Practical LDGM-based Multiple Description Coding

PRACTICAL LDGM-BASED MULTIPLE DESCRIPTION CODING

BY

YING ZHANG, M.Sc.

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2011)	McMaster University
(Electrical & Computer Engineering)	Hamilton, Ontario, Canada

TITLE:	Practical LDGM-based Multiple Description Coding
AUTHOR:	Ying Zhang M.Sc., (Electrical Engineering) Tianjin University, Tianjin, P.R.China
SUPERVISOR:	Dr. Sorina Dumitrescu
NUMBER OF PAGES:	x, 113

Abstract

This thesis presents two practical coding schemes based on low-density generator matrix (LDGM) codes, for two cases of the multiple description (MD) problem. The first one is for the two description problem for finite-alphabet sources with Hamming distortion measure. The proposed MD code targets the Zhang-Berger region, which is the best inner bound known so far for the corresponding MD rate-distortion region. The coding scheme can be regarded as a practical implementation of a theoretical sequential coding system for the corner points of a related rate-region, where the random codebooks are replaced by multilevel LDGM codebooks and the encoding at each stage is performed via a message passing algorithm. This coding system is further applied in three notable cases: 1) no excess sum-rate case for binary sources; 2) successive refinement for general finite-alphabet sources, 3) no excess marginal rate for uniform binary sources. Furthermore, in order to assist the code design in the no-excess sum-rate case for binary sources, the exact expression of the distortion region and of the auxiliary variables needed to achieve its boundary, are derived, which is another important contribution of the thesis.

The second proposed MD code is for the case of L descriptions with individual and central distortion constraints, for the memoryless Gaussian source with squared distortion measure. It is shown first that the coding problem for an arbitrary point on the dominant face of an L-description El Gamal-Cover (EGC) rate region, can be converted to that for a vertex of a K-description EGC rate region for some $K \leq 2L-1$, where the latter problem can be solved via successive coding. The practical coding scheme reduces each successive coding step to a Gaussian quantization operation, and implements this operation using multilevel LDGM codes.

The LDGM-based coding schemes are extensively tested in practice for all aforementioned cases. The experimental results show very good performance, verifying that the proposed schemes can approach the theoretical rate-distortion bounds.

Acknowledgements

I would like to take this chance to thank everyone who supports me, and loves me, and whom I love, in my personal and academic life.

My deepest gratitude goes first and foremost to my supervisor Dr. Sorina Dumitrescu who gave me lots of instructive advice and suggestions. I feel grateful for her help by spending tremendous time to walk me through all the stages in the past four years. In the meantime, she showed me the attitude of a professional and most importantly, of a hard working individual. Without her I would not have been able to experience this journey, a journey in which she guided me from knowing nothing to the completion of this thesis.

Another important person in my academic path is Dr. Jun Chen. He revealed to me the magic of information theory through his formal and informal lectures, which enlightened my understanding. Thanks for his encouragement that has helped me to get through my difficult times as well.

Thanks also goes to Dr. Xiaolin Wu and to the external examiner Dr. Vladimir Stankovic, who have spent their valuable time reviewing this thesis and providing suggestions.

Last but not least, I would like to thank my parents and my fiance, for their unconditional love, sacrifice, and support. Their respect for my choice made my way through these years. And thanks to all my friends who share their life with me, through happiness or sadness. I will miss the time we had together.

Contents

List of Figures

Chapter 1

Introduction

In the multiple description (MD) problem, a source sequence is compressed into two or more descriptions, which are constructed in such a way that an adequate reconstruction of the source is possible based on each description while any larger number of descriptions combined together can lead to better reconstruction quality.

In this thesis, we present two practical coding schemes based on low-density generator matrix (LDGM) codes for the MD problem. Interestingly, although the MD coding problem itself is a source coding problem, it was first posed at Bell Laboratories in the 1970s in the context of a communication problem due to unavoidable outages of transmission links over the telephone network [1]. As time went by, the circuit-switched telephone network was replaced by the packet-switched network—the Internet. The Internet protocol provides best-effort service, in other words it does make the best effort to transfer the packets, but it takes no action when packets are lost. Therefore, in order to achieve reliable transmission, several approaches were established to combat the packet loss, with MD being one of them. In the first section of this chapter, we compare these approaches, and highlight the benefit of MD coding over the others. The following section briefly reviews previous work on practical MD codes and the use of LDGM codes for compression. Section 1.3 presents the information theoretical formulation of the MD problem for the case of two descriptions and L descriptions with individual and central distortion constraints, and relevant theoretical results. The contribution of this thesis is described in section 1.4, and the last section presents the thesis organization.

1.1 Why use Multiple Description Coding?

In practical multimedia communication applications, the reliable data transmission over best-effort packet-switching networks such as the Internet, is quite challenging. The data messages are first segmented into packets of fixed size before transmission. A header which is used by the network to route the packet is added to the front of each packet. The packets are sent over the network and stored at intermediate nodes before being forwarded to other nodes through internal network links. If a packet is not received within a specific time duration, then it is assumed to be lost. Packet loss is the main factor of unreliable delivery occurrence on the Internet. Some packets have to be dropped in the following scenarios: 1) the transmission rate does not match the link capacity; 2) router buffers with limited size at a time, overflow and are marked as congested by Transmission Control Protocol (TCP); 3) the dynamic routing mechanism causes dynamic transmission delay. Due to network topology complexity, packet loss may happen at any part of the communication process and at different points in the network. Packet loss becomes unpredictable and random in this case. To alleviate the effect of packet loss, several well-known technologies incorporated in error control transmission systems aim to protect media data.

