) \sin \varphi(x) \end{bmatrix}. \quad (5.24)$$

At the receiver side, we use ML or MMSE decoding for signal recovery. Following the analysis for $K : 1$ bandwidth reduction in Chapter 2, the overall distortion can be approximated as follows

$$\begin{aligned} D_{K:1 \text{ Parametric}} &\approx \mathbb{E}_F[(\bar{\varepsilon}_a + \bar{\varepsilon}_{\text{ch}})|f] \\ &\approx \beta_{K:1} \Delta^{2-\theta_{K:1}} + P\sigma_W^2 \mathbb{E}[\tilde{X}]^2 \mathbb{E}_F \left[\frac{f^2}{(Pf^2 + \sigma_W^2)^2} \right] \end{aligned} \quad (5.25)$$

where $\theta_{3:1} = 1$, $\beta_{3:1} = 0.6312$, $\theta_{4:1} = 1.3$, and $\beta_{4:1} = 1.6244$ for $\sigma_V^2 = 1$, are found using a nonlinear curve fitting. Note that Δ , which has a different meaning than in the $2 : 1$ spiral mapping, is found by minimizing (5.25).

5.2.4 Nonparametric Mappings

Using the Lagrange multiplier method [30], the constrained optimization problem of minimizing the MSE distortion $\mathbb{E}[||V^K - \hat{V}^K||^2]/K$ subject to (5.2) can be recast into an unconstrained minimization problem via the Lagrange cost function

$$\min_{\alpha, \gamma} J(\alpha, \gamma) \quad (5.26)$$

where the cost function $J(\alpha, \gamma)$ is given by

$$J(\alpha, \gamma) = \frac{\mathbb{E}[||V^K - \hat{V}^K||^2]}{K} + \lambda \mathbb{E}[\alpha^2(V^K)] \quad (5.27)$$

and the Lagrange multiplier λ is used to control the average power. Note that if the (optimal) solution to the constrained problem lies on the convex hull of all feasible solutions to the problem, then the (optimal) solution to the unconstrained problem

coincides to the former; otherwise, the two solutions differ and in effect the optimal solution to the latter is suboptimal for the former problem [30]. The unconstrained minimization in (5.26) is still hard to solve due to the interdependencies between the optimized components, and since the encoder/decoder mappings are, in general, nonlinear functions. To overcome these challenges, we proceed in a similar way to classical design problems (e.g., vector quantizer design [49]) by formulating the necessary conditions for optimality. This is done by finding the optimal encoder α given the decoder γ , and vice versa. In what follows, we assume DCSI to derive the necessary conditions for optimality.

Necessary Conditions for Optimality

The problem of finding the optimal source mapping α^* (assuming γ is fixed) is given by

$$\alpha^* = \arg \min_{\alpha} \left\{ \mathbb{E}[\|V^K - \hat{V}^K\|^2]/K + \lambda \mathbb{E}[\alpha^2(V^K)] \right\}. \quad (5.28)$$

Using Bayes' rule, the distortion $\mathbb{E}[\|V^K - \hat{V}^K\|^2]$ can be expressed as follows

$$\mathbb{E}[\|V^K - \hat{V}^K\|^2] = \iiint p(v^k)p(f)p(y|\alpha(v^k), f)\|v^k - \hat{v}^k\|^2 dv^k df dy. \quad (5.29)$$

Note that this factorization follows from the fact that channel noise, source, and fading are assumed to be independent of each other. The mapping average power is given by

$$\int p(v^k)\alpha^2(v^k)dv^k. \quad (5.30)$$

Since $p(v^k)$ in (5.29) and (5.30) is nonnegative, the source mapping α can be optimized for each v^k individually according to

$$\alpha^*(v^k) = \arg \min_{x \in \mathbb{R}} \left\{ \iint p(f)p(y|x, f) \frac{\|v^k - \hat{v}^k\|^2}{k} df dy + \lambda x^2 \right\}. \quad (5.31)$$

Hence, (5.31) is a necessary condition for α to be the optimal mapping.

On the receiver side, the optimal decoder in the MSE sense (assuming α is fixed) is found using the conditional expectation of the source given the received symbol and the fading gain

$$\gamma^*(Y, F) = \mathbb{E}[V^K | y, f] = \frac{\int v^k p(y|v^k, f) p(v^k) dv^k}{\int p(y|v^k, f) p(v^k) dv^k}. \quad (5.32)$$

Design Algorithm

Based on the above necessary conditions for optimality, it is possible to optimize the mapping at the sensor nodes and the receiver using an iterative process. This is done by fixing one part while optimizing the other. The update equations (5.31) and (5.32) yield a lower distortion at each iteration step; hence, with a finite amount of training data, convergence is ensured. One common problem with such iterative technique is that the final solution will depend on the initialization of the algorithm and does not guarantee convergence to the global optimum solution. To get around these challenges, we use noisy channel relaxation [25]. This method suggests to design the system for a noisy channel, and uses the solution obtained as an initialization when designing the system for a less noisy channel. For a given CSNR, the design algorithm is stated as follows

1. Choose some initial mapping for the encoder α .

2. Find the optimal receiver γ according to (5.32).
3. Set the iteration index $i = 0$ and the cost $J^{(0)} = \infty$.
4. Set $i = i + 1$.
5. Find the optimal mapping α according to (5.31).
6. Find the optimal receiver γ according to (5.32).
7. Evaluate the cost function $J^{(i)}$. If the relative improvement of $J^{(i)}$ compared to $J^{(i-1)}$ is less than some positive threshold ε or $i > I_{max}$, stop iterating. Otherwise go to step 4.

The above algorithm is nested inside a “bracketing” Lagrange multiplier search. We first set $\lambda = \lambda_0$. If the designed α produces $\mathbb{E}[|\alpha(V^K)|^2] > P$, λ_0 is increased; else λ_0 is decreased. The search ends if $\mathbb{E}[|\alpha(V^K)|^2]$ is close enough to but $< P$. In our simulations, we used $\varepsilon = 10^{-4}$, $I_{max} = 15$ and a linear mapping for initializing the encoder α at low CSNR.

Implementation Aspects

For the actual implementation of (5.31) and (5.32), some modifications are required. By the fact that it is impossible to evaluate the formulas for all vector V^K in \mathbb{R}^K , we form as in [37] a set composed of Monte-Carlo (MC) samples drawn from the distribution of V^K . In our simulations, we use 10^4 samples to define this set. Since the channel input and output spaces are real valued, we discretize them using a pulse amplitude modulation (PAM) alphabet \mathcal{X}_d and \mathcal{Y}_d

$$\mathcal{X}_d = \mathcal{Y}_d = \left\{ -e\frac{L-1}{2}, -e\frac{L-3}{2}, \dots, e\frac{L-3}{2}, e\frac{L-1}{2} \right\} \quad (5.33)$$

where e and L determine the resolution and the cardinality of the discrete set, respectively. This set becomes close to analog by taking e to be small in relation to the standard deviation of the noise and by choosing a sufficiently large L . In our simulations, we use an L in the range [300 600], and set $e = 12/(L - 1)$.

Since complexity is one of our main concerns, it is important to note that the decoder side can be approximated with a table-lookup, thereby avoiding having to compute a numerical integration for each received symbol. This is done by first discretizing the fading gain F using a discrete set \mathcal{F} (e.g., using a quantizer) and mapping the discretized receiver input Y and the fading \mathcal{F} onto a decoded symbol $\hat{V}^K = \gamma(y, f)$ for $(Y, \mathcal{F}) = (y, f)$. Note that for a given cardinality $|\mathcal{F}|$, one can design an optimal quantizer for F . However, in our simulation results we use a uniform quantizer for simplicity and set the cardinality $|\mathcal{F}|$ to 256. In a similar way to the decoder side, the encoder can be also implemented via a table-lookup by quantizing the source input. Using this approach, the system complexity is reduced to that of table lookup.

The discretized versions of (5.31) and (5.32), which are used in the implementation of the design algorithm, are, respectively, expressed as follows

$$\alpha^*(v^k) = \arg \min_{x \in \mathcal{X}_d} \left\{ \sum_{f \in \mathcal{F}} \sum_{y \in \mathcal{Y}_d} P(f) P(y|x, f) \frac{\|v^k - \hat{v}^k\|^2}{k} + \lambda x^2 \right\} \quad (5.34)$$

and

$$\gamma^*(y, f) = \frac{\sum_{v^k} v^k P(y|\alpha(v^k), f)}{\sum_{v^k} P(y|\alpha(v^k), f)}. \quad (5.35)$$

Nonparametric Mapping with Full CSI

In this section, we assume that the CSI is available at both encoder and decoder (FCSI). The main change in the necessary conditions for optimality is in finding the optimal encoder mapping α^* . The encoder mapping is now optimized for each (v^k, f) according to

$$\alpha^*(v^k, f) = \arg \min_{x \in \mathcal{X}_d} \left\{ \sum_{y \in \mathcal{Y}_d} P(y|x, f) \frac{\|v^k - \hat{v}^k\|^2}{k} + \lambda x^2 \right\}. \quad (5.36)$$

Note that the optimal decoder is given by (5.35).

Nonparametric Mapping with No CSI

In this section, we assume that no CSI is available at both encoder and decoder. The main change in the necessary conditions for optimality from the one with CSI at the decoder, is in finding the optimal decoder mapping. The optimal decoder in this case can be expressed as follows

$$\gamma^*(y) = \frac{\sum_{v^k} \sum_{f \in \mathcal{F}} v^k P(y|\alpha(v^k), f)}{\sum_{v^k} \sum_{f \in \mathcal{F}} P(y|\alpha(v^k), f)}. \quad (5.37)$$

Note that the optimal encoder is given by (5.34).

