SLEPIAN-WOLF CODED NESTED QUANTIZATION FOR WYNER-ZIV CODING: HIGH-RATE PERFORMANCE ANALYSIS, CODE DESIGN, AND APPLICATION TO COOPERATIVE NETWORKS

A Dissertation

by

ZHIXIN LIU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2007

Major Subject: Electrical Engineering

SLEPIAN-WOLF CODED NESTED QUANTIZATION FOR WYNER-ZIV CODING: HIGH-RATE PERFORMANCE ANALYSIS, CODE DESIGN, AND APPLICATION TO COOPERATIVE NETWORKS

A Dissertation

by

ZHIXIN LIU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Zixiang Xiong
Committee Members,	Costas N. Georghiades
	Aniruddha Datta
	Dmitri Loguinov
Head of Department,	Costas . Georghiades

August 2007

Major Subject: Electrical Engineering

ABSTRACT

Slepian-Wolf Coded Nested Quantization for Wyner-Ziv Coding: High-rate Performance Analysis, Code Design, and Application to Cooperative Networks.

(August 2007)

Zhixin Liu, B.S., Tsinghua University, Beijing, P.R.China;

M.S., Tsinghua University, Beijing, P.R.China

Chair of Advisory Committee: Dr. Zixiang Xiong

Wyner-Ziv coding problem exploits the correlation between two signals (one is the source and the other is the side information) and thus makes it possible to encode the source signal alone and to decode it jointly with the help of the side information at the decoder. Nested lattice quantization provides a practical scheme for Wyner-Ziv coding. We examine the high-rate performance of nested lattice quantizers and give the theoretical performance for general continuous sources. Based on our analysis, a new practical Wyner-Ziv coding scheme called Slepian-Wolf coded nested lattice quantization (SWC-NQ) is proposed. Theoretical analysis shows that for the quadratic Gaussian case and at high rate, SWC-NQ performs the same as conventional entropy-coded lattice quantization with the side information available at both the encoder and the decoder. Practical designs of one- and two-dimensional nested lattice quantizers, together with multi-level LDPC codes for Slepian-Wolf coding, give performance close to the theoretical limits of SWC-NQ for the quadratic Gaussian case.

Furthermore, we apply our code design in cooperative networks as practical implementations of the cooperative strategies. In cooperative networks, relaying is an essential component to gain the cooperative diversity. Following the latest development in practical distributed source-channel coding, we studied Compress-forward

(CF) coding with BPSK modulation for the relay channel, where Wyner-Ziv coding is applied at the relay to exploit the joint statistics between signals at the relay and the destination. Practical issues such as quantizer design and selection of channel code parameters are discussed in detail. Simulation results show that, by using LDPC codes for error protection at the source and nested scalar quantization and IRA codes for Wyner-Ziv coding (or more precisely distributed joint source-channel coding) at the relay, our practical implementation comes within only 1.6-3 dB from the theoretical limit of CF for the Gaussian relay channel with BPSK modulation.

Summarily, my research work involves the development (theoretical analysis and practical design) of Wyner-Ziv coding, and its application in cooperative networks for cooperative diversity.

To My Parents

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Professor Zixiang Xiong, for his steadfast support, constant encouragement and expert guidance in my research. He has provided me an environment conducive to learning and quality research. This dissertation would never have been possible without his insight and support. I would like to especially thank Dr. Vladmir Stanković who has offered great help for my research. I would also like to thank Professor Costas Georghiades, Professor Aniruddha Datta, and Professor Dmitri Loguinov for serving as my committee members.

Furthermore, I owe my appreciation to the colleagues in the multimedia lab. It is a pleasure to have worked with all of you. In particular, I would like to thank Zhongmin Liu, Samuel Cheng, Shengjie Zhao, Tianli Chu, Yong Sun, Yang Yang, Qian Xu, Momin Uppal for their sincere help. In addition, I want to thank Dr. Tie Liu for his kind help in my research work. I learned a lot from the discussions with you.

Finally, I would like to express my profound gratitude to my parent, my loving wife, Ying Dai, and my sisters for their love and support.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

A. Slepian-Wolf Coded Nested Quantization for Wyner-Ziv Coding

Wyner-Ziv coding [1], or lossy source coding with side information at the decoder, is one of the main problems considered in network information theory [2, Chapter 14]. It generalizes lossless source coding with side information at the decoder – a special case of Slepian-Wolf coding [3][1]. The rate-distortion (R-D) function of Wyner-Ziv coding is known for both discrete and continuous alphabet cases of the source and the side information with a general distortion metric in [1, 4]. It is derived by using a technique called "binning" that divides the set of *jointly typical sequences* [2] into bins which are as *far apart* (in terms of the correlation statistics) as possible. The binning scheme for Wyner-Ziv coding can be applied to other related problems (e.g., Gelfand-Pinsker coding [5] and its special case of dirty-paper coding [6]) based on the duality between source coding and channel coding with side information [7, 8]. The theoretical analysis in [1, 4] are based on *random binning* which, due to its lack of structure, does not indicate how practical code design should be done.

In their information-theoretical work, Zamir *et al.* [9] outlined a structured *algebraic binning* scheme based on a pair of *nested* linear/lattice codes for Wyner-Ziv coding of binary symmetric/quadratic Gaussian sources, where the fine code in the nested pair plays the role of source coding while the coarse code does channel coding. The quadratic Gaussian case corresponds to when the correlation between

---

The journal model is *IEEE Transactions on Automatic Control*.

[1]Throughout this thesis, Slepian-Wolf coding means near-lossless source coding with side information at the decoder.

the source X and the side information Y can be modeled by an AWGN channel as $X = Y + Z$, $Z \sim N(0, \sigma_Z^2)$, with MSE distortion and arbitrarily distributed Y. Note that Wyner-Ziv coding in general suffers a rate loss when compared to coding with side information available both at the encoder and the decoder. The quadratic Gaussian case is special because there is no rate loss with Wyner-Ziv coding in this case[2]. Furthermore, it is shown in [9] that the Wyner-Ziv R-D function in this special case is asymptotically achievable using nested lattices, under the assumption that the lattices are ideally sphere-packed as the lattice dimensions go to infinity. However, high-dimensional lattice codes are difficult to implement in practice. Thus structured binning via nested lattice codes only facilitates high-dimensional asymptotic analysis [9].

In this thesis, we analyze the performance of *finite-dimensional* nested lattice quantizers for continuous sources under the high-rate assumption. Here the high-rate assumption is consistent with the one in classic quantization theory [10], meaning that the source is uniformly distributed inside the fine lattice cell of the quantizer. The distortion-rate (D-R) performance is analyzed for both the general and the quadratic Gaussian cases. For general continuous sources, the distortion under a specific rate consists of two parts: one from source coding and another from channel coding. For the quadratic Gaussian case, a tight lower bound of the D-R function is given, showing an increasing gap from the Wyner-Ziv limit, as the rate increases. Based on our analysis, we argue that this increasing gap is due to the decreasing boundary gain as the rate increases. Thus a practical approach to boosting the overall performance is to increase the boundary gain with a second stage of binning, which partitions the

[2]It was only shown in [1] that Wyner-Ziv coding of X suffers no rate loss when X and Y are zero mean and jointly Gaussian with MSE distortion. Pradhan *et al.* [8] recently extended this no rate loss result to the more general quadratic Gaussian case.

support region of the fine lattice (the Voronoi region of the coarse lattice for nested lattice quantizer) into k cosets. This way the volume of the support region decreases by a factor of k while the decoding error probability stays the same. According to the definition in [11], the boundary gain increases without changing the dimensionality of the lattices. Since various possible boundary gains are realizable using the second-stage of binning, there is only maximally 1.53dB granular gain left unexploited by the quantizer. Thus using Slepian-Wolf coding for second-stage binning allows us to show the theoretical performance limits at high rate.

Following this logic, we introduce a new framework for Wyner-Ziv coding of continuous i.i.d. sources based on Slepian-Wolf coded nested quantization (SWC-NQ). Slepian-Wolf coding [3] here refers to lossless source coding with side information at the decoder. Practical syndrome-based schemes for Slepian-Wolf coding using channel codes have been studied in [12, 13, 14, 15, 16]. The role of Slepian-Wolf coding in SWC-NQ is to exploit the correlation between the quantized source and the side information for further compression and by making the overall channel code stronger. SWC-NQ generalizes the classic source coding approach of entropy-coded quantization in the sense that the quantizer performs quite well alone and can exhibit further rate savings by employing a powerful Slepian-Wolf code. Moreover, it connects network information theory with the rich areas of practical lattice source code (e.g., [17]) and channel code (e.g., LDPC codes [18, 19]) designs, making it feasible to devise codes that can approach the Wyner-Ziv D-R function.

For the quadratic Gaussian case, we establish the high-rate performance of SWC-NQ with ideal Slepian-Wolf coding, assuming there is no channel decoding error in the latter. We show that SWC-NQ with finite dimensional nested lattice quantizer at high rate achieves the same performance of classic entropy-coded lattice quantization as if the side information is also available at the encoder. For example, with ideal

Slepian-Wolf coding, one-dimensional/two-dimensional SWC-NQ performs 1.53/1.36 dB away from the Wyner-Ziv D-R function for quadratic Gaussian sources at high rate.

We also implement one- and two-dimensional nested lattice quantizers in the rate range of 1.0-7.0 bits per sample (b/s), for the case when Y is also Gaussian (hence X and Y are jointly Gaussian), which is a special case of the quadratic Gaussian scenario. Our experiments using nested lattice quantizers together with irregular LDPC codes for Slepian-Wolf coding give performance close to the corresponding limit at high rate. Our work thus shows that SWC-NQ provides an efficient scheme for practical Wyner-Ziv coding with low-dimensional lattice quantizers at high rate.

Although our theoretical analysis assumes high rate, when a *non-linear* minimum MSE estimator is applied at the decoder, our simulated D-R performance of SWC-NQ at low rate is the same as that of classic entropy-coded quantization at low rate when the side information is also available at the encoder. At high rate, the non-linear estimator degenerates to the linear one used in our high-rate performance analysis.

We note that non-linear estimation at the decoder can yield significant gains only for low rate and for high rate it cannot help noticeably. This is confirmed by the agreement of the high rate analysis results, which assume that the linear estimation is used, with the high rate simulation results, for which the non-linear estimation method is always used.

1. Related Works

As mentioned earlier, Zamir *et al.* [9] studied the high-dimension asymptotic of nested lattice quantization for Wyner-Ziv coding. Practical approaches to Wyner-Ziv coding have recently been investigated in [16, 20, 21, 22, 23, 24, 25, 26, 27]. For example, in DISCUS [16], two source codes (scalar quantization and TCQ) and two

channel codes (scalar coset code and trellis-based coset code [28]) are used in source-channel coding for the Wyner-Ziv problem, resulting in four combinations. One of them (scalar quantization with scalar coset code) is nested scalar quantization and another one (TCQ with trellis-based coset code, also suggested in [20]) can effectively be considered as nested TCQ.

A recent work [29] starts with non-uniform quantization with index reuse and Slepian-Wolf coding and shows the same high-rate theoretical performance as ours when the quantizer becomes an almost uniform one without index reuse. This agrees with our finding in section V that at high rate, the nested quantizer asymptotically becomes a non-nested regular one so that strong channel coding is guaranteed.

Servetto [21] explored explicit nested lattice constructions based on similar sublattices [30]. But we point out that results presented here contradict those in [21, 31] in three aspects: 1) Whereas our analysis in section IV shows that finite-dimensional nested lattice quantization performs increasingly worse than the Wyner-Ziv limit as the rate increases[3], Fig. 3 of [21] seems to indicate that the performance of a low-dimensional nested lattice quantizer is a constant gap (in dB) away from the Wyner-Ziv limit in the 2.0-7.0 b/s rate range. 2) Our simulation results with two-dimensional nested lattice quantization shown in figure on page 63 are much worse than those in Fig. 4 of [21]. 3) In [31], the author attempts to apply nested lattice quantization to dense sensor networks with limited resource (e.g., fixed rate), where high correlation

[3]The intuitive explanation for this increasing performance gap from the Wyner-Ziv limit at high rate is given by a reviewer of [32, 21] as follows: Under the quadratic Gaussian correlation model $X = Y + Z$, with $N \sim N(0, \sigma_Z^2)$, the noise Z should in high probability be contained in the coarse lattice cell in order for the nested lattice coding scheme to function well. Otherwise the probability of decoding error will be too high, and will dominate the total distortion, in particular at high rate. Since a cubic cell does not match the spherical shape of a multi-dimensional Gaussian, we must take a large margin in the scaling of the coarse lattice to make the decoding error probability small enough. Moreover, the smaller the total distortion needs to be, the higher this margin must be.

among sensor outputs is achieved by increasing the number of sensors. In contrast, we show in section IV that, at fixed rate and dimensionality, the performance of a nested lattice quantizer is independent of the source correlation at high rate − hence there is no high-correlation asymptotic in Wyner-Ziv coding at high rate.

B. Compress-forward for Cooperative Networks Using Practical Wyner-Ziv Coding

Wyner-Ziv coding can be applied in a large number of real applications. In our work, we focus on the application of Wyner-Ziv coding in cooperative networks to approach the cooperative diversity, where the relay channel is an essential part. The relay channel, introduced by van der Meulen in [33], consists of three terminals: the source, the relay, and the destination. The source broadcasts message to both the relay and the destination. The relay processes the message it receives from the source and forwards the processed signal to the destination, which reconstructs the original message by decoding the signals received from both the source and the relay. The task of the relay is thus to facilitate joint decoding at the destination by means of spatial/temporal diversity.

The capacity of the general relay channel is still not known. Cover and El Gamal [34] derived the tightest upper and lower bounds of the general relay channel using random coding and converse arguments. These two bounds coincide only in few special cases (e.g., the degraded Gaussian relay channel). Since the optimum processing at the relay is unknown, several random coding schemes [35, 36, 34, 37, 38, 39, 40] have been proposed to obtain the lower bound on the achievable rate region.

In general, there are two classes of coding schemes for the relay channel: *decode-forward* (DF) and *observe-forward* [34], although hybrid schemes are also possible

[36, 34]. The main operation of DF is full decoding at the relay node. Upon receiving a noisy signal from the source, the relay node decodes it, and then re-encodes the reconstructed message before forwarding the resulting codeword to the destination. It should by emphasized that the relay might use a different codebook than the source. In any case, the source can completely predict what the relay will transmit, and full coherency is therefore possible. The destination attempts to reconstruct the message by combining the signals received from the source and the relay, using either successive list decoding [34], backward decoding [41, 42], or decoding based on parallel Gaussian channel arguments [43], which all result in the same achievable rate region. Although DF is efficient in some scenarios [44, 38], the achievable rates are bounded by the capacity of the channel between the source and relay since the relay must perfectly decode the source message. To alleviate this problem, a class of observe-forward schemes have been proposed, where the relay does not attempt to decode the signal from the source, but merely forwards a processed version of its received signal to the destination.

The simplest observe-forward scheme is *amplify-forward* (AF) [45, 46], in which the relay, sticking to its rudimentary role, just amplifies the received signal before forwarding. Although AF has low coding complexity, it has never been shown that it can outperform DF [37]. The CF scheme (also referred to as *estimate-forward* in [37]), which is rooted in the original work of Cover and El Gamal [34], is more sophisticated. As the name suggests, in CF, the relay compresses the signal it has received from the source within certain distortion. This can be achieved, for example, with a simple quantizer (the scheme referred to as *quantize-forward* in [37]). However, quantize-forward only quantizes the signal without exploiting the correlation between signals received at the relay and the destination, hence its compression efficiency is limited. A more powerful technique is compress-forward applying *source coding with*

side information (a.k.a., Wyner-Ziv coding).

When WZC is used in CF, the signal received at the destination from the source acts as the side information, and the signal received at the relay from the source is the signal to be compressed. Since these two received signals are different noisy versions of the same original source, they are correlated. The relay can exploit this correlation to compress the received signal without knowing the side information; thus, it does not attempt to decode. At the destination, the signal is recovered with the presence of side information.

When compressing its signal, the relay introduces distortion which determines the overall rate over the relay-destination channel. Intuitively, the signal at the relay should be compressed at the highest possible distortion level at which the destination is still able to accurately estimate it using the available side information. If the communication channels are cleaner, the signal to be Wyner-Ziv coded at the relay and the side information at the destination are more correlated enabling higher WZC efficiency and hence higher transmission rates.

DF and CF give the best known lower bound of the achievable rate region of the relay channel. Either CF or DF could be superior depending on the transmission parameters. In DF, the relay decodes and re-encodes its received signal, leading to coding gain. However, the relay forwards the decoded message in a hard-decision form. In CF, the relay uses soft information which resembles decoding error-correcting codes with soft decision rather than hard decision. Intuitively, DF performs well when the channel between the source and relay is clean (e.g., the relay is close to the source), whereas CF is desirable when the channel between the relay and destination is good. More importantly, the relay must perfectly decode the source message in DF, thus the achievable rates are bounded by the capacity of the channel between the source and the relay. On the other hand, CF always outperforms direct transmission. Thus,

even if the link between the source and the relay is poor, the relay can still help. Therefore, CF gives many rate points that are not achievable with any other coding strategy.

There are two operating modes in relaying: full-duplex and half-duplex. If the relay is able to transmit and receive signals simultaneously on the same frequency, then we say that it works in the full-duplex mode; otherwise, it works in the half-duplex mode. In this latter mode, the relay either works in a time-division (in which the relay receives and transmits signals in different time slots) or frequency-division manner (in which the relay receives and transmits at different frequencies). Half-duplex is a simpler and cheaper approach [47] because the microwave design challenge (e.g., due to the large difference in the transmitting and receiving signal power levels) associated with the full-duplex mode can be avoided. Therefore, we will focus on half-duplex relaying in our work. Since time-division and frequency-division are equivalent [2] from an information-theoretical viewpoint, we will assume time-division multiplexing in the sequel.

Following the WZC-based CF strategy of [44], we proposed the first practical limit-approaching CF design for the half-duplex Gaussian relay channel with BPSK modulation. In the relay channel, the signal to be compressed by the WZC and the side information are not jointly Gaussian as assumed in the literatures analyzing the theoretical achievable rates for CF, such as [44]. Instead, the source and the side information are Gaussian mixture generated from the BPSK modulation. Although the theoretical achievable rate of WZC for this model is unknown yet, a lower bound and an achievable upper bound are proposed. It is shown that, for the case where the relay is close to the destination thereafter CF outperforms DF, the two bounds are close enough, i.e. tight enough, so that the actual performance of ideal WZC can be well approximated. Our code design relies on practical WZC based on Slepian-Wolf

coded nested quantization (SWC-NQ) [48]. Practical SWC [49] is implemented by channel codes. Since the Slepian-Wolf compressed bitstream is to be transmitted over a noisy channel from the relay to the destination, channel coding is needed to protect them. This calls for distributed joint source-channel coding (DJSCC), i.e., joint Slepian-Wolf compression and channel protection. In our practical implementation, we use irregular repeat-accumulate (IRA) codes [50] by designing one multi-layer code to take care of two channels: one is the physical noisy channel between the relay and the destination; another is the "virtual" correlation channel [51], which characterizes the correlation between the quantized source at the relay and the decoder side information at the destination.

1. Related Works

Motivated by the wide application and information-theoretical importance of the relay channel, several research groups have recently proposed practical code designs. However, most of them employ DF. For example, practical user cooperative scheme is given in [45] using space-time coding. Rate-compatible punctured convolutional codes are employed in [52]. A DF scheme based on incremental redundancy designed in [53] exploits optimized convolutional codes in a Rayleigh slow fading environment; it was further shown that this scheme achieves full diversity. A practical turbo-based code design for DF was proposed in [43], in which two different recursive systematic convolutional codes are applied at the source and relay to provide spatial diversity; the coded bits are transmitted at different time slots, resulting in a distributed turbo code. A similar code design was also given in [54], and a more advanced turbo-based code design was proposed in [55] for both SISO and MIMO relay system, showing a gap typically 1.0-1.5 dB away from the theoretical limits.

2. Our CF Code Design

The main challenges in practical CF coding lie in quantizer design and DJSCC. The former involves optimization of the nesting ratio (in the case of NSQ) and the scalar quantization stepsize to minimize the end-to-end distortion. The latter lies in determining the optimal IRA code rates for all bit planes of the quantization indices and the soft input to the iterative joint decoder. Simulation results show that, by using low-density parity-check (LDPC) codes for error protection at the source and NSQ and IRA codes for DJSCC at the relay, our practical CF code design operates only 0.76-1.54 dB away from the theoretical limit of CF [44] for the Gaussian relay channel.

Since our publication of [56], two practical CF designs [57, 58] have appeared in the literature. The first [57] is a quantize-forward scheme that does not exploit WZC at the relay; instead, it uses simple scalar quantization, followed by LDPC coding for error protection. The second [58] is based on WZC at the relay and uses scalar quantization and convolution codes; however, it does not follow the best CF coding strategy of [44], hence it is not clear how its performance compares to the best achievable rate of CF given in [44].

C. Organization of the Dissertation

Although the theoretical limits of coding problems with side information are well-known, their implementations are not. Therefore, we focus on their practical Wyner-Ziv code designs and its application in the relay channels in this thesis, which is organized as follows. In Chapter II we review the theoretical background on distributed source coding including Wyner-Ziv coding (WZC) and Slepian-Wolf coding (SWC). In Chapter III we introduce the practical Wyner-Ziv code design, a.k.a. Slepian-Wolf

coded nested lattice quantization (SWC-NQ). The high rate performance of SWC-NQ for the quadratic Gaussian case assuming ideal SWC is analyzed, and practical code design issues are discussed. In Chapter IV we proposed a practical CF scheme for the relay channel, applying SWC-NQ with BPSK modulation and distributed joint source channel coding (DJSCC). The coding performance and the simulation results are provided in this chapter, showing that our scheme is efficient to approach the theoretical achievable bound.

D. Main Contributions on the Dissertation

1. High-rate performance analysis of nested lattice quantization for all dimensions, indicating for any finite dimension an increasing gap in distortion from the Wyner-Ziv D-R function as the rate increases.

2. Introduction of the SWC-NQ framework for Wyner-Ziv coding. The D-R performance of SWC-NQ for the quadratic Gaussian case is presented, showing agreement with the performance of entropy-coded lattice quantization in classic source coding.

3. Our proof that the performance of Wyner-Ziv coding of quadratic Gaussian sources with nested lattice quantization at a fixed high rate is independent of the source correlation, with or without Slepian-Wolf coding.

4. A non-linear minimum MSE estimator at the decoder of the nested lattice quantizer, which improves the quantizer performance at low rate.

5. Practical designs of one-dimensional scalar and two-dimensional hexagonal nested lattice quantizers and multi-level irregular LDPC codes for Slepian-Wolf coding,

confirming our high-rate performance analysis for both nested lattice quantization and SWC-NQ.

6. Discussion on the lower and achievable upper bound of SWC-NQ for the BPSK-modulated case, where the source to be coded and the side information are two noisy versions of the identical BPSK-modulated binary input to a additive noisy channel. This model is of special interest in the application of SWC-NQ to the communication systems, where the information is modulated before transmission.

7. Application of SWC-NQ for the BPSK-modulated case to the Compress-forward (CF) code design for the Gaussian half-duplex relay channel. Simulation results are presented in this thesis. As far as we know, our code design is the first practical CF code design utilizing distributed source coding.

CHAPTER II

THEORETICAL BACKGROUND ON SOURCE CODING WITH SIDE INFORMATION

A. Source Coding with Side Information

In the traditional point-to-point communication systems, information is processed by source coding to remove redundancy at the transmitter, and then transmitted to the receiver after being channel coded whose goal is to introduce useful redundancy with which the receiver can detect and potentially correct transmission errors caused by the channel noise. However, in many scenarios such as sensor networks and multiple-input-multiple-output (MIMO) communication systems, besides their regular inputs, the transmitter and/or the receiver are given some extra information regarding the source and the channel. For example, this "side information" can be the nature and the format of the source and the mean and the variance of the channel noise. To incorporate this side information in a communication system, it is thus necessary to study source coding and channel coding with side information.

Since side information can be given to the encoder and/or decoder, this results in four different cases. However, several of these cases are trivial in the sense that conventional source and channel coding techniques can be employed directly. For example, when side information is given to both the encoder and decoder, we can easily include this side information in the scheme design by using optimized coders for the different outcomes of the side information. Yet another example, consider source coding when side information is given to the encoder alone; it is shown in [59] that the side information is useless and thus can be ignored. The two most interesting cases are source coding with side information at the decoder, a.k.a. Wyner-Ziv coding

(WZC) [1], and channel coding with side information at the encoder, a.k.a. Gel'fand-Pinsker coding (GPC) [5]. In this thesis, we focus on the WZC problem (i.e., source coding with side information) and study the code design issues and applications to communication systems.

1. Wyner-Ziv Coding

Consider numerous heat sensors spreading over a region, measuring temperature, and sending it back to a base station. In order to save the production cost of these sensors and simplify the scheme design, we assume these sensors transmit measurements directly to the base station without the help of other sensors as relay. Hence, the transmitter in each sensor can only know its local measurement. However, in most cases, the measurements of all these sensors are correlated; so the question is: can we incorporate this correlation effectively to compress these measurements even though joint encoding is not permitted?

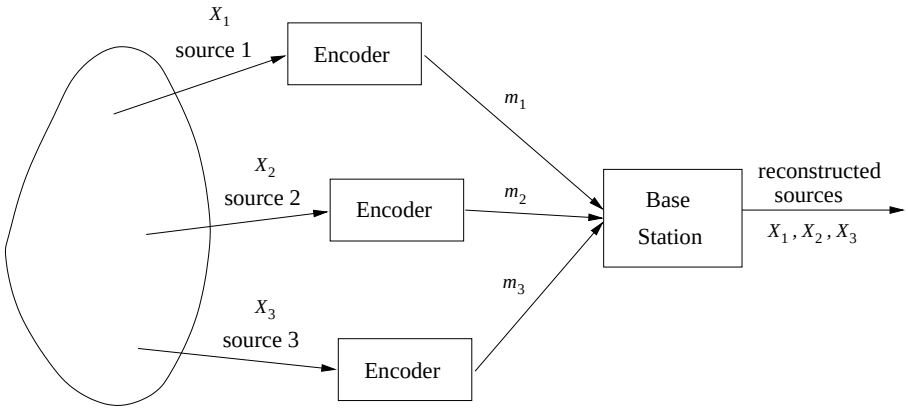


Fig. 1. Distributed source coding with three sources.

The above scenario is a typical example of distributed source coding in which

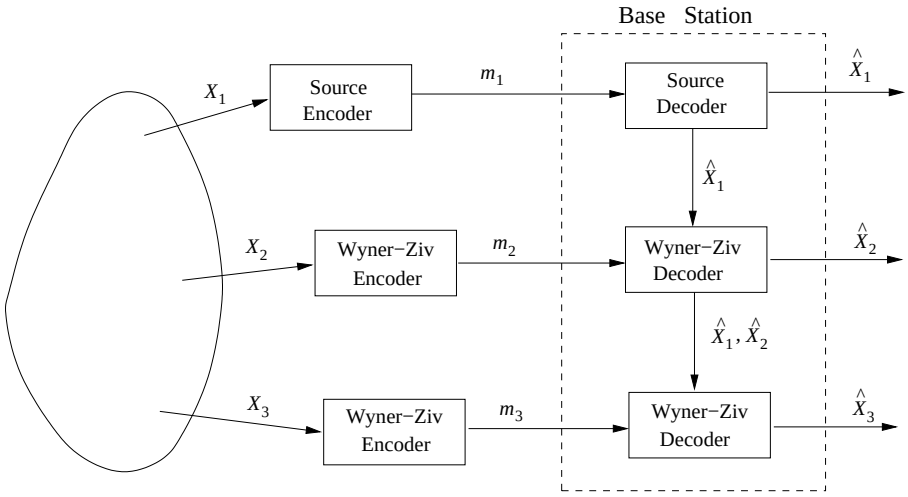


Fig. 2. Distributed source coding implemented by WZC.

several correlated sources are encoded separately but decoded jointly as shown in Fig. 1. A solution of this interesting problem can be implemented using WZC. As shown in Fig. 2, the first source X_1 will be coded using conventional source coding. At the base station, X_1 will be the first to be decoded and used as side information for the subsequent decoding of all other sources. Knowing the reconstructed $\hat{X}_1$ at the base station, the second source X_2 is coded using WZC. And just as $\hat{X}_1$, the reconstructed $\hat{X}_2$ is also treated as side information for the subsequent decoding stages. Similar decoding procedure with all the previous decoded sources as side information continues until all sources are reconstructed.

In general, Wyner-Ziv coding [1, 4] deals with the problem of R-D with side information at the decoder. It asks the question of how many bits are needed to encode X under the constraint that $E\{d(X, \hat{X})\} \leq D$, assuming the side information Y is available at the decoder but not at the encoder. This problem generalizes the setup of [3] in that coding of X is lossy with respect to a fidelity criterion rather

than lossless. For both discrete and continuous alphabets of $\mathcal{X}$ and general distortion metrics $d(\cdot)$, Wyner and Ziv [1] gave the R-D function $R_{WZ}(D)$ for this problem as $R_{WZ}(D) = \inf I(X; A|Y)$, where the infimum is taken over all auxiliary random variables A such that $Y \to X \to A$ is a Markov chain and there exists a function $\hat{X} = \hat{X}(A, Y)$ satisfying $E\{d(X, \hat{X})\} \leq D$. According to [1],

$$R_{WZ}(D) \geq R_{X|Y}(D) = \inf_{\{\hat{X} \in \mathcal{X}: E\{d(X,\hat{X})\} \leq D\}} I(X; \hat{X}|Y),$$

where $R_{X|Y}(D)$ is the classic R-D function of coding X with Y available at the encoder (and the decoder). This means that, compared to coding of X when the side information Y is also available at the encoder, there is in general a rate loss with Wyner-Ziv coding. Zamir quantified this loss in [60], showing a <0.22 bit loss for binary sources with Hamming distance and a <0.5 b/s loss for continuous sources with MSE distortion.

When D is very small and the source is discrete-valued, the Wyner-Ziv problem degenerates to the Slepian-Wolf problem with $R_{WZ}(D) = R_{X|Y}(D) = H(X|Y)$. Another interesting setup is the quadratic Gaussian case with the source model being $X = Y + Z$ and $Z \sim N(0, \sigma_Z^2)$, then $R_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log^+\left[\frac{\sigma_Z^2}{D}\right]$, where $\log^+ x = \max\{\log x, 0\}$, i.e., there is no rate loss in this case. Note that Y is arbitrarily distributed [8]. When Y is also Gaussian (then X and Y are jointly Gaussian memoryless sources), let the covariance matrix of (X_i, Y_i) be $\Lambda = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$ with $|\rho| < 1$ for all n, then $R_{WZ}(D) = R_{X|Y}(D) = \frac{1}{2}\log^+\left[\frac{\sigma_X^2(1-\rho^2)}{D}\right]$. This case is of special interest in practice because many image and video sources can be modeled as jointly Gaussian and Wyner-Ziv coding suffers no rate loss. Another interesting case is the WZC with BPSK modulation, where $X = aS + N_1$, $Y = bS + N_2$. Here S is the BPSK-modulated binary input to the channel, a and b are the channel coefficients

due to fading or channel gains, N_1 and N_2 are the additive channel noises which can be modeled as white Gaussian in most scenarios. The code design and performance analysis for both cases will be addressed in the following chapters.

2. Slepian-Wolf Coding

A special case of WZC is called Slepian-Wolf coding where the fidelity criterion $D = 0$. Slepian-Wolf coding is concerned with near-lossless source coding with side information at the decoder. For lossless compression of a pair of correlated, discrete random variables X and Y, a rate of $R_X + R_Y = H(X, Y)$ is sufficient if they are encoded jointly [2]. However, Slepian and Wolf [3] showed that the rate $R_X + R_Y = H(X, Y)$ is almost sufficient even for separate encoding (with joint decoding) of X and Y. Specifically, the Slepian-Wolf theorem says that the achievable region for coding X and Y is given by

$$R_X \geq H(X|Y), R_Y \geq H(Y|X), R_X + R_Y \geq H(X, Y). \tag{2.1}$$

This result shows that there is no loss of coding efficiency with separate encoding when compared to joint encoding as long as joint decoding is performed. When the side information (e.g. Y) is perfectly available at the decoder, then the aim of Slepian-Wolf coding is to compress X to the rate limit $H(X|Y)$.

CHAPTER III

SLEPIAN-WOLF CODED NESTED QUANTIZATION FOR WYNER-ZIV CODING

A. Lattices and Nested Lattices

1. Lattices

For a set of n independent basis vectors $\{\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_n\}$, an unbounded n-dimensional lattice Λ is defined by

$$\Lambda = \{\mathbf{l} = M\mathbf{i} : \mathbf{i} \in \mathbb{Z}^n\} \tag{3.1}$$

and its generator matrix $M = [\mathbf{m}_1 \mathbf{m}_2 \cdots \mathbf{m}_n]$. The nearest neighbor quantizer $Q_\Lambda(\cdot)$ associated with Λ is given by

$$Q_\Lambda(\mathbf{x}) = \arg\min_{\mathbf{l}\in\Lambda} |\mathbf{x} - \mathbf{l}|. \tag{3.2}$$

The basic Voronoi cell of Λ, which specifies the shape of the nearest-neighbor decoding region, is

$$\mathcal{V} = \{\mathbf{x} : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}. \tag{3.3}$$

Associated with the Voronoi cell $\mathcal{V}$ are several important quantities: the cell volume V, the second moment σ^2 and the normalized second moment $G(\Lambda)$, defined by

$$V = \int_{\mathcal{V}} d\mathbf{x}, \tag{3.4}$$

$$\sigma^2 = \frac{1}{nV}\int_{\mathcal{V}} |\mathbf{x}|^2 d\mathbf{x}, \tag{3.5}$$

$$G(\Lambda) = \frac{\sigma^2}{V^{\frac{2}{n}}}, \tag{3.6}$$

respectively. The minimum of $G(\Lambda)$ over all lattices in $\mathbb{R}^n$ is denoted as G_n. By [17],

$$G_n \geq \frac{1}{2\pi e}, \forall n \tag{3.7}$$

$$\lim_{n\to\infty} G_n = \frac{1}{2\pi e}. \tag{3.8}$$

2. Nested Lattices

A pair of n-dimensional lattices (Λ_1, Λ_2) with corresponding generator matrices M_1 and M_2 is *nested*, if there exists an $n \times n$ integer matrix P such that $M_2 = M_1 \times P$ and $|\det P| > 1$. In this case $\frac{V_2}{V_1}$ is called the *nesting ratio*, and Λ_1 and Λ_2 are called the *fine* and *coarse* lattice, respectively.

For a pair of nested lattices (Λ_1, Λ_2), the points in the set $\Lambda_1/\Lambda_2 \triangleq \{\Lambda_1 \bigcap \mathcal{V}_2\}$ are called the *coset leaders* of Λ_2 relative to Λ_1, where $\mathcal{V}_2$ is the basic Voronoi cell of Λ_2. For each $\mathbf{v} \in \Lambda_1/\Lambda_2$ the set of shifted lattice points $C(\mathbf{v}) \triangleq \{\mathbf{v} + \mathbf{l}, \forall \mathbf{l} \in \Lambda_2\}$ is called a *coset* of Λ_2 relative to Λ_1. The j-th point of $C(\mathbf{v})$ is denoted as $\mathbf{c}_j(\mathbf{v})$. Then

$$C(\mathbf{0}) = \{\mathbf{c}_j(\mathbf{0}), \forall j \in \mathbb{Z}\} = \Lambda_2, \tag{3.9}$$

and

$$\bigcup_{\mathbf{v}\in\Lambda_1/\Lambda_2} C(\mathbf{v}) = \Lambda_1. \tag{3.10}$$

Since

$$\mathbf{c}_j(\mathbf{v}) \in \Lambda_1, \forall j \in \mathbb{Z}, \tag{3.11}$$

we further define $R_j(\mathbf{v}) = \{\mathbf{x} : Q_{\Lambda_1}(\mathbf{x}) = \mathbf{c}_j(\mathbf{v})\}$ as the Voronoi region associated with $\mathbf{c}_j(\mathbf{v})$ in Λ_1, and $R(\mathbf{v}) = \bigcup_{j=-\infty}^{\infty} R_j(\mathbf{v})$, then

$$\bigcup_{j=-\infty}^{\infty} \bigcup_{\mathbf{v}\in\Lambda_1/\Lambda_2} R_j(\mathbf{v}) = \bigcup_{\mathbf{v}\in\Lambda_1/\Lambda_2} R(\mathbf{v}) = \mathbb{R}^n. \tag{3.12}$$

An examples of $\mathbf{v}$, $C(\mathbf{v})$ and $R(\mathbf{v})$ for $n = 2$ are shown in Fig. 3.