5.2.5 Numerical Results

In this section, we assume a Gaussian source vector V^K with unit variance $\sigma_V^2 = 1$ and a Rayleigh fading gain F with $\mathbb{E}[F^2] = 1$. Fig. 5.2 shows the SDR performance for parametric spiral and nonparametric 2 : 1 mapping under DCSI and FCSI. We can notice that for most CSNR values, the spiral and the non-parametric mappings

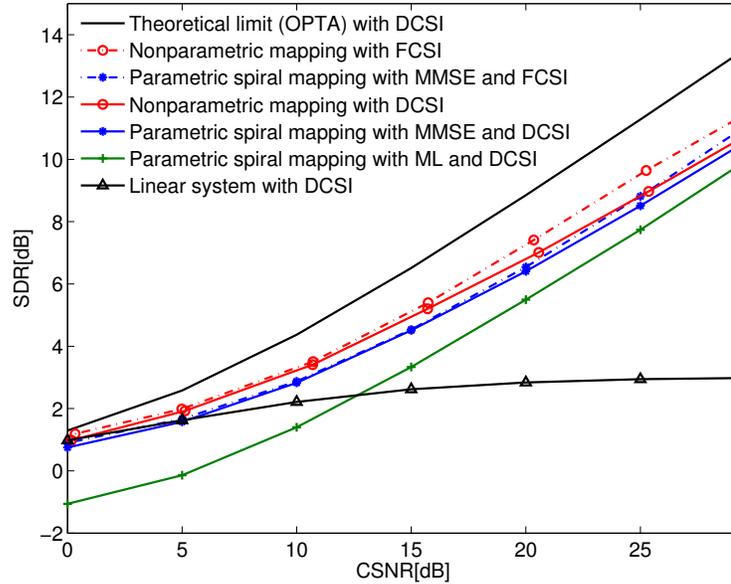


Figure 5.2: Performance of the parametric and the nonparametric mappings for 2:1 bandwidth reduction over Rayleigh fading channel. The performance of the linear system and the theoretical limit (OPTA) with CSI at the decoder are also included. Note that for FCSI, the performance of the linear system and the theoretical limit improves over the DCSI case by at most 0.2 dB in SDR.

outperform the linear system; there is around 2 dB gap from OPTA which is achievable asymptotically using a highly complex long block codes. For low to moderate CSNRs, using MMSE decoding with spiral mapping gives a substantial gain over ML decoding. However as the CSNR increases, the gap between ML and MMSE decoder diminishes. Moreover, using the optimal scaling factor with ML decoder gives a few dBs SDR gain over the one with suboptimal scaling at low CSNRs. Fig. 5.2 shows that the nonparametric mapping gives around 0.5 dB gain over the spiral mapping with MMSE decoder. Moreover, at low CSNRs, the nonparametric mapping does not underperform the linear system, unlike the parametric spiral mapping. Assuming

FCSI, we can notice around 0.5 dB gain in SDR over the DCSI case. This gain is numerically observed to reach 1 dB for CSNR = 40 dB.

Fig. 5.3 shows the performance of the spiral and the nonparametric 2 : 1 mappings when no CSI is available at both transmitter and receiver. It is clear that the non-

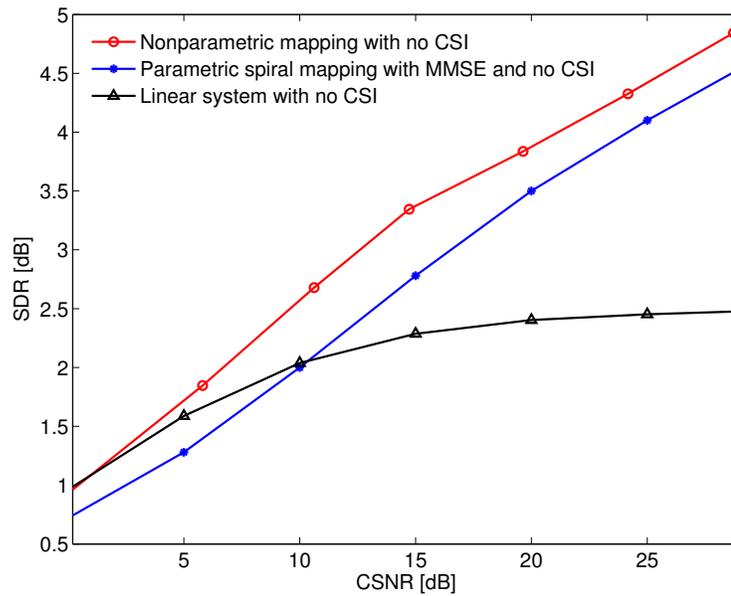


Figure 5.3: Performance of the parametric and the nonparametric mappings for 2:1 bandwidth reduction over Rayleigh fading channel. The performance of the linear system is also included. This figure is for no CSI at both encoder and decoder.

parametric mapping outperforms the spiral mapping and overcomes the saturation effect which is inevitable with the linear system.

Figs. 5.4 and 5.5 show the SDR performance of the nonparametric and the parametric mappings for 3 : 1 and 4 : 1 bandwidth reductions, respectively. In these figures we consider only the DCSI case. Similar to the 2 : 1 system, we can notice that the nonparametric mappings outperform the other systems and give a graceful

performance. More precisely, the nonparametric mappings are shown to give around 1 dB gain in SDR over the parametric mappings. This gain is due to the fact that the nonparametric mappings have a higher degree of freedom in placing points in space without being restrained to a specific structure.

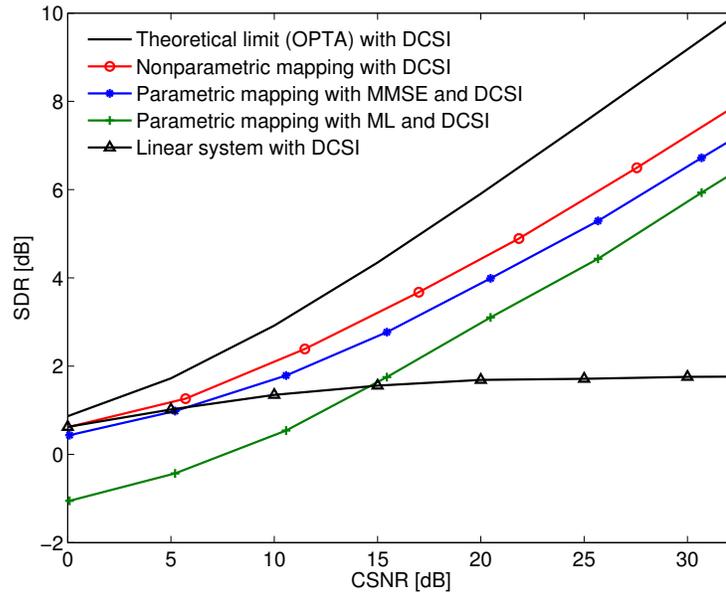


Figure 5.4: Performance of the parametric and the nonparametric mappings with 3 : 1 bandwidth reduction over Rayleigh fading channel. The performance of the theoretical limit and the linear system are also plotted.

Motivated by the broadcast scenario, we next optimize the encoder for a fixed-design CSNR level and assume that the true CSNR is known by the decoder. For this case, we can notice that both nonparametric and parametric mappings exhibit various degrees of robustness against mismatch in noise level; the nonparametric mappings still outperform the parametric mappings for most design and true CSNR levels. However, for a low design CSNR level, the parametric mappings give better performance when the true CSNR is very high. This can be explained from the fact that

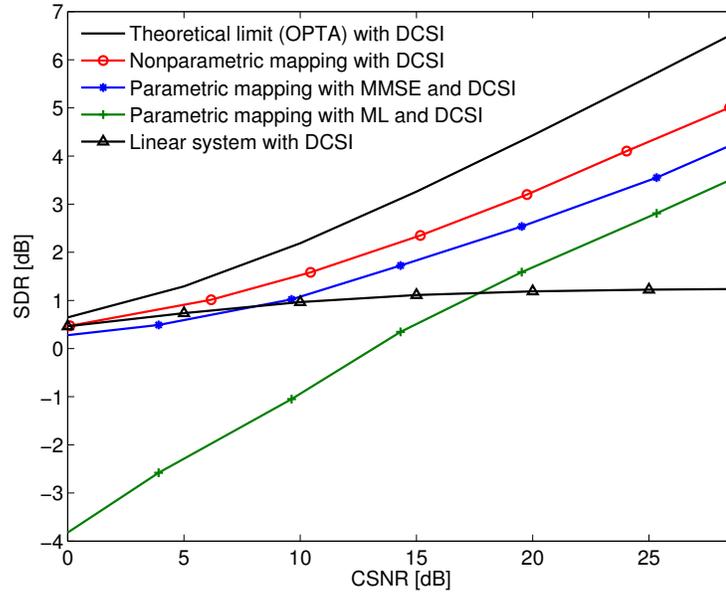


Figure 5.5: Performance of the parametric and the nonparametric mappings with 4 : 1 bandwidth reduction over Rayleigh fading channel. The performance of the theoretical limit and the linear system are also plotted.

the nonparametric mappings have a different structure at low CSNR than at high CSNR and look more like a linear mapping.

Fig. 5.6 shows the optimized decoder mappings for a 2 : 1 bandwidth reduction ratio. We can notice that the structure of the nonparametric mapping is similar to the Archimedes' spiral. Moreover, the length of the mapping curves increase as the fading gain increase (i.e., less noisy channel). This makes the mapping better fill the power-delimited space, in order to lower the approximation error without increasing much the distortion due to channel noise. Note also that the radial distance between the mapping arms is not uniform as in the case of a spiral mapping with ML decoder.