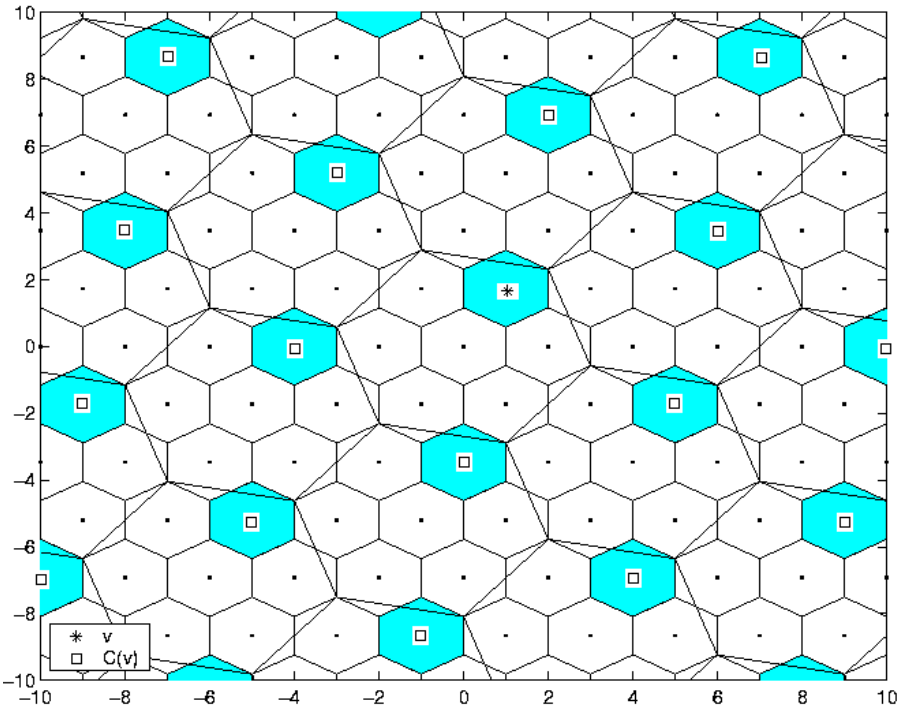


Fig. 3. An example of $\mathbf{v}$, $C(\mathbf{v})$ and $R(\mathbf{v})$ for $n = 2$, where the shaded regions correspond to $R(\mathbf{v})$.

B. Nested Lattice Quantization

Throughout this chapter, we use the correlation model of $X = Y + Z$, where X is the source to be coded, Y is the side information, and Z is the noise. Y and Z are independent. In this section we discuss the performance of nested lattice quantization for general sources where Y and Z are arbitrarily distributed continuous sources with zero mean, and for the quadratic Gaussian case with $Z \sim N(0, \sigma_Z^2)$. For both cases, MSE is used as the distortion measure.

Zamir *et al.*'s nested lattice quantization scheme [9] works as follows: Let the pseudo random vector U (the dither), known to both the quantizer encoder and the

decoder, be uniformly distributed over the basic Voronoi cell $\mathcal{V}_1$ of the fine lattice Λ_1. For a given target average distortion D, denote $\alpha = \sqrt{1 - \frac{D}{\sigma_Z^2}}$ as the estimation coefficient. Given the n-dimensional realizations of the source, the side information and the dither as $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{u}$, respectively, then according to [9], the nested quantizer encoder quantizes $\alpha\mathbf{x} + \mathbf{u}$ to the nearest point $\mathbf{x}_{Q_{\Lambda_1}} = Q_{\Lambda_1}(\alpha\mathbf{x} + \mathbf{u})$ in Λ_1, computes $\mathbf{s} = \mathbf{x}_{Q_{\Lambda_1}} - Q_{\Lambda_2}(\mathbf{x}_{Q_{\Lambda_1}})$ which is the coset shift of $\mathbf{x}_{Q_{\Lambda_1}}$ with respect to Λ_2, and transmits the index corresponding to this coset shift.

The nested quantizer decoder receives $\mathbf{s}$, forms $\mathbf{t} = \mathbf{s} - \mathbf{u} - \alpha\mathbf{y}$, and reconstructs $\mathbf{x}$ as $\hat{\mathbf{x}} = \mathbf{y} + \alpha(\mathbf{t} - Q_{\Lambda_2}(\mathbf{t}))$ using linear combination and dithering in estimation.

It is shown in [9] that the Wyner-Ziv D-R function $D_{WZ}(R) = \sigma_Z^2 2^{-2R}$ is achievable with infinite dimensional nested lattice quantization for quadratic Gaussian case.

In this chapter, we analyze the high-rate performance of finite-dimensional nested lattice quantizers. Our analysis is based on the high-resolution assumption, which means that D is small compared to σ_Z^2. Consequently, V_1 is small enough so that the pdf of X, $f(\mathbf{x})$, is approximately constant over each Voronoi cell of Λ_1. Under the high-rate assumption, $\alpha = 1$. In addition, dithering is not needed in our high-rate analysis. With $\alpha = 1$ and $\mathbf{u} = \mathbf{0}$, the encoder/decoder described above simplifies to

- The encoder quantizes $\mathbf{x}$ to $\mathbf{x}_{Q_{\Lambda_1}} = Q_{\Lambda_1}(\mathbf{x})$, computes $\mathbf{s} = \mathbf{x}_{Q_{\Lambda_1}} - Q_{\Lambda_2}(\mathbf{x}_{Q_{\Lambda_1}})$, and transmits an index corresponding to the coset leader $\mathbf{s}$.
- Upon receiving $\mathbf{s}$, the decoder forms $\mathbf{t} = \mathbf{s} - \mathbf{y}$ and reconstructs $\mathbf{x}$ as $\hat{\mathbf{x}} = \mathbf{y} + \mathbf{t} - Q_{\Lambda_2}(\mathbf{t}) = \mathbf{s} + Q_{\Lambda_2}(\mathbf{y} - \mathbf{s})$.

In the performance analysis, we limit ourselves to this simplified nested lattice quantization scheme for high rate, which is shown in Fig. 4 and was also used in [21].

1. High-rate Performance for General Sources with Arbitrary Distribution

Theorem 1: If a pair of n-dimensional nested lattices (Λ_1, Λ_2) with nesting ratio $N = \frac{V_2}{V_1}$ is used for nested lattice quantization, the distortion per dimension in Wyner-Ziv coding of X (with decoder side information Y) at high rate is the sum of the lattice source coding loss D_s and the lattice channel coding loss D_c, as follows.

$$\begin{aligned} &D_n = D_s + D_c, \\ &D_s = G(\Lambda_1)V_1^{\frac{2}{n}}, D_c = \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]. \end{aligned} \tag{3.13}$$

Proof. Since

$$\mathbb{R}^n = \bigcup_{j=-\infty}^{\infty} \bigcup_{\mathbf{v}\in\Lambda_1/\Lambda_2} R_j(\mathbf{v}), \tag{3.14}$$

the average distortion for a given realization of the side information $\mathbf{y}$ is

$$\begin{aligned} D(\mathbf{y}) &= \int_{\mathbb{R}^n} f(\mathbf{x}|\mathbf{y})\|\mathbf{x}-\hat{\mathbf{x}}\|^2 d\mathbf{x} \\ &= \sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty}\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|\mathbf{x}-\mathbf{c}_j(\mathbf{v})+\mathbf{c}_j(\mathbf{v})-\hat{\mathbf{x}}\|^2 d\mathbf{x} \\ &= \sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty}\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\big[\|\mathbf{x}-[\mathbf{c}_j(\mathbf{v})\|^2+\|\mathbf{c}_j(\mathbf{v})-\hat{\mathbf{x}}\|^2+2<\mathbf{x}-\mathbf{c}_j(\mathbf{v}),\mathbf{c}_j(\mathbf{v})-\hat{\mathbf{x}}>\big] d\mathbf{x} \\ &\overset{(a)}{=} \sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty}[f(\mathbf{c}_j(\mathbf{v})|\mathbf{y})\int_{\mathbf{x}\in R_j(\mathbf{v})}\|\mathbf{x}-\mathbf{c}_j(\mathbf{v})\|^2 d\mathbf{x}+\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|\mathbf{c}_j(\mathbf{v})-\hat{\mathbf{x}}\|^2 d\mathbf{x}] \\ &\overset{(b)}{=} \sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty}[f(\mathbf{c}_j(\mathbf{v})|\mathbf{y})nG(\Lambda_1)V_1^{1+\frac{2}{n}} \\ &+ \int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v}))-Q_{\Lambda_2}\big(\mathbf{y}-\mathbf{c}_j(\mathbf{v})+Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v}))\big)\|^2 d\mathbf{x}] \\ &\overset{(c)}{=} nG(\Lambda_1)V_1^{\frac{2}{n}}+\sum_{j=-\infty}^{\infty}\sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}(\mathbf{x}-\mathbf{y})\|^2 d\mathbf{x} \\ &= nG(\Lambda_1)V_1^{\frac{2}{n}}+\int_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}(\mathbf{x}-\mathbf{y})\|^2 d\mathbf{x}, \end{aligned} \tag{3.15}$$

where (a) comes from the high rate assumption and $\int_{\mathbf{x}\in R_j(\mathbf{v})} < \mathbf{x}-\mathbf{c}_j(\mathbf{v}), \mathbf{c}_j(\mathbf{v})-\hat{\mathbf{x}} > d\mathbf{x} = 0$, which is due to the fact that $\mathbf{x}-\mathbf{c}_j(\mathbf{v})$ is odd spherical symmetric for $\mathbf{x} \in R_j(\mathbf{v})$ and both $\mathbf{c}_j(\mathbf{v})$ and $\hat{\mathbf{x}}$ are fixed for $\mathbf{x} \in R_j(\mathbf{v})$ with given $\mathbf{v}$ and $\mathbf{y}$, (b) is due to $\mathbf{c}_j(\mathbf{v}) = Q_{\Lambda_1}(\mathbf{x})$ for $\mathbf{x} \in R_j(\mathbf{v})$ and $\hat{\mathbf{x}} = \mathbf{c}_j(\mathbf{v}) - Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v})) + Q_{\Lambda_2}(\mathbf{y} - \mathbf{c}_j(\mathbf{v}) + Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v})))$, and (c) is because

$$\sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty} f(\mathbf{c}_j(\mathbf{v})|\mathbf{y})V_1 = \sum_{\mathbf{v}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty}\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})d\mathbf{x} = \int_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}|\mathbf{y})d\mathbf{x} = 1, \tag{3.16}$$

and $Q_{\Lambda_2}(\mathbf{a}+Q_{\Lambda_2}(\mathbf{b})) = Q_{\Lambda_2}(\mathbf{a}) + Q_{\Lambda_2}(\mathbf{b}), \forall\, \mathbf{a},\mathbf{b} \in \mathbb{R}^n$, which leads to

$$\begin{aligned}
&\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v}))-Q_{\Lambda_2}\big(\mathbf{y}-\mathbf{c}_j(\mathbf{v})+Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v}))\big)\|^2 d\mathbf{x}\\
=&\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v}))-Q_{\Lambda_2}\big(\mathbf{y}-\mathbf{c}_j(\mathbf{v})\big)-Q_{\Lambda_2}(\mathbf{c}_j(\mathbf{v}))\|^2 d\mathbf{x}\\
=&\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}\big(\mathbf{y}-\mathbf{c}_j(\mathbf{v})\big)\|^2 d\mathbf{x}\\
=&\int_{\mathbf{x}\in R_j(\mathbf{v})} f(\mathbf{x}|\mathbf{y})\|Q_{\Lambda_2}(\mathbf{x}-\mathbf{y})\|^2 d\mathbf{x}.
\end{aligned} \tag{3.17}$$

Therefore, the average distortion per dimension over all realizations of $\mathbf{y}$ is

$$\begin{aligned}
D_n &= \frac{1}{n}E_Y[D(\mathbf{y})]\\
&= G(\Lambda_1)V_1^{\frac{2}{n}} + \frac{1}{n}\int_{\mathbf{x}}\int_{\mathbf{y}} f(\mathbf{x},\mathbf{y})\|Q_{\Lambda_2}(\mathbf{x}-\mathbf{y})\|^2 d\mathbf{x}d\mathbf{y}\\
&= G(\Lambda_1)V_1^{\frac{2}{n}} + \frac{1}{n}\int_{\mathbf{y}} f(\mathbf{y})\int_{\mathbf{z}} f(\mathbf{z})\|Q_{\Lambda_2}(\mathbf{z})\|^2 d\mathbf{z}d\mathbf{y}\\
&= G(\Lambda_1)V_1^{\frac{2}{n}} + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2].
\end{aligned} \tag{3.18}$$

□

Remarks

(a) For a fixed pair of the nested lattices (Λ_1, Λ_2), D_n only depends on Z, i.e, the correlation between X and Y. It is independent of the marginal distribution of X (or

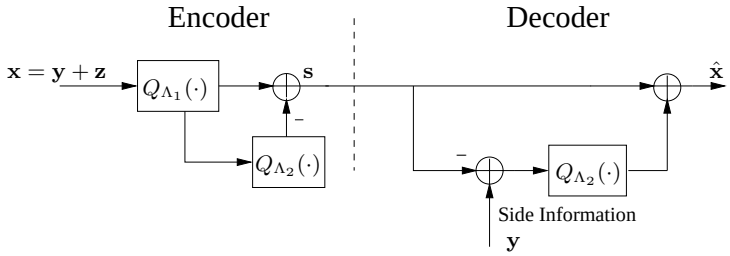


Fig. 4. The simplified nested lattice quantizer for Wyner-Ziv coding.

Y).

(b) $D_s = G(\Lambda_1)V_1^{\frac{2}{n}}$ is determined by the geometric structure and V_1. It is the same as the MSE for classic lattice quantizers [11]. $D_C = \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]$ depends on V_2 and the distribution of Z, and is characterised by the error probability of the lattice channel code.

2. The Quadratic Gaussian Case When $n \to \infty$

Corollary 1: In the quadratic Gaussian case,

$$\lim_{n\to\infty} D_n = D_{WZ} = \sigma_Z^2 2^{-2R}. \tag{3.19}$$

Proof. Since the nested lattice quantizer is a fixed-rate quantizer with rate $R = \frac{1}{n}\log(\frac{V_2}{V_1})$ per dimension, then (3.13) can be rewritten as

$$D_n = G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R} + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^{\mathbf{2}}]. \tag{3.20}$$

For the quadratic Gaussian case, according to [9, (3.14)],

$$\lim_{n\to\infty} \frac{1}{n}\log V_2 = \frac{1}{2}\log(2\pi e\sigma_Z^2) \tag{3.21}$$

if we assume Λ_2 is a *good* AWGN channel σ_Z^2-code [17], meaning $\mathcal{V}_2$ approximates a

Euclidean ball of radius $\sqrt{n}\sigma_Z$. Then

$$\lim_{n\to\infty} G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R} = \frac{1}{2\pi e}2\pi e\sigma_Z^2 2^{-2R} = D_{WZ}(R). \tag{3.22}$$

At the same time, according to [9, (3.12)], for any $\varepsilon > 0$ and sufficiently large n, $P_r\{\mathbf{Z} \notin \mathcal{V}_2\} < \varepsilon$ for the *good* Λ_2, hence

$$\lim_{n\to\infty} \frac{1}{n} E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2] = 0. \tag{3.23}$$

Consequently,

$$\lim_{n\to\infty} D_n = D_{WZ} = \sigma_Z^2 2^{-2R} \tag{3.24}$$

for the quadratic Gaussian case. □

The limit (3.19) we obtain under the high rate assumption is consistent with results in [9], which assert that nested lattice quantization can achieve the Wyner-Ziv limit asymptotically as the dimensionality n goes to infinity for all rates.

3. A Lower Bound on the D-R Performance with Finite n in the Quadratic Gaussian Case

The source coding loss $D_S = G(\Lambda_1)V_1^{\frac{2}{n}}$ in (3.13) is in an explicit form, while the channel coding loss $D_C = \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]$ not so clear. In the quadratic Gaussian case with $Z \sim (0, \sigma_Z^2)$, we obtain from Theorem 1 a lower bound on the high-rate D-R performance of finite-dimensional nested lattice quantizers.

Corollary 2: For $X = Y + Z$, $Z \sim N(0, \sigma_Z^2)$, the operational D-R function $D_n(R)$ of Wyner-Ziv coding of X (with decoder side information Y) using n-dimensional nested lattice quantizers is lower-bounded at high rate by

$$\overline{D}_n(R) \triangleq \min_{V_2>0}(D_s(R) + \overline{D}_c(R)), \quad \text{for} \quad n > 1, \tag{3.25}$$

with

$$D_s(R) = G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R}, \tag{3.26}$$

and

$$\overline{D}_c(R) = \frac{(n-1)}{n\Gamma(\frac{n+1}{2})2^{\frac{n}{2}}\pi^{\frac{1}{2}}}\sum_{\mathbf{l}\in\Lambda_2} l^2 \int_{l-r}^{l+r}\int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} \sin^{n-2}\theta d\theta \frac{u^{n-1}}{\sigma_Z^n}\exp(-\frac{u^2}{2\sigma_Z^2})du, \tag{3.27}$$

where $l \triangleq ||\mathbf{l}||$, $r = \pi^{-\frac{1}{2}}(\Delta\Gamma(\frac{n}{2}+1)V_2)^{\frac{1}{n}}$ is the packing radius [17, p. 6] of Λ_2, Δ its density [17]. When $n = 1$, the exact best possible high-rate D-R performance is

$$D_1(R) = \min_{V_2>0}\{G(\Lambda_1)V_2^2 2^{-2R} + \frac{1}{2}V_2^2\sum_{i=0}^{\infty}(2i+1)\mathrm{erfc}\big(\frac{V_2}{\sqrt{2}\sigma_Z}(i+\frac{1}{2})\big)\}. \tag{3.28}$$

Proof. 1) Source coding loss: The nested lattice quantizer is a fixed rate quantizer with $R = \frac{1}{n}\log_2(\frac{V_2}{V_1})$. Therefore $D_s(R) = G(\Lambda_1)V_1^{\frac{2}{n}} = G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R}$.
2) Channel coding loss: For the quadratic Gaussian case, the second term in (3.13) can be evaluated as

$$\begin{aligned}
\frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2] &= \frac{1}{n}\sum_{\mathbf{l}\in\Lambda_2}\int_{\mathbf{z}\in\mathcal{V}_2(\mathbf{l})}\|Q_{\Lambda_2}(\mathbf{Z})\|^2 f(\mathbf{z})d\mathbf{z} \\
&= \frac{1}{n}\sum_{\mathbf{l}\in\Lambda_2}\|\mathbf{l}\|^2\int_{\mathbf{z}\in\mathcal{V}_2(\mathbf{l})} f(\mathbf{z})d\mathbf{z} \\
&= \frac{1}{n}\sum_{\mathbf{l}\in\Lambda_2} l^2 P_r(\mathbf{z}\in\mathcal{V}_2(\mathbf{l})),
\end{aligned} \tag{3.29}$$

where $\mathcal{V}_2(\mathbf{l})$ is the Voronoi cell associated with the lattice point $\mathbf{l}\in\Lambda_2$.

For the one-dimensional ($n = 1$) case, $P_r(\mathbf{z}\in\mathcal{V}_2(\mathbf{l}))$ can be expressed in terms of the erfc function, then $E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]$ becomes [61]

$$\frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2] = \frac{1}{2}V_2^2\sum_{i=0}^{\infty}(2i+1)\mathrm{erfc}\big(\frac{V_2}{\sqrt{2}\sigma_Z}(i+\frac{1}{2})\big). \tag{3.30}$$

We hence have

$$D_1(R) = \min_{V_2>0}\{G(\Lambda_1)V_2^2 2^{-2R} + \frac{1}{2}V_2^2 \sum_{i=0}^{\infty}(2i+1)\mathrm{erfc}\big(\frac{V_2}{\sqrt{2}\sigma_Z}(i+\frac{1}{2})\big)\}. \tag{3.31}$$

For the case when $n > 1$, since the distribution of $\mathbf{z}$ is spherically symmetric, without loss of generality, we can assume that $\mathbf{l} \in \Lambda_2$ lies on an axis (e.g., the horizontal axis, as shown in Fig. 5 for $n = 2$) and use polar coordinate systems when computing $P_r\big(\mathbf{z} \in \mathcal{V}_2(\mathbf{l})\big)$. Let $\mathcal{S}$ be the packing sphere [17, p. 6] of the Voronoi region $\mathcal{V}_2(\mathbf{l})$. The volume of $\mathcal{S}$ is

$$V_{\mathcal{S}} = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}r^n = \Delta V_2. \tag{3.32}$$

When $n = 2$, from the (u, θ) polar coordinate system in Fig. 5, we have

$$\begin{aligned} P_r(\mathbf{z} \in \mathcal{V}_2(\mathbf{l})) &\geq P_r(\mathbf{z} \in \mathcal{S}) \\ &= \int_{l-r}^{l+r}\int_0^{2\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} u\frac{1}{2\pi\sigma_Z^2}\exp(-\frac{u^2}{2\sigma_Z^2})d\theta du \\ &= \int_{l-r}^{l+r}\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})\frac{u}{\pi\sigma_Z^2}\exp(-\frac{u^2}{2\sigma_Z^2})du. \end{aligned} \tag{3.33}$$

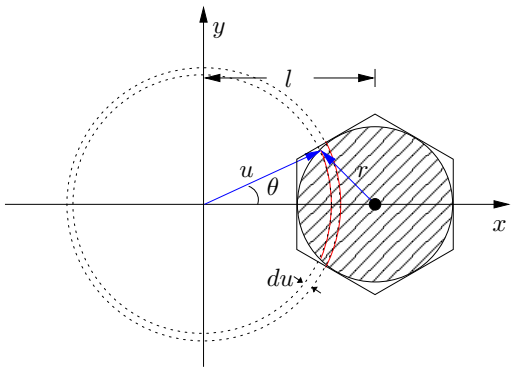


Fig. 5. Geometry used in evaluting $P_r(\mathbf{z} \in \mathcal{V}_2(\mathbf{l}))$ for the two-dimensional case.

Similarly, when $n = 3$, from the polar coordinate system (u, θ, θ_1), we have

$$\begin{aligned} P_r(\mathbf{z} \in \mathcal{V}_2(\mathbf{l})) &\geq P_r(\mathbf{z} \in \mathcal{S}) \\ &= \int_{l-r}^{l+r} \int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} \sin\theta d\theta \int_0^{2\pi} d\theta_1 u^2 \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_Z^3} \exp(-\frac{u^2}{2\sigma_Z^2}) du \\ &= \frac{1}{\sqrt{2\pi}} \int_{l-r}^{l+r} \int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} \sin\theta d\theta \frac{u^2}{\sigma_Z^3} \exp(-\frac{u^2}{2\sigma_Z^2}) du. \end{aligned} \tag{3.34}$$

For $n > 3$, we generalize (3.34) by using the polar coordinate system $(u, \theta, \theta_1, \cdots, \theta_{n-2})$ and its Jacobian determinant [62, p. 904] to get

$$\begin{aligned} & P_r(\mathbf{z} \in \mathcal{V}_2(\mathbf{l})) \geq P_r(\mathbf{z} \in \mathcal{S}) \\ =& \int_{l-r}^{l+r} \int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} \sin^{n-2}\theta d\theta \int_0^{\pi} \sin^{n-3}\theta_1 d\theta_1 \cdots \int_0^{\pi} \sin\theta_{n-3} d\theta_{n-3} \int_0^{2\pi} d\theta_{n-2} \\ & \frac{u^{n-1}}{(2\pi)^{\frac{n}{2}}\sigma_Z^n} \exp(-\frac{u^2}{2\sigma_Z^2}) du \\ =& \frac{n-1}{\Gamma(\frac{n+1}{2}) 2^{\frac{n}{2}} \pi^{\frac{1}{2}}} \int_{l-r}^{l+r} \int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} \sin^{n-2}\theta d\theta \frac{u^{n-1}}{\sigma_Z^n} \exp(-\frac{u^2}{2\sigma_Z^2}) du, \end{aligned} \tag{3.35}$$

because $\int_0^{\pi} \sin^{n-3}\theta_1 d\theta_1 \cdots \int_0^{\pi} \sin\theta_{n-3} d\theta_{n-3} \int_0^{2\pi} d\theta_{n-2} = \frac{(n-1)\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n+1}{2})}$.

Combining (3.33)-(3.35) with (3.29), we obtain the lower bound $\overline{D}_C$ in (3.27) for D_C when $n > 1$. Hence we have (3.25). □

Fig. 4 (a) shows distortions with different V_2's using nested A_2 lattices in two dimensions with $\sigma_Z^2 = 0.01$. The lower bound $\overline{D}_2(R)$ is the lower convex hull of all operational D-R points with different V_2, as shown in Fig. 4 (b). We observe from Fig. 4 (b) that the gap (in dB) from $\overline{D}_n(R)$ to $D_{WZ}(R)$ keeps increasing as the rate increases. This is due to the fact that the source coding loss $D_S = G(\Lambda_1)V_1^{\frac{2}{n}} = \frac{1}{12g(\Lambda_1)}(\frac{V_2}{N})^{\frac{2}{n}} = \frac{1}{12g(\Lambda_1)} V_2^{\frac{2}{n}} 2^{-2R}$ is bounded away from 2^{-2R} with increasing V_2, where $g(\Lambda_1) \triangleq \frac{1}{12G(\Lambda_1)}$ is the granular gain [11] of lattice Λ_1, and $R = \frac{1}{n}\log N$, with N being the nesting ratio.

Fig. 7 plots $\overline{D}_n(R)$ for n=1, 2, 4, 8 and 24 with $\sigma_Z^2 = 0.01$. We see that for fixed but finite n, the gap (in dB) between $\overline{D}_n(R)$ and the Wyner-Ziv D-R function $D_{WZ}(R) = \sigma_Z^2 2^{-2R}$ is an increasing function of R, and that for fixed R, it is a decreasing function of n. The following corollary asserts that this gap is independent of σ_Z^2 (or the correlation between X and Y).

From (3.13), (3.29), (3.33) and (3.35), we see that the gap between $D_n(R)$ and our lower bound $\overline{D}_n(R)$ for $n > 1$ is

$$\frac{1}{n} \sum_{\mathbf{l} \in \Lambda_2} l^2 [P_r(\mathbf{z} \in \mathcal{V}_2(\mathbf{l})) - P_r(\mathbf{z} \in \mathcal{S})]. \tag{3.36}$$

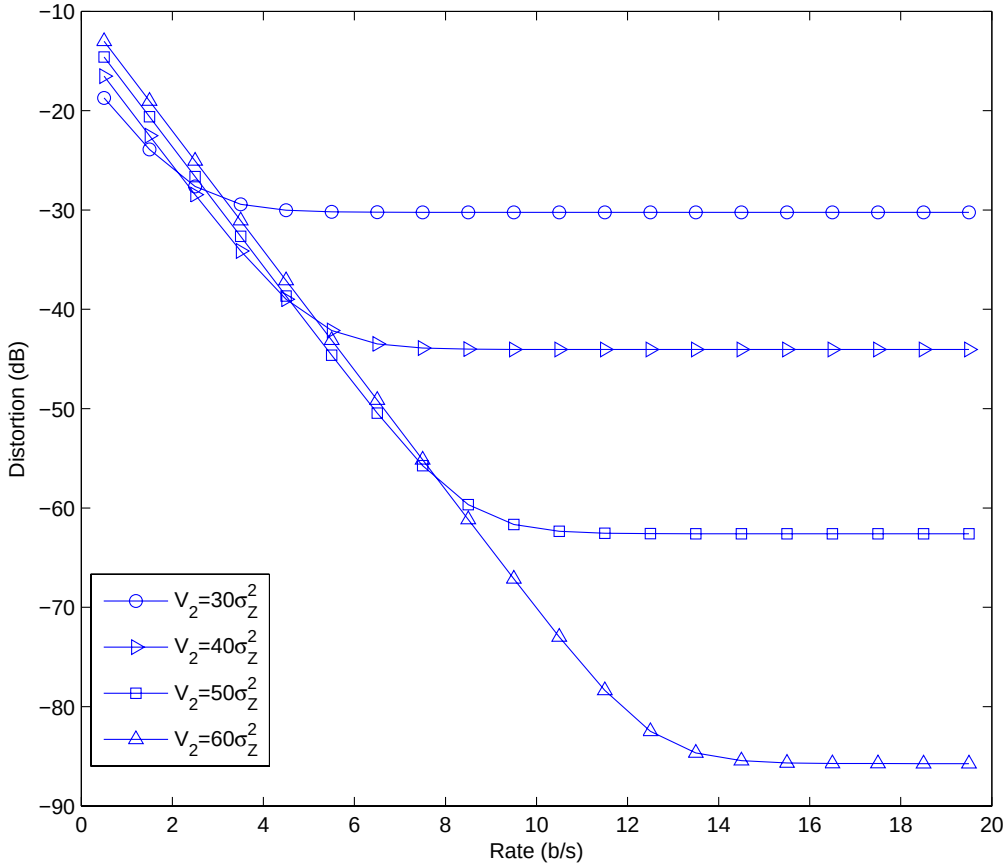
Since $\mathcal{V}_2(\mathbf{l})$ will be more and more spherical like as n increases, our lower bound is asymptotically tight as n goes to infinity. When $n = 2$, as we shall see from Fig. 16 in Section 4, this gap is 0.6 dB at high rate.

4. Performance Under Varying Source Correlation in the Quadratic Gaussian Case

With the source model $X = Y + Z$ in the quadratic Gaussian case, we have $\sigma_Z^2 = \sigma_X^2(1-\rho^2) = (\sigma_Y^2 + \sigma_Z^2)(1-\rho^2)$. If σ_Y^2 is fixed, then the correlation coefficient ρ varies with σ_Z^2.

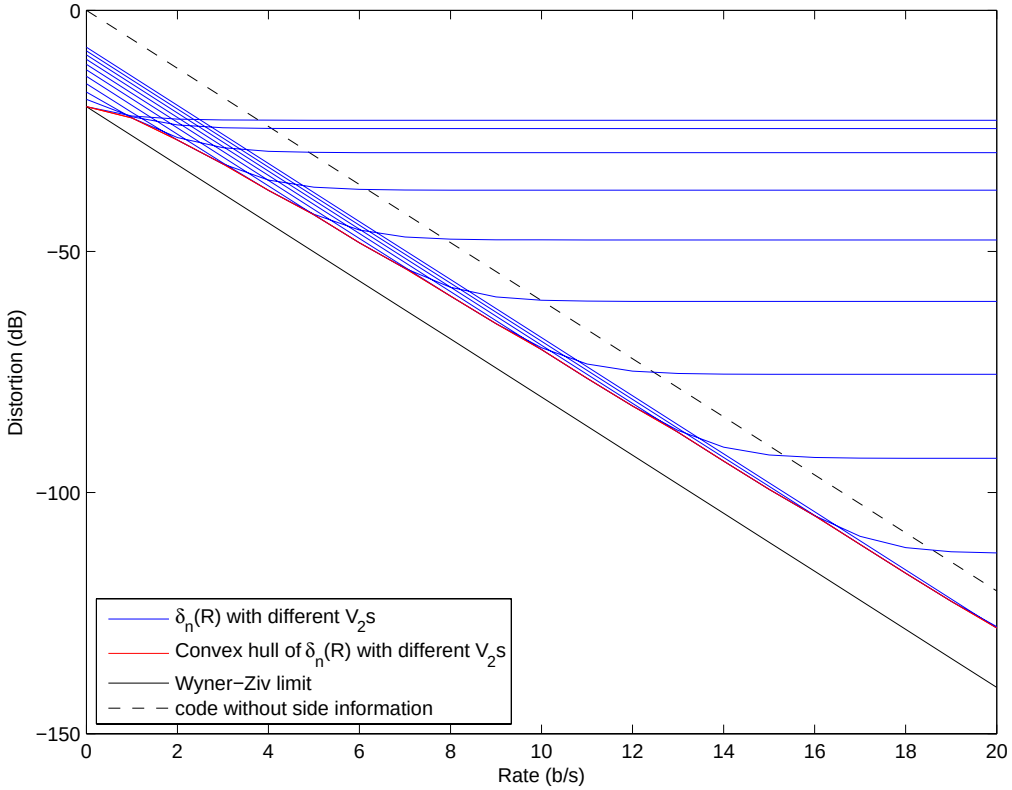
Corollary 3: For fixed rate R and dimensionality n, the lower bound $\overline{D}_n(R)$ in (3.25) remains a constant gap (in dB) from the Wyner-Ziv limit $D_{WZ}(R)$ for all $\sigma_Z^2 > 0$.

Proof. For any point $\mathbf{l} \in \Lambda_2$, according to (3.1), $\mathbf{l} = M_2\mathbf{i}$ for some $\mathbf{i} \in \mathbb{Z}^n$, then $l^2 = ||\mathbf{l}||^2 = \mathbf{i}^T M_2^T M_2 \mathbf{i}$ and $V_2 = [\det(M_2^T M_2)]^{\frac{1}{2}}$ [17, p. 4]. Consider a *similar* lattice [30] of Λ_2, denoted as Λ_2', with packing radius $r' = 1$ and generator matrix M_2', then Λ_2 is a scaled version of Λ_2' with scaling factor r, which is also the packing radius of Λ_2 defined as $r = \pi^{-\frac{1}{2}}(\Delta\Gamma(\frac{n}{2}+1)V_2)^{\frac{1}{n}}$ in Corollary 2. Thus $M_2 = rM_2'$, and $\mathbf{l} = rM_2'\mathbf{i}$.



(a)

Fig. 6. (a) Distortions with different V_2's in the two-dimensional case. (b) $\overline{D}_n(R)$ is the convex hull of distortions for different V_2's.



(b)

Fig. 6. Continued.

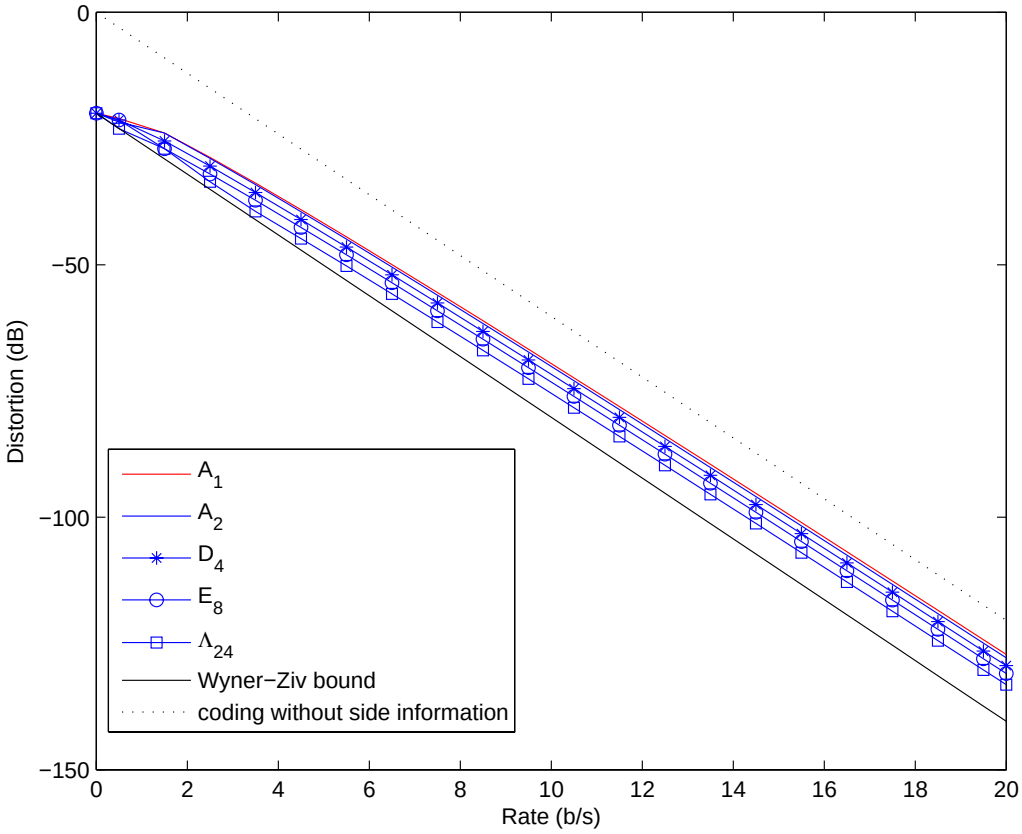


Fig. 7. The lower bound $\overline{D}(R)$ of $D_n(R)$ for different dimensions, with $\sigma_Z^2 = 0.01$, in the quadratic Gaussian case.

Let $\beta = \sqrt{\mathbf{i}^T M_2'^T M_2' \mathbf{i}}$, $\tilde{u} = \frac{u}{\sigma_Z}$, and $\tilde{r} = \frac{r}{\sigma_Z}$, then $l = \beta r = \beta \sigma_Z \tilde{r}$. Starting from (3.27), we get

$$D_s(R) = G(\Lambda_1) V_2^{\frac{2}{n}} 2^{-2R} = G(\Lambda_1)\pi(\Delta\Gamma(\frac{n}{2}+1))^{-\frac{2}{n}} \sigma_Z^2 \tilde{r}^2 2^{-2R}, \tag{3.37}$$

and

$$\begin{aligned}\overline{D}_c(R) &= \frac{n-1}{n\Gamma(\frac{n+1}{2})2^{\frac{n}{2}}\pi^{\frac{1}{2}}}\sum_{\mathbf{l}\in\Lambda_2} l^2\int_{l-r}^{l+r}\int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})}\sin^{n-2}\theta d\theta\frac{u^{n-1}}{\sigma_Z^n}\exp(-\frac{u^2}{2\sigma_Z^2})du \\ &= \frac{n-1}{n\Gamma(\frac{n+1}{2})2^{\frac{n}{2}}\pi^{\frac{1}{2}}}\sigma_Z^2\sum_{\mathbf{i}\in\mathbb{Z}^n:\beta=\sqrt{\mathbf{i}^T M_2'^T M_2'\mathbf{i}}}\beta^2\tilde{r}^2\int_{(\beta-1)\tilde{r}}^{(\beta+1)\tilde{r}}\int_0^{\cos^{-1}(\frac{(\beta^2-1)\tilde{r}^2+\tilde{u}^2}{2\beta\tilde{r}\tilde{u}})}\sin^{n-2}\theta d\theta\tilde{u}^{n-1}\exp(-\tilde{u}^2)d\tilde{u} \quad (3.38)\end{aligned}$$

Thus the distortion can be written as

$$D_s(R)+\overline{D}_c(R)=\sigma_Z^2\Psi_{\Lambda_1,\Lambda_2}(\tilde{r},R). \quad (3.39)$$

with the function $\Psi_{\Lambda_1,\Lambda_2}(\tilde{r},R)$ depending on σ_Z only through $\tilde{r}=\frac{r}{\sigma_Z}$. Therefore, for fixed R, both the optimal $\tilde{r}$, denoted as $\tilde{r}^*$, which minimizes $\Psi_{\Lambda_1,\Lambda_2}(\tilde{r},R)$ and the resulting minimal $\Psi_{\Lambda_1,\Lambda_2}(\tilde{r}^*,R)$ are independent of σ_Z. Then the lower bound of distortion is

$$\overline{D}_n(R)=\min_{V_2>0}(D_s(R)+\overline{D}_c(R))=\min_{\tilde{r}>0}\sigma_Z^2\Psi_{\Lambda_1,\Lambda_2}(\tilde{r},R)=\sigma_Z^2\Psi_{\Lambda_1,\Lambda_2}(\tilde{r}^*,R), \quad (3.40)$$

which is proportional to σ_Z^2. Since the Wyner-Ziv limit is $D_{WZ}(R)=\sigma_Z^2 2^{-2R}$, then the gap (in dB) between $\overline{D}_n(R)$ and $D_{WZ}(R)$ for fixed R and n is

$$10\log_{10}\frac{\overline{D}_n(R)}{D_{WZ}(R)}=10\log_{10}\frac{\Psi_{\Lambda_1,\Lambda_2}(\tilde{r}^*,R)}{2^{-2R}}, \quad (3.41)$$

which is a constant that is independent of σ_Z^2 (or the correlation between X and Y). □

In addition to the above result about the constant gap (in dB) between $\overline{D}_n(R)$ and $D_{WZ}(R)$ for varying source correlation, we have a similar result when $\overline{D}_n(R)$ is replaced by $D_n(R)$.

Corollary 4: For $X = Y + Z$, $Z \sim N(0, \sigma_Z^2)$, the operational D-R function

$$D_n(R) = \min_{V_2>0}\{G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R} + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]\} \tag{3.42}$$

of Wyner-Ziv coding of X (with decoder side information Y) using n-dimensional nested lattice quantizers at high but fixed rate R remains a constant gap (in dB) from the Wyner-Ziv limit $D_{WZ}(R)$ for all $\sigma_Z^2 > 0$.

Proof. It suffices to show that $D_n(R)$ is proportional to σ_Z^2. Let $\tilde{\mathbf{z}} = \frac{\mathbf{z}}{\sigma_Z}$ and define lattice $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ with generator matrix $\tilde{M}_1 = \frac{1}{\sigma_Z}M_1$ and $\tilde{M}_2 = \frac{1}{\sigma_Z}M_2$, respectively, then for any lattice point $\tilde{\mathbf{l}} \in \tilde{\Lambda}_2$, $\tilde{\mathbf{l}} = \frac{1}{\sigma_Z}\mathbf{l}$, where $\mathbf{l}$ is the corresponding lattice point of Λ_2. In addition, $\tilde{l} = \|\tilde{\mathbf{l}}\| = \frac{1}{\sigma_Z}l$, and the volume of the Voronoi cell of $\tilde{\Lambda}_2$ is $\tilde{V}_2 = \frac{1}{\sigma_Z^n}V_2$. Since the normalized second moment remains unchanged with respect to lattice scaling and rotation, we have

$$\begin{aligned} D_n(R) &= \min_{V_2>0}\{G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R} + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^{\mathbf{2}}]\} \\ &= \min_{V_2>0}\left\{G(\Lambda_1)V_2^{\frac{2}{n}}2^{-2R} + \frac{1}{n}\sum_{\mathbf{l}\in\Lambda_2} l^2 \int_{\mathbf{z}\in\mathcal{V}_{\mathbf{2}}(\mathbf{l})} \frac{1}{(2\pi\sigma_Z^2)^{\frac{n}{2}}}\exp(-\frac{\|\mathbf{z}\|^2}{2\sigma_Z^2})d\mathbf{z}\right\} \\ &= \sigma_Z^2 \min_{\tilde{V}_2>0}\left\{G(\tilde{\Lambda}_1)\tilde{V}_2^{\frac{2}{n}}2^{-2R} + \frac{1}{n}\sum_{\tilde{\mathbf{l}}\in\tilde{\Lambda}_2} \tilde{l}^2 \int_{\tilde{\mathbf{z}}\in\tilde{\mathcal{V}}_2(\tilde{l})} \frac{1}{(2\pi)^{\frac{n}{2}}}\exp(-\frac{\|\tilde{\mathbf{z}}\|^2}{2})d\tilde{\mathbf{z}}\right\}, \end{aligned} \tag{3.43}$$

which is proportional to σ_Z^2. □

C. Slepian-Wolf Coded Nested Lattice Quantization (SWC-NQ)

1. Motivation of SWC-NQ

Recall from theorem 1 that the distortion per dimension of the nested lattice quantizer is $D_n = D_S + D_C$, where $D_S = G(\Lambda_1)V_1^{\frac{2}{n}}$ is the source coding loss, characterized by the granular gain $g(\Lambda_1)$ of Λ_1 and the boundary gain $b(\Lambda_2)$ of Λ_2, whereas the channel

coding loss $D_C \triangleq \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]$ is characterized by the error probability of lattice channel decoding. Suppose the coarse lattice Λ_2 with Voronoi region $\mathcal{V}_2$ in the n-dimensional space has the same overload probability as a cubic support region of side E centered at the origin, then, according to [11], $b(\Lambda_2)$ is defined as the ratio of the normalized volume E^2 of the cubic support region to the normalized volume $V_2^{\frac{2}{n}}$. That is

$$b(\Lambda_2) = \frac{E^2}{V_2^{\frac{2}{n}}}. \tag{3.44}$$

Then

$$D_S = G(\Lambda_1)V_1^{\frac{2}{n}} = \frac{1}{12g(\Lambda_1)}V_2^{\frac{2}{n}}N^{-\frac{2}{n}} = \frac{1}{12g(\Lambda_1)}\frac{E^2}{b(\Lambda_2)}N^{-\frac{2}{n}}, \tag{3.45}$$

where N is the nesting ratio. If N is fixed, D_S will decrease with increasing $b(\Lambda_2)$; but D_C remains unaffected because the overload probability stays unchanged.

To increase the boundary gain $b(\Lambda_2)$, a second-stage of binning can be applied to the quantization indices. The essence of binning is using a channel code to partition the support region into cosets. Assume ideal channel code is employed to partition the support region $\mathcal{V}_2$ into k cosets without decoding errors and denote the set consisting of the coset leaders as $\mathcal{L}$, then $|\mathcal{L}| = k$ and the support region for the quantization indices (or the nested quantizer), which is also the set of the quantization cells associated with $\mathcal{L}$, has volume $\frac{V_2}{k}$. Thus the *effective* volume of the support region decreases by a factor of k after the second stage of binning, and therefore the boundary gain $b(\Lambda_2)$ increases by a factor of $k^{\frac{2}{n}}$.

We thus propose a framework for Wyner-Ziv coding of i.i.d. sources based on SWC-NQ, which follows NQ by Slepian-Wolf coding to perform second-stage binning. Despite the fact that there is almost no correlation among the nested quantization indices that identify the coset leaders $\mathbf{s} = \mathbf{x}_{Q_{\Lambda_1}} - Q_{\Lambda_2}(\mathbf{x}_{Q_{\Lambda_1}}) \in \Lambda_1/\Lambda_2$ of nested lattice

pair (Λ_1, Λ_2), there still remains correlation between $\mathbf{s}$ and the side information $\mathbf{y}$, especially at high rate. Write $\mathbf{W} = Q_{\Lambda_1}(\mathbf{X})$ and $\mathbf{S} = \mathbf{W} - Q_{\Lambda_2}(\mathbf{W})$. Ideal Slepian-Wolf coding can be used to compress $\mathbf{S}$ to the rate of $R = \frac{1}{n}H(\mathbf{S}|\mathbf{Y})$ per dimension. In practice, state-of-the-art channel codes, such as LDPC codes, can be used to approach the Slepian-Wolf limit $\frac{1}{n}H(\mathbf{S}|\mathbf{Y})$ [14]. The role of Slepian-Wolf coding in SWC-NQ is thus to exploit the correlation between $\mathbf{S}$ and $\mathbf{Y}$ for further compression.

2. High-rate Performance for the Quadratic Gaussian Case

Lemma 1: For the quadratic Gaussian case, a lower bound for the high-rate performance of SWC-NQ with a pair of nested lattices (Λ_1, Λ_2) is given as

$$\begin{aligned} D_n(R) &\geq D_s(R) + \overline{D}_c(R), \\ D_s(R) &= G(\Lambda_1)2^{\frac{2}{n}h'(\mathbf{X},\Lambda_2)}\sigma_Z^2 2^{-2R} \\ \overline{D}_c(R) &= \frac{(n-1)}{n\Gamma(\frac{n+1}{2})2^{\frac{n}{2}}\pi^{\frac{1}{2}}} \sum_{\mathbf{l}\in\Lambda_2} l^2 \int_{l-r}^{l+r} \int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})} \sin^{n-2}\theta d\theta \frac{u^{n-1}}{\sigma_Z^n} \exp(-\frac{u^2}{2\sigma_Z^2}) du, \end{aligned} \quad (3.46)$$

where

$$h'(\mathbf{X}, \Lambda_2) \triangleq -\int_{\mathbf{x}\in\mathbb{R}^n} \overline{f}(\mathbf{x}) \log_2[\sum_{j=-\infty}^{\infty} \overline{f}(\mathbf{x} + \frac{\mathbf{c}_j(\mathbf{0})}{\sigma_Z})] d\mathbf{x}, \quad (3.47)$$

$\overline{f}(\cdot)$ is the pdf of an n-dimensional i.i.d. Gaussian source with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_{n\times n}$, and $\mathbf{c}_j(\mathbf{0})$ is defined in section III.B as lattice points of Λ_2.

Proof. See appendix A. □

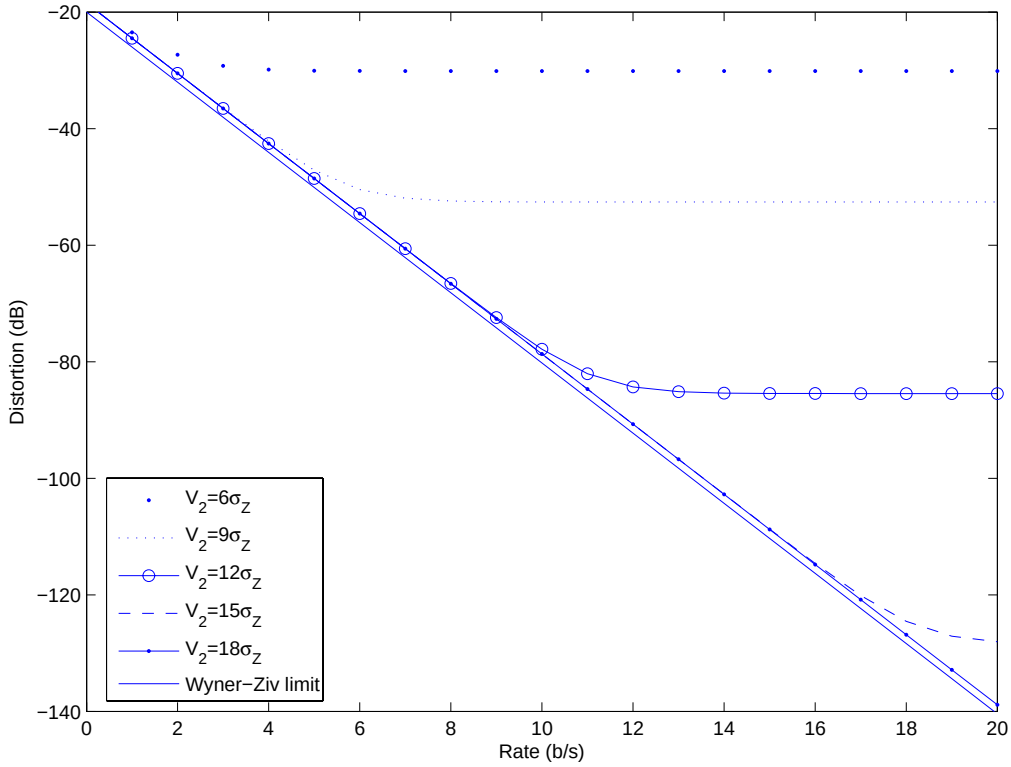


Fig. 8. Lower bounds of $D_1(R)$ with different V_2's in the one-dimensional case.

The lower bounds of $D_1(R)$ for the one-dimensional case with different V_2 are plotted in Fig. 8. It indicates that the *best* performance of a Slepian-Wolf coded nested scalar quantizer remains a constant gap (in dB) from the Wyner-Ziv limit at high rate. Here the *best* means that the minimal achievable distortion D over all possible V_2 for a given rate R. Before rigorously stating our main result in a theorem, we give the following lemma.

Lemma 2: For nested lattice quantization with $\mathbf{W} = Q_{\Lambda_1}(\mathbf{X})$ and $\mathbf{S} = \mathbf{W} - Q_{\Lambda_2}(\mathbf{W})$, $H(\mathbf{S}|\mathbf{Y}) = H(\mathbf{W}|\mathbf{Y})$ at high rate.

Proof. See appendix B. □

Theorem 2: For the quadratic Gaussian case, the optimal D-R performance of SWC-NQ using n-dimensional nested lattices at high rate is

$$D_n^*(R) \triangleq \min_{V_2} D_n(R) = 2\pi e G(\Lambda_1)\sigma_Z^2 2^{-2R}. \tag{3.48}$$

Proof. 1) By Lemma 1

$$\begin{aligned} nR &= H(\mathbf{S}|\mathbf{Y}) = H(\mathbf{W}|\mathbf{Y}) \\ &= H(Q_{\Lambda_1}(\mathbf{X})|\mathbf{Y}) \\ &= h(\mathbf{X}|\mathbf{Y}) - \log V_1 \\ &= \frac{n}{2}\log(2\pi e\sigma_Z^2) - \log V_1 \end{aligned} \tag{3.49}$$

and $D_n(R) = D_S + D_C = D_S = G(\Lambda_1)V_1^{\frac{2}{n}}$ since $D_C = 0$ under ideal Slepian-Wolf coding. Combine R and D_n through V_1 and we get the D-R function as

$$\lim_{V_2\to\infty} D_n(R) = 2\pi e G(\Lambda_1)\sigma_Z^2 2^{-2R}. \tag{3.50}$$

Since

$$D_n^*(R) = \min_{V_2} D_n(R) \le \lim_{V_2\to\infty} D_n(R),$$

we have

$$D_n^*(R) \le 2\pi e G(\Lambda_1)\sigma_Z^2 2^{-2R}. \tag{3.51}$$

2) Denote $\mathbf{w} \triangleq Q_{\Lambda_1}(\mathbf{x})$, and $\mathcal{S}_1 \triangleq \{(\mathbf{X}, \hat{\mathbf{X}}) : \frac{1}{n}E[d(\mathbf{X}, \hat{\mathbf{X}})] \le D_n\}$. The rate of

Wyner-Ziv coding with respect to D_n is [1]

$$\begin{aligned} nR^*(D_n) &= \min_{p(\mathbf{s}),p(\hat{\mathbf{x}}|\mathbf{s},\mathbf{y}),(\mathbf{X},\hat{\mathbf{X}})\in\mathcal{S}_1} I(\mathbf{X};\mathbf{S}|\mathbf{Y}) \\ &\overset{(a)}{=} \min_{p(\mathbf{s}),p(\hat{\mathbf{x}}|\mathbf{s},\mathbf{y}),(\mathbf{X},\hat{\mathbf{X}})\in\mathcal{S}_1} H(\mathbf{S}|\mathbf{Y}) \\ &\overset{(b)}{=} \min_{p(\mathbf{s}),p(\hat{\mathbf{x}}|\mathbf{s},\mathbf{y}),(\mathbf{X},\hat{\mathbf{X}})\in\mathcal{S}_1} H(\mathbf{W}|\mathbf{Y}). \end{aligned} \tag{3.52}$$

where (a) comes from $H(\mathbf{S}|\mathbf{X},\mathbf{Y}) = 0$ and (b) is due to Lemma 1.

Define $\mathcal{S}_2 \triangleq \{(\mathbf{X},\mathbf{W}) : E[d(\mathbf{X},\mathbf{W})] \le D_n\}$. From Theorem 1,

$$\begin{aligned} \frac{1}{n}E[d(\mathbf{X},\hat{\mathbf{X}})] &= G(\Lambda_1)V_1^{\frac{2}{n}} + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2] \\ &= \frac{1}{n}E[d(\mathbf{X},\mathbf{W})] + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2] \\ &\ge \frac{1}{n}E[d(\mathbf{X},\mathbf{W})], \end{aligned} \tag{3.53}$$

then for $\forall(\mathbf{X},\hat{\mathbf{X}}) \in \mathcal{S}_1$,

$$D_n \ge \frac{1}{n}E[d(\mathbf{X},\hat{\mathbf{X}})] \ge \frac{1}{n}E[d(\mathbf{X},\mathbf{W})], \tag{3.54}$$

which means $(\mathbf{X},\hat{\mathbf{X}}) \in \mathcal{S}_2$. Hence $\mathcal{S}_1 \subseteq \mathcal{S}_2$, and

$$\begin{aligned} nR^*(D_n) &= \min_{p(\mathbf{s}),p(\hat{\mathbf{x}}|\mathbf{s},\mathbf{y}),(\mathbf{X},\hat{\mathbf{X}})\in\mathcal{S}_1} H(\mathbf{W}|\mathbf{Y}) \\ &\ge \min_{p(\mathbf{s}),p(\hat{\mathbf{x}}|\mathbf{s},\mathbf{y}),(\mathbf{X},\hat{\mathbf{X}})\in\mathcal{S}_2} H(\mathbf{W}|\mathbf{Y}). \end{aligned} \tag{3.55}$$

Since $H(\mathbf{W}|\mathbf{Y}) = \frac{n}{2}\log(2\pi e\sigma_Z^2) - \log V_1$ and $E[d(\mathbf{X},\mathbf{W})] = G(\Lambda_1)V_1^{\frac{2}{n}}$,

(3.55) becomes

$$\begin{aligned} nR^*(D_n) &\geq \min_{p(\mathbf{s}),p(\hat{\mathbf{x}}|\mathbf{s},\mathbf{y}),(\mathbf{X},\hat{\mathbf{X}})\in\mathcal{S}_2} \frac{n}{2}\log(2\pi e\sigma_Z^2) - \log V_1 \\ &= \min_{E[d(\mathbf{X},\mathbf{W})]\leq D_n} \Big\{\frac{n}{2}\log(2\pi e\sigma_Z^2) - \frac{n}{2}\log\frac{E[d(\mathbf{X},\mathbf{W})]}{G}(\Lambda_1)\Big\} \\ &= \min_{E[d(\mathbf{X},\mathbf{W})]\leq D_n} \frac{n}{2}\log\frac{2\pi eG(\Lambda_1)\sigma_Z^2}{E[d(\mathbf{X},\mathbf{W})]} \\ &= \frac{n}{2}\log\frac{2\pi eG(\Lambda_1)\sigma_Z^2}{D_n}. \end{aligned} \tag{3.56}$$

We thus have

$$D_n^*(R) \geq 2\pi eG(\Lambda_1)\sigma_Z^2 2^{-2R}. \tag{3.57}$$

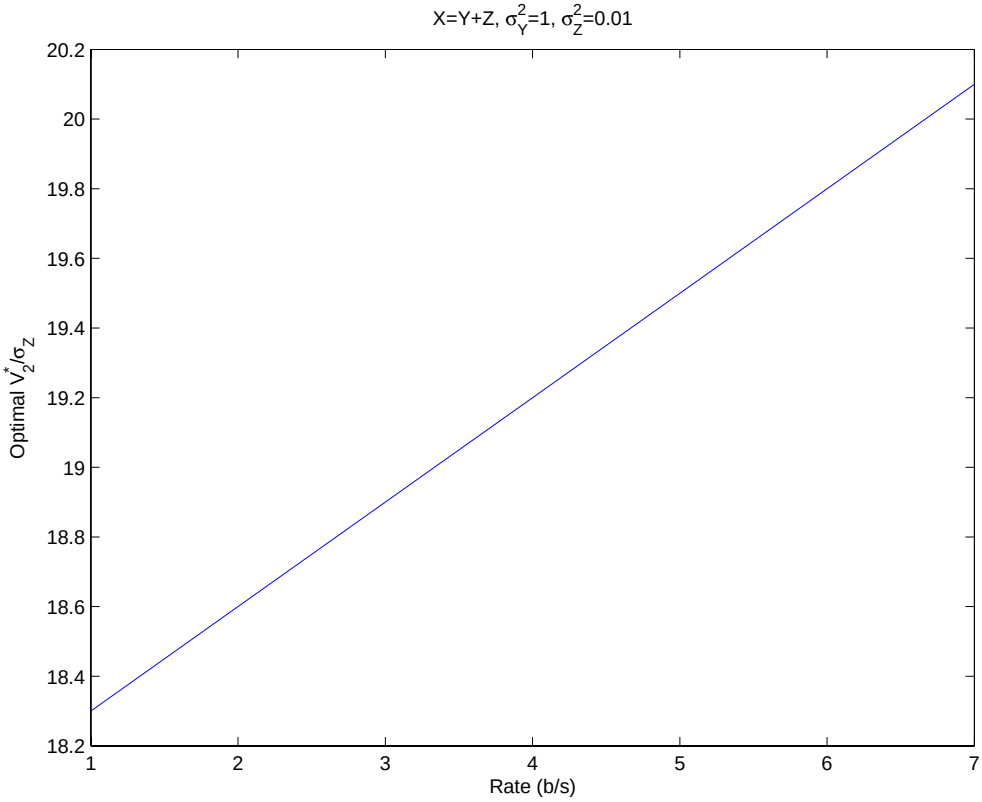
Combining (3.51) and (3.57), we conclude that, at high rate, the *best* D-R performance of the quadratic Gaussian SWC-NQ using n-dimensional lattices is

$$D_n^*(R) = 2\pi eG(\Lambda_1)\sigma_Z^2 2^{-2R}. \tag{3.58}$$

□

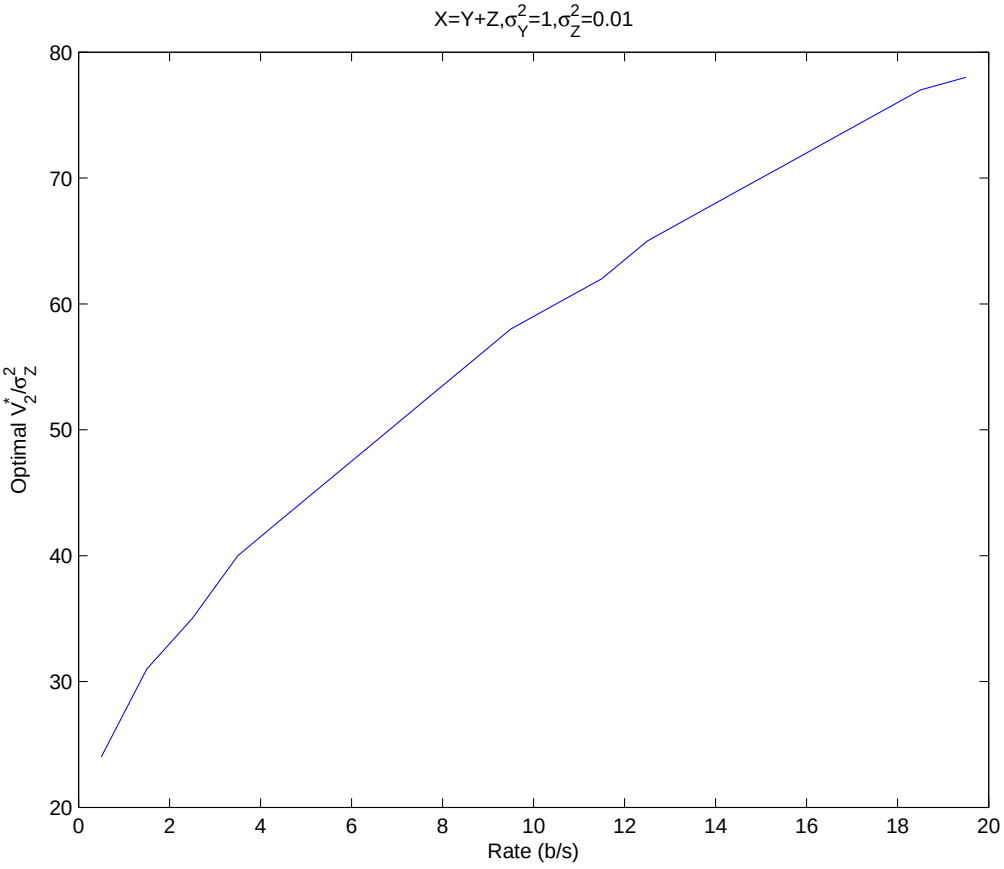
Thus at high rate and for the quadratic Gaussian case, SWC-NQ performs the same as classic entropy-coded lattice quantization with the side information available at both the encoder and decoder. Specifically, the D-R functions with one-dimensional (scalar) lattice and two-dimensional (hexagonal) nested lattices are 1.53 dB and 1.36 dB away from the Wyner-Ziv D-R function, respectively. This interesting and important fact is highlighted in Table I.

We found that for finite rate R and small n (e.g., $n = 1$ and 2), the optimal V_2, denoted as V_2^*, that minimizes the distortion $D_n(R)$ is also finite. Fig. 2 (a) and (b) plot the optimal V_2^* (scaled by σ_Z) as functions of R for $n = 1$ and $n = 2$. We see that as R goes to infinity, V_2^* also goes to infinity. We also observe that for fixed R and n, $D_n(R)$ stays roughly unchanged for $V_2 > V_2^*$.



(a)

Fig. 9. $\frac{V_2^*}{\sigma_Z}$ vs. R for (a) $n = 1$ and (b) $n = 2$.



(b)

Fig. 9. Continued.

Table I. High-rate classic source coding vs. high-rate SWC-NQ for Wyner-Ziv coding.

Classic source coding		WZC	
Coding scheme	Gap to $D_X(R)$	Coding scheme	Gap to $D_{WZ}(R)$
ECSQ [10]	1.53 dB	SWC-NSQ	1.53 dB
ECLQ (2-D) [63]	1.36 dB	SWC-NQ (2-D) [48]	1.36 dB
ECTCQ [64]	0.2 dB	SWC-TCQ [65]	0.2 dB

Remarks

(a) Since SWC-NQ has its root in variable-rate quantization, it is not surprising to see (3.58) as an elegant generalization of the result in entropy-coded quantization from classic source coding to Wyner-Ziv coding.

(b) For practical lossless source coding, if one regards lossless source coding as a special case of Slepian-Wolf coding without side information at the decoder, then channel coding techniques can also used for source coding based on syndromes [66]. In this light, the SWC component in SWC-NQ can be viewed as the counterpart of entropy coding in classic source coding. Although the idea of using channel codes for source coding dates back to the Shannon-MacMillan theorem [67, 68] and theoretical results appeared in [69, 70], practical turbo/LDPC code based noiseless data compression scheme did not appear until very recently [71, 72].

(c) SWC-NQ relies on conditional entropy coding (or Slepian-Wolf coding implemented via channel coding) to achieve rate savings after nested lattice quantization. That fact that (3.58) holds means two things: 1) Just like entropy coding can achieve all the boundary gain in classic source coding of Gaussian sources [10], Slepian-Wolf coding in SWC-NQ can realize all the remaining boundary gain left unexploited by the coarse lattice channel code of the nested lattice quantizer. 2) Ideal Slepian-Wolf

coding also renders the channel code loss D_C in $D_n = D_S + D_C$ to zero. Thus, with ideal Slepian-Wolf coding, the only remaining loss in D_n is the granular loss portion of D_S, which is maximally 1.53 dB [11].

Similar to Corollary 4, we give (without proof) the following corollary of Theorem 2 for high-rate SWC-NQ.

Corollary 5: For fixed dimensionality n, the optimal performance of high-rate SWC-NQ in the quadratic Gaussian case is $10\log_{10} 2\pi e G(\Lambda_1)$ dB from the Wyner-Ziv limit $D_{WZ}(R)$ for all $\sigma_Z^2 > 0$ (or any correlation between X and Y).

The above result is stronger than that in Corollary 3 because the $10\log_{10} 2\pi e G(\Lambda_1)$ dB gap is also independent of the rate R. We thus conclude that the high-rate performance of Wyner-Ziv coding of quadratic Gaussian sources with nested lattice quantization is independent of the source correlation, regardless of whether Slepian-Wolf coding is used or not. There are only high-dimension asymptotics with nested lattice quantization (studied in [9]) and high-rate asymptotics with SWC-NQ (presented in Theorem 2).

D. Code Design and Simulation Results

1. Design of Nested Lattice Quantizer

The nested lattice quantizer design problem involves optimizing the nesting in the quantizer encoder (under a fixed nesting ratio N) and devising the minimum MSE estimator at the decoder.

a. Optimizing the Nesting Scheme of Lattices

Lattices with the densest packing (i.e. the best channel code) and the thinnest covering (i.e. the best source code) are introduced in [17]. For example, the hexagon

lattice A_2 is the best lattice source code and the best lattice channel code at $n = 2$. To optimize the nesting scheme, Λ_2 should be *clean* and geometrically similar to Λ_1 [30, 21], where the former means that the Voronoi cell boundaries for Λ_2 do not intersect Λ_1. We follow the scheme suggested by Conway *et al.* in [30] to search for clean and similar nested lattice pairs. Fig. 10 illustrates the nesting lattice pair for A_2 with nesting ratio $N = 31$.

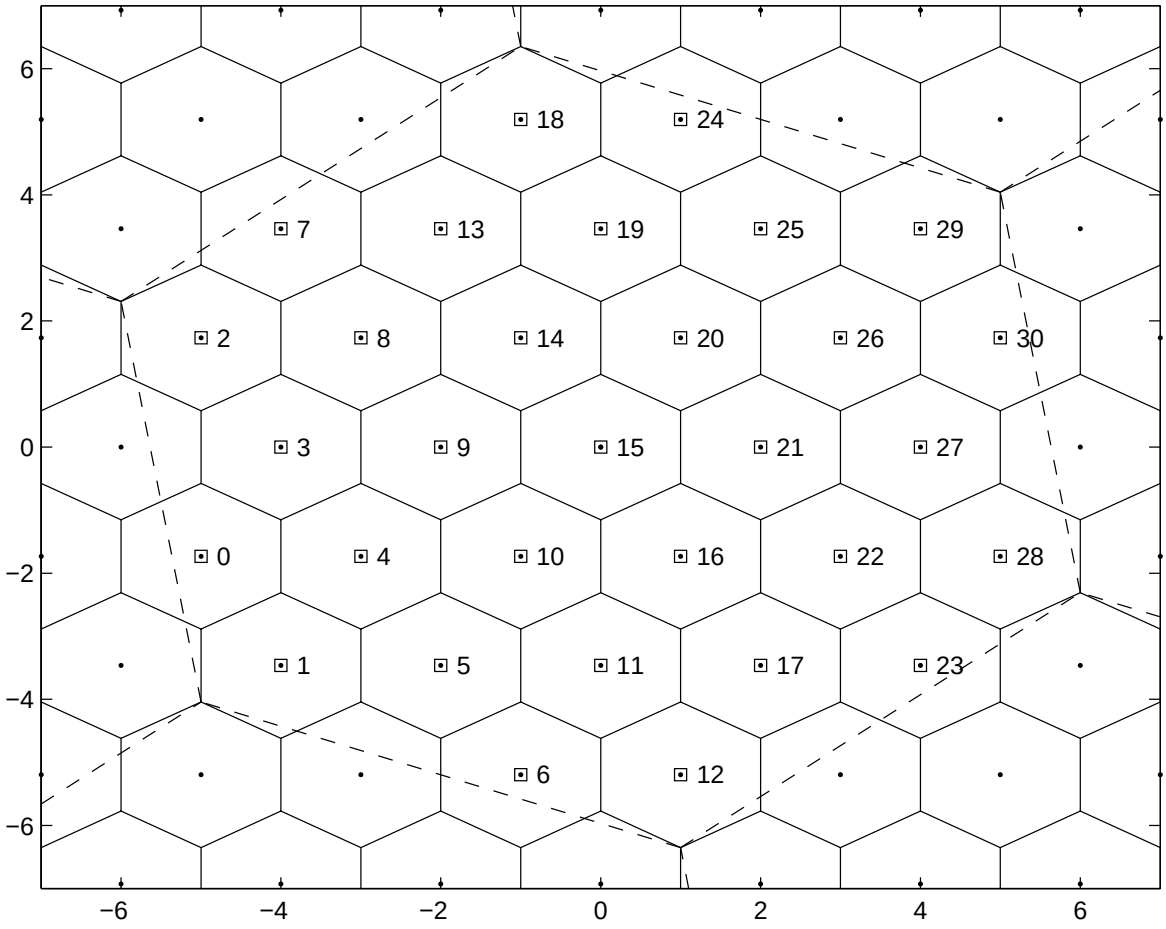


Fig. 10. A clean and geometrically similar nested hexagonal lattice pair with nesting ratio $N = 31$.

b. The Optimal Decoder

The optimal decoder for the nested lattice quantizer is the one that minimizes the MSE between X and the reconstructed $\hat{X}$.

Theorem 3: The minimum MSE decoder for the nested lattice quantizer is

$$\hat{x} = E[\mathbf{X}|\mathbf{s},\mathbf{y}] = \frac{\int_{R(\mathbf{s})} \mathbf{x} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\int_{R(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}, \tag{3.59}$$

where $\mathbf{s} = \mathbf{x}_{Q_{\Lambda_1}} - Q_{\Lambda_2}(\mathbf{x}_{Q_{\Lambda_1}}) \in \Lambda_1/\Lambda_2$ is the received coset leader for $\mathbf{x}$ at the decoder.

Proof.

$$\begin{aligned}
\hat{\mathbf{x}} &= E[\mathbf{X}|\mathbf{s},\mathbf{y}] \\
&= \int_{\mathbb{R}^n} \mathbf{x} f(\mathbf{x}|\mathbf{s},\mathbf{y}) d\mathbf{x} \\
&= \int_{\mathbb{R}^n} \mathbf{x} \frac{f(\mathbf{x},\mathbf{s}|\mathbf{y})}{p(\mathbf{s}|\mathbf{y})} d\mathbf{x} \\
&= \frac{\int_{\mathbb{R}^n} \mathbf{x} f(\mathbf{x},\mathbf{s}|\mathbf{y}) d\mathbf{x}}{p(\mathbf{s}|\mathbf{y})} \\
&= \frac{\int_{\mathbb{R}^n} \mathbf{x} f(\mathbf{x}|\mathbf{y}) p(\mathbf{s}|\mathbf{x},\mathbf{y}) d\mathbf{x}}{\int_{R(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}.
\end{aligned} \tag{3.60}$$

Since $\mathbf{S}$, $\mathbf{X}$ and $\mathbf{Y}$ form a Markov chain $\mathbf{Y} \longleftrightarrow \mathbf{X} \longleftrightarrow \mathbf{S}$,

$$p(\mathbf{s}|\mathbf{x},\mathbf{y}) = p(\mathbf{s}|\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} \notin R(\mathbf{s}), \\ 1 & \text{if } \mathbf{x} \in R(\mathbf{s}). \end{cases} \tag{3.61}$$

Thus

$$\begin{aligned}
\hat{\mathbf{x}} &= \frac{\int_{\mathbb{R}^n} \mathbf{x} f(\mathbf{x}|\mathbf{y}) p(\mathbf{s}|\mathbf{x}) d\mathbf{x}}{\int_{R(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}} \\
&= \frac{\int_{R(\mathbf{s})} \mathbf{x} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\int_{R(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}.
\end{aligned} \tag{3.62}$$

□

Note that this optimal quantizer decoder (3.62), which in the form of a non-linear estimator, is consistent with the centroid condition in classic minimum MSE lattice (or vector) quantizer design [10].

As an example, in the quadratic Gaussian case with $X = Y + Z$, $Y \sim N(0, \sigma_Y^2)$ and $Z \sim N(0, \sigma_Z^2)$, when $n = 1$, $f(x|y) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{(x-\rho y)^2}{2\sigma_Z^2}}$ with $\rho = \frac{\sigma_Y}{\sqrt{\sigma_Y^2+\sigma_Z^2}}$, the optimal decoder for a nested scalar quantizer can be expressed as

$$\hat{x} = \frac{\sum_{j=-\infty}^{\infty} \int_{sq_1+jq_2}^{(s+1)q_1+jq_2} x e^{-\frac{(x-\rho y)^2}{2\sigma_Z^2}} dx}{\sum_{j=-\infty}^{\infty} \int_{sq_1+jq_2}^{(s+1)q_1+jq_2} e^{-\frac{(x-\rho y)^2}{2\sigma_Z^2}} dx}, \tag{3.63}$$

where q_1 and q_2 are the stepsizes of the nested scalar quantizers Λ_1 and Λ_2 with $q_2 = Nq_1$ and N being the nesting ratio.

Non-linear estimation at the decoder plays an important role at low rate. Fig. 11 shows the improvement gained at low rate by using the non-linear minimum MSE estimator of (3.62) vs. the linear estimator $\mathbf{s} + Q_{\Lambda_2}(\mathbf{y} - \mathbf{s})$ valid for high rate (see section IV) for $n = 2$ with $\sigma_Y^2 = 1$ and $\sigma_Z^2 = 0.01$.