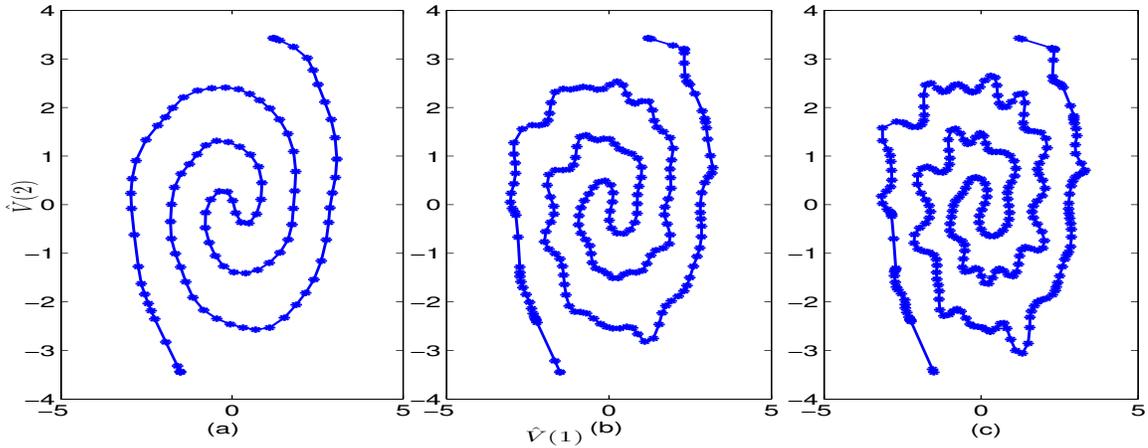


Figure 5.6: Decoded pairs for 2 : 1 bandwidth reduction, and fading gain f : (a) 0.3, (b) 0.6, and (c) 0.9. This graph is made for CSNR = 30 dB and DCSI. The lines that are drawn between asterisk points correspond to neighbours in the one dimensional mapped source (i.e., channel space).

5.3 Low Delay Coding for Fading Channels in the Presence of Correlated Interference

5.3.1 Problem Formulation and Main Contributions

We consider the transmission of a memoryless Gaussian source V^K over a Rayleigh fading channel in the presence of Gaussian interference S^N known to the transmitter. We studied this problem in Chapter 3 (Fig. 3.1) under high delay coding constraint. In this section, we aim to find an encoder α and decoder γ that minimize the MSE distortion $D = \mathbb{E}[||V^K - \hat{V}^K||^2]/K$ under low delay coding constraint (K and N take on small values). Our main contributions can be summarized as follows

- We show that for matched bandwidth between the source and the channel (i.e., $r = 1$), the uncoded scheme is optimal among all single-letter codes when the source and the interference are fully correlated.

- To benefit from nonlinearity whenever possible, we derive the necessary conditions for optimality and use an iterative algorithm based on joint optimization between the encoder and the decoder.
- To lower the complexity of our algorithm, we use importance sampling (at the decoder) technique. We also use a targeted (focused) search method in order to make our algorithm more scalable to larger dimensions (K and N) as well as more computationally efficient.

5.3.2 Optimality of the Uncoded Scheme

Lemma 5.2. *Among all single-letter (symbol-by-symbol) codes with $K = N = 1$, the uncoded scheme is optimal for full correlation between the source and the interference ($\rho_{VS} = 1$).*

Proof. In a similar way as we prove Lemma 3.1 in Chapter 3, we obtain a lower bound on the system's distortion for $K = N = 1$. From the rate distortion theorem, we have the following

$$\frac{1}{2} \log \frac{\text{Var}(V|\tilde{S})}{\mathbb{E}[(V - \hat{V})^2]} \leq I(V; \hat{V}|\tilde{S}). \quad (5.38)$$

Moreover, using the data processing inequality for the mutual information, the facts that conditioning reduces differential entropy and that the Gaussian distribution maximizes differential entropy, $I(V; \hat{V}|\tilde{S})$ can be upper bounded for any memoryless single-letter codes (α, γ) with $X = \alpha(V, S)$ and $\hat{V} = \gamma(Y, f)$ by

$$I(V; Y|\tilde{S}) = h(Y|\tilde{S}) - h(W) \leq \frac{1}{2} \log \frac{\text{MSE}(Y; \tilde{S})}{\sigma_W^2}. \quad (5.39)$$

Combining (5.38) and (5.39), we have $\mathbb{E}[(V - \hat{V})^2|f] \geq \frac{\text{Var}(V|\tilde{S})\sigma_W^2}{\text{MSE}(Y;\tilde{S})}$. As a result,

$$\mathbb{E}[(V - \hat{V})^2] \geq \mathbb{E}_F \left[\frac{\text{Var}(V|\tilde{S})\sigma_W^2}{\text{MSE}(Y;\tilde{S})} \right]. \quad (5.40)$$

The bound in (5.40) can be tightened by maximizing the right hand side of (5.40) over η_1 and η_2 . If we choose $\eta_1 = 1$ and $\eta_2 = 0$, (5.40) reduces to

$$\mathbb{E}[(V - \hat{V})^2] \geq \mathbb{E}_F \left[\frac{\sigma_V^2}{\left(1 + \frac{|f|^2(\sqrt{P} + \rho_{VS}\sigma_S)^2}{\sigma_W^2}\right)} \right]. \quad (5.41)$$

The MSE of the uncoded scheme (using D_{linear}) in Sec. 3.3.2 is given by

$$\begin{aligned} D_{uncoded} &= \mathbb{E}_F \left[\frac{|f|^2\sigma_V^2\sigma_S^2(1 - \rho_{VS}^2)}{|f|^2(P + \sigma_S^2 + 2\sqrt{P}\rho_{VS}\sigma_S) + \sigma_W^2} \right] \\ &+ \mathbb{E}_F \left[\frac{\sigma_V^2}{1 + \frac{|f|^2(P + \sigma_S^2 + 2\sqrt{P}\rho_{VS}\sigma_S)}{\sigma_W^2}} \right]. \end{aligned} \quad (5.42)$$

It is easy to see that equality occurs in (5.41) using the uncoded scheme for $\rho_{VS} = 1$ (see (5.42)). \square

5.3.3 Design Algorithm for Low Delay Coding

In a similar way as done in Sec. 5.2.4, we present a scheme based on joint optimization between the encoder and the decoder through an iterative algorithm. Using the Lagrange multiplier method, the problem is to find the encoder and decoder that minimize the Lagrange cost function $J(\alpha, \gamma)$

$$J(\alpha, \gamma) = \frac{\mathbb{E}[||V^K - \gamma(Y^N, F^N)||^2]}{K} + \lambda \frac{\mathbb{E}[||\alpha(V^K, S^N)||^2]}{N}. \quad (5.43)$$

Necessary Conditions for Optimality

The optimal encoder mapping α^* (assuming γ is fixed) is given by

$$\arg \min_{\alpha} \left\{ \frac{\mathbb{E}[\|V^K - \gamma(Y^N, F^N)\|^2]}{K} + \lambda \frac{\mathbb{E}[\|\alpha(V^K, S^N)\|^2]}{N} \right\}. \quad (5.44)$$

Using Bayes' rule, the distortion $\mathbb{E}[\|V^K - \hat{V}^K\|^2]$ is given by

$$\iiint p(v^k, s^n) p(y^n | \alpha(v^k, s^n), s^n, f^n) p(f^n) \|v^k - \hat{v}^k\|^2 dv^k ds^n dy^n df^n. \quad (5.45)$$

The average power is given by

$$\frac{1}{N} \iint p(v^k, s^n) \|\alpha(v^k, s^n)\|^2 dv^k ds^n. \quad (5.46)$$

Since $p(v^k, s^n)$ in (5.45)–(5.46) is nonnegative, α can be optimized “pointwise” for each v^k and s^n according to

$$\alpha^*(v^k, s^n) = \arg \min_{x^n \in \mathbb{R}^n} \left\{ \frac{1}{K} \iint p(y^n | x^n, s^n, f^n) \|v^k - \hat{v}^k\|^2 p(f^n) dy^n df^n + \frac{\lambda}{N} \|x^n\|^2 \right\}. \quad (5.47)$$

Thus, (5.47) is a necessary condition for an optimal encoder.

On the receiver side, the optimal decoder in the MSE sense (assuming α is fixed) is given by $\mathbb{E}[V^K | y^n, f^n]$ as follows

$$\gamma^*(y^n, f^n) = \frac{\iint v^k p(y^n | v^k, s^n, f^n) p(v^k, s^n) dv^k ds^n}{\iint p(y^n | v^k, s^n, f^n) p(v^k, s^n) dv^k ds^n}. \quad (5.48)$$

Design Algorithm and Implementation Aspects

Based on the above necessary conditions for optimality, we optimize α and γ using an iterative process based on (5.47) and (5.48). The design algorithm is similar to the one in Sec. 5.2.4.

For the implementation of (5.47) and (5.48), some modifications are required. Since it is intractable to evaluate the formulas for all real-valued (V^K, S^N) , we form a set of pairs $(\mathcal{V}, \mathcal{S})$ composed of samples drawn from $p(v^k, s^n)$. We also discretize the channel input and output using a PAM alphabets \mathcal{X}_d and \mathcal{Y}_d .