Corollary 6: The non-linear minimum MSE estimator of (3.62) degenerates to the linear one $\hat{\mathbf{x}} = \mathbf{s} + Q_{\Lambda_2}(\mathbf{y} - \mathbf{s})$ at high rate.

Proof. At high rate,

$$\begin{aligned}
\hat{\mathbf{x}} &= \frac{\int_{R(\mathbf{s})} \mathbf{x} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\int_{R(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}} \\
&= \frac{\sum_{j=-\infty}^{\infty} \int_{R_j(\mathbf{s})} \mathbf{x} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\sum_{j=-\infty}^{\infty} \int_{R_j(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}} \\
&\overset{(a)}{=} \frac{\sum_{j=-\infty}^{\infty} c_j(\mathbf{s}) \int_{R_j(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\sum_{j=-\infty}^{\infty} \int_{R_j(\mathbf{s})} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}} \\
&\overset{(b)}{=} \frac{[\mathbf{s} + Q_{\Lambda_2}(\mathbf{y} - \mathbf{s})] \int_{\mathcal{V}_1(\mathbf{s}+Q_{\Lambda_2}(\mathbf{y}-\mathbf{s}))} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}}{\int_{\mathcal{V}_1(\mathbf{s}+Q_{\Lambda_2}(\mathbf{y}-\mathbf{s}))} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}} \\
&= \mathbf{s} + Q_{\Lambda_2}(\mathbf{y} - \mathbf{s}),
\end{aligned} \tag{3.64}$$

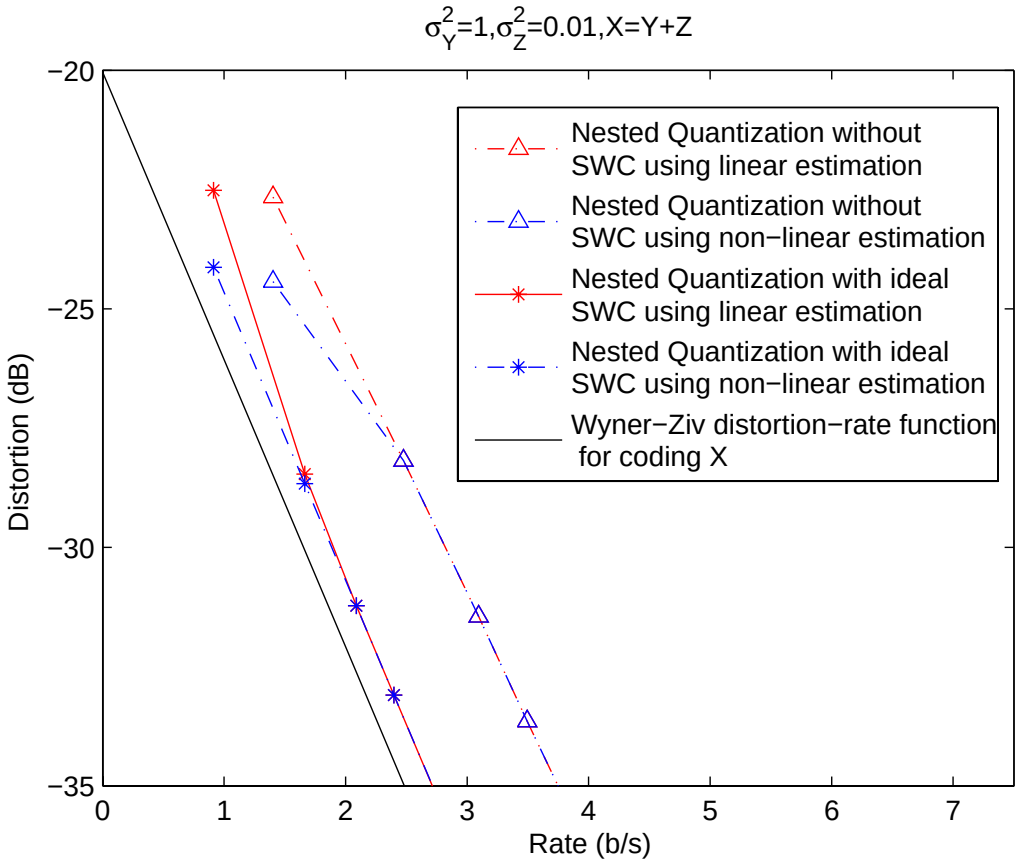


Fig. 11. Improvement gained at low rate by using the non-linear minimum MSE estimator vs. the linear estimator for $n = 2$, $\sigma_Y^2 = 1$ and $\sigma_Z^2 = 0.01$.

where (a) and (b) are due to the high rate assumption and $\mathcal{V}_1(\mathbf{s}+Q_{\Lambda_2}(\mathbf{y}-\mathbf{s})) = \{\mathbf{x}: Q_{\Lambda_1}(\mathbf{x}) = \mathbf{s}+Q_{\Lambda_2}(\mathbf{y}-\mathbf{s})\}$. □

Recall that the linear estimator of (3.64) is the one we use for high-rate performance analysis in section IV. But non-linear estimation is employed for all rate in our simulations. Since (3.62) involves integration over a disconnected region consisting of many isolated Voronoi cells, we use Monte Carlo method to compute this integration in our simulations.

2. Slepian-Wolf Code Design

a. Random Binning

Let V and S be the source and side information, respectively. Since we are considering lossless coding, V has a finite alphabet in general.

We will consider a block code of length-n here. The idea of random binning [2, pp. 410-413] is to partition all the length-n sequences of V randomly into bins and only the indices of these bins are transmitted to the decoder. For an i.i.d. discrete source V, the set of all length-n sequences generated by V is randomly partitioned into 2^{nR} bins. Hence, if we compress V at rate R, there should be 2^{nR} bins.

Knowing the bin index and the sequence of side information S^n, the decoder reconstructs $\hat{V}^n$ as the sequence that is jointly typical[1] with S^n and lies inside the desired bin. We can interpret the above reconstruction process as a channel decoding procedure and S as the output of a hypothetical channel with input V. Therefore, for a sufficiently large n, V can be reconstructed with arbitrarily small error probability as long as the rate of transmission via this hypothetical channel is less than $I(V;S)$, or in other words, each bin can have maximally $\approx 2^{nI(V;S)}$ elements to have lossless reconstruction. Since the total number of typical sequence of V with length n is approximately $2^{nH(V)}$, the number of bins required is $2^{nH(V)}/2^{nI(V;S)} = 2^{nH(V|S)}$. Hence, we can compress V at a rate $H(V|S)$ with this random binning scheme.

Assume now side information S is also given to the encoder. For the instance when $S = s$, we can optimally compress V at rate $H(V|s)$ using classic source coding. Hence, the optimal average compression rate is $\sum_s H(V|s)p(s) = H(V|S)$. Comparing this rate with that obtained by random binning scheme in SWC, we can draw

[1] This may involve joint typicality of a continuous random variable and a discrete random variable since S may be continuous. However, such joint typicality can be easily obtained by generalizing the classic case.

two important conclusions. First, the random binning scheme must attain maximum possible compression since it cannot outperform the optimal scheme in the better equipped setup when side information is also provided to the encoder. Second, contrary to the fact that WZC setup has rate loss in general (see Section II), SWC setup has no rate loss comparing with this better equipped setup when side information is also given to the encoder.

b. Structure Binning

Unfortunately, the random binning scheme is not friendly to implement. The main difficulty is to assign a random binning that yet can facilitate decoding with low computational complexity. However, a more detail observation of our previous discuss concludes that purely random assignment of codewords is not necessary; it is more important instead to have each bin to behave like a good channel code so as to approach the hypothetical channel capacity $I(V;S)$.

An interesting approach that was first suggested by Wyner [73] and was rediscovered and first implemented by Pradhan and Ramchandran [8] is to use an arbitrary linear channel code to partition the set of all v^n's into cosets or bins with different syndromes. Since all cosets of a linear channel code share the same distance properties, all bins (cosets) now are indeed good channel code as desired provided that the linear channel code itself is good. Note that the syndrome now acts as the bin index to be transmitted at the encoder. Hence, for the (n,k)-channel code and thus with a $(n-k)\times n$ parity matrix $\mathbf{H}$, Slepian-Wolf encoding is to compute and output the syndrome

$$w^{n-k} = v^n \mathbf{H}^T.$$

Since the length of the syndrome is $n-k$, the compression ratio is $n : n-k$.

In order to perform Slepian-Wolf decoding, channel decoding is modified in such a way that $\hat{v}^n$ is reconstructed as a code vector inside the coset with the desired syndrome instead of a codeword of the channel code. More precisely, receiving s^n, the decoder should select from the bin that maximize the a posteriori probability, i.e.,

$$\hat{v}^n = \arg \max_{v \in \{v' | w^m = v'^n \mathbf{H}^T\}} p(v^n | s^n). \tag{3.65}$$

c. LDPC Code Based Slepian-Wolf Coding

The low-density parity-check (LDPC) code [74] is suitable for implementing SWC because: first, the LDPC code is capacity approaching with good performance; second, it allows flexible code designs to adapt any kind of channel [75, 76, 77]. The second benefit is especially appealing since the hypothetical channel between V and S can be weird in the sense of conventional channel coding.

A LDPC code is a linear block code. As the name suggested, the parity check matrix is sparse such that the number of non-zero elements in the parity matrix is relatively small. A LDPC code is best represented using a Tanner graph [78]. As an example, the Tanner graph of a binary (6,2)-LDPC code is shown in Fig. 12. The circles in the left are called the variable nodes and the squares on the right are called the check nodes. Each check node corresponds to a parity check equation of the LDPC code. A valid codeword of the LDPC code needs to have all parity checks equal to 0. The number of branch enumerated from a variable/check node is called the degree of that variable/check node. Note that each branch in a Tanner graph corresponds to a non-zero elements in the parity check matrix. Hence, the "low-density" property of LDPC codes translates to small average degrees of the variable and check nodes. If all variable nodes have the same degree and so are all the check nodes, then the LDPC code is called regular. Otherwise, the LDPC code is irregular.

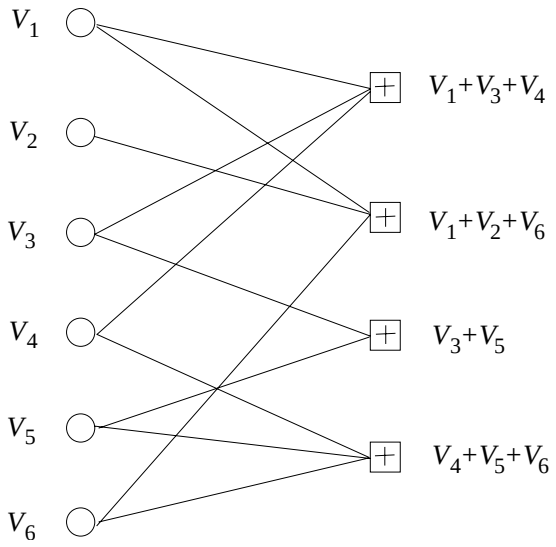


Fig. 12. The Tanner graph of a binary (6,2)-LDPC code.

To perform Slepian-Wolf encoding, the encoder computes and output the values of all check nodes, which are equivalent to the syndrome bits W^{n-k}. Given the side information S^n, the Slepian-Wolf decoder should reconstruct V^n as the best estimate out of all code vectors with syndrome W^{n-k}. MAP decoder (3.65) is optimum but is not realistic to implement for large code length n. Alternatively, a very good estimate $\hat{V}^n$ can be obtained using message-passing algorithm [79] as in conventional LDPC decoding.

As the name suggested, messages are exchanged between two ends of each branch in a message-passing algorithm. The message going into or out of a variable node possesses the "belief" of the value of that variable node. For binary LDPC codes, these messages are typically in the form of log-likelihood ratios (i.e., $\log \frac{p(\text{observation}|V_i=1)}{p(\text{observation}|V_i=-1)}$ for the messages passing into or out of the variable node V_i). Upon receiving the messages, both variable and check nodes update the messages by combining the beliefs

of the messages, and send the new messages to the other ends. To avoid the belief in a message is doubly counted, the message originated from the same branch is not included in the update (see Fig. 13).

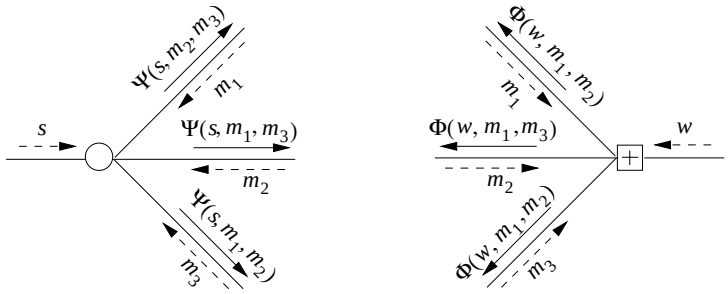


Fig. 13. Message updates of a variable node and a check node.

Recall that V, W, and S represent the value of a variable node, the value of a check node, and side information, respectively. Denote $\mathcal{V}, \mathcal{W}$, and $\mathcal{S}$ as their corresponding alphabets. Use $\mathfrak{m}$ to represent the value of a message and $\mathcal{M}$ to represent its alphabet. For a variable node i, denote the initial message mapping as $\Psi_i^{(0)} : \mathcal{S} \to \mathcal{M}$, the variable node message mapping as $\Psi_i : \mathcal{S} \times \mathcal{M}^{d_i - 1} \to \mathcal{M}$, and the final message mapping as $\Psi_i^{(f)} : \mathcal{S} \times \mathcal{M}^{d_i} \to \mathcal{M}$, where d_i is the degree of the variable node i and the final message mapping combine all received messages to facilitate estimation of the actual values of the variable node i. Similarly, for a check node j, denote the check node message mapping as $\Phi_j : \mathcal{W} \times \mathcal{M}^{d_j - 1} \to \mathcal{M}$, where d_j is the degree of the check node j. It is understood that the $d_i - 1$ ($d_j - 1$) input messages are those connected to the variable (check) node excluding to the message coming from the same branch as the output message. Now, the message-passing algorithm can be more precisely summarized as follows:

1. Initialization: for every variable node i, generate message using initial message

mapping $\Psi_i^{(0)}$ and pass it to every connected check node.

2. Loop:

   - Check node update: for every check node j and for every branch in that check node, update message using check node message mapping Φ_j and pass it back to the connected variable node.
   - Variable node update: for every variable node i and for every branch in that variable node, update message using variable node message mapping Ψ_i and pass it back to the connected check node.
   - Exit conditions: for every variable node i, use the final message mapping $\Psi_i^{(f)}$ to estimate the value of the variable node i. Exit if 1). the estimated variable nodes possess the desired syndrome, or 2). the maximum number of iterations is reached.

It is generally impossible to combine the beliefs of the messages exactly. However, if we assume all received message are independent of the others, then the mappings Ψ_i, $\Psi_i^{(f)}$, and Φ_j have relatively simple forms that [14]

$$\Psi_i(s, \mathfrak{m}_1, \mathfrak{m}_2, ..., \mathfrak{m}_{d_i-1}) = \log \frac{p(s|V_i = 1)}{p(s|V_i = -1)} + \sum_{i'=1}^{d_i-1} \mathfrak{m}_{i'} \tag{3.66}$$

$$\Psi_i^{(f)}(s, \mathfrak{m}_1, \mathfrak{m}_2, ..., \mathfrak{m}_{d_i}) = \log \frac{p(s|V_i = 1)}{p(s|V_i = -1)} + \sum_{i'=1}^{d_i} \mathfrak{m}_{i'}, \tag{3.67}$$

and

$$\Phi_j(w, \mathfrak{m}_1, \mathfrak{m}_2, ..., \mathfrak{m}_{d_j-1}) = 2 \operatorname{atanh} \left(w \prod_{i=1}^{d_j-1} \tanh \left(\frac{\mathfrak{m}_i}{2} \right) \right), \tag{3.68}$$

when the messages are in the form of log-likelihood ratios that

$$\Psi_i^{(0)}(s) = \log \frac{p(s|V_i = 1)}{p(s|V_i = -1)}.$$

The resulting message-passing algorithm is commonly known as the belief propagation algorithm [80] or the sum-product decoding algorithm [81]. When the information of a message passes back to itself, the assumption that the received messages are independent will obviously fail. This happens if there exists cycles in the Tanner graph and the number of iterations is larger than or equal to half of the length of the shortest cycle. For long block length n and small average node degree (low-density), the average length of cycles is large and the belief-propagation algorithm has good performance.

For LDPC coding in conventional channel coding, the decoding error probability is independent of the transmitting codeword provided that the channel satisfies certain symmetry condition. Hence, we can assume any codeword to be sent when we analyze the LDPC code performance. In specific, by assuming all-one codeword is sent and by tracking the density distribution of the average beliefs of the variable nodes, we could estimate the probability of decoding error after any number of iterations in theory. However, this cannot be easily done for a specific LDPC code since each variable/check node can have different degree. Nonetheless, if we consider an ensemble of codes which bear the same degree profile in the sense that the fraction of nodes with any particular degree is the same, then the problem become tractable and this technique is commonly known as density evolution. Density evolution can be employed for LDPC code design. The basic idea is to adjust the degree profile interactively such that the decoding error probability predicted by density evolution is smallest.

3. Practical LDPC Code Design for Slepian-Wolf Coding

We compress $\mathbf{S}$ using multi-level Slepian-Wolf coding with $\mathbf{Y}$ as the decoder side information. Denote J $(0 \leq J \leq N-1)$ as the index of $\mathbf{S}$ and write J as $B_m B_{m-1} ... B_2 B_1$ in its binary representation, where B_m is the most significant bit of J, and B_1 the

least significant bit. At first B_1 is compressed using the first Slepian-Wolf code to rate $R_1 = H(B_1|\mathbf{Y})$, then B_2 is compressed with the second Slepian-Wolf code to rate $R_2 = H(B_2|\mathbf{Y}, B_1)$, and so on. Finally B_m is compressed with the m-th Slepian-Wolf code to rate $R_m = H(B_m|\mathbf{Y}, B_1, ...B_{m-1})$. By the chain rule, $R_1 + R_2 + \cdots + R_m = H(J|\mathbf{Y}) = H(\mathbf{S}|\mathbf{Y})$. The rate per dimension is $R = \frac{1}{n}(R_1+R_2+\cdots+R_m) = \frac{1}{n}H(\mathbf{S}|\mathbf{Y})$.

By splitting $\mathbf{S}$ into multiple bit planes, well-studied binary channel codes can be used to implement Slepian-Wolf coding of each of them. The idea is to treat B_i ($1 \leq i \leq m$) as an input into some "hypothetical" channel with output $(\mathbf{Y}, B_1, B_2, \cdots, B_{i-1})$. At the encoder, the syndrome of the designed channel code for a sequence of realizations of B_i is computed and passed to the decoder. Therefore, the compression rate is one minus the channel code rate and thus the ideal code rate is $1 - R_i = 1 - H(B_i|\mathbf{Y}, B_1, \cdots, B_{i-1})$. Decoding the i-th bit plane B_i is similar to conventional channel decoding except one thing: instead of decoding into a channel codeword, the decoder estimates the sequence of realizations of B_i using the received syndrome in conjunction with $(\mathbf{Y}, B_1, B_2, \cdots, B_{i-1})$. Fig. 14 depicts SWC-NQ with multi-level Slepian-Wolf coding.

In practice, LDPC codes have been used to implement the Slepian-Wolf codes for their near-capacity performance [26]. To design the Slepian-Wolf code for the i-th bit plane B_i, it is essential to estimate the desired channel code rate, i.e., $1 - H(B_i|\mathbf{Y}, B_1, \cdots, B_i - 1)$, via gathering the statistics for the hypothetical channel with input B_i and output $(\mathbf{Y}, B_1, \cdots, B_i - 1)$. In the following, we illustrate this Slepian-Wolf code design process, starting from estimating $Pr(B_1 = b_1, \cdots, B_i = b_i|\mathbf{Y} = \mathbf{y})$, for all $\mathbf{y}$ and $1 \leq i \leq m$.

Because B_i is a deterministic function of $\mathbf{X}$, with a slight abuse of notation, we

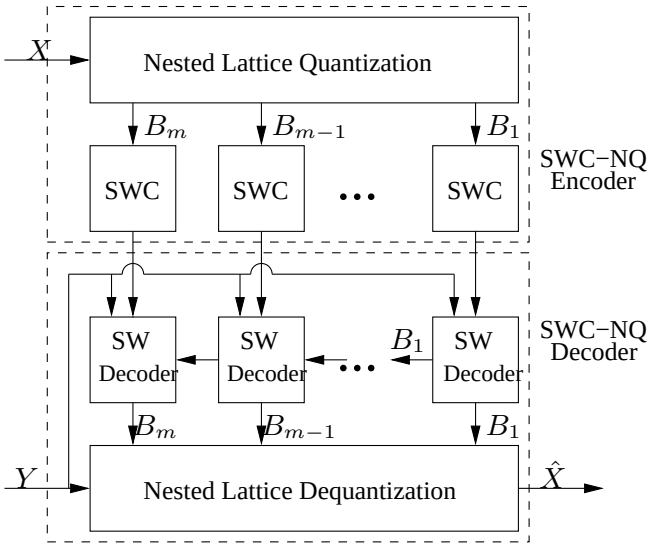


Fig. 14. SWC-NQ with multi-level Slepian-Wolf coding.

define $b_i(\cdot)$ as this function. In other words, $B_i = b_i(\mathbf{X})$. Then we have

$$\begin{aligned} & Pr(B_1 = b_1, \cdots, B_i = b_i | \mathbf{Y} = \mathbf{y}) \\ = & \int_{b_1(\mathbf{x})=b_1, \cdots, b_i(\mathbf{x})=b_i} f(\mathbf{x}|\mathbf{y}) d\mathbf{x} \qquad (3.69) \\ = & \int_{b_1(\mathbf{z}+\mathbf{y})=b_1, \cdots, b_i(\mathbf{z}+\mathbf{y})=b_i} f(\mathbf{z}) d\mathbf{z}. \qquad (3.70) \end{aligned}$$

In the one-dimensional case (with $n = 1$), the integration interval $\{z | b_1(z + y) = b_1, \cdots, b_i(z+y) = b_i\}$ in (3.70) corresponds to a union of infinite number of intervals. However, since $f(z)$ decays exponentially from the origin, we can approximate the integral, thus $Pr(B_1 = b_1, \cdots, B_i = b_i | \mathbf{Y} = \mathbf{y})$, accurately as a sum of a few Gaussian tail probabilities.

In higher dimensional cases (with $n \geq 2$), however, $Pr(B_1 = b_1, \cdots, B_i = b_i | \mathbf{Y} = \mathbf{y})$ cannot be obtained analytically. For example, the integration region $\{\mathbf{z} | b_1(\mathbf{z} + \mathbf{y}) = b_1, \cdots, b_i((\mathbf{z} + \mathbf{y}) = b_i\}$ in the two-dimensional case corresponds to a union of hexagons and hence no simple analytical solution can be found. Although

numerical integration may be used to evaluate (3.70), we use Monte Carlo simulations because they are more flexible. Specifically, we first quantize the real axis into K intervals and partition all possible values of $\mathbf{Y}$ into K^n regions $\mathbb{Y}_j$ for $j = 1, \cdots, K^n$. Denote $\mathbb{Y}(\mathbf{y})$ as the region that contains $\mathbf{y}$ and define $1(\cdot)$ as an indicator function that equals to one if its argument is true and zero otherwise. We then draw L independent samples of $\mathbf{Y}' \sim N(\mathbf{0}, \sigma_Y^2 \mathbf{I}_n)$ and $\mathbf{Z}' \sim N(\mathbf{0}, \sigma_Z^2 \mathbf{I}_n)$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix, and let the l-th samples be $\bar{\mathbf{y}}_l$ and $\bar{\mathbf{z}z}_l$, respectively. Then $\bar{\mathbf{x}}_l = \bar{\mathbf{y}}_l + \bar{\mathbf{z}}_l, 1 \le l \le L$, will be statistically equivalent to samples of $\mathbf{X}$. Finally, by simple number counting, we approximate the probability $Pr(B_1 = b_1, \cdots, B_i = b_i | \mathbf{Y} = \mathbf{y})$ in (3.70) as

$$\begin{aligned} & Pr(B_1 = b_1, \cdots, B_i = b_i | \mathbf{Y} = \mathbf{y}) \\ \approx\ & Pr(B_1 = b_1, \cdots, B_i = b_i | \mathbf{Y} \in \mathbb{Y}(\mathbf{y})) \\ \approx\ & \frac{\sum_{l=1}^{L} 1(\bar{\mathbf{y}}_l \in \mathbb{Y}(\mathbf{y}), b_1(\bar{\mathbf{x}}_l) = b_1, \cdots, b_i(\bar{\mathbf{x}}_l) = b_i)}{\sum_{l=1}^{L} 1(\bar{\mathbf{y}}_l \in \mathbb{Y}(\mathbf{y}))}. \end{aligned} \tag{3.71}$$

a. Desired Channel Code Rate Computation

Armed with (3.71), we estimate the LDPC code rate for the i-th bit plane B_i as

$$\begin{aligned} & 1 - H(B_i | \mathbf{Y}, B_1, \ldots B_{i-1}) \\ =& 1 - \sum_{j=1}^{K^n} \sum_{\substack{b_1 \in \{0,1\} \\ \cdots \\ b_{i-1} \in \{0,1\}}} Pr(\mathbf{Y} \in \mathbb{Y}_j, B_1 = b_1, \cdots, B_{i-1} = b_{i-1}) \mathcal{H}(Pr(B_i = 1 | \mathbf{Y} \in \mathbb{Y}_j, B_1 = b_1, \cdots, B_{i-1} = b_{i-1})) \\ =& 1 - \sum_{j=\mathbf{b}_1 \in \{0,1\}}^{K^n} \sum_{\substack{\cdots \\ b_{i-1} \in \{0,1\}}} Pr(\mathbf{Y} \in \mathbb{Y}_j, B_1 = b_1, \cdots, B_{i-1} = b_{i-1}) \mathcal{H}\left(\frac{Pr(B_1 = b_1, \cdots, B_i = 1 | \mathbf{Y} \in \mathbb{Y}_j)}{Pr(B_1 = b_1, \cdots, B_{i-1} = b_{i-1} | \mathbf{Y} \in \mathbb{Y}_j)} \right), \end{aligned}$$

where $Pr(B_1 = b_1, \cdots, B_i = 1 | \mathbf{Y} \in \mathbb{Y}_j)$ and $Pr(B_1 = b_1, \cdots, B_{i-1} = b_{i-1} | \mathbf{Y} \in \mathbb{Y}_j)$ are obtained directly from (3.71), $\mathcal{H}(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$, and $Pr(\mathbf{Y} \in \mathbb{Y}_j, B_1 = b_1, \cdots, B_{i-1} = b_{i-1})$ estimated by using similar Monte Carlo simulations as

in (3.71) with

$$Pr(\mathbf{Y} \in \mathbb{Y}_j, B_1 = b_1, \cdots, B_{i-1} = b_{i-1}) = \frac{1}{L} \sum_{l=1}^{L} 1(\bar{\mathbf{y}}_l \in \mathbb{Y}_j, b_1(\bar{\mathbf{x}}_l) = b_1, \cdots, b_{i-1}(\bar{\mathbf{x}}_l) = b_{i-1}). \tag{3.72}$$

b. Channel Estimation

Recall that at the i-th bit plane, Slepian-Wolf decoding can be viewed as channel decoding over a hypothetical channel with input B_i and output $(\mathbf{Y}, B_1, B_2, \cdots, B_{i-1})$. Assuming that $\mathbf{Y} = \mathbf{y}, B_1 = b_1, \cdots, B_{i-1} = b_{i-1}$, the channel statistics, which is needed for decoding B_i, is entirely captured by the log-likelihood ratio (LLR)

$$\begin{aligned}
\mathrm{L_{ch}} &= \log \frac{P_r(\mathbf{Y} = \mathbf{y}, B_1 = b_1, \cdots, B_{i-1} = b_{i-1} | B_i = 1)}{P_r(\mathbf{Y} = \mathbf{y}, B_1 = b_1, \cdots, B_{i-1} = b_{i-1} | B_i = 0)} \\
&= \log \frac{P_r(\mathbf{Y} = \mathbf{y}, B_1 = b_1, \cdots, B_{i-1} = b_{i-1}, B_i = 1)}{P_r(\mathbf{Y} = \mathbf{y}, B_1 = b_1, \cdots, B_{i-1} = b_{i-1}, B_i = 0)} - \log \frac{P_r(B_i = 1)}{P_r(B_i = 0)} \\
&= \log \frac{P_r(B_1 = b_1, \cdots, B_{i-1} = b_{i-1}, B_i = 1 | \mathbf{Y} = \mathbf{y})}{P_r(B_1 = b_1, \cdots, B_{i-1} = b_{i-1}, B_i = 0 | \mathbf{Y} = \mathbf{y})} - \log \frac{P_r(B_i = 1)}{P_r(B_i = 0)} \\
&\triangleq \log \frac{P_r(B_1 = b_1, \cdots, B_{i-1} = b_{i-1}, B_i = 1 | \mathbf{Y} = \mathbf{y})}{P_r(B_1 = b_1, \cdots, B_{i-1} = b_{i-1}, B_i = 0 | \mathbf{Y} = \mathbf{y})} - \mathrm{L_{ap}},
\end{aligned}$$

where $\mathrm{L_{ap}} = \log \frac{P_r(B_i=1)}{P_r(B_i=0)}$ is the *a priori* LLR, which can be estimated as $\log \frac{\sum_{l=1}^{L} 1(B_i(\bar{\mathbf{x}}_l)=1)}{\sum_{l=1}^{L} 1(B_i(\bar{\mathbf{x}}_l)=0)}$.

In practice, $\mathrm{L_{ch}}$ is used to design the LDPC code degree profiles with the help of density evolution [82, 77]. It also served as an initial estimate during decoding. $\mathrm{L_{ch}}$ is updated after each decoding iteration. After the final iteration, it is added to the *a priori* LLR $\mathrm{L_{ap}}$ for the estimation of B_i. Finally, the estimate $\hat{b}_i$ will be one if the sum is positive and zero otherwise.

4. Simulation Results

We carry out one-dimensional nested lattice/scalar quantizer design when X and Y are jointly Gaussian with $X = Y + Z$. Fig. 15 shows results with nested lattice

quantization alone and SWC-NQ with $\sigma_Y^2 = 1.0$ and $\sigma_Z^2 = 0.01$, One-dimensional nested lattice quantization exhibits a 3.95-9.60 dB performance gap from $D_{WZ}(R)$ for R =1.0-7.0 b/s, which agrees with the lower bound given in Corollary 2 at high rate. We observe that the gap between our simulation results with ideal Slepian-Wolf coding (with rate computed as $H(\mathbf{s}|\mathbf{Y})$) and $D_{WZ}(R)$ is indeed 1.53 dB at high rate. With practical Slepian-Wolf coding based on irregular LDPC codes of length 10^6 bits, this gap is 1.66-1.80 dB for R =0.93-5.0 b/s.

For two-dimensional nested lattice quantization, we use the A_2 hexagonal lattices. Fig. 16 shows results with nested lattice quantization alone and SWC-NQ with $\sigma_Y^2 = 1$, $\sigma_Z^2 = 0.01$, and Table (II) shows the conditional entropies and practical LDPC code rates for each bit plane when the nesting ratio $n = 7$ and 31, along with the profiles of the practical LDPC codes. Two-dimensional nested lattice quantization exhibits a 4.06-8.48 dB performance gap from $D_{WZ}(R)$ for R =1.40-5.0 b/s, again in agreement with the lower bound given in Corollary 2 at high rate. We observe that the gap between our results with ideal Slepian-Wolf coding and $D_{WZ}(R)$ is 1.36 dB at high rate. With practical Slepian-Wolf coding based on irregular LDPC codes (of length 10^6 bits), this gap is 1.67-1.84 dB for R=0.95-3.29 b/s.

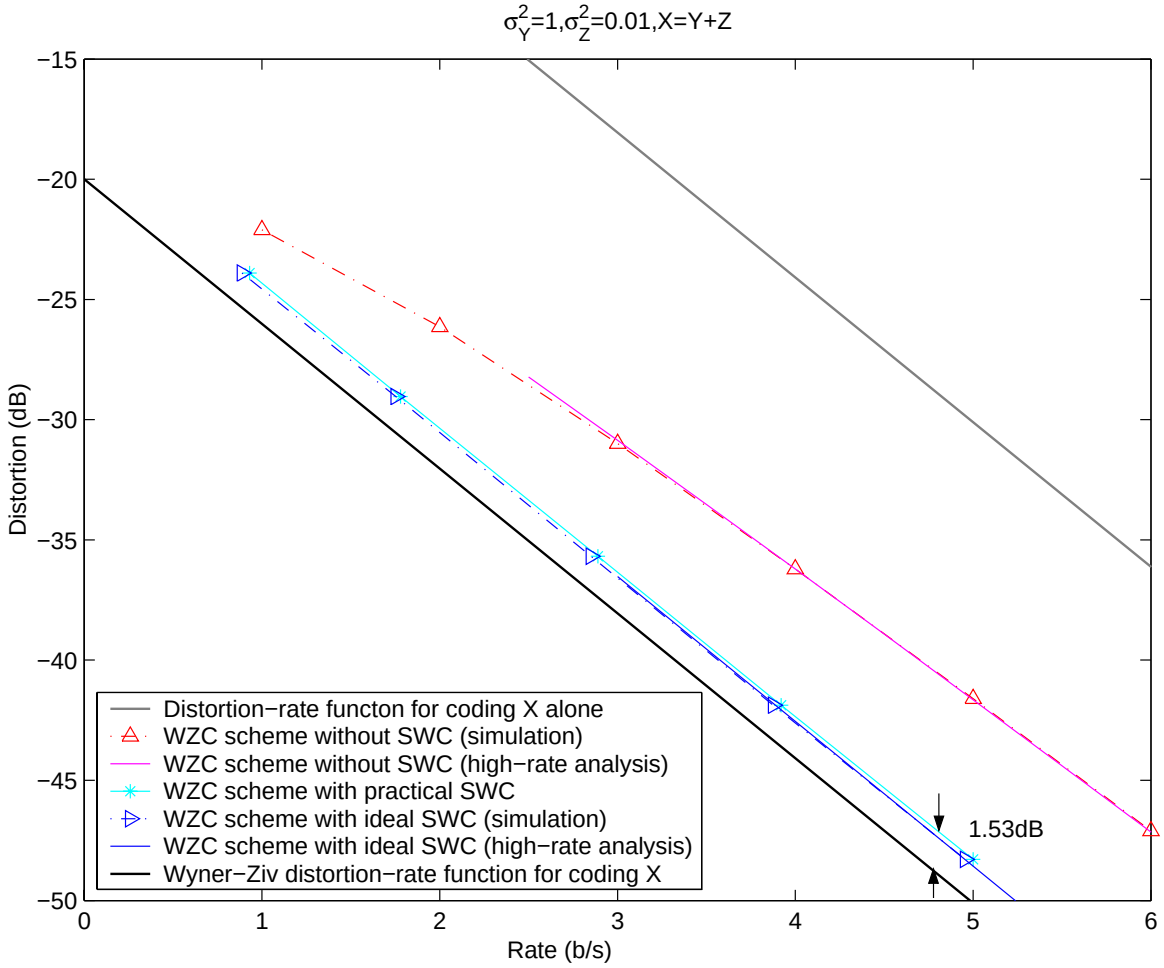


Fig. 15. Results based on one-dimensional nested lattice quantization with and without SWC.

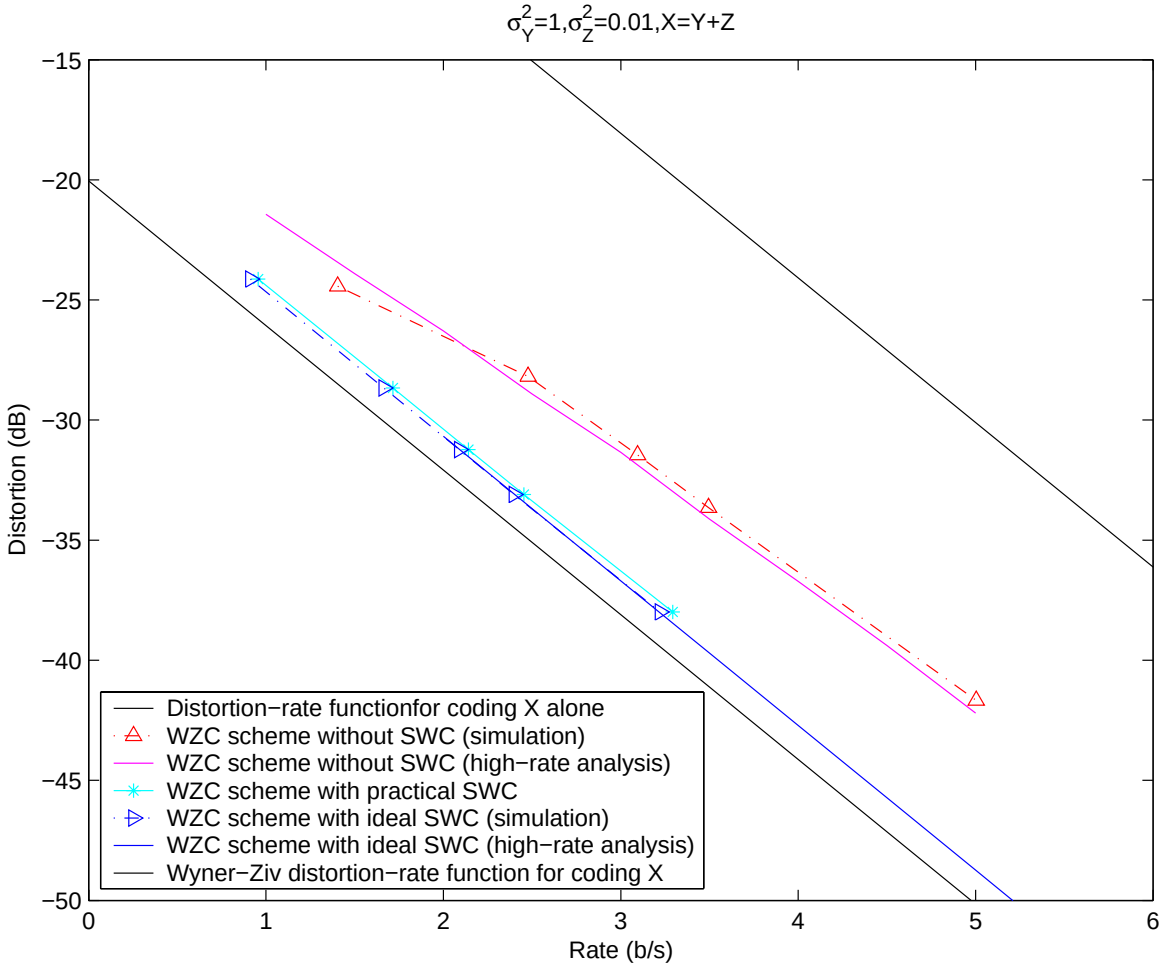


Fig. 16. Results based on two-dimensional nested A_2 lattice quantization with and without SWC.

Table II. The conditional entropy, LDPC code rate, and the corresponding degree distribution polynomials $\lambda(x)$ and $\rho(x)$ for each bit plane of two-dimensional Slepian-wolf coded nested lattice quantization with nesting ratio (a) N=7 and (b) N=31.

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
1	0.7760298	0.204	$0.142139x + 0.250041x^2 + 0.027853x^7 + 0.116382x^8 + 0.463585x^{39}$	$0.1x^5 + 0.9x^6$
2	0.6187541	0.342	$0.188401x + 0.432910x^2 + 0.107433x^{11} + 0.271255x^{12}$	$0.3x^4 + 0.7x^5$
3	0.4307767	0.494	$0.154196x + 0.607571x^2 + 0.191598x^8 + 0.046635x^9$	$0.5x^5 + 0.5x^6$

(a)

Bit plane	Conditional entropy	LDPC code rate	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
1	0.7536632	0.218	$0.255287x + 0.418496x^2 + 0.300630x^{11} + 0.025587x^{12}$	$0.6x^3 + 0.4x^4$
2	0.7097249	0.255	$0.229576x + 0.416782x^2 + 0.142361x^{11} + 0.211281x^{12}$	$0.2x^3 + 0.8x^4$
3	0.7903277	0.178	$0.149076x + 0.247592x^2 + 0.039490x^7 + 0.108175x^8 + 0.455668x^{39}$	$0.4x^5 + 0.6x^6$
4	0.4695057	0.502	$0.147216x + 0.603324x^2 + 0.122048x^8 + 0.127413x^9$	$0.3x^5 + 0.7x^6$
5	0.6043050	0.349	$0.193289x + 0.450023x^2 + 0.356688x^{11}$	$0.4x^4 + 0.6x^5$

(b)

CHAPTER IV

COMPRESS-FORWARD FOR COOPERATIVE NETWORKS USING PRACTICAL WYNER-ZIV CODING

In this Chapter, we applied our Wyner-Ziv code design in cooperative networks as practical implementations of the cooperative strategies. In cooperative networks, relaying is an essential component to gain the cooperative diversity. Cover and El Gamal derived the tightest bounds on the capacity of the relay channel using random coding and suggested two coding strategies, namely, decode-forward (DF) and compress-forward (CF), to provide the best known lower bound of the achievable rate region. Several practical code designs have been proposed in recent literature to approach the lower bound. Most of these designs only exploit DF. Following the latest development in practical distributed source-channel coding, we studied CF coding with BPSK modulation for the relay channel. In CF scheme, Wyner-Ziv coding is applied at the relay to exploit the joint statistics between signals at the relay and the destination.

In the following, we will give an overview of half-duplex relay channel in Section A. Then we introduce the CF relaying with BPSK modulation in Section B. In order to get the WZC performance of CF relaying and fill the gap between the theory and practical CF code design, Section C studies WZC for Gaussian mixture signals with BPSK modulation and proposed a lower and an upper bounds. Then DJSCC is addressed in Section D. Section E details our practical code design. Section F presents simulation results.

A. The Channel Model, Capacity Bounds, and Achievable Rates for Half-duplex Gaussian Relay Channel

1. The Channel Model

A simple three-node relay channel is shown in Fig. 17, where c_{sd}, c_{sr}, and c_{rd} denote the channel gains/coefficients of the links from the source to destination, source to relay, and relay to destination, respectively. In this chapter, we focus on the Gaussian relay channel where the channel coefficients are fixed and the noises are Gaussian, and the term "relay channel" thereafter refers to Gaussian relay channel.

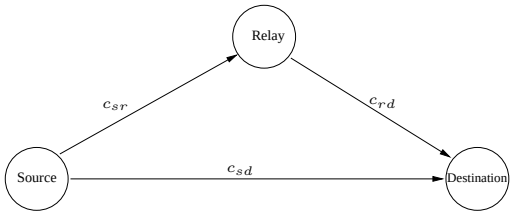


Fig. 17. The relay channel with three nodes: the source, the relay and the destination.

In time-division half-duplex relaying, the message m at the source is split into two non-overlapping parts m_1 and m_2. Then, m_1 is encoded into the $n\alpha$-length codeword $x_{s1}(m_1)$ as the first block and m_2 into the $n(1-\alpha)$-length codeword $x_{s2}(m_2)$ as the second block. At the frame level, the time interval T is divided into the relay-receive period T_1 and the relay-transmit period T_2 with $T = T_1 + T_2$. During the relay-receive period, the received signals at the relay and the destination are

$$y_r[i] = c_{sr}x_{s1}(m_1)[i] + z_r[i], \tag{4.1}$$

$$y_{d1}[i] = c_{sd}x_{s1}(m_1)[i] + z_{d1}[i], \quad i = 1, \ldots, n\alpha, \tag{4.2}$$

respectively, where z_r and z_{d1} are white Gaussian noises with unit power. During the relay-transmit period, the relay and the source, respectively, send the $n(1-\alpha)$-length

codewords $x_r(m_1)$ and $x_{s2}(m_2)$ to the destination, which receives

$$y_{d2}[i] = c_{rd}x_r(m_1)[i] + c_{sd}x_{s2}(m_2)[i] + z_{d2}[i], \quad i = 1, \ldots, n(1-\alpha), \tag{4.3}$$

where z_{d2} is again white Gaussian noise with unit power.

When the relay channel works in the asynchronous mode, the phases of the channel coefficients are unknown to all nodes, and the receivers know only the magnitudes of all channel coefficients. Thus, c_{sr}^2, c_{sd}^2, and c_{rd}^2 characterize the channel conditions. For Gaussian relay channel, c_{sr}^2, c_{sd}^2, and c_{rd}^2 are constant values when the geographical locations of all the nodes are fixed. In this chapter we focus on the Gaussian relay channel, therefore the term "relay channel" refers to Gaussian relay channel in general.

Throughout this chapter, we use the same setup as in [43] to examine the rate bounds and code design for the CF relaying with BPSK modulation, where the relay is located along a straight line from the source to the destination, which are 10 m apart, as shown in Fig. 18. The channel coefficient of the link from sender i to the receiver j (sender i could be the source or relay, and receiver j could be the relay or destination) is $c_{ij}^2 = K_o d_{ij}^{-n}$ [43], where d_{ij} is the distance from the sender i to the receiver j, n is the path loss coefficient, $K_o = (c/4\pi d_o f_c)^2$, c is the light speed, d_o is the free-space reference distance, f_c is the transmission frequency. The experimental setup is fixed with $f_c = 2.4$ GHz carrier frequency, path loss coefficient $n = 3$, and free-space reference distance $d_o = 1$ m. Therefore the channel coefficients are: $c_{sd}^2 = 10^{-7}$, $c_{sr}^2 = d^{-3} \times 10^{-4}$, and $c_{rd}^2 = (10-d)^{-3} \times 10^{-4}$. Note that c_{sd} is fixed and c_{sr} and c_{rd} are functions of d, thus for each particular d, there is a set of the coefficients.

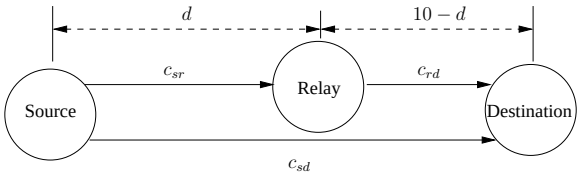


Fig. 18. The relay is located along the straight line between the source and destination.

2. Capacity Bound and Achievable Rates

A general upper bound on the capacity of the relay channel is given in [34] as

$$C \leq \max_{f(x_s,x_r)} \min\{I(X_s, X_r; Y_d), I(X_s; Y_r, Y_d|X_r)\}, \tag{4.4}$$

where X_s and X_r are signals transmitted from the source and the relay, respectively, Y_r and Y_d are signals received at the relay and the destination, respectively, and $f(x_s, x_r)$ is the joint probability density function (pdf) of random variable X_s and X_r. The lower bound, also being viewed as the achievable rate limit, is given in [34] as

$$C \geq \max_{f(x_s,x_r)} \min\{I(X_s, X_r; Y_d), I(X_s; Y_r|X_r)\}. \tag{4.5}$$

The upper bound on the capacity of the half-duplex relay channel is derived in [44, 47] from the max-flow-min-cut theorem [2] and given as

$$C_{ub} = \max_{0\leq\beta\leq1, 0\leq\alpha\leq1} \min\{C_{ub1}, C_{ub2}\}, \tag{4.6}$$

where

$$C_{ub1} = \frac{\alpha}{2}\log(1 + (c_{sr}^2 + c_{sd}^2)P_{s1}) + \frac{1-\alpha}{2}\log(1 + (1-\beta^2)c_{sd}^2 P_{s2}), \tag{4.7}$$

and

$$C_{ub2} = \frac{\alpha}{2}\log(1 + c_{sd}^2 P_{s1}) + \frac{1-\alpha}{2}\log(1 + c_{sd}^2 P_{s2} + c_{rd}^2 P_r + 2\sqrt{\beta^2 c_{sd}^2 c_{rd}^2 P_{s2} P_r}), \tag{4.8}$$

respectively. Note that the parameter β, which reflects the correlation between signals at the source and relay, can be written in closed form [44, 47].

In Fig. 19, for the setup with $P_r = P_s = 5$ dB, $c_{sd}^2 = 0$ dB, and $c_{rd}^2 = 10$ dB, we compare the upper capacity bound (4.6), the lower rate bound achievable with DF, the lower rate bound (4.14) achievable with CF, and the rate bound of multi-hop transmission, which is given by the minimum between the capacity of the source-relay link and that of the relay-destination link. All rates are relative to that of direct transmission and plotted as functions of c_{sr}^2 (increasing c_{sr}^2 is equivalent to decreasing the distance between the source and relay). It is seen that CF outperforms DF at low c_{sr}^2, i.e., when the relay is close to the destination, and that DF is worse than direct transmission when $c_{sr}^2 < c_{sd}^2 = 0$ dB. See [44] for more discussions under different transmission conditions.

The upper bound given by (4.7) and (4.8) is derived under the assumption that all signals being conveyed through the links are Gaussian. Unfortunately, this assumption is no longer true for practical CF implementations where all signals to be transmitted are BPSK modulated before they are sent out.Therefore the channel model is the binary-input AWGN memoryless channel, with its capacity given by [83] as

$$C(\text{snr}) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\tau^2/2}}{\sqrt{s\pi}} \log(1 + e^{-2\sqrt{\text{snr}}\tau - 2\text{snr}}) d\tau, \tag{4.9}$$

where snr is the signal-to-noise ratio.

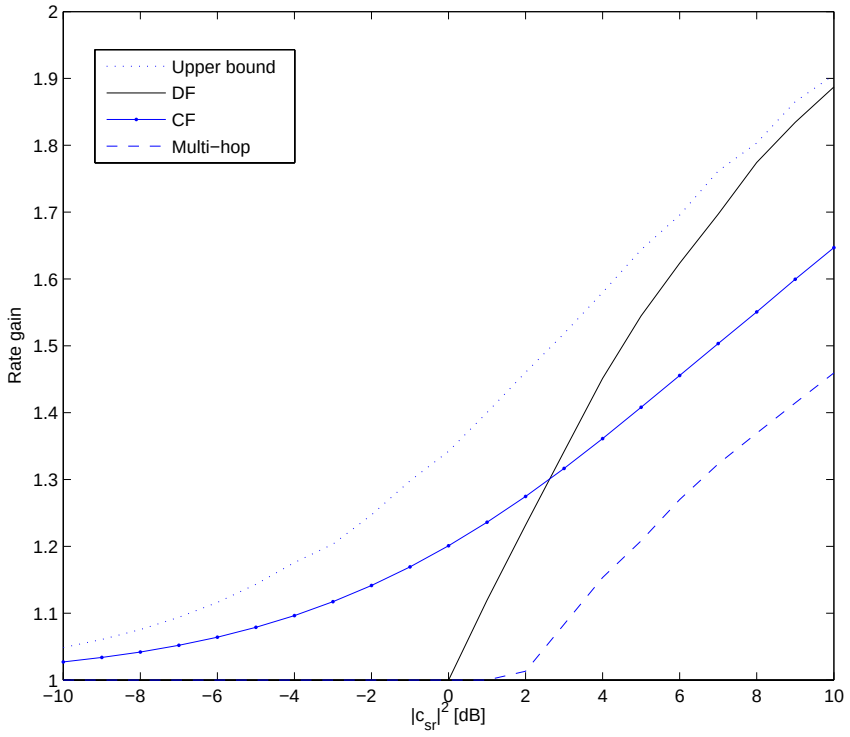


Fig. 19. The upper capacity bound (4.6), the lower rate bound achievable with DF, the lower rate bound (4.14) achievable with CF, and the rate bound of multi-hop transmission, as functions of c_{sr}^2 when $P_r = P_s = 5$ dB, $c_{sd}^2 = 0$ dB, and $c_{rd}^2 = 10$ dB for the Gaussian half-duplex relay channel. Each rate is plotted relative to that of direct transmission.

For the sake of simplicity, we call the relay channel where all signals to be transmitted are BPSK modulated before they are sent out as the relay channel with BPSK modulation. Then the upper bound on the capacity of the relay channel with BPSK modulation is

$$C_{ub}^{\mathrm{BPSK}} = \max_{0\le\beta\le1, 0\le\alpha\le1} \min\{C_{ub1}^{\mathrm{BPSK}}, C_{ub2}^{\mathrm{BPSK}}\}, \tag{4.10}$$

where

$$C_{ub1}^{\mathrm{BPSK}} = \alpha C((c_{sr}^2 + c_{sd}^2)P_{s1}) + (1-\alpha)C((1-\beta^2)c_{sd}^2 P_{s2}), \tag{4.11}$$

and

$$C_{ub2}^{\mathrm{BPSK}} = \alpha C(c_{sd}^2 P_{s1}) + (1-\alpha)C(c_{sd}^2 P_{s2} + c_{rd}^2 P_r + 2\sqrt{\beta^2 c_{sd}^2 c_{rd}^2 P_{s2} P_r}), \tag{4.12}$$

where $C(\cdot)$ is given in (4.9). The achievable rate of DF for the relay channel with BPSK modulation is derived from (4.5) as [44, 47]

$$R_{DF}^{\mathrm{BPSK}} \leq \max_{0\leq\rho\leq 1, 0\leq\alpha\leq 1} \min\{R_{DF1}^{\mathrm{BPSK}}, R_{DF2}^{\mathrm{BPSK}}\}, \tag{4.13}$$

where $R_{DF1}^{\mathrm{BPSK}} = \alpha C(c_{sd}^2 P_{s1}) + (1-\alpha)C[(1-\beta^2)c_{sd}^2 P_{s2}]$ and $R_{DF2}^{\mathrm{BPSK}} = C_{ub2}^{\mathrm{BPSK}}$.

The achievable rate of CF for the relay channel with BPSK modulation is derived as (similar to [44])

$$R_{CF}^{\mathrm{BPSK}} \leq \max_{0\leq\alpha\leq 1} \{R_r^{\mathrm{BPSK}}(\alpha) + R_d^{\mathrm{BPSK}}(\alpha)\}, \tag{4.14}$$

where

$$R_r^{\mathrm{BPSK}}(\alpha) = \alpha C\left(c_{sd}^2 P_{s1} + \frac{c_{sr}^2 P_{s1}}{1 + D_{\mathrm{WZ}}}\right), \tag{4.15}$$

$$R_d^{\mathrm{BPSK}}(\alpha) = (1-\alpha)C(c_{sd}^2 P_{s2}), \tag{4.16}$$

with D_{WZ} being the distortion of the Wyner-Ziv coding for two correlated signals driven by one BPSK modulated source signal. Similarly, we name this problem as WZC with BPSK modulation. The theoretical performance of WZC with BPSK modulation is not known yet. In the following section, we will examine the rate-distortion function of WZC with BPSK modulation, and get the achievable rate of CF for the relay channel with BPSK modulation.

B. CF Relaying with BPSK Modulation

Fig. 20 depicts the overall CF coding scheme. During T_1, message m_1 is channel coded and then BPSK modulated into αn binary symbols $X_{s1}[1], \ldots, X_{s1}[\alpha n]$ with $\frac{1}{\alpha n}\sum_{i=1}^{\alpha n} X_{s1}[i]^2 \leq P_{s1}$ and broadcasted to the relay and the destination. The received versions are $Y_r = c_{sr}X_{s1} + Z_r$ at the relay and $Y_{d1} = c_{sd}X_{s1} + Z_{d1}$ at the destination. We thus have a broadcast channel in the relay-receive period.

During T_2, Y_r is compressed into S using Wyner-Ziv encoder by treating Y_{d1} at the destination as the decoder side information. Then, the relay encodes S into binary channel codeword X_r of length $(1-\alpha)n$ with $\frac{1}{(1-\alpha)n}\sum_{i=1}^{(1-\alpha)n} X_r[i]^2 \leq P_r$ and sends it to the destination. At the same time, the source encodes m_2 into $(1-\alpha)n$ binary symbols $X_{s2}[1],\ldots, X_{s2}[(1-\alpha)n]$ with $\frac{1}{(1-\alpha)n}\sum_{i=1}^{(1-\alpha)n} X_{s2}[i]^2 \leq P_{s2}$ and sends them to the destination as well. The signal received at the destination from the source and relay is $Y_{d2} = c_{rd}X_r + c_{sd}X_{s2} + Z_{d2}$. We hence have a multiple-access channel (MAC) in the relay-receive period.

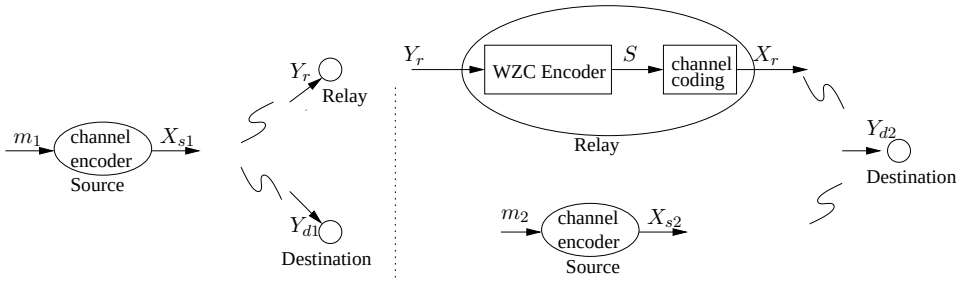


Fig. 20. The CF coding scheme for half-duplex relaying based on WZC.

At the destination, m_1 and m_2 are recovered sequentially. First, Y_r is reconstructed into Y_r' via Wyner-Ziv decoding at the destination (with the help of the side information Y_{d1}), yielding an average distortion of $D_{WZ}(R)$, where R is the WZC rate. According to [2, 1], we can write $Y_r' = Y_r + N$, where N is the quantization noise

with its variance as the Wyner-Ziv distortion limit $D_{WZ}(R)$, and the corresponding WZC rate R is the capacity of the link between the relay and the destination (with both $c_{sd}X_{s2}$ and Z_d being treated as noise) and given by

$$R = \frac{1-\alpha}{\alpha} C\left(\frac{c_{rd}^2 P_r}{1+c_{sd}^2 P_{s2}}\right), \tag{4.17}$$

where the normalization factor $\frac{1-\alpha}{\alpha}$ is due to half-duplex relaying.

With both $Y_r' = Y_r + N = c_{sr}X_{s1} + Z_r + N$ and $Y_{d1} = c_{sd}X_{s1} + Z_{d1}$ available at the destination as corrupted versions of X_{s1}, we can recover m_1 with the information provided by Y_r' and Y_{d1} jointly. The joint log-likelihood-ratio (LLR) is

$$\begin{aligned} L_{ch}(m_1|y_r', y_{d1}) &= \log\left(\frac{f(m_1=0|y_r', y_{d1})}{f(m_1=1|y_r', y_{d1})}\right) = \log\left(\frac{f(y_r', y_{d1}|m_1=0)P(m_1=0)}{f(y_r', y_{d1}|m_1=1)P(m_1=1)}\right) \\ &= -2\sqrt{P_{s1}}\left(\frac{c_{sr}}{1+D_{WZ}(R)} y_r' + c_{sd}y_{d1}\right) + \log\frac{P(m_1=0)}{P(m_1=1)} \\ &= -2\sqrt{P_{s1}}\tilde{y} + \log\frac{P(m_1=0)}{P(m_1=1)}, \end{aligned} \tag{4.18}$$

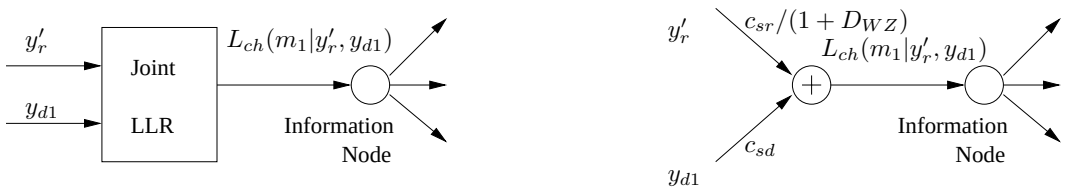
where $\tilde{y} = \frac{c_{sr}}{1+D_{WZ}(R)} y_r' + c_{sd}y_{d1}$. Assume $P(m_1=0) = P(m_1=1) = 0.5$, then $L_{ch}(m_1|y_r', y_{d1}) = -2\sqrt{P_{s1}}\tilde{y}$ which is the LLR from the binary-input AWGN channel whose output is $\tilde{Y}$ with modulation power as P_{s1} and unit noise variance. Therefore this LLR is equivalent to the LLR of the combination of y_{d1} and y_r' with the same coefficients as maximum ratio combining (MRC) [84, 85]. Then we can decode m_1 using joint decoding similar to MRC with rate

$$R_r^{\mathrm{BPSK}}(\alpha) \le \alpha C\left(c_{sd}^2 P_{s1} + \frac{c_{sr}^2 P_{s1}}{1+D_{\mathrm{WZ}}(R)}\right). \tag{4.19}$$

The equivalence of joint LLR calculation and MRC decoding is illustrated in Fig. 21.

Once m_1 is recovered, X_r can be reconstructed and $c_{rd}X_r$ eliminated from $Y_{d2} = c_{rd}X_r + c_{sd}X_{s2} + Z_{d2}$. Then, m_2 can be decoded with rate $R_d^{\mathrm{BPSK}}(\alpha) = (1-\alpha)C\big(c_{sd}^2 P_{s2}\big)$.

Consequently, the overall achievable rate of CF for the half-duplex relay channel



(a) Joint LLR Calculation (b) MRC Decoding

Fig. 21. The comparison of (a) joint LLR calculation and (b) MRC decoding. They are shown to be equivalent according to (4.18).

with specific α is

$$\begin{aligned} R_{\mathrm{CF}}^{\mathrm{BPSK}}(\alpha) &= R_r^{\mathrm{BPSK}}(\alpha) + R_d^{\mathrm{BPSK}}(\alpha) \\ &\leq \alpha C\left(c_{sd}^2 P_{s1} + \frac{c_{sr}^2 P_{s1}}{1 + D_{\mathrm{WZ}}(R)}\right) + (1-\alpha) C\big(c_{sd}^2 P_{s2}\big), \end{aligned} \tag{4.20}$$

Therefore, the bound for the achievable rate with CF can be written as

$$\begin{aligned} R_{\mathrm{CF}}^{\mathrm{BPSK}} &\leq \max_{0\leq\alpha\leq 1} R_{\mathrm{CF}}^{\mathrm{BPSK}}(\alpha) \\ &\leq \max_{0\leq\alpha\leq 1} \left(\alpha C\big(c_{sd}^2 P_{s1} + \frac{c_{sr}^2 P_{s1}}{1 + D_{\mathrm{WZ}}(R)}\big) + (1-\alpha) C\big(c_{sd}^2 P_{s2}\big)\right). \end{aligned} \tag{4.21}$$

Note that the above achievable rate is given under the transmitting power constraints P_{s1}, P_{s2}, and P_r. We now consider the rates under the average power constraints P_s and P_r. Since the relay only transmits during the relay-transmit period T_2 with block length $n(1-\alpha)$, the *normalized* transmitting power at the relay is $\frac{P_r}{1-\alpha}$. Similarly, the normalized transmitting power at the source during the relay-receive period T_1 and the relay-transmit period T_2 is $P_{s1} = \frac{kP_s}{\alpha}$ and $P_{s2} = \frac{(1-k)P_s}{1-\alpha}$, respectively, where $k(0 \leq k \leq 1)$ and α determine the power allocation at the transmitter.

C. Performance Bounds of WZC at the Relay

1. WZ Distortion at Zero Rate

In the practical relay system, the signals from the source are BPSK modulated into two constellations, then transmitted independently through two AWGN channels to the relay and destination, respectively. Therefore, for the relay channel with BPSK modulation, the signals Y_r and Y_{d1} are given by (4.1) and (4.2), and X_{s1} is a BPSK-modulated signal, taking values at $\sqrt{P_{s1}}$ and $-\sqrt{P_{s1}}$ with the probabilities p and $1-p$, respectively. Z_r and Z_d are i.i.d. Gaussian noise with zero mean and unit variance. Without loss of generality, we assume $p = 0.5$. Because $Y_r \leftrightarrow X_{s1} \leftrightarrow Y_{d1}$ forms a Markov chain, the conditional pdf $f(y_r|y_{d1})$ is

$$\begin{aligned}
f(y_r|y_{d1}) &= \frac{f(y_r, y_{d1})}{f(y_{d1})} = \frac{\sum_{x_{s1}} f(y_r|y_{d1}, x_r) f(y_{d1}|x_{s1}) P(x_{s1})}{\sum_{x_{s1}} f(y_{d1}|x_{s1}) P(x_{s1})} \\
&= \frac{\sum_{x_{s1}} f(y_r|x_r) f(y_{d1}|x_{s1}) P(x_{s1})}{\sum_{x_{s1}} f(y_{d1}|x_{s1}) P(x_{s1})} \\
&= \frac{\exp(-\frac{(y_{d1}-c_{sd}\sqrt{P_{s1}})^2}{2\sigma_d^2})}{\exp(-\frac{(y_{d1}-c_{sd}\sqrt{P_{s1}})^2}{2\sigma_d^2}) + \exp(-\frac{(y_{d1}+c_{sd}\sqrt{P_{s1}})^2}{2\sigma_d^2})} f(y_r|x_{s1} = \sqrt{P_{s1}}) \\
&+ \frac{\exp(-\frac{(y_{d1}+c_{sd}\sqrt{P_{s1}})^2}{2\sigma_d^2})}{\exp(-\frac{(y_{d1}-c_{sd}\sqrt{P_{s1}})^2}{2\sigma_d^2}) + \exp(-\frac{(y_{d1}+c_{sd}\sqrt{P_{s1}})^2}{2\sigma_d^2})} f(y_r|x_{s1} = -\sqrt{P_{s1}}) \\
&= \zeta \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y_r - c_{sr}\sqrt{P_{s1}})^2}{2}) + (1-\zeta) \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y_r + c_{sr}\sqrt{P_{s1}})^2}{2}) \quad (4.22)
\end{aligned}$$

where $\zeta = \frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})}$.

It is seen from (4.22) that the conditional probability of Y_r given Y_{d1} is the weighted superposition (mixture) of two Gaussian distributions centered at $c_{sr}\sqrt{P_{s1}}$ and $-c_{sr}\sqrt{P_{s1}}$, respectively, with the identical unit variance. Several examples of $f(y_r|y_{d1})$ with specific values of y_{d1} at d=9 m are shown in Fig. 22, where $P_{s1} = 69.4$ dB, $c_{sr}^2 = 1.4 \times 10^{-7}$, and the centers of two Gaussian distributions are $y_r = 1.2$ and

$y_r = -1.2$, respectively.

For Wyner-Ziv coding of Y_r with side information Y_{d1} at rate zero, only Y_{d1} is available at the decoder, thus Y_r is reconstructed into the minimum mean square error (MMSE) estimate of Y_r given Y_{d1} as [86]

$$\begin{aligned} \hat{y_r}^{\text{MMSE}} &= E[Y_r|Y_{d1} = y_{d1}] \\ &= \zeta c_{sr}\sqrt{P_{s1}} - (1-\zeta)c_{sr}\sqrt{P_{s1}} \\ &= c_{sr}\sqrt{P_{s1}}\tanh(c_{sd}\sqrt{P_{s1}}y_{d1}). \end{aligned} \tag{4.23}$$

Define $E = Y_r - \hat{Y_r}^{\text{MMSE}} = c_{sr}(X_{s1} - \sqrt{P_{s1}}\tanh(c_{sd}\sqrt{P_{s1}}Y_{d1})) + Z_r$, then E is the decoding error of WZC at rate zero. Its conditional pdf $f(e|y_{d1})$ is given in the following proposition, and shown in Fig. 23 w.r.t. e and y_{d1}, with $d = 9$ m, $c_{sr}^2 = 1.4 \times 10^{-7}$, and $P_{s1} = 69.4$ dB.

Proposition 1: The conditional pdf $f(e|y_{d1})$ is

$$f_E(e|y_{d1}) = \frac{(1-\zeta)}{\sqrt{2\pi}}\exp(-\frac{(e+2c_{sr}\sqrt{P_{s1}}\zeta)^2}{2}) + \frac{\zeta}{\sqrt{2\pi}}\exp(-\frac{(e-2c_{sr}\sqrt{P_{s1}}(1-\zeta))^2}{2}). \tag{4.24}$$

Proof: Proof is provided in Appendix A.

The conditional pdf $f(e|Y_{d1})$ also tells us about the WZC distortion at zero rate. When the WZC rate is zero, there's no information provided by the source Y_r at the decoder side, hence Y_r is reconstructed purely based on the information provided by the side information Y_{d1}. Note that Y_{d1} is independent of Z_r, then we get the distortion of Wyner-Ziv coding for Y_r with side information Y_{d1} at rate zero as

$$\begin{aligned} D_{\text{WZ}}(0) &= E[(Y_r - \hat{Y}_r^{\text{MMSE}})^2|Y_{d1} = y_{d1}] \\ &= E_{Y_{d1}}[E_E[E^2|y_{d1}]] \\ &= \sigma_E^2, \end{aligned} \tag{4.25}$$

where σ_E^2 is the variance of E, which is given by [86] as follows,

$$\sigma_E^2 = 1 + c_{sr}^2 P_{s1} E_{Y_{d1}}[\cosh^{-2}(c_{sd}\sqrt{P_{s1}} y_{d1})]. \tag{4.26}$$

Based on the conditional pdf of Y_r given Y_{d1} in (4.22) and the conditional pdf of E given Y_{d1} in (4.24), we obtain the following lower and upper bounds for WZC performance with BPSK modulation.

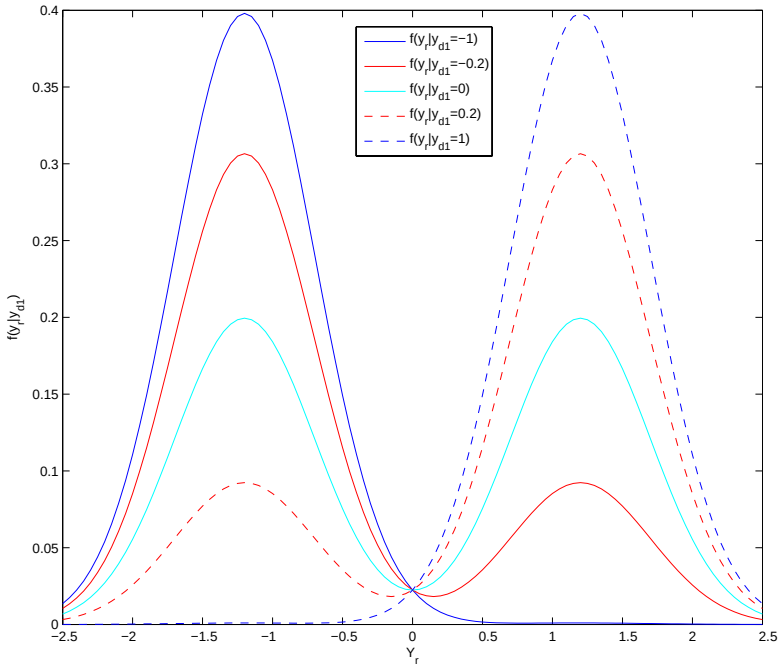


Fig. 22. An example of the conditional distribution of Y_r given particular values of Y_{d1}, with d=9 m, $P_{s1} = 69.4$ dB.

2. Lower Bound On the WZ RD Function

Let $\hat{Y}_r$ denote the reconstructed Y_r at the decoder given side information Y_{d1}, and $\hat{E}$ denote the reconstructed E at the decoder, then the Shannon lower bound is given

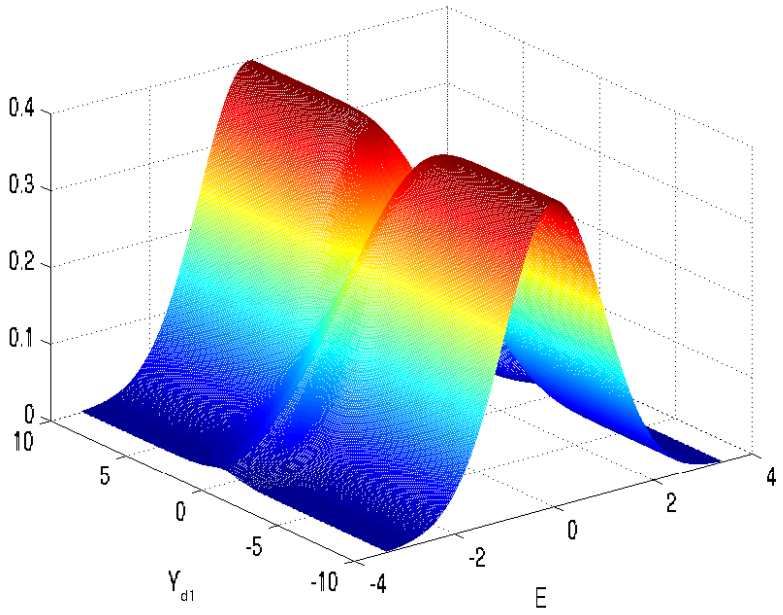


Fig. 23. The conditional pdf $f_E(e|y_{d1})$ w.r.t. e and y_{d1}, with $d = 9$ m, $c_{sr}^2 = 1.4 \times 10^{-7}$, and $P_{s1} = 69.4dB$.

by [87] as

$$\begin{aligned} R_{\mathrm{WZ}}(D) &\geq \inf_{\{\hat{Y}_r : E[d(Y_r, \hat{Y}_r)] \leq D\}} I(Y_r; \hat{Y}_r | Y_{d1}) \\ &= \inf_{\{\hat{Y}_r : E[d(Y_r, \hat{Y}_r)] \leq D\}} I(Y_r - \hat{Y}_r^{\mathrm{MMSE}}; \hat{Y}_r - \hat{Y}_r^{\mathrm{MMSE}} | Y_{d1}) \\ &= \inf_{\{\hat{E} : E[d(E, \hat{E})] \leq D\}} I(E; \hat{E} | Y_{d1}) \\ &\overset{(a)}{\geq} h(E|Y_{d1}) - \frac{1}{2}\log(2\pi e D). \end{aligned} \tag{4.27}$$

where (a) is the Shannon lower bound and denoted as $R_{\mathrm{WZ}}^{\mathrm{SLB}}(D)$. $h(E|Y_{d1})$ can be calculated directly from $f_E(e|y_{d1})$ given by (4.24).

3. Upper Bound On the WZ RD Function

Zamir [60] proposed an upper bound for RD function of WZC, given as

$$R_{\mathrm{WZ}}(D) \leq R^{\mathrm{add}}_{Y_r|Y_{d1}}(D), \tag{4.28}$$

where $R^{\mathrm{add}}_{Y_r|Y_{d1}}(D)$, namely the *addtive noise rate distortion function* is defined as

$$R^{\mathrm{add}}_{Y_r|Y_{d1}}(D) = \text{lower convex envelop}\{f_{Y_r|Y_{d1}}(D)\}, \tag{4.29}$$

and $f_{Y_r|Y_{d1}}(D)$ is defined as

$$f_{Y_r|Y_{d1}}(D) = \inf_{N^* \in \mathcal{Y}_r, N^* \nparallel (Y_r, Y_{d1}), \varepsilon_d(Y_r|Y_r+N^*, Y_{d1}) \leq D} I(Y_r; Y_r + N^*|Y_{d1}), \tag{4.30}$$

which is the capacity of an auxiliary additive noise channel with constraints on the distortion in reconstructing Y_r from $Y_r + N^*$ and Y_{d1}. In (4.30), $\nparallel$ means statistical independence, and $\varepsilon_d(Y_r|Y_r + N^*, Y_{d1})$ means the minimal possible average error in estimating Y_r from $Y_r + N^*$ and Y_{d1}. In our set-up, all signals are continuous, therefore the optimal estimate is the MMSE estimate given as $\hat{Y_r} = E[Y_r|y_{d1}, y_r + n^*]$.