Even after discretizing the channel input and output using \mathcal{X}_d and \mathcal{Y}_d , the (offline) design algorithm is still computationally expensive. This is due to the fact that our problem has both fading and interference on top of AWGN. To lower the complexity, we resort to MC and importance sampling techniques. Using MC technique, (5.47) is given by

$$\alpha^* = \arg \min_{x^n \in \mathcal{X}_d^n} \sum_{i=1}^l \frac{\|v^k - \gamma(y_i^n, f_i^n)\|^2}{K} + \frac{\lambda}{N} \|x^n\|^2 \tag{5.49}$$

and (5.48) can be expressed as follows

$$\gamma^* = \frac{\sum_{i=1}^l v_i^k p(y_i^n | \alpha(v_i^k, s_i^n), s_i^n, f_i^n)}{\sum_{i=1}^l p(y_i^n | \alpha(v_i^k, s_i^n), s_i^n, f_i^n)} \tag{5.50}$$

where (v_i^k, s_i^n) and f_i^n are MC samples that follow the distributions $p(v^k, s^n)$ and $p(f^n)$, respectively, and y_i^n represents the discretized version of the corresponding MC samples using \mathcal{Y}_d ; equality in (5.49) and (5.50) are valid by the strong law of large numbers when $l \rightarrow \infty$. Note that the discretization of y_i^n is necessarily to couple (5.49) and (5.50); this is how we implement the “fixing” of decoder mapping when updating the encoder (5.49). To increase the convergence rate of (5.50) (i.e.,

lower l and the complexity), an alternative to sampling from $p(v^k)$ is to use importance sampling. This is conducted by using samples from another distribution $q(v^k)$. After some manipulation, (5.50) is given by

$$\gamma^*(y^n, f^n) = \frac{\sum_{i=1}^l \frac{v_i^k p(y^n | \alpha(v_i^k, s_i^n), s_i^n, f^n) p(v_i^k, s_i^n)}{q(v_i^k)}}{\sum_{i=1}^l \frac{p(y^n | \alpha(v_i^k, s_i^n), s_i^n, f^n) p(v_i^k, s_i^n)}{q(v_i^k)}}. \quad (5.51)$$

$q(v^k)$ has to be chosen to improve the convergence speed. Since suboptimal linear decoding gives some information about the estimate, we choose $q(v^k) \sim \mathcal{N}(v_{subopt}^k, D_{linear})$, where v_{subopt}^k and D_{linear} are the source estimate and the MSE distortion from applying the linear decoder, respectively. One major issue that stands against the scalability of our algorithm is that for each pair of (v^k, s^n) one has to search over a set of cardinality $|\mathcal{X}_d|^n$ which scales exponentially with the number of channel dimensions n . To reduce the search complexity, we use a heuristic targeted search approach. Instead of fixing the N dimensional PAM alphabet \mathcal{X}_d^n and blindly searching over the whole region (fixed-search method), this targeted method starts by mapping each pair (v^k, s^n) to a channel input $(x^n)^{(0)}$ using a parametric mapping and then searching over a region that varies for each pair and iteration i . For each source-interference pair, we choose the region to have a span of $\mathcal{X}_d^{(i)} = (x)^{(i-1)} \pm 4(\sigma_{span})^{(i)}$ in each dimension, where $(x^n)^{(i-1)}$ is the channel input found for that pair in iteration $i - 1$. We have chosen $(\sigma_{span}^2)^{(i)}$ to decrease with i as follows

$$(\sigma_{span}^2)^{(i)} = \tilde{c} \frac{(L-1)e}{4} \left(1 - \frac{i}{I_{max}}\right)^3 \quad i = 1, \dots, I_{max}$$

where \tilde{c} is a constant. This method limits the search (at iteration i) to some small neighbourhood of the current state $(x^n)^{(i-1)}$, with neighbourhoods diminishing in size

as the algorithm progresses. This method is not used at the channel output and we still need to use a fixed \mathcal{Y}_d . We run the algorithm twice with the second round initialized with the result of the first; this helps “escaping” from a local minimum at the beginning of the second round when $(\sigma_{span}^2)^{(1)}$ is large. In our simulation, we use 10^5 pairs to define $(\mathcal{V}, \mathcal{S})$, $\varepsilon = 10^{-4}$, $I_{max} = 14$, $\tilde{c} = 1/8$ and $l = 250$ for $K, N \leq 3$. Note that after the design process of the encoder and decoder ends, both elements can be implemented using a table-lookup; for the encoder, this is done by quantizing the source-interference pairs and mapping each pair to a channel input.

5.3.4 Numerical Results and Discussion

In this section, we consider a source with $\sigma_V^2 = 1$ that is correlated to the interference as considered in Chapter 3 (only the first $\min(K, N)$ pairs $\{V(i), S(i)\}_{i=1}^{\min(K, N)}$ are correlated) and Rayleigh fading with $\mathbb{E}[F^2] = 1$.

Matched Bandwidth Case

Table 5.1 shows the SDR of the optimized mapping along with the linear and the uncoded schemes for $K = N = 1$. We first design our system for $\rho_{VS} = 0$ (linear scheme is used as initialization); the solution obtained is then used as initialization to design the system for higher values of ρ_{VS} . We perform a numerical search over the coding matrices for the linear scheme so that D_{linear} is minimized under the power constraint P . We notice that as ρ_{VS} increases, the SDR gap between the optimized and the linear scheme decreases (also the upper bound on all single-letter codes). For $\rho_{VS} = 1$, the numerical result confirms Lemma 5.2; all schemes reduce to the uncoded one and achieve optimality.

Table 5.1: SDR in dB versus ρ_{VS} for $P = 0.6$, $\sigma_S^2 = 1$, $\sigma_W^2 = 0.01$ and $r = 1$.

ρ_{VS}	0	0.2	0.4	0.6	0.8	1
Uncoded	1.9	2.9	4.1	5.7	8.4	17.8
Linear	3.5	5.6	8.7	12.4	15.8	17.8
Optimized mapping	6.4	6.8	9.1	12.5	15.9	17.8
SDR upper bound	12.2	12.4	12.9	14.2	16.9	17.8

Mismatched Bandwidth Case

Fig. 5.7 shows the SDR of our proposed mapping for 2 : 1 and 3 : 2 bandwidth reductions versus CSNR. We notice that the optimized mapping outperforms the other reference schemes for all CSNRs (for the same $K : N$). For the 2:1 system, we initialize the algorithm with a parametric mapping based on a spiral curve [32] and the search range $\mathcal{X}_d^{(i)}$ is set to 100 points. As a reference, we also plot the performance of an (optimized) parametric mapping based on the combination of spiral and sawtooth (modulo technique used for interference cancelation [16]). For the 3 : 2 case, a linear mapping is used for initialization and $\mathcal{X}_d^{(i)}$ is set to 15 points in each direction. For reference, a parametric mapping that uses a linear transmission on the first symbol and 2 : 1 spiral mapping on the last two symbols to achieve a 3 : 2 bandwidth reduction is combined linearly with the interference and simulated. For the alphabet \mathcal{Y}_d , we use $e = 8/(L - 1)$, $L = 500$ (for the 2 : 1 system) and $L = 20$ (for the 3:2 system). Fig. 5.7 shows also the SDR upper bound; it is important to note that the bound is an asymptotic result in the sense of infinite source and coding block lengths, hence the gap to our low delay scheme is not surprising. We also design an optimized mapping for 1 : 2 expansion system; we notice up to 1.6 dB gain in SDR over the “best” parametric system (linear in this case) for CSNR = 20 and $\rho_{VS} = 0.3$. As initialization to our algorithm, we use the resulting optimized mapping assuming no

fading.

From other simulations, comparing the targeted search method to the fixed one using the same alphabets cardinality (with similar and tractable complexity), we notice that the proposed search method gives up to 1 dB gain in SDR over the fixed search method for moderate to high CSNRs (for 2 : 1 system). For low CSNRs, not much gain is observed; this is due to the fact that the (required) number of discrete points L depends on the noise levels (in our case, L is fixed). For high noise levels, L can be made smaller with no performance loss.

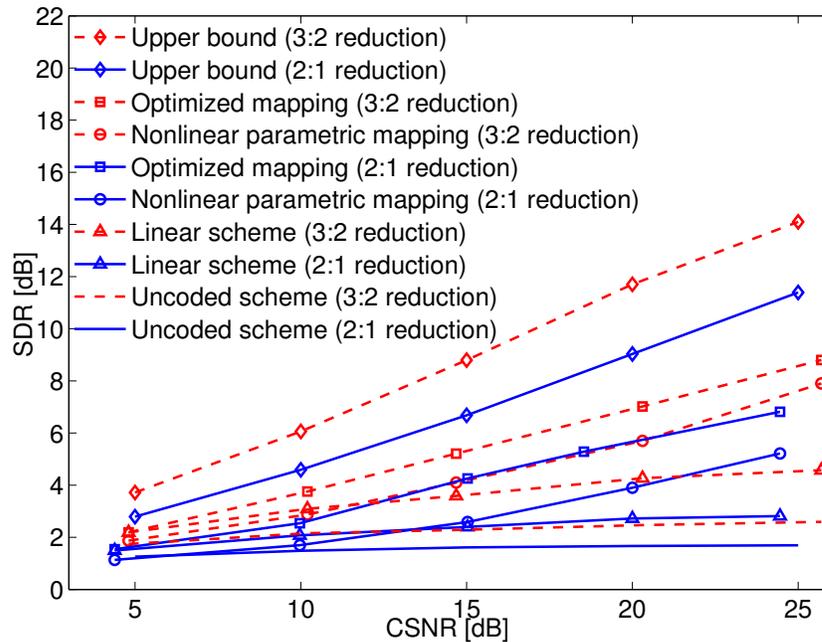


Figure 5.7: SDR Performance versus CSNR for $P = 1$, $\sigma_S^2 = 1$, $\rho_{VS} = 0.3$.

5.4 Low Delay Coding for Source and State Interference Estimation

5.4.1 Problem Formulation and Main Contributions

In this section, we consider the transmission of analog source and channel state interference over an AWGN channel with additive interference that is known to the transmitter. The system model, which is a special case of the one considered in Sec. 3.5 (no fading and $K = N = 1$), is shown in Fig. 5.8. We aim to find an encoder

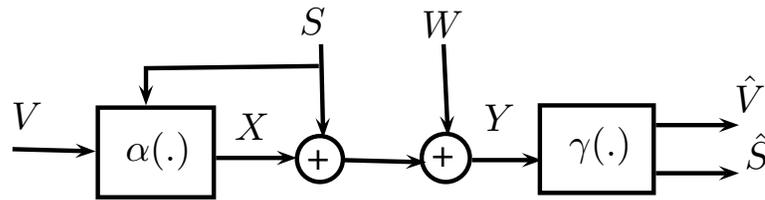


Figure 5.8: System model for joint source and state interference estimation.

α and a decoder $\gamma = (\gamma_v, \gamma_s)$ that minimize the MSE distortion defined by

$$D \triangleq \underbrace{\mathbb{E}[(V - \hat{V})^2]}_{D_v} + \underbrace{\mathbb{E}[(S - \hat{S})^2]}_{D_s}. \tag{5.52}$$

More precisely, we focus on zero delay analog joint source-channel coding techniques by studying parametric and nonparametric (nonlinear) mappings. Our main contributions can be summarized as follows

- To benefit from nonlinearity whenever possible, we study a parametric analog mapping based on the sawtooth (modulo) function which has a low design complexity. We derive an upper bound on the system’s distortion by assuming a suboptimal decoder at the receiver side. To optimize the system parameters, we use two suboptimal methods; the first one is partially numerical in which part

of the parameters are derived by minimizing a (partial) distortion expression that assumes no sawtooth mapping is used at the encoder. The other method minimizes the derived upper bound expression.

- Whenever storage and offline design complexity are not an issue, we design a nonlinear mapping; this is done by deriving the necessary conditions for optimality and proposing an iterative algorithm based on joint optimization between the transmitter and the receiver.

5.4.2 Distortion Lower Bound

In Lemma 3.4, we have derived an outer bound on the distortion region (D_v, D_s) for the same problem over fading channels. For the AWGN channel, the outer bound simplifies as follows

$$\begin{aligned} D_v &\geq D_v^{ob} \triangleq \frac{\text{Var}(V|S)\sigma_W^2}{\zeta P + \sigma_W^2}, \\ D_s &\geq D_s^{ob} \triangleq \frac{\sigma_S^2(\zeta P + \sigma_W^2)}{P + \sigma_S^2 + 2\sqrt{(1-\zeta)P\sigma_S^2} + \sigma_W^2}. \end{aligned} \quad (5.53)$$

As a result, the lower bound on the system's distortion can be expressed as follows

$$D \geq \inf_{\zeta} \{D_v^{ob} + D_s^{ob}\}. \quad (5.54)$$

5.4.3 Parametric Mapping

Recall that since V and S are correlated, we can write the source as $V = \frac{\rho_{VS}\sigma_V}{\sigma_S}S + N_{\rho_{VS}}$, where $N_{\rho_{VS}} \sim \mathcal{N}(0, (1 - \rho_{VS}^2)\sigma_V^2)$ is independent of S . Motivated by the high delay scheme that uses analog and hybrid layers in Sec. 3.5, we herein propose

a layered scheme based on linear coding and sawtooth mapping (sawtooth coding is used in [73] for the relay channel). The sawtooth mapping can be seen as one dimensional lattice coding; high delay lattice coding is widely studied for AWGN channel with side information [18, 40].

System Structure

The proposed scheme is composed of two superposed layers and outputs

$$X = c(X_1 + X_2) \quad (5.55)$$

where c is a gain factor related to the power constraint (defined later). The first layer, which outputs $X_1 = \sqrt{P_s/\sigma_S^2}S$, simply scales the interference S , where $P_s \leq P$ represents the power consumed by this layer. The second layer, starts by forming a linear combination of the partial information of the source $N_{\rho_{VS}}$ and the interference S ; this is given by $X_a = \alpha_1 N_{\rho_{VS}} + \alpha_2 S$, where α_1, α_2 are real parameters. We then use a sawtooth mapping $\mathcal{M}(\cdot)$ on X_a to output X_2 as follows

$$X_2 \triangleq \mathcal{M}(X_a) = (X_a - 2\Lambda m) \text{ for } X_a \in [\Lambda(2m - 1), \Lambda(2m + 1)) \quad (5.56)$$

where m is an integer and Λ is a nonnegative parameter dependent on the channel condition. The gain factor c in (5.55) is given by $c = \sqrt{P/(P_s + \mathbb{E}[X_2^2] + 2\sqrt{P_s/\sigma_S^2}\mathbb{E}[SX_2])}$, where $\mathbb{E}[X_2^2]$ can be written as follows

$$\mathbb{E}[X_a^2] + \sum_m -4\Lambda m \underbrace{\int_{D_m} x_a p(x_a) dx_a}_{I_1} + 4\Lambda^2 m^2 \underbrace{\int_{D_m} p(x_a) dx_a}_{I_2} \quad (5.57)$$

and $\mathbb{E}[SX_2]$ is given by

$$\alpha_2 \sigma_S^2 - \sum_m 2\Lambda m \int \int_{D_m} sg(x_a - \alpha_2 s) p(s) dx_a ds \quad (5.58)$$

where $D_m = [\Lambda(2m - 1), \Lambda(2m + 1))$ is the m^{th} domain region of $\mathcal{M}(\cdot)$ and $g(\cdot)$ is the pdf of $\alpha_1 N_{\rho_{VS}} \sim \mathcal{N}(0, \alpha_1^2 \sigma_V^2 (1 - \rho_{VS}^2))$. Note that the integrals in (5.57) can be simplified as follows

$$\begin{aligned} I_1 &= \frac{\sqrt{\mathbb{E}[X_a^2]}}{\sqrt{2\pi}} \left[-\exp\left(\frac{-(\Lambda(2m+1))^2}{2\mathbb{E}[X_a^2]}\right) + \exp\left(\frac{-(\Lambda(2m-1))^2}{2\mathbb{E}[X_a^2]}\right) \right], \\ I_2 &= \frac{1}{2} \left[\operatorname{erf}\left(\frac{\Lambda(2m+1)}{\sqrt{2\mathbb{E}[X_a^2]}}\right) - \operatorname{erf}\left(\frac{\Lambda(2m-1)}{\sqrt{2\mathbb{E}[X_a^2]}}\right) \right] \end{aligned} \quad (5.59)$$

where $\operatorname{erf}(\cdot)$ is the Gaussian error function. At the decoder side, to obtain an estimate of the source and the interference, we use the optimal MMSE estimator ($\hat{V} = \mathbb{E}[V|Y]$, $\hat{S} = \mathbb{E}[S|Y]$). The use of an optimal decoder comes at the expense of computational and design complexity. To lower the design complexity, we resort to two suboptimal methods for choosing the system parameters as described next.

System Optimization

Method 1

In this method, the optimized parameters α_1 and α_2 are found by assuming that no sawtooth mapping is used. In such case, the parameters α_1 and α_2 are found by minimizing the MSE distortion $D_{N_{\rho_{VS}}}$ from reconstructing $N_{\rho_{VS}}$ using an LMMSE estimator. This distortion is given by (assuming no sawtooth mapping) $D_{N_{\rho_{VS}}} =$

$\sigma_V^2(1 - \rho_{VS}^2) - \frac{\mathbb{E}[N_{\rho_{VS}}Y]^2}{\mathbb{E}[Y^2]}$. The sawtooth parameter Λ and P_s , however, are found numerically to minimize the overall MSE distortion D by performing a grid search. This is done by generating a large set of (V, S, W) triplets and computing D empirically for each possible (P_s, Λ) in the search space.

Method 2

In this method, we optimize the system parameters by minimizing an upper bound on the system's distortion. To get a closed form expression on the upper bound, we propose the use of a suboptimal decoder. Let us first note that the sawtooth mapping, which uses the symmetric modulo function (5.56) over the interval $[-\Lambda, \Lambda]$, can be written as $\mathcal{M}(X_a) = X_a \bmod \Lambda$. To reconstruct the interference, we simply use an LMMSE estimator based on the received signal Y . The distortion from reconstructing S is given by

$$\begin{aligned} (D_s)_{\text{parametric}} &= \sigma_S^2 - \frac{\mathbb{E}[SY]^2}{\mathbb{E}[Y^2]} \\ &= \sigma_S^2 - \frac{(c(\mathbb{E}[S(X_a \bmod \Lambda)] + \sqrt{P_s}\sigma_S) + \sigma_S^2)^2}{P + \sigma_S^2 + \sigma_W^2 + 2c(\sqrt{P_s}\sigma_S + \mathbb{E}[S(X_a \bmod \Lambda)])} \end{aligned} \quad (5.60)$$

where $\mathbb{E}[S(X_a \bmod \Lambda)]$ can be written as in (5.58).

To get an estimate of the source V , we first use a modulo function on the received signal and then apply an LMMSE estimator. More precisely, we first obtain

$$\begin{aligned} \tilde{Y} &= (Y/c) \bmod \Lambda \\ &= \underbrace{\left(\alpha_1 N_{\rho_{VS}} + \left(\alpha_2 + \sqrt{\frac{P_s}{\sigma_S^2} + \frac{1}{c}} \right) S + \frac{W}{c} \right)}_{\tilde{Z}} \bmod \Lambda \end{aligned} \quad (5.61)$$

where the last equality follows from the fact that the modulo operation satisfies the “distributive law” (i.e., $[x \bmod \Lambda + y] \bmod \Lambda = [x + y] \bmod \Lambda$). We then decode V using an LMMSE estimator based on \tilde{Y} . The resulting distortion is

$$(D_v)_{\text{parametric}} = \sigma_V^2 - \frac{\mathbb{E}[V(\tilde{Z} \bmod \Lambda)]^2}{\mathbb{E}[(\tilde{Z} \bmod \Lambda)^2]} \quad (5.62)$$

where \tilde{Z} is as defined in (5.61) and $\mathbb{E}[(\tilde{Z} \bmod \Lambda)^2]$ is given by

$$\mathbb{E}[\tilde{Z}^2] - 4\Lambda \sum_m m \int_{\tilde{D}_m} \tilde{z} p(\tilde{z}) d\tilde{z} + 4\Lambda^2 \sum_m m^2 \int_{D_m} p(\tilde{z}) d\tilde{z} \quad (5.63)$$

and $\mathbb{E}[V(\tilde{Z} \bmod \Lambda)]$ can be expressed as follows

$$\mathbb{E}[V\tilde{Z}] - \sum_m 2\Lambda m \int_{\tilde{D}_m} \iint vp(v|s)p(\tilde{z}|v, s)p(s)dvdsd\tilde{z} \quad (5.64)$$

where $\tilde{D}_m = [\Lambda(2m - 1), \Lambda(2m + 1))$ is the m^{th} domain region of $\mathcal{M}(\cdot)$. Note that integrals in (5.63) can be simplified in a similar way as in (5.59) and distributions in (5.64) are Gaussian. The upper bound on the system’s distortion D_{upper} for parametric mapping is then given by

$$D_{\text{upper}} = (D_v)_{\text{parametric}} + (D_s)_{\text{parametric}}. \quad (5.65)$$

The system’s parameters $(\alpha_1, \alpha_2, P_s, \Lambda)$ are found by minimizing D_{upper} .