From $I(Y_r; Y_r + N^*|Y_{d1}) = h(Y_r|Y_{d1}) - h(Y_r|Y_r + N^*, Y_{d1})$, and $f(y_r|y_r + n^*, y_{d1}) = \frac{f(y_r, y_r+n^*|y_{d1})}{\int f(y_r, y_r+n^*|y_{d1})dy_r}$, we need to calculate $f(y_r, y_r + n^*|y_{d1})$. Since (4.30) takes the infima over all possible distributions of N^*, we assume that N^* is Gaussian and get an upper bound for $f_{Y_r|Y_{d1}}(D)$, hence an upper bound for $R_{\mathrm{WZ}}(D)$. Furthermore, [60] claims that the Gaussian random variable N^* is actually the quantization noise of the dithered quantization assuming infinite-dimensional lattice quantization. Since dithering is always desired at low rate, N^* is essentially the quantization noise N introduced in section III, and this upper bound is achievable by practical CF scheme assuming infinite lattice dimension. Thereafter we use the notation N for the Gaussian random variable N^*, and denote $Z = Y_r + N$, with σ_n^2 being its variance, then we

have the achievable upper bound on the WZ RD function as $R^{\text{add}}_{Y_r|Y_{d1}}(D) = I(Y_r; Z|Y_{d1})$ with constraint $\varepsilon_d(Y_r|Z, Y_{d1}) \leq D$ where $Z = Y_r + N$, $N \sim (0, \sigma_n^2)$.

- Rate computation of the achievable upper bound:

  Since

$$\begin{aligned}
& f(y_r, z|y_{d1}) = f(z|y_r, y_{d1}) f(y_r|y_{d1}) \\
\overset{(a)}{=} \;& f(z|y_r) f(y_r|y_{d1}) \\
\overset{(b)}{=} \;& \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp(-\frac{(z-y_r)^2}{2\sigma_n^2})[\zeta \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y_r - c_{sr}\sqrt{P_{s1}})^2}{2}) \\
+ \;& (1-\zeta)\frac{1}{\sqrt{2\pi}} \exp(-\frac{(y_r + c_{sr}\sqrt{P_{s1}})^2}{2})] \\
= \;& \frac{\zeta}{2\pi\sigma_n} \exp(-\frac{(1+\sigma_n^2)(y_r - \frac{z+\sigma_n^2 c_{sr}\sqrt{P_{s1}}}{1+\sigma_n^2})^2}{2\sigma_n^2}) \exp(-\frac{(z - c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)}) \\
+ & \frac{(1-\zeta)}{2\pi\sigma_n} \exp(-\frac{(1+\sigma_n^2)(y_r - \frac{z-\sigma_n^2 c_{sr}\sqrt{P_{s1}}}{1+\sigma_n^2})^2}{2\sigma_n^2}) \exp(-\frac{(z + c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)}),
\end{aligned} \tag{4.31}$$

  where (a) comes from the fact that $Y_{d1} \longleftrightarrow Y_r \longleftrightarrow Y_r + N$ forms a Markov chain, (b) comes from (4.22).

  Then

$$\begin{aligned}
f(z|y_{d1}) &= \int f(y_r, z|y_{d1}) dy_r \\
&= \frac{\zeta}{\sqrt{2\pi(1+\sigma_n^2)}} \exp(-\frac{(z - c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)}) \\
&+ \frac{(1-\zeta)}{\sqrt{2\pi(1+\sigma_n^2)}} \exp(-\frac{(z + c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)}).
\end{aligned} \tag{4.32}$$

  It shows that Z, given y_{d1}, is a composite Gaussian random variable with variance $1 + \sigma_n^2$ and means $c_{sr}\sqrt{P_{s1}}$ and $-c_{sr}\sqrt{P_{s1}}$, at probabilities ζ and $1 - \zeta$ respectively.

From (4.31) and (4.32), the conditional entropy $h(Y_r|Z, Y_{d1})$ can be calculated as

$$h(Y_r|Z, Y_{d1}) = -\int f(y_{d1}) \int f(z|y_{d1}) \int f(y_r|z, y_{d1}) \log(f(y_r|z, y_{d1}))dy_r dz dy_{d1}, \tag{4.33}$$

and the rate of the achievable upper bound can be calculated as

$$\begin{aligned} & R^{\text{add}}_{Y_r|Y_{d1}} = I(Y_r; Z|Y_{d1}) = h(Y_r|Y_{d1}) - h(Y_r|Z, Y_{d1}) \\ = \; & h(Y_r|Y_{d1}) + \int f(y_{d1}) \int f(z|y_{d1}) \int f(y_r|z, y_{d1}) \log(f(y_r|z, y_{d1}))dy_r dz dy_{d1}. \end{aligned} \tag{4.34}$$

- Distortion computation of the achievable upper bound:

  The reconstructed version of Y_r at the WZ decoder given Z and Y_{d1} is

$$\begin{aligned} \hat{y_r} &= E[Y_r|y_{d1}, z] \\ &= \frac{\int y_r f(y_r, z|y_{d1})dy_r}{\int f(y_r, z|y_{d1})dy_r} \\ &= \frac{\zeta \frac{z+\sigma_n^2 c_{sr}\sqrt{P_{s1}}}{1+\sigma_n^2} \exp(-\frac{(z-c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)}) + (1-\zeta)\frac{z-\sigma_n^2 c_{sr}\sqrt{P_{s1}}}{1+\sigma_n^2} \exp(-\frac{(z+c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)})}{\zeta \exp(-\frac{(z-c_{sr}A)^2}{2(1+\sigma_n^2)}) + (1-\zeta)\exp(-\frac{(z+c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)})}. \end{aligned} \tag{4.35}$$

  Therefore the distortion D is given as

$$\begin{aligned} & D = E[(Y_r - \hat{Y_r})^2|Y_{d1}, Z] \\ = \; & \int f(y_{d1}) \int f(z|y_{d1}) \int f(y_r|z, y_{d1})(y_r - \hat{y_r})^2 dy_r dz dy_{d1} \\ = & \iint f(y_{d1}) \frac{\frac{1}{\sqrt{2\pi(1+\sigma_n^2)}}\zeta(1-\zeta)\exp(-\frac{(z-c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)})\exp(-\frac{(z+c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)})}{\zeta \exp(-\frac{(z-c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)}) + (1-\zeta)\exp(-\frac{(z+c_{sr}\sqrt{P_{s1}})^2}{2(1+\sigma_n^2)})} \\ & (\frac{2\sigma_n^2 c_{sr}\sqrt{P_{s1}}}{1+\sigma_n^2})^2 dz dy_{d1} + \frac{\sigma_n^2}{1+\sigma_n^2}. \end{aligned} \tag{4.36}$$

From (4.31), (4.32), (4.33), (4.34), and (4.36), we can calculate the upper bound with rate $R^{\text{add}}_{Y_r|Y_{d1}}$ and distortion D, both taking σ_n^2 as a parameter. For any arbitrary WZC rate R, the parameter σ_n^2 is picked up to make $R^{\text{add}}_{Y_r|Y_{d1}}$ from (4.33) equals to R, then the corresponding distortion D is calculated from (4.36) based on the specific σ_n^2. The achievable rate given by (4.34) and (4.33) and distortion given by (4.36) are shown in Fig. 24, where $c_{sd}^2 = 10^{-7}$, and $c_{sr} = 1.4 \times 10^{-7}$ (i.e., the relay is 9 m away from the source). From the figure it can be seen that the rate is decreasing as σ_n^2 increases, while the distortion increases as σ_n^2 increases. Both rate and distortion are monotonic functions of σ_n^2.

Remark

(a) When $\sigma_n^2 \to 0$, N appears to be a constant, therefore $\hat{y_r} = E[Y_r|y_{d1}, z] \to y_r$, which results in $D \to 0$, and $R^{\text{add}}_{Y_r|Y_{d1}} = I(Y_r; Z|Y_{d1}) \to I(Y_r, Y_r|Y_{d1}) = h(Y_r|Y_{d1})$.

(b) When $\sigma_n^2 \to \infty$, Z and Y_r are independent, therefore $R^{\text{add}}_{Y_r|Y_{d1}} = I(Y_r; Z|Y_{d1}) \to 0$, with $\hat{y_r} = E[Y_r|y_{d1}, z] = \int y_r f(y_r|y_{d1}, z) dy_r \to \int y_r f(y_r|y_{d1}) dy_r = E[Y_r|y_{d1}] = c_{sr}\sqrt{P_{s1}} \tanh(c_{sd}\sqrt{P_{s1}} y_{d1})$, and $D = E[(Y_r - \hat{Y_r})^2|Y_{d1}] = E[E^2|Y_{d1}]$. In this case, (4.36) becomes

$$\lim_{\sigma_n^2 \to \infty} D = 1 + c_{sr}^2 P_{s1} E_{Y_{d1}}[\cosh^{-2}(c_{sd}\sqrt{P_{s1}} y_{d1})], \tag{4.37}$$

which is consistent with (4.26).

Fig.25 shows the lower and upper bounds of the WZC RD performance for half duplex CF relaying with (a)$c_{sr}^2 = 8.0 \times 10^{-7}$ (i.e., the relay is 5m away from the source), and (b)$c_{sr}^2 = 1.4 \times 10^{-7}$ (i.e., the relay is 9m away from the source), respectively. The ideal WZC performance should reside between these two bounds. From Fig.25 one can conclude that the lower and upper bounds are close to each other, hence tight, for the scenarios where Y_r and Y_{d1} are closely correlated, and consequently provide a good approximation for the theoretical WZC performance.

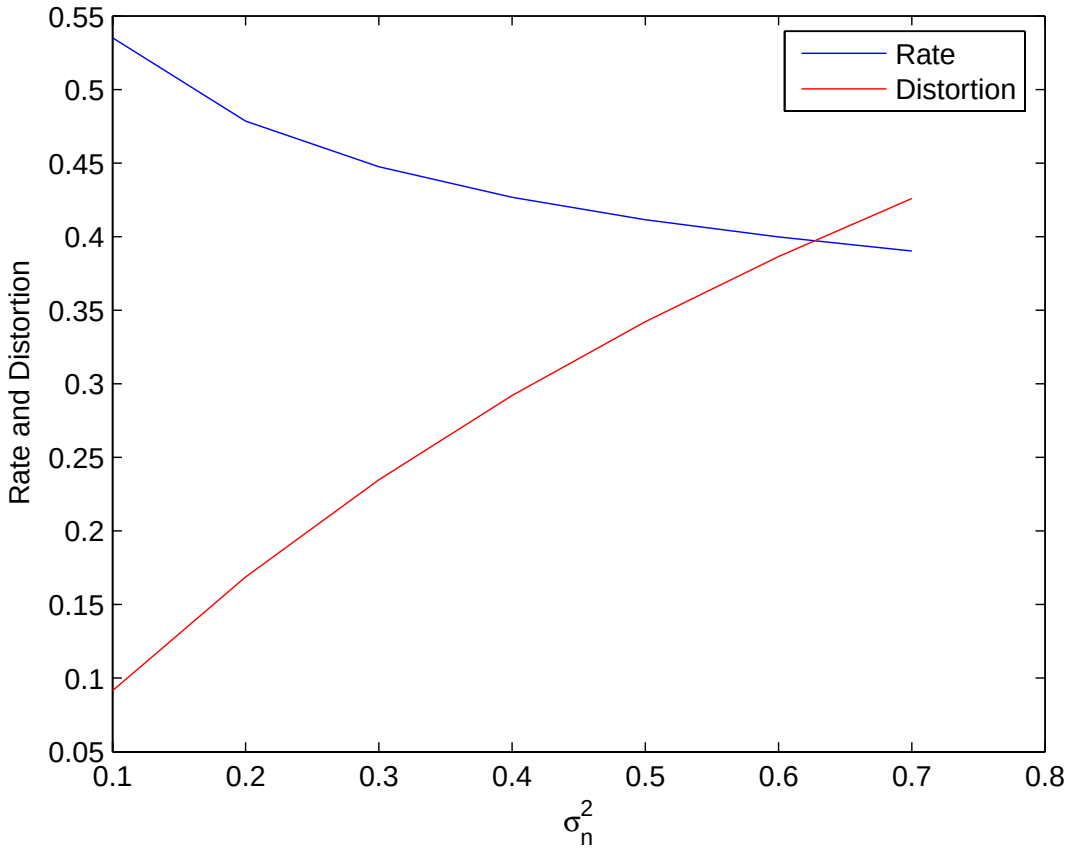
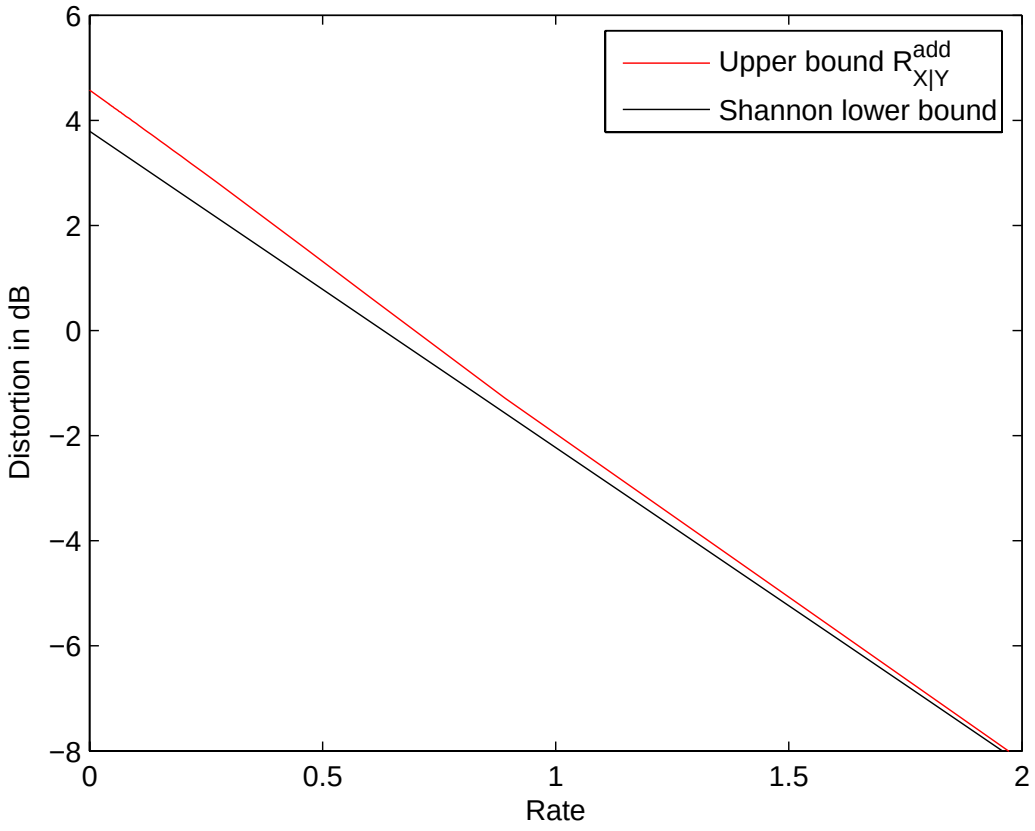
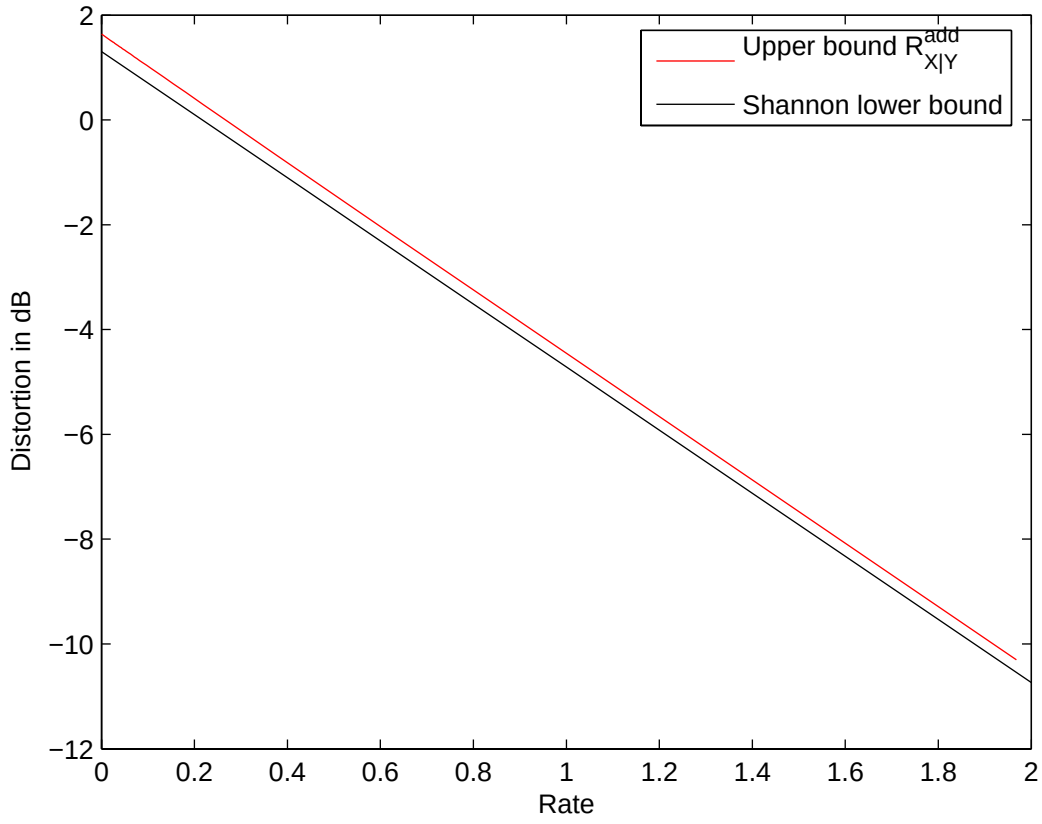


Fig. 24. The *additive noise* upper bound of WZC rate and distortion, both as functions of σ_n^2, where $c_{sr}^2 = 1.4 \times 10^{-7}$ (i.e., the relay is 9m away from the source).



(a)

Fig. 25. The upper and lower bounds for half-duplex CF relaying with (a)$c_{sr}^2 = 8.0 \times 10^{-7}$ (i.e., the relay is 5m away from the source), and (b)$c_{sr}^2 = 1.4 \times 10^{-7}$ (i.e., the relay is 9m away from the source).



(b)

Fig. 25. Continued.

4. Upper and Lower Bounds of the Achievable Rate for CF Relaying with BPSK Modulation

From (4.27) and (4.36) we get the lower and upper bounds for the WZC rate with specific distortion, denoted as $R_{\mathrm{WZ}}^{\mathrm{SLB}}(D)$ and $R_{Y_r|Y_{d1}}^{\mathrm{add}}(D)$, respectively. Conversely, we get the lower and upper bounds for the WZC distortion with specific rate, as

$$D_{\mathrm{WZ}}^{\mathrm{SLB}}(R) = \{D : R_{\mathrm{WZ}}^{\mathrm{SLB}}(D) = R\}, \tag{4.38}$$

and

$$D_{\mathrm{WZ}}^{\mathrm{add}}(R) = \{D : R_{Y_r|Y_{d1}}^{\mathrm{add}}(D) = R\}. \tag{4.39}$$

Based on the lower and upper bounds of WZC distortion with BPSK modulation, we get the corresponding lower and upper bounds on the achievable rate of CF scheme for relay channel with BPSK modulation.

Then the lower and upper bounds for the achievable CF rate are

$$\underline{R}_{CF} \leq \max_{0\leq\alpha\leq 1}\{\underline{R}_r(\alpha) + R_d(\alpha)\}, \tag{4.40}$$

and

$$\overline{R}_{CF} \leq \max_{0\leq\alpha\leq 1}\{\overline{R}_r(\alpha) + R_d(\alpha)\}, \tag{4.41}$$

where

$$\underline{R}_r(\alpha) = \alpha C\left(c_{sd}^2 P_{s1} + \frac{c_{sr}^2 P_{s1}}{1 + D_{WZ}^{\mathrm{SLB}}}\right), \tag{4.42}$$

and

$$\overline{R}_r(\alpha) = \alpha C\left(c_{sd}^2 P_{s1} + \frac{c_{sr}^2 P_{s1}}{1 + D_{WZ}^{\mathrm{add}}}\right), \tag{4.43}$$

$C(\cdot)$ is given by (4.9). Since $D_{\mathrm{WZ}}^{\mathrm{add}}(R)$ is an achievable upper bound, $\overline{R}_{CF}$ is also an achievable upper bound.

Fig. 26(a) shows the upper and lower bounds on the achievable rate of CF

scheme for the relay channel with BPSK modulation. We also include in Fig. (26)(a) the capacity bounds and achievable rates of DF. The max-flow-min-cut bound on the capacity of the relay channel with BPSK modulation is given in (4.10). The achievable rate of DF for the relay channel with BPSK modulation is given by (4.13). In Fig. 26(b), we compare the max-flow-min-cut bounds for BPSK modulation and for Gaussian modulation (which means that the input signals to all the channels are Gaussian) given by [44]. Fig. 26(c) compares the achievable rates for DF with BPSK modulation and Gaussian modulation [44], respectively; Fig. 26(d) shows the upper and lower bounds for the achievable CF rate with BPSK modulation, together with the achievable CF rate with Gaussian modulation [44]. In these figures, the locations of the source and destination nodes are fixed with a distance of 10 m; the relay node is moving along the line between the source and destination nodes. The overall rate is fixed at 0.5 b/s, and the transmitting power from the relay P_r is fixed to be 70 dB. The x-axis is the distance from the source to the relay node, and the y-axis is the transmitting power from the source P_s. From Fig. 26(a) and (d), one can conclude that the upper and lower bounds on the achievable rate of CF scheme for the relay channel with BPSK modulation are very close (with a gap from 0.06 dB to 0.22 dB).

D. DJSCC at the Relay

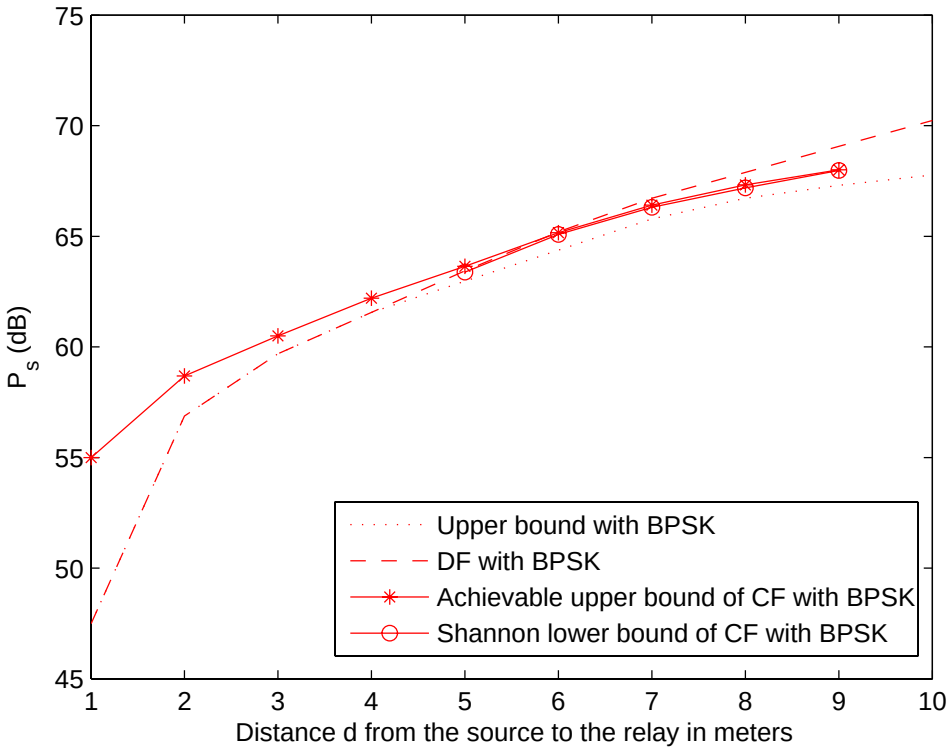
In the practical CF scheme for relay channel, Slepian-Wolf coded nested quantization (SWC-NQ) [48] is employed for WZC. Specifically, Y_r is quantized to $W \in \{1, \ldots, N\}$ by nested lattice quantization, where N is the nesting ratio [48, 9], and W further compressed into S using SWC by treating Y_{d1} at the destination as the decoder side information, with rate limit $H(W|Y_{d1})$ bit/sample. Accordingly a noisy channel WZC problem arises at the relay because the Slepian-Wolf compressed bitstream S has to be

protected by extra channel coding before being transmitted through the noisy channel between the relay and the destination. Instead of applying two *separate* channel codes for Slepian-Wolf coding and error-protecting separately, only one channel code can be utilized as distributed joint source-channel coding (DJSCC).

1. Source-channel Coding with Side Information

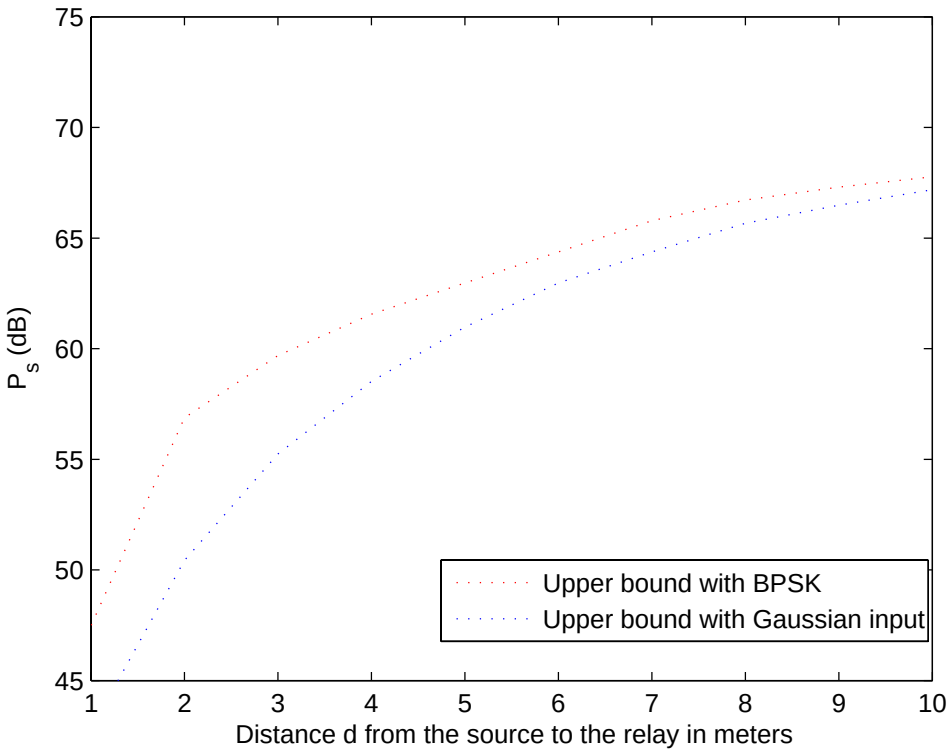
The single-user separation theorem of source-channel coding put forth by Shannon in 1948 [88] has two parts: a direct part that states that if the minimum achievable source coding rate of a given source is strictly below the capacity of the channel, then the source can be transmitted reliably by appropriate encoding-decoding; and a converse part stating that if the source coding rate is strictly greater than the channel capacity, then reliable transmission is impossible. Implicit in the theorem is the fact that reliable transmission can be accomplished by separate source and channel coding.

When the transmission channel is noisy in the distributed source coding problems, error protection is needed. In the noisy channel SWC case, the separation theorem is proved in [89] where it is shown that if the receiver has side information Y of the uncoded source X, then the entropy of the source $H(X)$ in the standard separation theorem is replaced by $H(X|Y)$. Equivalently, the Slepian-Wolf limit in this noisy channel case is $\frac{H(X|Y)}{C}$, where $C \leq 1$ is the channel capacity. A separation theorem for lossy source-channel coding with side information, i.e., the noisy channel WZC case, is given in [83]. It replaces the conditional entropy $H(X|Y)$ in the separation theorem for noisy channel SWC [89] by the Wyner-Ziv rate-distortion limit $R_{WZ}(D)$, i.e., the limit becomes $\frac{R_{WZ}(D)}{C}$.



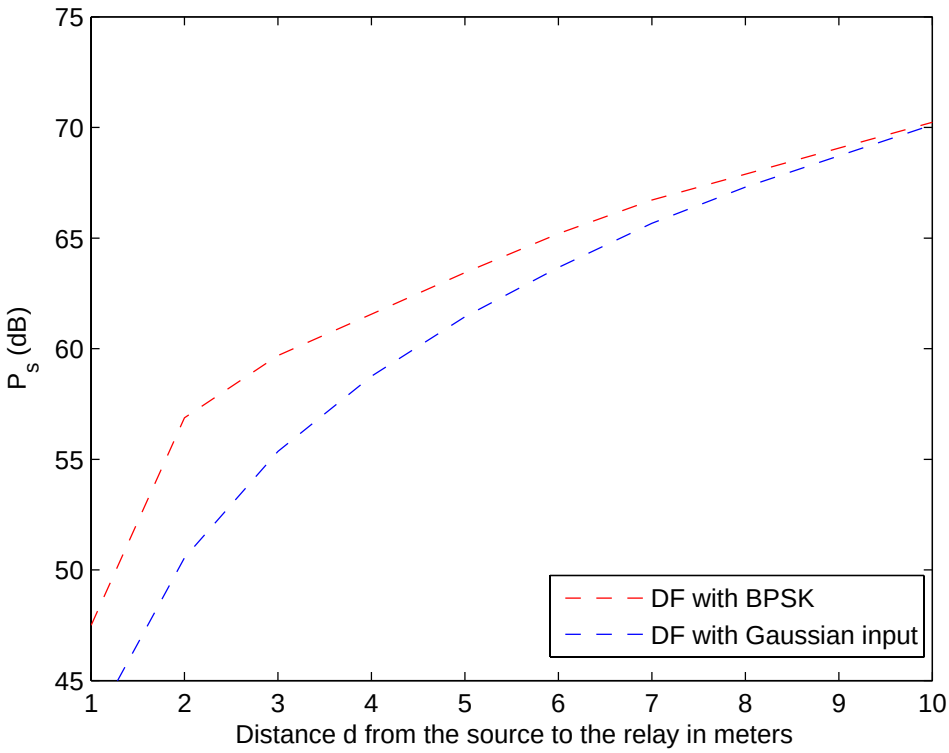
(a)

Fig. 26. (a) The upper bound, DF rate, achievable upper bound for CF and the Shannon lower bound for CF for half-duplex relaying with BPSK modulation. (b) The upper bounds for half-duplex relaying with BPSK modulation and with Gaussian input. (c) The DF rates for half-duplex relaying with BPSK modulation and with Gaussian input. (d) The achievable upper bound and Shannon lower bound for half-duplex CF relaying with BPSK modulation, and the CF rate with Gaussian input. The locations of the source and the destination are fixed with a distance of 10 m. The overall rate is 0.5 b/s, and the transmitting power from the relay is 70 dB.



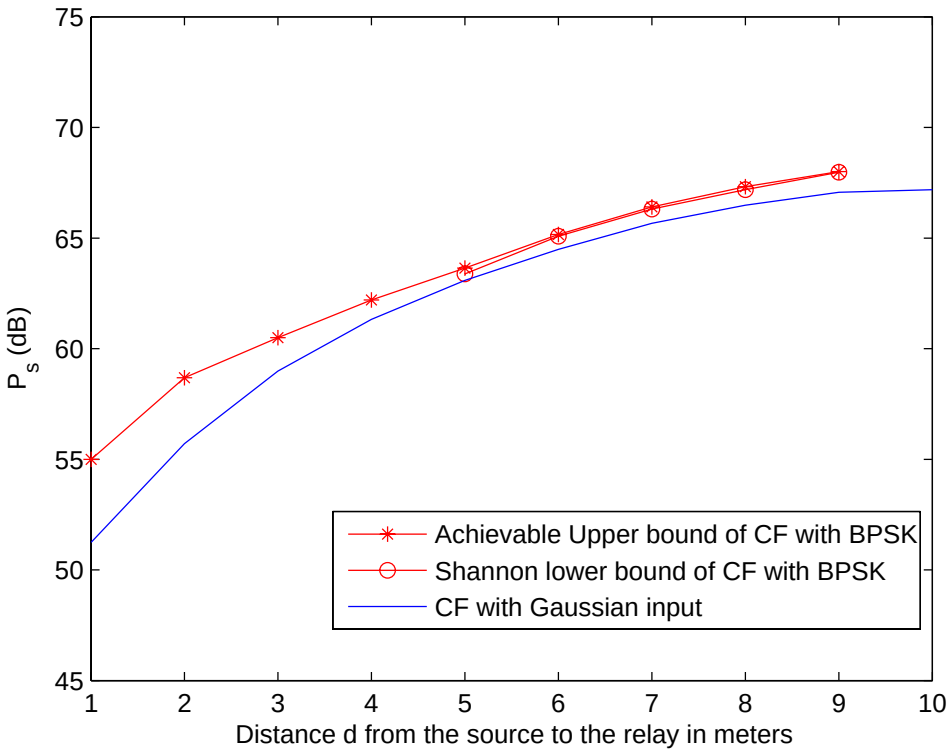
(b)

Fig. 26. Continued.



(c)

Fig. 26. Continued.



(d)

Fig. 26. Continued.

2. DJSCC at the Relay

For SWC-NQ as practical WZC for CF relaying, since SWC is implemented by channel codes, *separate* source-channel coding at the relay (with side information Y_{d1} at the destination) requires two channele codes: one for SWC (or source coding) and another for forward error protection (or channel coding). However, just like Shannon's classic separation principle, the separation principle for the noisy channel SWC/WZC problem only holds asymptotically (i.e., with infinite code length). In practical designs with $finite$ code length, $joint$ source-channel coding with side information (or DJSCC) should outperform a separate design. Indeed, simulation results in [90, Fig. 6] confirm the advantage of DJSCC over a separate design (with LDPC codes of length 10^5 bits for SWC followed by IRA codes of length 2×10^5 bits for error protection) when transmitting a binary i.i.d source X over an AWGN channel to the decoder, whose side information Y is correlated to X in a binary symmetric channel fashion.

The basic idea of DJSCC, as depicted in Fig. 27, is to use *one* channel code for both Slepian-Wolf compression and forward error protection. This is possible because a) in addition to the optimal syndrome-based approach [73] for SWC, parity bits of a systematic channel code can also be used for SWC, and b) if the number of parity bits exceeds the Slepian-Wolf limit, the added redundancy can be exploited for protection. In the following, we briefly explain the so-called *parity-based* approach for SWC before moving on to parity-based DJSCC.

The parity-based SWC scheme for binary i.i.d. sources employs an $(n + r, n)$ linear systematic channel code. To compress an n-bit vector from the source X, the encoder outputs r parity bits of the underlying systematic channel code as its compressed version, meaning $r \leq n$. In addition, $r \geq nH(X|Y)$ by the Slepian-Wolf theorem [3]. Thus the rate $\frac{n}{n+r}$ of the employed systematic channel code must be

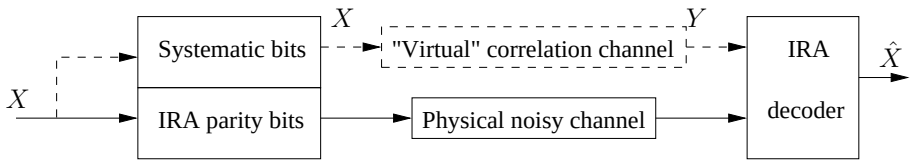


Fig. 27. DJSCC of binary source X with decoder side information Y using systematic IRA codes that are designed for both the physical noisy channel and the "virtual" correlation channel between X and Y.

no greater than $\frac{1}{1+H(X|Y)}$, which is no less than $\frac{1}{2}$. The decoder concatenates the r parity bits and the corresponding n side information bits from Y to form the received $(n+r)$-bit codeword before attempting to reconstruct its original n-bit systematic part as the decoded source vector.

When $r = n - k$, the $(2n-k, n)$ systematic channel code in the the above parity-based SWC scheme can be designed to give the same performance [90] as the syndrome-based SWC scheme [73], which outputs $n-k$ syndrome bits of an (n,k) binary channel code for the "virtual" correlation channel between the two correlated sources X and Y. The syndrome-based approach is optimal in the sense that if the (n,k) binary channel code approaches the capacity of the the "virtual" correlation channel, it also provides limit-approaching performance in SWC.

Although a longer $(2n-k, n)$ code is needed in the parity-based approach to obtain the same SWC performance as an (n,k) code in the syndrome-based approach − the reason why the latter is preferred for SWC, the advantage of the former lies in the ease with its generalization to DJSCC. On the other hand, it is not clear if the latter can be extended to DJSCC. This is because in contrast to parity bits, syndrome bits cannot provide error protection.