Remark 5.1. For low correlation values between the source and the interference, we propose to use a slightly different decoder (in method 2) for estimating V that gives a better performance. The only modification is that $\tilde{Y} = \frac{-\alpha_2 Y}{c\kappa} \bmod \Lambda$, where

$\kappa = \left(\sqrt{\frac{P_u}{\sigma_S^2}} + \frac{1}{c} \right)$. After some manipulations, we can write $\tilde{Y} = (\alpha_1 V + W_{eq}) \bmod \Lambda$, where $W_{eq} = \frac{-\alpha_2 W}{c\kappa} - (1 + \frac{\alpha_2}{\kappa}) X_a \bmod \Lambda$. Note that W_{eq} can be regarded as an equivalent noise term.

5.4.4 Nonparametric Mapping

In a similar way as done in Sec. 5.2.4 and Sec. 5.3.3, we next present a scheme based on joint optimization between the encoder and the decoder through an iterative algorithm. The rest of this section is dedicated to the design of the source-channel mapping $\alpha(V, S)$ and the decoder $\gamma(Y) = (\gamma_v(Y), \gamma_s(Y))$. The Lagrange cost function $J(\alpha, \gamma)$ of the unconstrained minimization is given by

$$\mathbb{E}[(V - \gamma_v(Y))^2] + \mathbb{E}[(S - \gamma_s(Y))^2] + \lambda \mathbb{E}[\alpha^2(V, S)]. \quad (5.66)$$

The optimal encoder mapping α^* (assuming γ is fixed) can be expressed as follows

$$\arg \min_{\alpha} \left\{ \mathbb{E}[(V - \hat{V})^2] + \mathbb{E}[(S - \hat{S})^2] + \lambda \mathbb{E}[\alpha^2(V, S)] \right\}. \quad (5.67)$$

Using Bayes' rule, the distortion $\mathbb{E}[(V - \hat{V})^2]$ is given by

$$\iiint p(v, s) p(y | \alpha(v, s), s) (v - \hat{v})^2 dv ds dy. \quad (5.68)$$

Similarly, the distortion $\mathbb{E}[(S - \hat{S})^2]$ can be expressed as follows

$$\iiint p(v, s) p(y | \alpha(v, s), s) (s - \hat{s})^2 dv ds dy. \quad (5.69)$$

The average power consumed by the mapping is given by

$$\iint p(v, s)\alpha(v, s)^2 dv ds. \quad (5.70)$$

Since $p(v, s)$ in (5.68), (5.69) and (5.70) is nonnegative, the encoder α^* can be optimized “pointwise” for each (v, s) according to

$$\arg \min_{x \in \mathbb{R}} \left\{ \int p(y|x, s)[(v - \hat{v})^2 + (s - \hat{s})^2] dy + \lambda x^2 \right\}. \quad (5.71)$$

Thus, (5.71) is a necessary condition for an optimal encoder.

On the receiver side, the optimal decoder γ_v in the MSE sense (assuming α is fixed) is given by $\mathbb{E}[V|y]$ as follows

$$\gamma_v^*(y) = \frac{\iint v p(y|v, s) p(v, s) dv ds}{\iint p(y|v, s) p(v, s) dv ds}. \quad (5.72)$$

Similarly, the optimal decoder γ_s is given by $\mathbb{E}[S|y]$ as follows

$$\gamma_s^*(y) = \frac{\iint s p(y|v, s) p(v, s) dv ds}{\iint p(y|v, s) p(v, s) dv ds}. \quad (5.73)$$

Using the above necessary conditions for optimality, we optimize α and γ via an iterative process based on (5.71), (5.72) and (5.73) in a similar fashion as done in Sec. 5.2.4. For the implementation of the algorithm, we again use a PAM alphabets \mathcal{X}_d and \mathcal{Y}_d . In our simulations, we use $\varepsilon = 10^{-4}$, $I_{max} = 15$, $L = 700$ and $e = 12/(L - 1)$ for the algorithm and the alphabets \mathcal{X}_d and \mathcal{Y}_d .

5.4.5 Numerical Results

In this section, we consider source-interference pairs that are transmitted over an AWGN channel with Gaussian interference and power constraint $P = 1$. Fig. 5.9 shows the performance, defined as $10 \log \frac{1}{D}$, versus the correlation ρ_{VS} for channel signal-to-noise ratio CSNR = 25 dB.

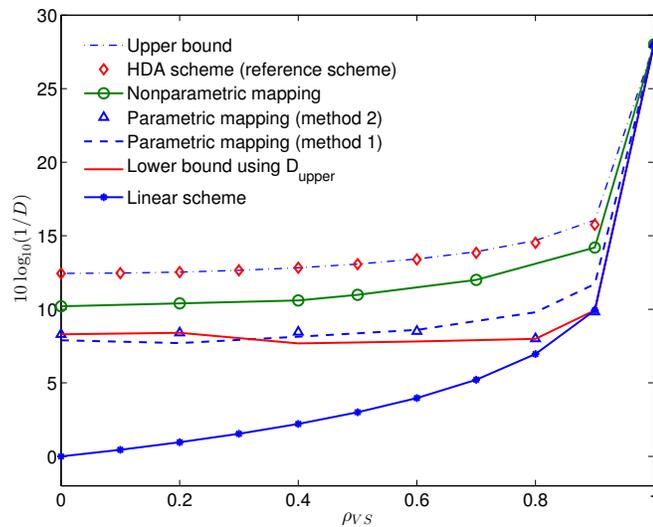


Figure 5.9: Performance versus source-interference correlation ρ_{VS} , for $\sigma_V^2 = \sigma_S^2 = 1$ and CSNR = 25 dB.

The nonparametric mapping outperforms other zero delay coding schemes; using parametric mapping as initialization for the algorithm gives 0.5 to 1 dB gain in performance (for $\rho_{VS} \leq 0.8$) over the case where we use a linear mapping to initialize the algorithm. The parametric mapping (with MMSE decoding) which is easier to design, outperforms the linear scheme and performs relatively close to the nonparametric mapping. Note that using the optimized parameters resulting from method 1, gives the ‘best’ performance for high correlation values; for low correlation values, however,

it is better to use the optimized parameters resulting from minimizing D_{upper} given by (5.65) (method 2). This behaviour comes from the fact that in method 2 we minimize a different objective function than the one used in method 1. As shown from Fig. 5.9, the lower bound obtained from the derived D_{upper} is close to the performance of the parametric mapping (optimized via method 1) with MMSE decoding for low to moderate correlation values; for high correlation values, we can notice some gap that is also manifested in the performance of method 2. Moreover, for the case of $\rho_{VS} = 1$, all schemes revert to the uncoded scheme which is optimal. It is worth mentioning that the upper bound and the HDA scheme (reference scheme) presented in Sec. 3.5 are asymptotic in the sense of requiring infinite source and coding block lengths, hence the gap to the proposed zero delay schemes is not surprising.

Fig. 5.10 shows the performance versus CSNR levels. We can notice that the non-

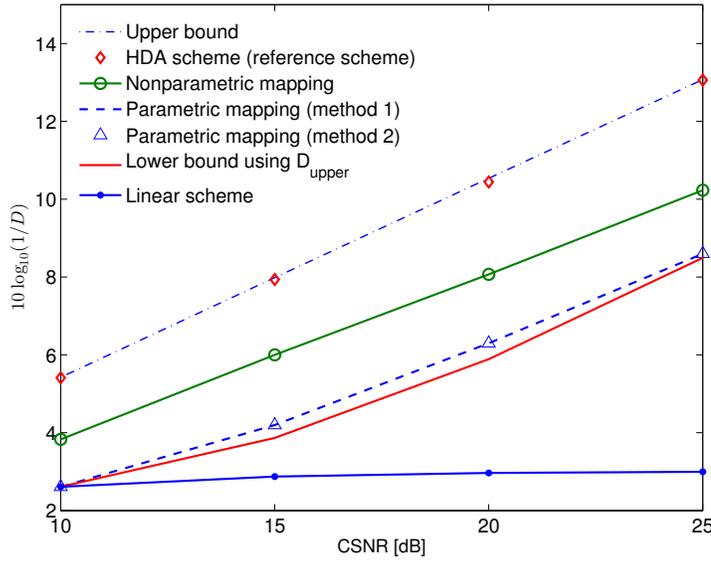


Figure 5.10: Performance versus CSNR levels, for $\rho_{VS} = 0.5$ and $\sigma_V^2 = \sigma_S^2 = 1$.

parametric mapping outperforms other zero delay schemes and that the lower bound found using D_{upper} is very close to the performance of the parametric mapping (optimized via method 1) with MMSE decoder. Moreover, the use of nonlinear mappings defeats the saturation effect which is inevitable with the use of the linear scheme.