Under the same encoding/decoding structure that employs an $(n+r, n)$ linear systematic channel code for parity-based SWC, the extension to parity-based DJSCC involves two steps. First, because the r parity bits generated by the encoder now

provides joint Slepian-Wolf compression and error protection, r is not upper bounded by n any more. In addition, $r \geq \frac{nH(X|Y)}{C}$ according to [89]. Since the capacity $C \leq 1$, the encoder generally outputs more parity bits than the Slepian-Wolf limit. It is this added redundancy that provides error protection. Second, because we are using one channel code in DJSCC to do two jobs (SWC and error protection), the code design now involves two channels: one is the "virtual" correlation channel between the correlated sources; another is the physical noisy channel through which the parity bits are transmitted. Finding the right class of linear systematic code whose design process can readily accommodate two such channels is the starting point of DJSCC.

3. DJSCC Using IRA Codes

Practical code designs for DJSCC have been proposed in [91, 51, 92, 93, 94] based on advanced channel codes. In particular, the work of [51] employs systematic IRA codes for DJSCC of binary source X with decoder side information Y. Its basic idea is depicted in Fig. 27, where the distributed joint source-channel (DJSC) encoder only generates IRA parity bits for transmission over the noisy channel, and the already existing side information Y at the decoder is viewed as a "noisy" version of the source X (or systematic part); the IRA/DJSC decoder combines Y and the received noisy parity bits to reconstruct $\hat{X}$.

IRA codes, introduced in [50], can perform close to capacity on the binary-input AWGN channel. In addition, systematic IRA codes have the advantages of both LDPC codes (with message-passing iterative decoding, code design, superior performance) and turbo codes (with linear-time encoding). They have a more constrained structure than LDPC codes and an ensemble of systematic IRA codes is described by the degree distribution polynomials $\lambda(x) = \sum_{i=2}^{J} \lambda_i x^{i-2}$ and $\rho(x) = x^{a-1}$. They are well suited for DJSCC because they can be designed using Gaussian approximation

to take into account the two different channels. Simulation results in [51] are better than that reported in [94] for transmission over binary, Gaussian and Rayleigh fading channels.

a. Encoding

To jointly source-channel encode an information sequence of length n using an IRA code with degree distribution $\lambda(x)$ and $\rho(x) = x^{a-1}$ and a given realization of the bipartite graph, we first determine the corresponding parity bit sequence $\mathbf{P}$ of length $r = n/\left(a\int_0^1 \lambda(x)dx\right)$ for an arbitrary input sequence $\mathbf{X}$ of length n. This is the output of the source-channel encoder sent through the channel if the source-channel code is nonsystematic (NSSCC). For a systematic source-channel code (SSCC) both $\mathbf{X}$ and $\mathbf{P}$ are transmitted through the actual channel.

b. Decoding

Denote the received noisy versions of the source $\mathbf{X}$ and the parity bits $\mathbf{P}$ at the receiver are $\mathbf{Y}$ and $\mathbf{U}$, respectively. The message-passing iterative decoding algorithm can be used where both the r parity nodes (corresponding to $\mathbf{U}$) and the n systematic nodes (corresponding to $\mathbf{Y}$) can be handled as variable (left) node. The only difference is in the initialization of the algorithm with the log-likelihood ratios (LLRs) from the correlation or the actual channel. The LLR ar the i^{th} systematic node is

$$q_{i,0}^{(sys)} = \log \frac{Pr[x_i = 0|y_i]}{Pr[x_i = 1|y_i]}, i = 1, 2, .., n, \tag{4.44}$$

and the LLR at the i^{th} parity node is

$$q_{i,0}^{(par)} = \log \frac{Pr[p_i = 0|u_i]}{Pr[p_i = 1|u_i]}, i = 1, 2, .., n. \tag{4.45}$$

c. Code Design

The possibility of designing systematic IRA codes with different channel conditions for the parity part is the main advantage of using IRA codes in joint source-channel coding with side information. Proposed in [51] and following the Gaussian approximation approach of [50], the linear optimization of the systematic node degree distribution $\lambda(x) = \sum_{i=2}^{J} \lambda_i x^{i-1}$ for a given check node degree distribution $\rho(x) = x^{a-1}$ with the maximal allowable systematic node degree J is done by maximizing $\sum_{i=2}^{J} \frac{\lambda_i}{i}$ subject to the conditions [50]

$$\lambda(1) = 1, \tag{4.46}$$

$$F(x) > x, \forall x \in [x_0, 1], \tag{4.47}$$

where $F(x)$ is defined as [50]

$$F(x) = \sum_{i=1}^{J} \lambda_i \phi\left(\mu_{\text{sys}} + (i-1)\phi^{-1}\left(\frac{\phi^2(f(x))}{x^{a+1}}\right)\right), \tag{4.48}$$

and $x_0 = \phi(\mu_{\text{sys}})$. The function $f(x)$ is determined frm the equation $\phi(f(x)) = x^a \phi(\mu_{\text{par}} + f(x))$ [50].

E. Practical CF Code Design

1. System Overview

The block diagram of our proposed practical CF code design is shown in Fig. 28. At the relay node of Fig. 28, the signal $Y_r = c_{sr} X_{s1} + Z_r$ received from the source during the relay-receive period is first quantized to W by nested lattice quantization, followed by parity-based DJSCC of W. Its output X_r is BPSK modulated and transmitted to the destination with power P_s during the relay-transmit period. In our implementation, systematic IRA codes [50] are employed for DJSCC.

At the destination node of Fig. 28, the received signal is $Y_{d1} = c_{sd}X_{s1}+Z_{d1}$ during the relay-receive period, and $Y_{d2} = c_{rd}X_r + c_{sd}X_{s2} + Z_{d2}$ during the relay-transmit period, which is the superposition of the signals from the source X_{s2} and the relay X_r due to the fact that the channel is a MAC. Successive decoding is applied at the MAC decoder to recover messages m_1 and m_2. First, the nested quantization index W is reconstructed into $\hat{W}$ by the DJSC decoder with side information Y_{d1} while assuming $c_{sd}X_{s2}+Z_d2$ as background noise. Then $\hat{W}$ is used to generate $\hat{X}_r$, which leads to the subtraction of $c_{rd}\hat{X}_r$ from Y_{d2}, yielding an estimate of $c_{sd}X_{s2} + Z_{d2}$. This estimate is subsequently given to the rate-$R_d(\alpha)/(1-\alpha)$ channel decoder to recover m_2. In the meanwhile, $\hat{W}$ helps estimate Y_r into $\hat{Y}_r$, which is fed to the joint decoder together with Y_{d1} to reconstruct m_1 via MRC.

From the above, we see that CF coding at the relay involves nested lattice quantization and DJSCC. Details on these two parts are given below for Gaussian relay channels.

2. Quantizer Design

Nested scalar quantizer design for Y_r targets at finding the optimal nesting ratio N and scalar quantization stepsize q to minimize the distortion while subjecting to the rate constraint

$$R \leq \frac{1-\alpha}{\alpha}C_{rd} = \frac{1-\alpha}{\alpha}C\left(\frac{c_{rd}^2 P_r}{1+c_{sd}^2 P_{s2}}\right), \tag{4.49}$$

where the rate $R = H(W|Y_{d1})$ due to Slepian-Wolf coding of the nested quantization index W and $C_{rd} = C\left(\frac{c_{rd}^2 P_r}{1+c_{sd}^2 P_{s2}}\right)$ is the capacity of the channel between the relay and the destination with BPSK modulation. Due to BPSK modulation, NSQ has to operate at the low rate. We hence resort to simulations to generate the operational distortion-rate function $\tilde{D}_{WZ}(R)$ of SWC-NSQ by varying N and q.

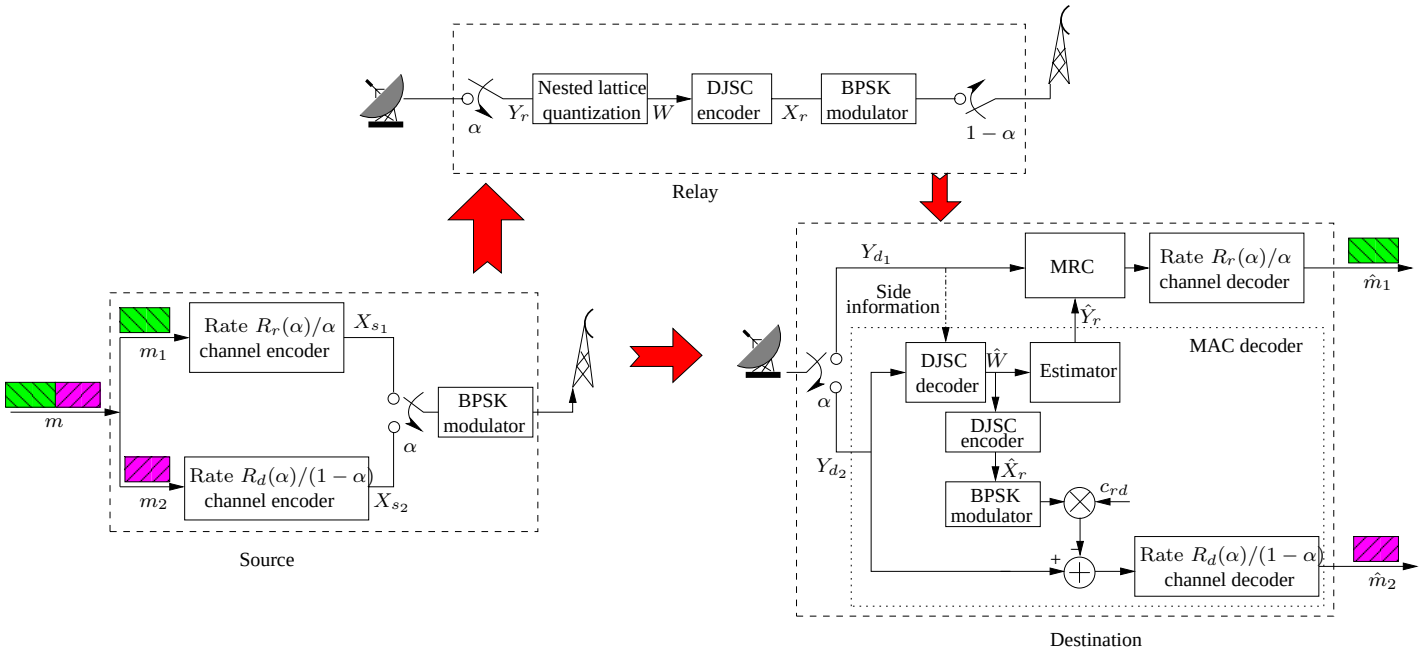


Fig. 28. Block diagram of our of our proposed CF code design.

Based on $\tilde{D}_{WZ}(R)$, the operational point at R that is slightly less than the target rate $\frac{1-\alpha}{\alpha}C_{rd}$ is picked and its corresponding N and q identified as the optimal paramters for NSQ.

We draw L (e.g., 10^5) samples of Y_r' and Y_{d1}' offline (here we use Y_r' and Y_{d1}' to distinguish them from Y_r and Y_{d1} because Y_{d1} is not available at the relay) according to the joint distribution of Y_r and Y_{d1}, quantize Y_r' into W', decode $\hat{Y}_r'$ jointly from W' and Y_{d1}', and compute the corresponding rate $R' = H(W'|Y_{d1}')$ and distortion $D_{WZ}'(R') = \frac{1}{L}\sum_{i=1}^{L}|Y_r'[i] - \hat{Y}_r'[i]|^2$ with different N and q. For $c_{sr}^2 = 1.4^{-7}, c_{sd}^2 = 10^{-7}$, and $P_{s1} = 69.4$ dB, Fig. 3(a)shows the distortion-rate curves for several different nesting ratios $N = 4, 8, 16, \infty$, where each curve is generated by varying q while fixing N. The lower envelope of these curves is the operational distortion-rate function $\tilde{D}_{WZ}(R)$ of SWC-NSQ, which is 1.5 dB away from the upper bound $D_{WZ}^{\text{add}}(R)$ *at high rate*. Fig. 3 (b) shows the operational distortion-rate curve of SWC-TCQ with TCQ rate being 7. At high rate, the operational rate-distortion performance is 0.2 dB away from the upper bound $D_{\text{WZ}}^{\text{add}}(R)$.

When reconstructing W' into $\hat{Y}_r'$, *non-linear estimation* [48] is applied to reduce the distortion, especially at low rate. Denote $J(W')$ as the index of W', $0 \le J \le N-1$, then the $\hat{Y}_r'$ is reconstructed into [48]

$$\hat{Y}_r' = \frac{\sum_k \int_{(kN+J(W'))q}^{(kN+J(W')+1)q} y_r f(y_r|y_{d1})dy_r}{\sum_k \int_{(kN+J(W'))q}^{(kN+J(W')+1)q} f(y_r|y_{d1})dy_r}, \tag{4.50}$$

where $f(y_r|y_{d1})$ is given by (4.22).

3. DJSCC Based on IRA Codes

When the nesting ratio $N = 2$ in NSQ, using a binary systematic $(n, n\alpha)$ IRA code of rate α, we apply parity-based DJSCC at the relay and encode the binary quantization

index W (of length $n\alpha$) into parity bits X_r of length $n(1-\alpha)$ for transmission to the destination. The destination receives $Y_{d2} = c_{rd}X_r + c_{sd}X_{s2} + Z_{d2}$ where $c_{sd}X_{s2} + Z_{d2}$ is treated as the additive noise. Meanwhile, the side information Y_{d1} at the destination plays the role of the "noisy" systematic part of the IRA codeword. Then W is decoded from $[Y_{d1}, Y_{d2}]$ by the IRA/DJSC decoder, resulting in $\hat{W}$. Since in the optimal NSQ design, we pick its rate such that

$$R = H(W|Y_{d1}) < \frac{1-\alpha}{\alpha} C_{rd}, \tag{4.51}$$

we have

$$n(1-\alpha) > \frac{n\alpha H(W|Y_{d1})}{C_{rd}}, \tag{4.52}$$

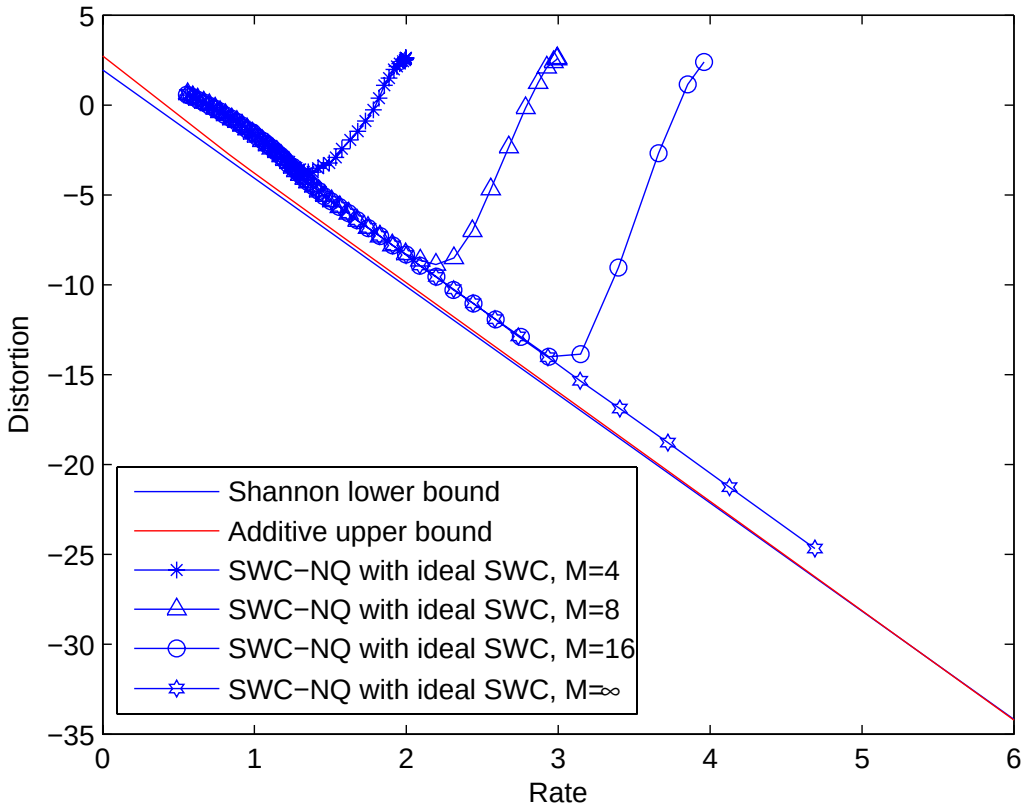
in DJSCC, which fulfills the requirement for successful decoding of W according to [89].

When $N > 2$ in NSQ, we employ a multi-level systematic IRA code for DJSCC, where each of the $\lceil \log N \rceil$ levels is used for one bit plane of W. Denote $J (0 \leq J \leq N-1)$ as the index of W and write J as $B_{\lceil \log N \rceil}, \cdots, B_1$ in its binary representation, where B_1 is the least significant bit of W and $B_{\lceil \log N \rceil}$ its most significant bit. The first-level binary systematic $(n\alpha + r_1, n\alpha)$ IRA code with

$$r_1 > \frac{n\alpha H(B_1|Y_{d1})}{C_{rd}} \tag{4.53}$$

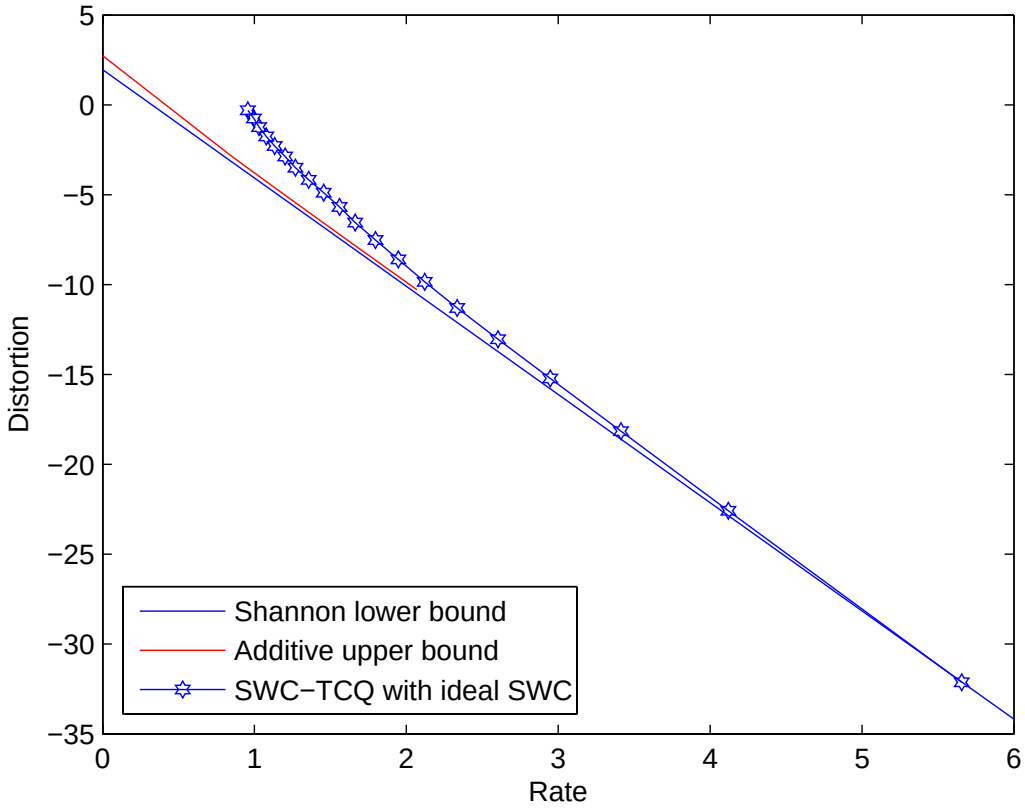
outputs r_1 parity bits after DJSCC of B_1, and the j-th level $(2 \leq j \leq \lceil \log N \rceil)$ binary systematic $(n\alpha + r_j, n\alpha)$ IRA code with

$$r_j > \frac{n\alpha H(B_j|Y_{d1}, B_1^{j-1})}{C_{rd}} \tag{4.54}$$



(a)

Fig. 29. Operational distortion-rate curves of (a) SWC-NSQ (assuming ideal SWC after NSQ) and (b) SWC-TCQ (assuming ideal SWC after TCQ) of Y_r with decoder side information Y_{d1}. (a) is generated by varying the quantization stepsize q with several different nesting ratio N. The lower envelope of these curves is the operational distortion-rate function of SWC-NSQ. In (b), the TCQ rate is 7. The relay is 9 m away from the source, with $c_{sr}^2 = 1.4 \times 10^{-7}$, $c_{sd}^2 = 10 \times -7$, $P_{s1} = 69.4$ dB.



(b)

Fig. 29. Continued.

outputs r_j parity bits after DJSCC of B_j. Here B_1^k means $B_1, \cdots, B_k$. In addition, the r_j's are chosen so that

$$\sum_{j=1}^{\lceil \log N \rceil} r_j = n(1-\alpha). \tag{4.55}$$

By the chain rule,

$$\sum_{j=1}^{\lceil \log N \rceil} H(B_j|Y_{d1}, B_1^{j-1}) = H(W|Y_{d1}), \tag{4.56}$$

then (4.53)-(4.56) lead to

$$n(1-\alpha) = \sum_{j=1}^{\lceil \log N \rceil} r_j > \frac{n\alpha H(W|Y_{d1})}{C_{rd}}, \tag{4.57}$$

which is again guaranteed by our choice of rate in (4.51) for NSQ.

a. Rate Computation for Each Bit Plane

From (4.53) and (4.54), we see that knowing the "sum-rate" $H(W|Y_{d1})$ after NSQ is not enough for multi-level IRA code design in DJSCC, the conditional entropy of each bit plane of W is also needed. We start from estimate $P_r(B_1^j = b_1^j|Y_{d1} = y_{d1})$, where $b_1^j = b_1, \ldots, b_j$ and y_{d1} are specific realizations of B_1^j and Y_{d1}, respectively. Since B_j is determined by Y_r, we denote $B_j = b_j(Y_r)$ as a function of Y_r. Therefore we have

$$P_r(B_1^j = b_1^j, |Y_{d1} = y_{d1}) = \int_{b_1(y_r)=b_1,\ldots,b_j(y_r)=b_j} f(y_r|y_{d1})dy_r. \tag{4.58}$$

When NSQ is applied for quantization, the integration region $\{y_r|b_1(y_r) = b_1, \ldots, b_j(y_r) = b_j\}$ is a union of infinite number of disjoint intervals, and (4.58) can be calculated analytically using the *erfc* function. Since $f(y_r|y_{d1})$ decays exponentially from the origin, the sum up of a few Gaussian tail probabilities could be a good approximation of (4.58).

For the general quantization such as non-uniform quantization or high-dimensional

quantization, however, $P_r(B_1^j = b_1^j | Y_{d1} = y_{d1})$ cannot be calculated analytically. Instead, we use Monte Carlo simulations. At first, the real axis is divided into M intervals, partitioning all possible Y_{d1} into M regions $\mathbb{Y}_m$ for $m = 1, \ldots, M$. Denote $\mathbb{Y}(y_{d1})$ as the region containing y_{d1}, and define $I(\cdot)$ as the indicator function taking value one if its argument is true, or zero otherwise. We then calculate $P_r(B_1^j = b_1^j | Y_{d1} = y_{d1})$ offline again by relying on the same L sample of (Y_r', Y_{d1}') we collect during the optimal NSQ design that results in

$$\begin{aligned} & P_r(B_1^j = b_1^j | Y_{d1} = y_{d1}) \\ \approx\ & P_r(B_1^j = b_1^j | Y_{d1} \in \mathbb{Y}(y_{d1})) \\ \approx\ & \frac{\sum_{i=1}^{L} I(Y_{d1}'[i] \in \mathbb{Y}(y_{d1}), b_1(Y_r'[i]) = b_1, \ldots, b_j(Y_r'[i]) = b_j)}{\sum_{i=1}^{L} I(Y_{d1}'[i] \in \mathbb{Y}(y_{d1}))}. \end{aligned} \tag{4.59}$$

Based on (4.59), the j-th level ($1 \le j \le \lceil \log N \rceil$) binary systematic $(n\alpha + r_j, n\alpha)$ IRA code can be designed with

$$\begin{aligned} & r_j > \frac{n\alpha}{C_{rd}} H(B_j | Y_{d1}, B_1^{j-1}) \\ =\ & \frac{n\alpha}{C_{rd}} \sum_{m=1}^{M} \sum_{\substack{b_1 \in \{0,1\}, \\ \cdots, \\ b_{j-1} \in \{0,1\}}} P_r(Y_{d1} \in \mathbb{Y}_m, B_1^{j-1} = b_1^{j-1}) \mathcal{H}(P_r(B_j = 1 | Y_{d1} \in \mathbb{Y}_m, B_1^{j-1} = b_1^{j-1})) \\ =\ & \frac{n\alpha}{C_{rd}} \sum_{m=1}^{M} \sum_{\substack{b_1 \in \{0,1\}, \\ \cdots, \\ b_{j-1} \in \{0,1\}}} P_r(Y_{d1} \in \mathbb{Y}_m, B_1^{j-1} = b_1^{j-1}) \mathcal{H}\left(\frac{P_r(B_1^{j-1} = b_1^{j-1}, B_j = 1 | Y_{d1} \in \mathbb{Y}_m)}{P_r(B_1^{j-1} = b_1^{j-1} | Y_{d1} \in \mathbb{Y}_m)} \right) \end{aligned} \tag{4.60}$$

where $P_r(B_1^{j-1} = b_1^{j-1}, B_j = 1 | Y_{d1} \in \mathbb{Y}_m)$ and $P_r(B_1^{j-1} = b_1^{j-1} | Y_{d1} \in \mathbb{Y}_m)$ are obtained directly from (4.59), $\mathcal{H}(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$, and $P_r(Y_{d1} \in \mathbb{Y}_m, B_1^{j-1} = b_1^{j-1})$

is estimated by using similar Monte Carlo simulations as (4.59) with

$$P_r(Y_{d1} \in \mathbb{Y}_m, B_1^{j-1} = b_1^{j-1}) = \frac{1}{L}\sum_{i=1}^{L} I(Y'_{d1}[i] \in \mathbb{Y}_m, b_1(Y'_r[i]) = b_1, \ldots, b_j(Y'_r[i]) = b_j). \tag{4.61}$$

b. Soft Threshold Decoding

In the iterative decoding procedure at the j-th bit plane, the information about the j-th bit from the channel is expressed in term of the log-likelihood-ratio, as follows,

$$\begin{aligned} & \mathrm{L_{ch}}^{(j)}(y_{d1}|b_1^{j-1}) = \log \frac{P(y_{d1}, B_1^{j-1} = b_1^{j-1}|B_j = 1)}{P(y_{d1}, B_1^{j-1} = b_1^{j-1}|B_j = 0)} \\ = & \log \frac{P(B_1^{j-1} = b_1^{j-1}, B_j = 1|y_{d1})/P(B_j = 1)}{P(B_1^{j-1} = b_1^{j-1}, B_j = 0|y_{d1})/P(B_j = 0)} \\ = & \log \frac{P(B_1^{j-1} = b_1^{j-1}, B_j = 1|y_{d1})}{P(B_1^{j-1} = b_1^{j-1}, B_j = 0|y_{d1})} - \log \frac{P(B_j = 1)}{P(B_j = 0)} \\ \triangleq & \tilde{\mathrm{L}}_{\mathrm{ch}}^{(j)}(y_{d1}|b_1^{j-1}) + \mathrm{L_{ext}}(B_j). \end{aligned} \tag{4.62}$$

where b_1^{j-1} are the specific realizations of the bits B_1^{j-1}, $\mathrm{L_{ch}}^{(j)}(y_{d1}|b_1^{j-1})$ characterizes the information about the j-th bit plane given previously decoded bits $b_1, \cdots, b_{j-1}$, and it is a function of y_{d1}. $\tilde{\mathrm{L}}_{\mathrm{ch}}^{(j)}(y_{d1}|b_1^{j-1})$ denotes the information about the j-th bit from the "virtual" channel, and $\mathrm{L_{ext}}(B_j)$ denotes the information provided by the distribution of the j-th bit itself. For NSQ, due to the symmetric property of $f(y_r|y_{d1})$ as shown in (4.22) and Fig. 22, $P(B_j = 0) = P(B_j = 1) = \frac{1}{2}$, thus $\mathrm{L_{ext}}(B_j) = 0$, and $\mathrm{L_{ch}}^{(j)}(y_{d1}|b_1, \cdots, b_{j-1}) = \tilde{\mathrm{L}}_{\mathrm{ch}}^{(j)}(y_{d1}|b_1, \cdots, b_{j-1})$.

The conditional probabilities of each quantization index given the side information Y_{d1} when $d = 8$ m, $c_{sr}^2 = 2 \times 10^{-7}$, $P_{s1} = 68.2$ dB, and the corresponding $\mathrm{L_{ch}}^{(j)}$ for the same d, are shown in Fig. F (a) and (b), respectively.

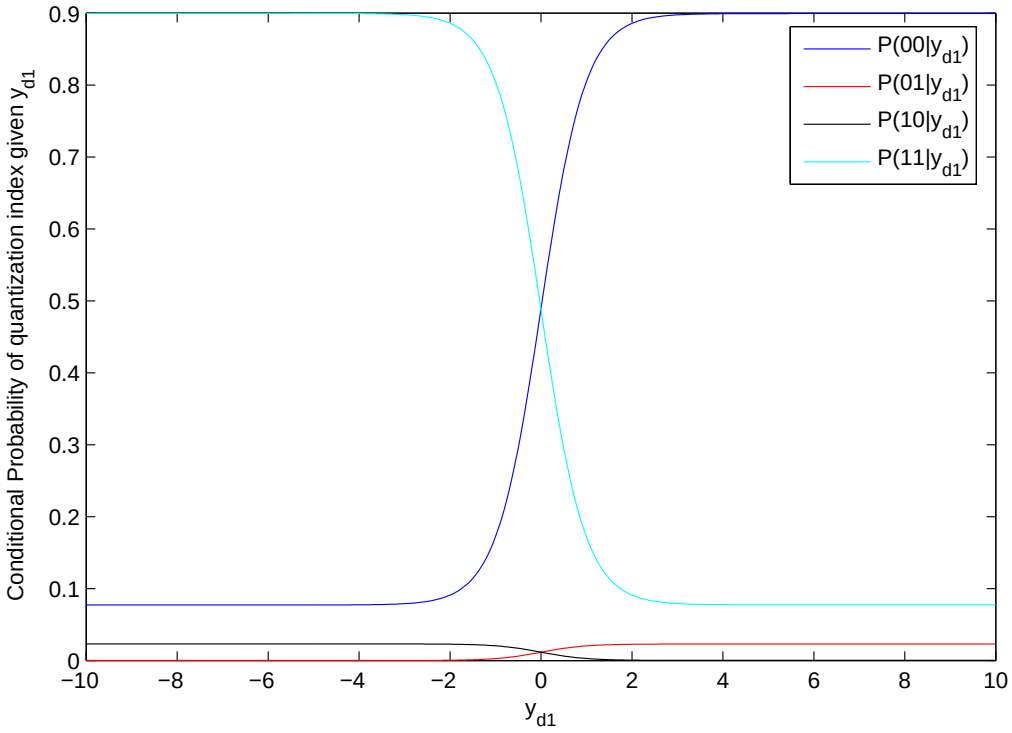
F. Simulations

In this section we present our results for the half-duplex relay channel using nested scalar quantization as the practical quantization followed by DJSCC, and compare them with those from our implementation of the best DF design [55]. In all our simulations we assume that the distribution of source messages is uniform. Prior to transmission, the pdf's needed for decoding are stored in look-up tables at the receiver nodes. In each experiment, BPSK modulation is always assumed.

Our experimental setup is as introduced in section II and Fig. 18. In our experiments, we set the target transmission rate at 0.5 bit per channel use and the average relay power $P_r = 70$ dB, and examine the average transmitting power from the source with different distance d from the source to relay.

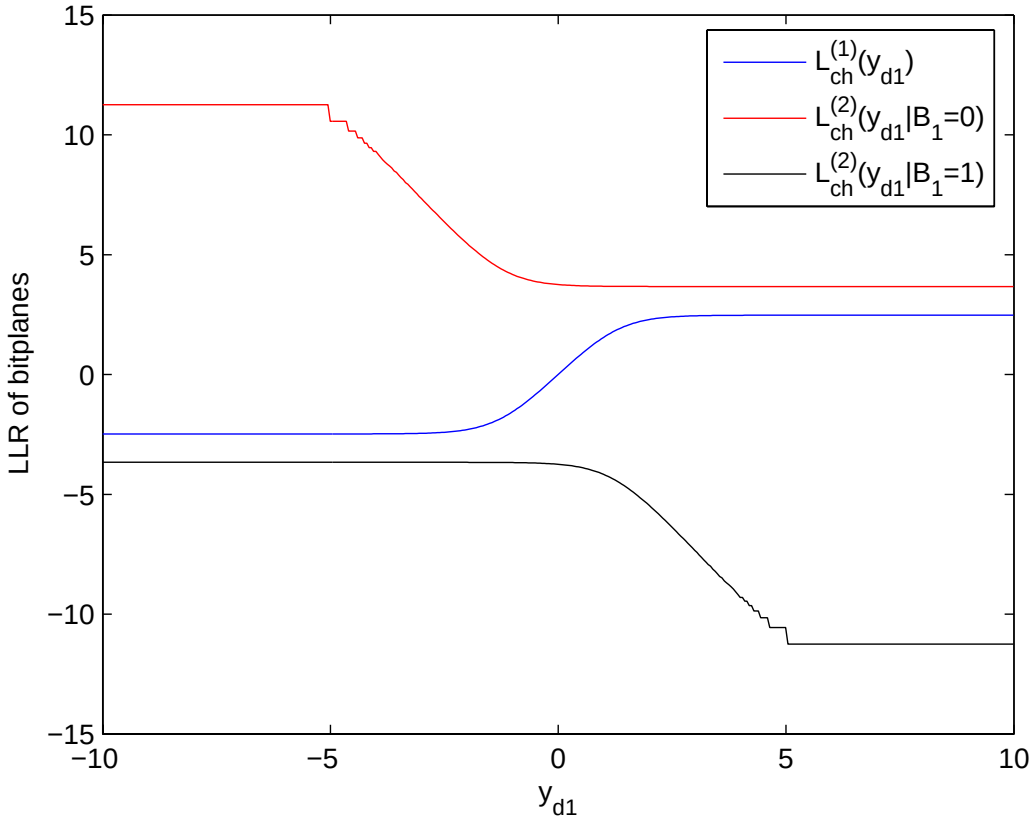
The DJSCC rate for each bit plane and the soft information for iterative decoding are collected off-line according to (4.60) and (4.62). The rates and IRA code profiles for each bit plane using NSQ for quantization when $d = 7$ m and $d = 9$ m are listed in Table I, with nesting ration $N = 4$ for both cases.

Simulation results for the half-duplex Gaussian relay channel with BPSK modulation are shown in Fig. 31 and Fig. 32 using NSQ and IRA for DJSCC with $Pr = 70$dB and 65 dB, respectively, together with the theoretical capacity bound and achievable rates given by [44] assuming Gaussian modulation, and the bounds on the capacity and achievable rates assuming BPSK modulation given by (4.40), (4.41), (4.10), and (4.13). The x axis of Fig. 31 and Fig. 32 are the distance d from the source to relay in meters, and y axis are the average transmitting power P_s from the source. The additive white Gaussian noises over the transmitting channels are all set to be with unit variance.



(a)

Fig. 30. (a) The conditional probabilities of different NSQ indices given the side information Y_{d1} when the nesting ratio is $N = 4$ in the Gaussian relay setup with $d = 8$ m, and (b) Soft input for iterative decoding of DJSCC as $\mathrm{L_{ch}}^{(1)}(y_{d1})$ for the first bit plane. For the second bit plane, since the IRA code rate is approximately 1, there is no need to evaluate the information for iterative decoding. Both (a) and (b) are functions of y_{d1}.



(b)

Fig. 30. Continued.

For coding two parts of the message, m_1 and m_2, we employ two different LDPC codes designed via density evolution. In the experiment, we independently simulate transmission of $100,000$ bits, increase the source power level until the bit error rate falls below the target error rate of 10^{-4}. When $d = 9$ m, our practical result is 1.6 dB away from the CF limit (and 1.8 dB away from the upper bound of half-duplex relaying), and 0.73 dB away from the CF limit assuming nested scalar quantization and ideal DJSCC. When $d = 5$ m, our practical CF design performs 3.03 dB away by using nested scalar quantization and IRA code for practical DJSCC, and 0.73 dB away assuming ideal DJSCC. We re-implemented the DF scheme proposed in [55], using the turbo code with generator (33/31) and block length $100,000$ [55, p. 1901]. Fig. 31 indicates that for $Pr = 70$dB, when $d > 8$ m, CF outperforms DF theoretically, and the proposed practical CF code is preferable to the practical DF code. Note that, when 7.5 m$< d < 8$ m where DF is superior in theory, our scheme still performs better than the practical DF code. In Fig. 32 for $Pr = 65$dB, the simulation results of practical CF and DF codes are consistent with the theoretical bound, i.e., when the CF outperforms DF theoretically, the practical CF code gain a better performance than the practical DF code, and vice versa. The code profiles for DJSCC are shown in Table III.

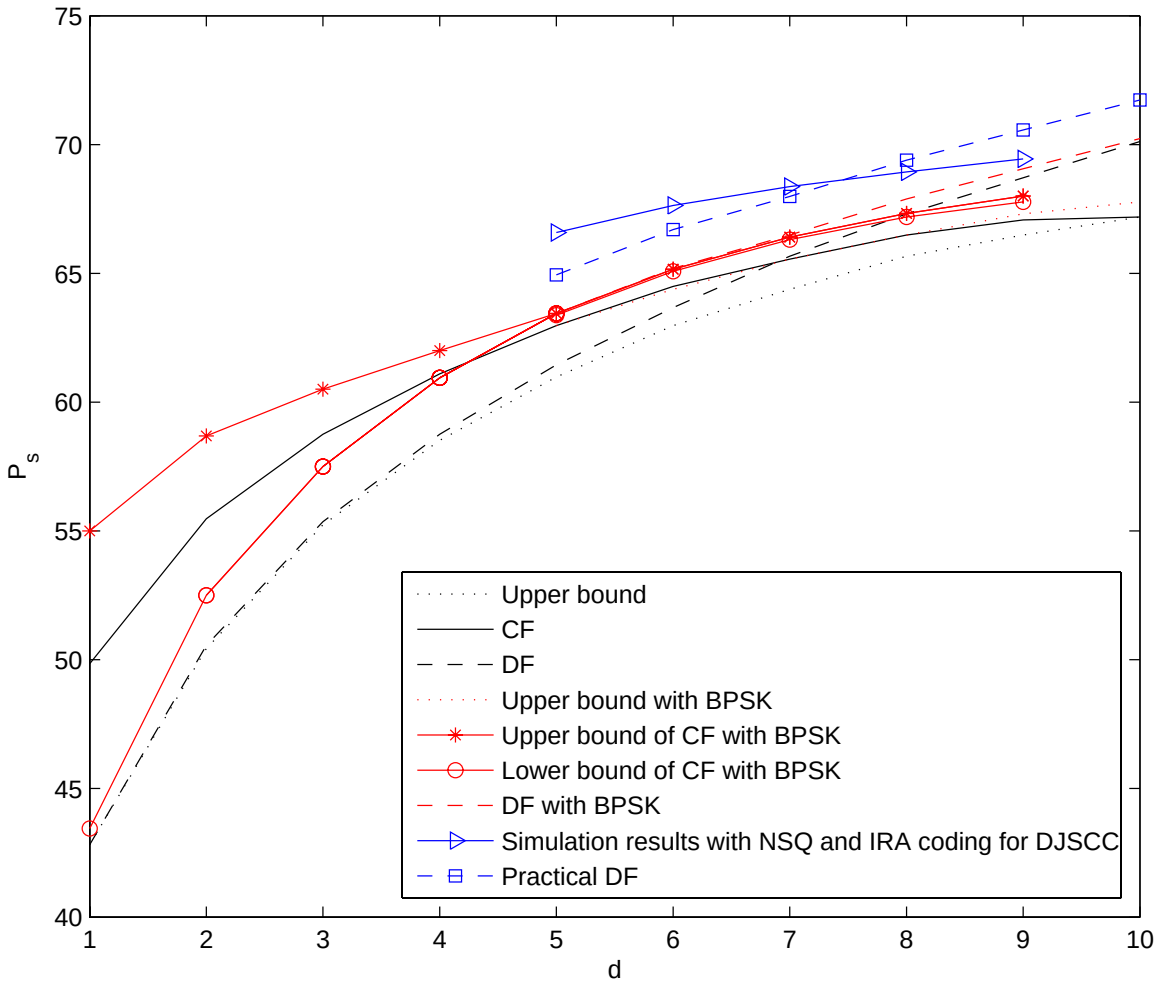


Fig. 31. Half-duplex Gaussian relay channel using NSQ for quantization and IRA code for DJSCC. With fixed overall transmitting rate 0.5 and average transmitting power from the relay as $P_r = 70$ dB, the average transmitting power from the source P_s is examined with different distance d from the source to relay, while the relay is moving along the line from the source to destination. BPSK signaling is assumed.

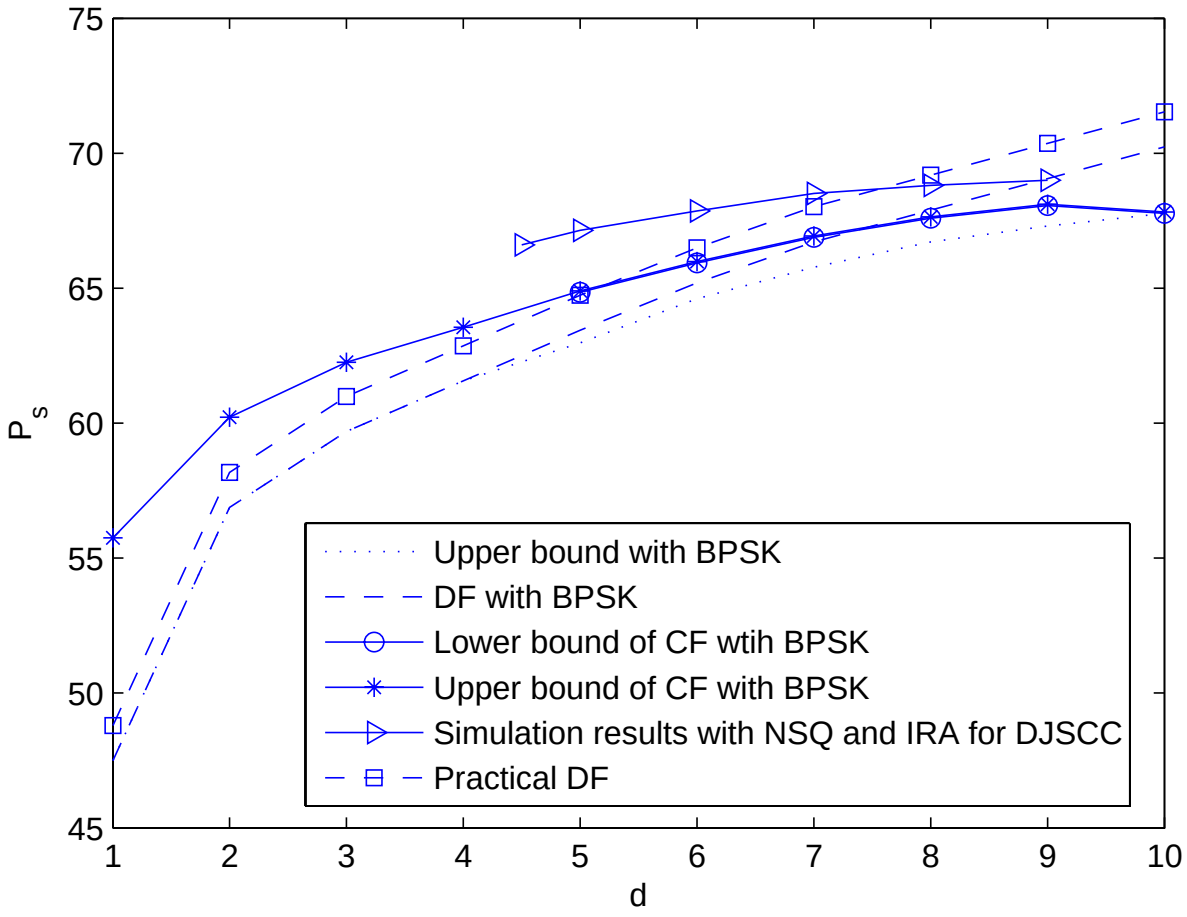


Fig. 32. Half-duplex Gaussian relay channel using NSQ for quantization and IRA code for DJSCC. With fixed overall transmitting rate 0.5 and average transmitting power from the relay as $P_r = 65$ dB, the average transmitting power from the source P_s is examined with different distance d from the source to relay, while the relay is moving along the line from the source to destination. BPSK signaling is assumed.

Table III. The conditional entropy and the corresponding degree distribution polynomials $\lambda(x)$ and $\rho(x)$ for each bit plane of CF for Gaussian relay channels using nested scalar quantization when (a) $d = 7$ m and (b) $d = 9$ m.

Bit plane j	Conditional entropy	$r_j/n\alpha$	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
1	0.765635	0.77	$0.246779x^2 + 0.160927x^3$ $+0.0000215323x^4 + 0.0000265542x^5$ $+0.158502x^6 + 0.0366871x^8$ $+0.0390752x^{12} + 0.0354665x^{19}$ $+0.0125102x^{20} + 0.116069x^{23}$ $+0.0000125083x^{98} + 0.0000207854x^{99}$ $+0.1939040000x^{100}$	x^6
2	0.178334	0.2	$0.124799x^2 + 0.17686x^3$ $+0.176933x^7 + 0.0424318x^8$ $+0.00796064x^9 + 0.0113302x^{19}$ $+0.134332x^{20} + 0.0389052x^{23}$ $+0.00866671x^{25} + 0.0330878x^{28}$ $+0.00109343x^{37} + 0.0517676x^{69}$ $+0.122933x^{74} + 0.0287246x^{77}$ $+0.0401743x^{80}$	x^{30}

(a) $d = 7$ m

Bit plane j	Conditional entropy	$r_j/n\alpha$	Degree polynomials	
			$\lambda(x)$	$\rho(x)$
1	0.818696	0.82	$0.303792x^2 + 0.173188x^3$ $+0.0671337x^5 + 0.0123568x^6$ $+0.134132x^7 + 0.0314767x^{13}$ $+0.0108393x^{15} + 0.025639x^{17}$ $+0.0910351x^{20} + 0.0400076x^{40}$ $+0.0000240473x^{46} + 0.0117242x^{52}$ $+0.0189157x^{58} + 0.0112433x^{63}$ $+0.0684922x^{77}$	x^5
2	0.140643	0.19	$0.114444x^2 + 0.179976x^3$ $+0.14779x^7 + 0.0806827x^8$ $+0.0126197x^{17} + 0.0208893x^{19}$ $+0.132564x^{21} + 0.0336681x^{22}$ $+0.00152895x^{38} + 0.0606354x^{42}$ $+0.0197357x^{69} + 0.0334799x^{72}$ $+0.123705x^{79} + 0.0382818x^{87}$	x^{44}

(b) $d = 9$ m

CHAPTER V

SUMMARY

A. SWCNQ for Wyner-Ziv Coding

we have analyzed the high-rate performance of nested lattice quantization for Wyner-Ziv coding, showing an increasing gap from the theoretical limit as the rate increases. The reason for the increase of the gap is mainly because the boundary loss is an increasing function of the rate. To compensate for the boundary loss, a SWC-NQ framework has been proposed for Wyner-Ziv coding, where Slepian-Wolf coding plays the role of second-stage binning to save rate after nested lattice quantization. Assuming ideal Slepian-Wolf coding, SWC-NQ is shown to perform a constant gap (in dB) away from the Wyner-Ziv D-R function at high rate. This result mirrors that from entropy-coded quantization in classic source coding. A non-linear minimum MSE estimator at the decoder is introduced and used in simulations that degenerates to the linear estimator we use in our high-rate performance analysis. Simulations with one- and two-dimensional nested lattice quantization and SWC-NQ (with ideal Slepian-Wolf coding) for quadratic Gaussian sources show agreement with our high-rate analytical results. Using irregular LDPC codes for practical Slepian-Wolf coding in SWC-NQ exhibits a roughly constant gap from the Wyner-Ziv limit for a wide range of rates.

We have also proved that the performance of Wyner-Ziv coding of quadratic Gaussian sources with nested lattice quantization at a fixed high rate is independent of the source correlation, with or without Slepian-Wolf coding.

B. An Application of SWCNQ: Practical CF Code Design for Half-Duplex Gaussian Relay Channels with BPSK Modulation

Based on the recent information-theoretical results, we have developed the first practical CF code design for the half-duplex Gaussian relay channel with BPSK modulation using WZC. The WZC with BPSK modulation is studied at first, with a lower bound and an upper bound being proposed, which are derived from the Shannon lower bound and additive noise upper bound, respectively, concerning the correlation of the signal to be compressed and the side information under the assumption of BPSK modulation. Based on the upper and lower bounds of WZC, code design issues are discussed including the quantizer design and distributed source-channel coding (DJSCC). Simulation results with nested scalar quantization for quantization and IRA code for DJSCC exhibit a gap of 1.6 to 3.03 dB from the CF limit for the half-duplex Gaussian relay channel with BPSK moculation, focusing on the scenario where CF outperforms DF theoretically. We believe that our code design is paving the way for realizing the potential gains of cooperative diversity in wireless ad hoc and sensor networks, as promised by the theory.

REFERENCES

[1] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, Jan. 1976.

[2] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, NY, Wiley Interscience, 1991.

[3] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.

[4] A. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: general sources," *Inform. Contr.*, vol. 38, pp. 60–80, 1978.

[5] S.I. Gelfand and M.S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. and Inform. Theory*, vol. 9, pp. 19–31, 1980.

[6] M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. 29, pp. 439–441, May 1983.

[7] R.J.Barron, B.Chen, and G.W.Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1159–1180, May 2003.

[8] S.S. Pradhan, J. Chou and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1181–1203, May 2003.

[9] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250–1276, June 2002.

[10] R. Gray and D. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[11] M.V. Eyuboglu and G.D. Forney, Jr., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks – high-rate theory for memoryless sources," *IEEE Trans. Information Theory*, vol. 39, pp. 46–59, Jan. 1993.

[12] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. DCC'02*, Snowbird, UT, Mar. 2002, pp. 252–261.

[13] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Comm. Letters*, vol. 5, pp. 417–419, Oct. 2001.

[14] A. Liveris, Z. Xiong and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440–442, Oct. 2002.

[15] A. Liveris, Z. Xiong and C. Georghiades, "Distributed compression of binary sources using conventional parallel and serial concatenated convolutional codes," in *Proc. DCC'03*, Snowbird, UT, Mar. 2003.

[16] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, Mar. 2003.

[17] J.H. Conway and N.J.A. Sloane, *Sphere Packings. Lattices and Groups*, New York, NY, Springer.

[18] R. Gallager, *Low Density Parity Check Codes*, Cambridge, MA, MIT Press, 1963.

[19] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Information Theory*, vol. 45, pp. 399–431, Mar. 1999.

[20] X. Wang and M. Orchard, "Design of trellis codes for source coding with side information at the decoder," in *Proc. DCC'01*, Snowbird, UT, Mar. 2001.

[21] S. Servetto, "Lattice quantization with side information," in *Proc. DCC'00*, Snowbird, UT, Mar. 2000.

[22] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbo-like codes," in *Proc. ISIT'02*, Lausanne, Switzerland, June 2002.

[23] H. Feng, Q. Zhao, and M. Effros, "Network source coding using entropy constrained dithered quantization," in *Proc. DCC'03*, Snowbird, UT, Mar. 2003, pp. 427–436.

[24] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Proc. DCC'03*, Snowbird, UT, Mar. 2003.

[25] J. Chou, S. Pradhan, and K. Ramchandran, "Turbo and trellis-based constructions for source coding with side information," in *Proc. DCC'03*, Snowbird, UT, Mar. 2003.

[26] A. Liveris, Z. Xiong and C. Georghiades, "Nested convolutional/turbo codes for the binary Wyner-Ziv problem," in *Proc. ICIP'03*, Barcelona, Spain, Sept. 2003.

[27] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," *submitted to IEEE Trans. Signal Processing*, June 2007.

[28] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, pp. 55–67, Jan. 1982.

[29] D. Rebollo-Monedero, A. Aaron, and B. Girod, "Transforms for high-rate distributed source coding," in *Proc. 37th Asilomar Conf.*, Pacific Grove, CA, Nov. 2003.

[30] J. Conway, E. Rains, and N. Sloane, "On the existence of similar sublattices," *Canad. J. Math.*, vol. 51, pp. 1300–1306, 1999.

[31] S.D.Servetto, "On the feasibility of large-scale wireless sensor networks," in *Proc. 40th Allerton Conf. on Communication, Control, and Computing*, Urbana,IL, Oct. 2002.

[32] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Nested quantization and Slepian-Wolf coding for Wyner-Ziv coding: performance analysis and code design," in *Proc. DCC-2004*, Snowbird, UT, Mar. 2004.

[33] E.C. Meulen, "Three-terminal communication channels," *Advanced Applied Probability*, vol. 3, pp. 120–154, 1971.

[34] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, Sept. 1979.

[35] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 4152–4172, Dec. 2005.

[36] H. Chong, M. Motani, and H. Garg, "New coding strategies for the relay channel," in *Proc. ISIT-2005*, Adelaide, Australia, Sept. 2005.

[37] M. Khojastepour, A. Sabharwal, and B. Aazhang, "Lower bounds on the capacity of Gaussian relay channel," in *Proc. CISS-2004*, Princeton, NJ, Mar. 2004.

[38] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3037–3063, Sept. 2005.

[39] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[40] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity part I and part II," *IEEE Trans. Comm.*, vol. 51, pp. 1927–1948, Nov. 2003.

[41] F. Willems and E. van der Meulen, "The discrete memoryless multi-access channel with cribbing encoders," *IEEE Trans. Inform. Theory*, vol. 31, pp. 313–327, May 1985.

[42] C. Zeng, F. Kuhlmann, and A. Buzo, "Achievablility proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1160–1165, Nov. 1989.

[43] M. Valenti and B. Zhao, "Distributed turbo codes: towards the capacity of the relay channel," in *Proc. VTC-2003*, Orlando, FL, Oct. 2003.

[44] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for the wireless relay channel," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2020–2040, June 2005.

[45] J. Laneman and G. Wornell, "Distributed space-time coding protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2415–2425, Oct. 2003.

[46] R. Nabar, H. Bölcskei, and F. Kneubühler, "Fading relay channels: performance limits and space-time signal design," *IEEE JSAC*, vol. 22, pp. 1099–1109, Aug. 2004.

[47] M. Khojastepour, A. Sabharwal, and B. Aazhang, "On capacity of Gaussian 'cheap' relay channel," in *Proc. Globecom-2003*, San Francisco, CA, Dec. 2003.

[48] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Slepian-Wolf coded nested quantization for Wyner-Ziv coding: performance analysis and code design," *IEEE Trans. Inform. Theory*, vol. 52, pp. 4358–4379, Oct. 2006.

[49] A. Liveris, Z. Xiong and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, pp. 440–442, Oct. 2002.

[50] H. Jin, A. Khndekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd Int. Symp. Turbo Codes and the Related Topics*, Sept. 2000.

[51] A.D. Liveris, Z. Xiong, and C.N. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using IRA codes," in *Proc. MMSP-2002*, St. Thomas, US Virgin Islands, Dec. 2002.

[52] T. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. Wireless Comm.*, vol. 5, pp. 283–289, Feb. 2006.

[53] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. Communications*, vol. 52, pp. 1603–1612, Sept. 2004.

[54] M. Janani, A. Hedayat, T. Hunter, and A. Nosratinia, "Coded cooperative in wireless communications: space-time transmission and iterative decoding," *IEEE Trans. Signal Processing*, vol. 52, pp. 362–371, Feb. 2004.

[55] Z. Zhang and T. Duman, "Capacity approaching turbo coding and iterative decoding for relay channels," *IEEE Trans. Comm.*, vol. 53, pp. 1895–1905, Nov. 2005.

[56] Z. Liu, V. Stanković, and Z. Xiong, "Practical compress-forward code design for the half-duplex relay channel," in *Proc. CISS'05*, Baltimore, MD, Mar. 2005.

[57] A. Chakrabarti, A. de Baynast, A. Sabharwal, and B. Aazhang, "Half-duplex estimate-and-forward relaying: bounds and code design," in *Proc. ISIT-2006*, Seattle, WA, July 2006.

[58] R. Hu and J. Li, "Practical compress-forward in user cooperation: Wyner-ziv cooperation," in *Proc. ISIT-2006*, Seattle, WA, July 2006.

[59] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, NJ, Prentice-Hall, 1971.

[60] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073–2084, Nov. 1996.

[61] L. A. Dalton, "Analysis of 1-D nested lattice quantization and Slepian-Wolf coding for Wyner-Ziv coding of i.i.d. sources," May 2002, Technical Report, Texas A&M University.

[62] J. Bronski1 and R. McLaughlin, "Rigorous estimates of the tails of the probability distribution function for the random linear shear model," *J. Statistical Physics*, vol. 98, pp. 897–915, Feb. 2000.

[63] J. Conway and N. Sloane, "A lower bound on the average error of vector quantizers," *IEEE Trans. Inform. Theory*, vol. 31, pp. 106–109, Jan. 1985.

[64] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards, and Practice*, Boston, MA, Kluwer.

[65] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," in *Proc. 37th Asilomar Conf.*, Pacific Grove, CA, 2003.

[66] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by ldpc codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2225–2229, Sept. 2003.

[67] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.

[68] B. MacMillan, "The basic theorems of information theory," *Ann. Math. Statist.*, vol. 24, pp. 196–219, June 1953.

[69] E. Weiss, "Compression and coding," *IRE Trans. Inform. Theory*, vol. 8, pp. 256–257, Apr. 1962.

[70] T. Ancheta, "Syndrome source coding and its universal generalization," *IEEE Trans. Inform. Theory*, vol. 22, pp. 432–436, July 1976.

[71] J. Garcia-Frias and Y. Zhao, "Compression of binary memoryless sources using punctured turbo codes," *IEEE Comm. Letters*, pp. 394–396, Sept. 2002.

[72] G. Caire, S. Shamai, and S. Verdu, "Lossless data compression with error correcting codes," in *Proc. ISIT'03*, Yokohama, Japan, July 2003.

[73] A. Wyner, "Recent results in the shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, pp. 2–10, Jan. 1974.

[74] R. G. Gallager, *Low-Density Parity-Check Codes*, Cambridge, MA, MIT Press, 1963.

[75] S. Y. Chung, D. Forney, T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Comm. Letters*, vol. 5, pp. 58–60, Feb. 2001.

[76] S. Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.

[77] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[78] M. R. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, pp. 533–547, Sept. 1981.

[79] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[80] J. Pearl, *Probability Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Francisco, CA, Morgan Kaufmann Publishers, Inc., 1988.

[81] F. R. Kschischang, B. J. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[82] S. Cheng and Z. Xiong, "Successive refinement for the wyner-ziv problem and layered code design," *IEEE Trans. Signal Processing*, vol. 53, pp. 3269–3281, Aug. 2005.

[83] S. Shamai, S. Verdu, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.

[84] W. Jakes, *Microwave Mobile Communications*, Piscataway, NJ, Wiley, 1974.

[85] J. Proakis, *Digital Communications*, New York, NY, McGraw-Hill, 2000.

[86] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. Information Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[87] R.Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Information Theory*, vol. 19, pp. 480–489, July 1973.

[88] C.E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[89] S. Shamai and S. Verdu, "Capacity of channels with side information," *European Trans. Telecommunications*, vol. 6, pp. 587–600, Sept. 1995.

[90] Q. Xu, V. Stanković, and Z. Xiong, "Layered wyner-ziv video coding for transmission over unreliable channels," *Signal Processing: Special Issue on Distributed Source Coding*, vol. 86, pp. 3212–3225, Nov. 2006.

[91] J. Garcia-Frias, "Joint source-channel decoding of correlated sources over noisy channels," in *Proc. DCC-2001*, Snowbird, UT, Mar. 2001, pp. 283–292.

[92] P. Mitran and J. Bajcsy, "Turbo source coding: A noise-robust approach to data compression," in *Proc. DCC-2002*, Snowbird, UT, Mar. 2002.

[93] M. Sartipi and F. Fekri, "Source and channel coding in wireless sensor networks using ldpc codes," in *Proc. 1st Annual IEEE Communications Society Conf. on Sensor C ommunications and Networks*, Santa Clara, CA, Oct. 2004.

[94] G.-C. Zhu and F. Alajaji, "Turbo codes for non-uniform memoryless sources over noisy channels," *IEEE Communications Letters*, vol. 6, pp. 64–66, Feb. 2002.

[95] R.G.Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

APPENDIX A

PROOF OF LEMMA 1 FOR THE LOWER BOUND ON THE D-R PERFORMANCE SWC-NQ IN THE QUADRATIC GAUSSIAN CASE.

1) Rate Computation: The rate for SWC-NQ is

$$R = \frac{1}{n} H(\mathbf{S}|\mathbf{Y}). \tag{A.1}$$

Since at high rate,

$$\begin{aligned} p(\mathbf{s}|\mathbf{y}) &= \sum_{j=-\infty}^{\infty} \int_{\mathbf{x}\in R_j(\mathbf{s})} f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}) d\mathbf{x} \\ &= \sum_{j=-\infty}^{\infty} \int_{\mathbf{x}\in R_0(\mathbf{s})} f_{\mathbf{Z}}(\mathbf{x}+\mathbf{c}_j(\mathbf{0})) d\mathbf{x} \\ &\overset{(a)}{=} \sum_{j=-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{s}+\mathbf{c}_j(\mathbf{0})) V_1 \\ &= g(\mathbf{s}) V_1, \end{aligned} \tag{A.2}$$

where (a) is due to the high rate assumption and $g(\mathbf{x}) \triangleq \sum_{j=-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{x}+\mathbf{c}_j(\mathbf{0}))$.

Then the achievable rate of SWC-NQ is

$$\begin{aligned} nR &= H(\mathbf{S}|\mathbf{Y}) = -\sum_{\mathbf{s}\in\Lambda_1/\Lambda_2} p(\mathbf{s}|\mathbf{y})\log_2[p(\mathbf{s}|\mathbf{y})] \\ &\overset{(b)}{=} -\sum_{\mathbf{s}\in\Lambda_1/\Lambda_2} p(\mathbf{s}|\mathbf{y})\log_2[g(\mathbf{s})V_1] \\ &= -\sum_{\mathbf{s}\in\Lambda_1/\Lambda_2}\sum_{j=-\infty}^{\infty}\int_{\mathbf{x}\in R_0(\mathbf{s})} f_{\mathbf{Z}}(\mathbf{x}+\mathbf{c}_j(\mathbf{0}))d\mathbf{x}\log_2 g(\mathbf{s}) - \log_2 V_1 \\ &\overset{(c)}{=} -\sum_{j=-\infty}^{\infty}\sum_{\mathbf{s}\in\Lambda_1/\Lambda_2}\int_{\mathbf{x}\in R_0(\mathbf{s})} f_{\mathbf{Z}}(\mathbf{x}+\mathbf{c}_j(\mathbf{0}))\log_2 g(\mathbf{x})d\mathbf{x} - \log_2 V_1 \\ &\overset{(d)}{=} -\int_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbf{Z}}(\mathbf{x})\log_2 g(\mathbf{x})d\mathbf{x} - \log_2 V_1 \\ &= -\int_{\mathbf{x}\in\mathbb{R}^n} f_{\mathbf{Z}}(\mathbf{x})\log_2[\sum_{j=-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{x}+\mathbf{c}_j(\mathbf{0}))]d\mathbf{x} - \log_2 V_1, \end{aligned}$$

where (b) and (c) use the high rate assumption and (d) is due to the periodic property of $g(\cdot)$, i.e., $g(\mathbf{x}-\mathbf{l}) = g(\mathbf{x}), \forall \mathbf{l}\in\Lambda_2$. Thus the achievable rate of SWC-NQ is

$$nR = H(\mathbf{S}|\mathbf{Y}) = h'(\mathbf{X},\Lambda_2) + \log_2\sigma_Z^n - \log_2 V_1. \tag{A.3}$$

2) Distortion Computation: From Theorem 4.1, the average distortion of nested lattice quantization over all realizations of $(\mathbf{X},\mathbf{Y})$ is $D_n = G(\Lambda_1)V_1^{\frac{2}{n}} + \frac{1}{n}E_{\mathbf{Z}}[\|Q_{\Lambda_2}(\mathbf{Z})\|^2]$, which can be lower bounded as

$$D_n \geq G(\Lambda_1)V_1^{\frac{2}{n}} + \frac{(n-1)}{n\Gamma(\frac{n+1}{2})2^{\frac{n}{2}}\pi^{\frac{1}{2}}}\sum_{\mathbf{l}\in\Lambda_2} l^2\int_{l-r}^{l+r}\int_0^{\cos^{-1}(\frac{l^2+u^2-r^2}{2lu})}\sin^{n-2}\theta d\theta\frac{u^{n-1}}{\sigma_Z^n}\exp(-\frac{u^2}{2\sigma_Z^2})du \tag{A.4}$$

according to (3.29) and (3.35).

Because Slepian-Wolf coding is near-lossless, the distortion of SWC-NQ is also D_n. Combining D_n and R through V_1 in (A.3) and (A.4), we obtain the D-R performance of SWC-NQ with a pair of n-dimensional nested lattices (Λ_1,Λ_2) as (3.46).

APPENDIX B

PROOF OF LEMMA 2

Proof: This proof closely follows remark 3) of [1, p. 3] with slight modifications. Let $\delta \triangleq \min_{\mathbf{w}\neq\hat{\mathbf{x}}} d(\mathbf{w},\hat{\mathbf{x}}) > 0$. Here δ is actually the minimum distance between points in Λ_2. Thus if $(\mathbf{X},\hat{\mathbf{X}}) \in \mathcal{S}_1$,

$$\lambda \triangleq P_r\{\mathbf{W} \neq \hat{\mathbf{X}}\} \leq \frac{1}{\delta}E[d(\mathbf{W},\hat{\mathbf{X}})] \overset{(a)}{\leq} \frac{1}{\delta}(E[d(\mathbf{W},\mathbf{X})] + E[d(\mathbf{X},\hat{\mathbf{X}})]), \tag{B.1}$$

where (a) is due to the triangle inequality. From theorem 4.1, $D = \frac{1}{n}E[d(\mathbf{X},\hat{\mathbf{X}})] = D_S + D_C$, where $D_S = \frac{1}{n}E[d(\mathbf{W},\mathbf{X})]$ is the source coding loss and D_C is the channel coding loss, then

$$\lambda \leq \frac{2nD}{\delta}. \tag{B.2}$$

Now since $\hat{\mathbf{X}}$ is a function of $\mathbf{S}$ and $\mathbf{Y}$, Fano's inequality [2, 95] implies that

$$H(\mathbf{W}|\mathbf{S},\mathbf{Y}) \leq -\lambda\log\lambda - (1-\lambda)\log(1-\lambda) + \lambda\log(|\mathcal{W}|) \triangleq \varepsilon(\lambda), \tag{B.3}$$

then

$$H(\mathbf{S}|\mathbf{Y}) \geq I(\mathbf{W};\mathbf{S}|\mathbf{Y}) = H(\mathbf{W}|\mathbf{Y}) - H(\mathbf{W}|\mathbf{S},\mathbf{Y}) \geq H(\mathbf{W}|\mathbf{Y}) - \varepsilon(\frac{2nD}{\delta}) \tag{B.4}$$

Meanwhile, from the data processing lemma [2], we have $H(\mathbf{S}|\mathbf{Y}) \leq H(\mathbf{W}|\mathbf{Y})$. At high rate, $D \to 0$ and $\varepsilon(\frac{2nD}{\delta}) \to 0$, thus $H(\mathbf{S}|\mathbf{Y}) = H(\mathbf{W}|\mathbf{Y})$.

APPENDIX C

CALCULATION OF CONDITIONAL PDF $F(E|Y_{D1})$.

Let us start from $f(\tilde{e}|y_{d1})$, where $\tilde{e} = x_{s1} - \sqrt{P_{s1}}\tanh(c_{sd}\sqrt{P_{s1}}y_{d1})$ and $e = c_{sr}\tilde{e} + z_r$. To calculate $f(\tilde{e}|y_{d1})$, let us first evaluate $P(\tilde{e} < \xi|y_{d1})$.

$$\begin{aligned}
P(\tilde{e} < \xi|y_{d1}) &= P(x_{s1} - \sqrt{P_{s1}}\tanh(c_{sd}\sqrt{P_{s1}}y_{d1}) < \xi|y_{d1}) \\
&= \sum_{x_{s1}} P(\tilde{e} < \xi|x_{s1}, y_{d1})P(x_{s1}|y_{d1}) \\
&= \zeta P(\tilde{e} < \xi|x_{s1} = \sqrt{P_{s1}}, y_{d1}) + (1-\zeta)P(\tilde{e} < \xi|x_{s1} = -\sqrt{P_{s1}}, y_{d1}) \\
&= \zeta P(1 - \tanh(c_{sd}\sqrt{P_{s1}}y_{d1}) < \frac{\xi}{\sqrt{P_{s1}}}|y_{d1}) \\
&\quad +(1-\zeta)P(-1 - \tanh(c_{sd}\sqrt{P_{s1}}y_{d1}) < \frac{\xi}{\sqrt{P_{s1}}}|y_{d1}) \\
&= \zeta P(\frac{1}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} < \frac{\xi}{2\sqrt{P_{s1}}}|y_{d1}) \\
&\quad +(1-\zeta)P(\frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} > -\frac{\xi}{2\sqrt{P_{s1}}}|y_{d1}),
\end{aligned}$$

Let us discuss $P(\tilde{e} < \xi|y_{d1})$ for 3 cases:

- When $\xi < -2\sqrt{P_{s1}}$: In this region,$P(\frac{1}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} < \frac{\xi}{2\sqrt{P_{s1}}}|y_{d1}) = 0$ because $\frac{\xi}{2\sqrt{P_{s1}}} < 0$ and $\frac{1}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} > 0$. Also $P(\frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} > -\frac{\xi}{2\sqrt{P_{s1}}}|y_{d1}) = 0$ because $-\frac{\xi}{2\sqrt{P_{s1}}} > 1$ and $\frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} < 1$. Then $P(\tilde{e} < \xi|y_{d1}) = 0$ when $\xi < -2\sqrt{P_{s1}}$.

- When $-2\sqrt{P_{s1}} < \xi < 0$: In this region, $P(\frac{1}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} < \frac{\xi}{2\sqrt{P_{s1}}}|y_{d1}) = 0$, then $P(\tilde{e} < \xi|y_{d1}) = \zeta P(\frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} > -\frac{\xi}{2\sqrt{P_{s1}}}|y_{d1})$. By dividing this region furthermore, we have $P(\tilde{e} < \xi|y_{d1}) = 0$ when $-2\sqrt{P_{s1}} < \xi < \frac{-2\sqrt{P_{s1}}}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})}$, and $P(\tilde{e} < \xi|y_{d1}) = 1 - \zeta$ when $\frac{-2\sqrt{P_{s1}}}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} < \xi < 0$.

- When $\xi > 0$: In this region, $P(\frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} > -\frac{\xi}{2c_{sd}\sqrt{P_{s1}}}|y_{d1}) = 1$ because $\frac{1}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} > 0$ and $-\frac{\xi}{2\sqrt{P_{s1}}} < 0$. Then $P(\tilde{e} < \xi|y_{d1}) = \zeta P(\frac{1}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} < \frac{\xi}{2\sqrt{P_{s1}}}|y_{d1}) + (1-\zeta)$. By dividing this region furthermore, we have $P(\tilde{e} < \xi|y_{d1}) = 1-\zeta$ when $0 < \xi < \frac{2\sqrt{P_{s1}}}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})}$, and $P(\tilde{e} < \xi|y_{d1}) = \zeta + (1-\zeta) = 1$ when $\xi > \frac{2\sqrt{P_{s1}}}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})}$.

Summarily, we get $P(\tilde{e} < \xi|y_{d1})$ as

$$P(\tilde{e} < \xi|y_{d1}) = \begin{cases} 0 & \text{if } \xi < \frac{-2\sqrt{P_{s1}}}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} \\ 1-\zeta & \text{if } \frac{-2\sqrt{P_{s1}}}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})} < \xi < \frac{2\sqrt{P_{s1}}}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} \\ 1 & \text{if } \xi > \frac{2\sqrt{P_{s1}}}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})} \end{cases} \tag{C.1}$$

From (C.1) we can get $f(\tilde{e}|y_{d1})$ as a function consists two pulses at locations $\frac{-2\sqrt{P_{s1}}}{1+\exp(-2c_{sd}\sqrt{P_{s1}}y_{d1})}$ and $\frac{2\sqrt{P_{s1}}}{1+\exp(2c_{sd}\sqrt{P_{s1}}y_{d1})}$, with magnitudes $1-\zeta = \frac{1}{1+\exp(2\gamma y_{d1})}$ and $\zeta = \frac{1}{1+\exp(-2\gamma y_{d1})}$, respectively. Then we can write $f_{\tilde{E}}(\tilde{e}|y_{d1})$ as

$$f_{\tilde{E}}(\tilde{e}|y_{d1}) = (1-\zeta)\delta(\tilde{e} + 2\sqrt{P_{s1}}\zeta) + \zeta\delta(\tilde{e} - 2\sqrt{P_{s1}}(1-\zeta)). \tag{C.2}$$

Since $e = c_{sr}\tilde{e} + z_r$, we get the conditional distribution of E given y_{d1} as

$$\begin{aligned} f_E(e|y_{d1}) &= \frac{1}{c_{sr}} f_{\tilde{E}}(\frac{u}{c_{sr}}|y_{d1}) * f_{Z_r}(u) \\ &= \frac{1-\zeta}{\sqrt{2\pi}} \exp(-\frac{(e + 2c_{sr}\sqrt{P_{s1}}\zeta)^2}{2}) + \frac{\zeta}{\sqrt{2\pi}} \exp(-\frac{(e - 2c_{sr}\sqrt{P_{s1}}(1-\zeta))^2}{2}). \end{aligned} \tag{C.3}$$

VITA

Zhixin Liu received the B.S. and M.S. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, P.R.China, in 1999 and 2001 respectively. In 2007, he received the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX.

From 2002 to 2006, he was a Research Assistant with the Department of Electrical and Computer Engineering at Texas A&M University. In August 2007, he was a Teaching Assistant with the same department. He once served as a reviewer for GlobalComm. His research interests include joint source-channel coding, signal processing, coding theory, digital communications, and information theory.

In his graduate career, he had 2 journal papers and 5 conference papers published.

The typist for this thesis was Zhixin Liu.