Figs. 5.11 and 5.12 show the encoder-decoder structure of the nonparametric mapping for two different correlation values. It is clear that the encoder and decoder map-

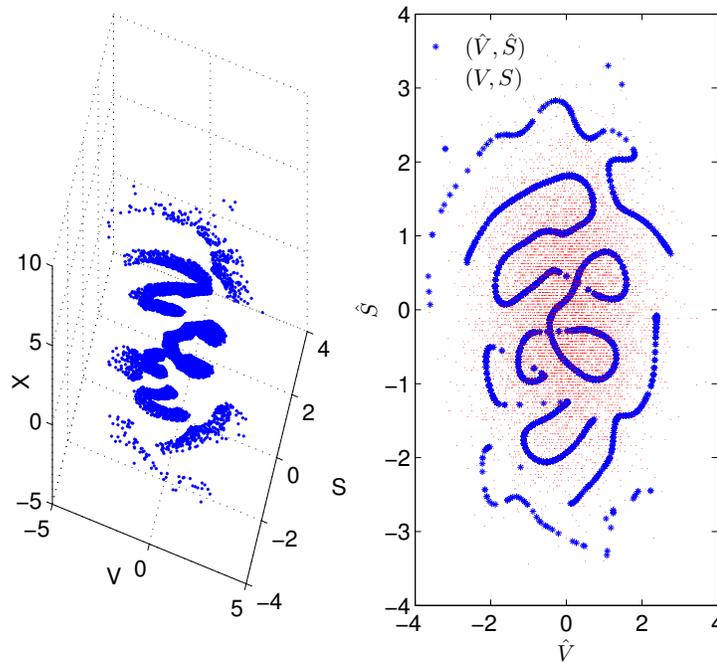


Figure 5.11: Encoder (left) and its corresponding decoder (right) mappings optimized using the proposed algorithm for $\text{CSNR} = 25$, $\rho_{VS} = 0$, $\sigma_V = \sigma_S = 1$ and $P = 1$; parametric mapping is used for the initialization of the algorithm. In the figure to the right, the asterisks show the reconstructed (\hat{V}, \hat{S}) and the small dots are samples from the distribution of (V, S) .

pings comprise a piecewise nonlinear function that combines hard and soft decision signalling. The proposed parametric mapping uses such combination; this explains

the good performance achieved using parametric mapping. There is always a gain from using the nonparametric mapping; this is due to the fact that the nonparametric mapping has a higher degree of freedom in placing points in space without being restrained to a specific structure. Such gain comes at the expense of higher storage and offline design complexity.

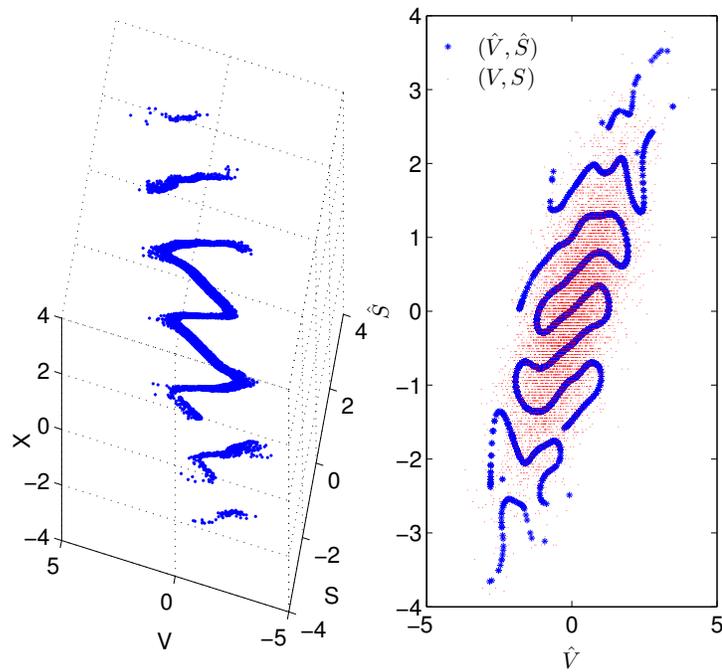


Figure 5.12: Encoder (left) and its corresponding decoder (right) mappings optimized using the proposed algorithm for $\text{CSNR} = 25$, $\rho_{VS} = 0.7$, $\sigma_V = \sigma_S = 1$ and $P = 1$; parametric mapping is used for the algorithm initialization. In the figure to the right, the asterisks show the reconstructed (\hat{V}, \hat{S}) and the small dots are samples from the distribution of (V, S) .

5.5 Summary and Conclusions

In this chapter, we present low delay lossy joint source-channel coding schemes for: 1) fading channels; 2) fading channels in the presence of correlated interference; and 3) source-channel-state (interference) estimation. A design algorithm for optimizing the source-channel mapping is presented based on the necessary conditions for optimality. Parametric mappings that use spiral and sawtooth curves are studied. Numerical results show that parametric and nonparametric mappings outperform the linear scheme. Moreover, these nonlinear mappings overcome the saturation effect which is inevitable with linear system.

Chapter 6

Conclusions

In this chapter, conclusions are drawn based on the principal results of the previous chapters. Some issues and recommendations for further research in this area are also discussed.

6.1 Summary and Conclusions

Chapter 3 considers the problem of sending a Gaussian source over a fading channel in the presence of Gaussian interference known non-causally to the transmitter. Joint source-channel coding schemes for the case of unequal bandwidth between the source and the channel and when the source and the interference are correlated are studied. An outer bound on the system's distortion is derived in Sec. 3.3. This is achieved by assuming partial knowledge of the interference at the decoder side. Sec. 3.4 presents a layered coding schemes based on proper combination of power splitting, bandwidth splitting, Wyner-Ziv and hybrid coding. More precisely, a hybrid layer, that uses the source and the interference, is concatenated (superimposed) with a purely digital layer to achieve bandwidth expansion (reduction). The achievable (square error) distortion regions of these schemes under matched and mismatched noise levels are

then analyzed. The proposed schemes are shown to perform close to the best derived bound and to be resilient to channel noise mismatch. As an extension to this communication problem, a joint source-state-interference transmission is studied over fading channels in Sec. 3.5; the receiver, in this case, aims to jointly estimate both the source signal as well as the channel-state (interference). Sec. 3.5.1 derives the outer bound on the source-channel-state distortion region. The bound on the source reconstruction distortion is found by assuming full knowledge of the interference at the decoder side; no additional knowledge, however, is assumed to get a bound on the distortion from reconstructing the interference. An inner (achievable distortion region) bound on the source-interference distortion region is derived for the fading channel in Sec. 3.5.2 by proposing HDA scheme based on VQ and Costa coding. Comparison of inner and outer bounds shows that the proposed HDA scheme performs close to the outer bound. Moreover, our setting contains several interesting limiting cases. In the absence of fading and/or correlation and for some source-channel bandwidths, our setting resorts to the scenarios considered in [35, 63, 67].

In Chapter 4 a generalization of the problem considered in Chapter 3 to the Gaussian degraded broadcast channel is studied. More precisely, the transmission of bivariate Gaussian sources over the two-user degraded Gaussian broadcast channel in the presence of interference that is correlated to the source and known non-causally to the transmitter is tackled. Each user is interested in estimating one of the sources. Sec. 4.3 derives an outer bound on the system's distortion region. The derived outer bound is obtained by assuming knowledge of one of the sources at the strong user and partial/full knowledge of the interference at both users; for the limiting case of single source transmission, a better bound for the case of source-channel bandwidth

expansion is derived with the help of auxiliary random variable. Hybrid digital-analog schemes are studied and their achievable (square-error) distortion regions under matched and expansion bandwidth regimes are analyzed in Sec. 4.4. These schemes, which use the idea of superposition coding for broadcast channels by employing different layers that are meant to different users, require proper combinations of power splitting, bandwidth splitting, rate splitting, vector-quantizer, Wyner-Ziv and Costa coding. The proposed schemes are shown to outperform tandem and linear schemes. For low correlation values between the sources, our proposed schemes perform close to the derived outer bounds. For high correlation values, however, the gap between the proposed schemes and the outer bounds increases. Interestingly, our schemes are always able to benefit from the interference whenever possible.

In Chapter 5, low delay analog source-channel coding for different noisy channels in the presence of interference is examined. In Sec. 5.2, low delay analog coding for Rayleigh fading channels is presented. In Sec. 5.3, the communication scenario of Sec. 3.2 is studied under low delay coding and complexity constraints. Sec. 5.4 investigates the use of nonlinear zero delay analog mappings for source and state interference estimation over AWGN channels. In this chapter, parametric mappings based mainly on spiral and sawtooth curves are studied. Nonparametric mappings are also designed. This is done by first deriving the necessary conditions for optimality and then proposing an iterative algorithm based on joint optimization between the encoder and the decoder. A reduced-complexity approach for the implementation of the design algorithm is also presented. The proposed (nonlinear) mappings are shown to outperform linear scheme and give a graceful performance over wide range of noise levels. Moreover, these nonlinear mappings, which are shown to fit well the channel

space, overcome the inevitable saturation effect of linear scheme.

6.2 Suggestions for Further Research

Presented here is a list of issues which merit further consideration.

- Relaying has emerged to be a major factor for establishing reliable wireless communication based on cooperation. Thus, it is interesting to consider communication scenarios based on relaying systems with interference management; one scenario is to study joint source-channel coding for relay channels with interference that is correlated to the source signal and known at the transmitter (but not at the relay). An easier problem in which the interference is known to the relay is not yet fully solved and is worth also considering.
- Two-way relay channels are getting more relevant than ever; in such scenario, two nodes exchange their correlated messages with the help of a relay. They comprise two communication channels (phases): 1) a multiple access channel and 2) a broadcast channel. Now assuming that during the second phase, the received signals are disturbed by some interference that is known to the relay. One can use the result of Chapter 3 and 4 to study the effect of using hybrid coding on this type of channel; information theoretic bounds can be also obtained for such communication system.
- Further investigation of Chapter 3 can be done to accommodate different constraint settings such as secrecy and privacy.

Bibliography

- [1] K. M. Abadir and J. R. Magnus, *Matrix Algebra*. New York: Cambridge University Press, 2005.
- [2] E. Akyol, K. Rose, and T. Ramstad, "Optimal mappings for joint source channel coding," in *IEEE Inform. Theory Workshop*, Cairo, Egypt, 2010.
- [3] E. Akyol, K. Rose, K. Viswanatha, and T. Ramstad, "On zero-delay source-channel coding," *IEEE Trans. Inform. Theory*, vol. 60, no. 12, pp. 7473–7489, Dec 2014.
- [4] B. Bandemer, C. Tian, and S. Shamai, "Gaussian state amplification with noisy state observations," in *Proc. IEEE Int. Symp. on Inform. Theory*, Istanbul, Turkey, July 2013.
- [5] H. Behroozi, F. Alajaji, and T. Linder, "On the performance of hybrid digital-analog coding for broadcasting correlated Gaussian sources," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3335–3342, Dec 2011.
- [6] T. Berger and D. W. Tufts, "Optimum pulse amplitude modulation part I: transmitter-receiver design and bounds from information theory," *IEEE Trans. Inform. Theory*, vol. IT-13, no. 2, pp. 196–208, Apr 1967.

-
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U. K.: Cambridge Univ. Press, 2003.
- [8] G. Brante, R. Souza, and J. Garcia-Frias, “Analog joint source-channel coding in Rayleigh fading channels,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Prague, Czech Republic, May 2011.
- [9] X. Chen and E. Tuncel, “Zero-delay joint source-channel coding using hybrid digital-analog schemes in the Wyner-Ziv setting,” *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 726–735, Feb 2014.
- [10] Y.-K. Chia, R. Soundararajan, and T. Weissman, “Estimation with a helper who knows the interference,” in *Proc. IEEE Int. Symp. on Inform. Theory*, Cambridge, MA, July 2012.
- [11] S.-Y. Chung, *On the Construction of some Capacity-Approaching Coding Schemes*. Ph.D. dissertation, Massachusetts Institute of Technology, 2000.
- [12] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.
- [14] H. Coward and T. Ramstad, “Hybrid digital-analog transmission of analog source signals,” in *Norwegian Signal Process. Symp.*, Norway, Sep 1999.
- [15] —, “Quantizer optimization in hybrid digital-analog transmission of analog source signals,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, 2000, pp. 2637–2640.

- [16] J. Du, E. G. Larsson, M. Xiao, and M. Skoglund, "Optimal symbol-by-symbol Costa precoding for a relay-aided downlink channel," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2274–2284, Aug 2011.
- [17] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2012.
- [18] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov 2005.
- [19] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inform. Theory*, vol. 37, no. 1, pp. 155–159, Jan 1991.
- [20] P. A. Floor, *On the Theory of Shannon-Kotel'nikov Mappings in Joint Source-Channel Coding*. Ph.D. dissertation, Norwegian University of Science and Technology, 2008.
- [21] P. A. Floor, A. N. Kim, N. Wernersson, T. Ramstad, M. Skoglund, and I. Balasingham, "Zero-delay joint source-channel coding for a bivariate Gaussian on a Gaussian MAC," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3091–3102, Oct 2012.
- [22] P. A. Floor, T. A. Ramstad, and N. Wernersson, "Power constrained channel optimized vector quantizers used for bandwidth expansion," in *Proc. IEEE Int. Symp. Wireless Commun. System*, Trondheim, Norway, Oct 2007.

-
- [23] A. Fuldseth, *Robust Subband Video Compression for Noisy Channels with Multi-level Signaling*. Ph.D. dissertation, Norwegian University of Science and Technology, 1997.
- [24] A. Fuldseth and T. A. Ramstad, "Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Munich, Germany, Apr 1997.
- [25] S. Gadkari and K. Rose, "Noisy channel relaxation for VQ design," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Atlanta, GA, May 1996.
- [26] Y. Gao and E. Tuncel, "Separate source-channel coding for transmitting correlated Gaussian sources over degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. 59, no. 6, pp. 3619–3634, June 2013.
- [27] —, "Wyner-Ziv coding over broadcast channels: Hybrid digital/analog schemes," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 5660–5672, Sept 2011.
- [28] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code: lossy source-channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [29] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Inform. Contr.*, vol. 9, no. 1, pp. 19–31, 1980.

- [30] H. Everett III, “Generalized Lagrange multiplier method for solving problems of optimum allocation of resources,” *Operations Research*, vol. 11, no. 3, pp. 399–417, 1963.
- [31] F. Hekland, *On the Design and Analysis of Shannon-Kotel’nikov Mappings for Joint Source-Channel Coding*. Ph.D dissertation, Norwegian University of Science and Technology, 2007.
- [32] F. Hekland, P. A. Floor, and T. A. Ramstad, “Shannon-Kotel’nikov mappings in joint source-channel coding,” *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 94–105, Jan 2009.
- [33] F. Hekland, G. E. Oien, and T. A. Ramstad, “Using 2:1 Shannon mapping for joint source-channel coding,” in *Proc. Data Compression Conf.*, Snowbird, Utah, Mar 2005.
- [34] Y. Hu, J. Garcia-Frias, and M. Lamarca, “Analog joint source channel coding using non-linear mappings and MMSE decoding,” *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3016–3026, Nov 2011.
- [35] Y.-C. Huang and K. R. Narayanan, “Joint source-channel coding with correlated interference,” *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1315–1327, May 2012.
- [36] H. Imai and S. Hirakawa, “A new multilevel coding method using error-correcting codes,” *IEEE Trans. Inform. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
- [37] J. (Karlsson) Kron, *Low-Delay Sensing and Transmission*. Ph.D. dissertation, Royal Institute of Technology, 2011.

- [38] J. (Karlsson) Kron and M. Skoglund, "Optimized low-delay source-channel-relay mappings," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1397–1404, May 2010.
- [39] M. Kleiner and B. Rimoldi, "Asymptotically optimal joint source-channel coding with minimal delay," in *Proc. Global Telecommun. Conference*, Honolulu, HI, Nov 2009, pp. 332–341.
- [40] Y. Kochman and R. Zamir, "Joint Wyner-Ziv/dirty-paper coding by modulo-lattice modulation," *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 4878–4889, Nov 2009.
- [41] E. Koken and E. Tuncel, "Gaussian HDA coding with bandwidth expansion and side information at the decoder," in *Proc. IEEE Int. Symp. on Inform. Theory*, Istanbul, Turkey, July 2013.
- [42] V. A. Kotel'nikov, *The Theory of Optimum Noise Immunity*. New York: McGraw-Hill, 1959.
- [43] J. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electronics and Engineering in Japan*, vol. 67-B, pp. 39–47, Jan 1984.
- [44] A. Kurtenbach and P. Wintz, "Quantizing for noisy channels," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 291–302, Apr 1969.
- [45] A. Lapidoth and S. Tinguely, "Broadcasting correlated Gaussians," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3057–3068, July 2010.
- [46] —, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 2714–2752, Jun 2010.

- [47] K. H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, 1976.
- [48] S. H. Lim, P. Minero, and Y.-H. Kim, "Lossy communication of correlated sources over multiple access channels," in *Proc. of the 48th Annual Allerton Conf. on Commun., Control and Computing*, Monticello, IL, Sep 2010.
- [49] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, Jan 1980.
- [50] U. Mittal and N. Phamdo, "Hybrid digital analog (HDA) joint source channel codes for broadcasting and robust communications," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.
- [51] J. W. Modestino and D. G. Daut, "Combined source-channel coding of images," *IEEE Trans. Commun.*, vol. 27, no. 11, pp. 1644–1659, Nov 1979.
- [52] J. Nayak, E. Tuncel, and D. Gunduz, "Wyner-Ziv coding over broadcast channels: digital schemes," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1782–1799, Apr 2010.
- [53] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec 1980.
- [54] V. M. Prabhakaran, R. Puri, and K. Ramachandran, "Hybrid analog-digital strategies for source-channel broadcast," in *Proc. 43rd Allerton Conf. Commun., Control and Computing*, Allerton, IL, Sep 2005.
- [55] T. A. Ramstad, "Shannon mappings for robust communication," *Teletronikk*, vol. 98, no. 1, pp. 114–128, 2002.

- [56] Z. Reznic, M. Feder, and R. Zamir, “Distortion bounds for broadcasting with bandwidth expansion,” *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3778–3788, Aug 2006.
- [57] D. Sakrison, *Communication Theory: Transmission of Waveforms and Digital Information*. John Wiley & Sons Inc, 1968.
- [58] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [59] —, “Communication in the presence of noise,” *Proc. IRE*, vol. 86, no. 2, pp. 447–457, 1949.
- [60] M. Skoglund, N. Phamdo, and F. Alajaji, “Design and performance of VQ-based hybrid digital-analog joint source-channel codes,” *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 708–720, Mar 2002.
- [61] —, “Hybrid digital-analog source-channel coding for bandwidth compression/expansion,” *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3757–3763, Aug 2006.
- [62] R. Soundararajan and S. Vishwanath, “Hybrid coding for Gaussian broadcast channels with Gaussian sources,” in *Proc. IEEE Int. Symp. on Inform. Theory*, Seoul, Korea, July 2009.
- [63] A. Sutivong, M. Chiang, T. Cover, and Y.-H. Kim, “Channel capacity and state estimation for state-dependent Gaussian channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1486–1495, Apr 2005.

-
- [64] C. Tian, S. N. Diggavi, and S. Shamai, “The achievable distortion region of sending a bivariate Gaussian source on the Gaussian broadcast channel,” *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6419–6427, Oct 2011.
- [65] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2005.
- [66] V. A. Vaishampayan, *Combined Source-Channel Coding for Bandwidth Waveform Channels*. Ph.D. dissertation, University of Maryland, 1989.
- [67] M. Varasteh and H. Behroozi, “Optimal HDA schemes for transmission of a Gaussian source over a Gaussian channel with bandwidth compression in the presence of an interference,” *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 2081–2085, April 2012.
- [68] Y. Wang, F. Alajaji, and T. Linder, “Hybrid digital-analog coding with bandwidth compression for Gaussian source-channel pairs,” *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 997–1012, April 2009.
- [69] —, “Hybrid digital-analog coding with bandwidth compression for Gaussian source-channel pairs,” *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 997–1012, Apr 2009.
- [70] M. P. Wilson, K. R. Narayanan, and G. Caire, “Joint source-channel coding with side information using hybrid digital analog codes,” *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 4922–4940, Oct 2010.

-
- [71] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.
- [72] J.-J. Xiao, Z.-Q. Luo, and N. Jindal, "Linear joint source-channel coding for Gaussian sources through fading channels," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, CA, Dec 2006.
- [73] S. Yao, M. Khormuji, and M. Skoglund, "Sawtooth relaying," *IEEE Commun. Letters*, vol. 12, no. 9, pp. 612–614, Sep 2008.
- [74] W. Zhang, S. Kotagiri, and J. Laneman, "Writing on dirty paper with resizing and its application to quasi-static fading broadcast channels," in *Proc. IEEE Int. Symp. on Inf. Theory*, Nice, France, June 2007.