One way to handle the packet loss is the automatic repeat request (ARQ) protocol, which is a peer-to-peer protocol providing reliable transmission. The simplest case of ARQ includes acknowledgements (ACK)s, which acknowledge the successful receipt of packets, and a time-out mechanism. The receiver transmits an ACK to inform the sender that it has received the packet, and it does nothing if it fails to receive the packet. If the ACK can not be received by the sender due to time-out or receipt failure by the receiver, then the packet will be retransmitted. ARQ is very effective in the case of low loss rate, but when the loss rate is relatively high, each retransmission can cause longer delay, at least of one round-trip time, and increase the congestion which will introduce large bandwidth overhead. Although ARQ is very useful in point-to-point transmission, it requires a feedback channel, therefore, it is infeasible for multicast, broadcast and point-to-point transmission without a feedback channel.

The second approach is forward error correction (FEC) with erasure correcting codes, such as Reed-Solomon (RS) codes. The basic idea of FEC is to exploit an error recovery mechanism to recover all transmitted data without sending a request for missing packets. In FEC with systematic codes, first k redundant data packets are generated from the n original data packets, then they are inserted into the original data packets to form a block/frame corresponding to one RS code. If the number of lost packets among a block is not greater than k, then the original data packets can be completely recovered by the RS code. FEC needs no retransmission, thus it is suitable for no feedback channel and short delay transmission cases. But if the number of lost packets among a block is greater than k, eg. burst packet loss, the whole block data can not be recovered, which results in awful final decoded quality. Because of this, the number of redundant packets usually is set for the worst case scenario, which in turn will increase the bandwidth overhead and cause longer delay.

The third approach is packet loss concealment (PLC), which estimates the lost data packets at the destination side, then provides data packets for the decoder if the actual packets are lost, via particular methods, such as repeat/replay last received packet, or implement interpolation to replace the lost packet. This is a best effort technique, which does not offer a quality guarantee.

Finally, another approach is multiple description coding (MDC). MDC was proposed as a source coding scheme which is very robust against high loss rate. In MDC, the source data is encoded into several streams, which are called descriptions, and each description is transmitted to the same destination through the network. The more descriptions received, the higher reconstruction quality achieved. The robustness of MDC lies in the fact that all descriptions have correlation with one another so that even if some of the descriptions are lost, the source can still be reconstructed with acceptable quality from the available descriptions. It is unlikely that all descriptions are lost, therefore it is impossible to have the cliff effect with MDC.

1.2 Previous Work

Due to its robustness, flexibility and wide range of applications for transmission networks, MDC has been intensively investigated. Some popular practical coding techniques are MD quantization, MD based on correlating transforms and PET-MD (MD based on priority encoding transmission). In MD quantization, separate descriptions are created by using quantizers, e.g., scalar quantizers [2–9], trellis-coded quantizers [10], vector or lattice vector quantizers [11–16], or Delta-Sigma quantizers [17]. In MD based on correlated transforms [18, 19], transform methods are used to introduce correlation between descriptions. PET-MD is obtained by combining a progressive source code with uneven erasure protection [20]. The related redundancy allocation problem is addressed in [21–24], while in [25, 26], an improvement to PET-MD is proposed by dissociating the single progressive stream into independent sub-streams.

As highlighted above, the existing practical code designs for MDC mostly focused on the case of continuous-alphabet sources with squared error distortion, while the case of discrete-alphabet sources with Hamming distortion measure has been neglected so far. Even for the former case, the performance of existing MD code designs can not approach the known MD rate-distortion inner bounds arbitrarily close. Motivated by this, this thesis addresses the problem of practical MD code construction to approach the known fundamental limits. Inspired by the success of low density generator matrix (LDGM) codes for the single description problem, we apply them in our proposed MD coding schemes. Next we present a brief review of the prior use of LDGM codes in source coding.

LDGM codes are dual code of low density parity check (LDPC) codes, which were first introduced by Gallager [27] in 1963, along with a message passing decoding algorithm. Later Tanner introduced in [28] a bipartite graph model to represent LDPC codes. This model facilitated a new way of interpreting the message passing algorithm by iteratively decomposing the entire process into a series of partial subprocesses at a type of nodes level. The success of sparse-graph codes like turbo and LDPC codes in channel coding, which were proven to approach the channel capacity under optimal decoding, has generated interest in applying sparse-graph codes for compression. The authors of [29, 30], construct LDPC, respectively, dual LDPC codes (i.e., LDGM codes), and prove that they approach the rate distortion limit for the binary symmetric source [29], respectively binary erasure source [30], by using the maximum likelihood channel decoder as the optimal source encoder. In [31], the authors propose a hybrid LDPC-LDGM code and prove that it approaches the rate distortion bounds for the uniformly distributed source if optimum encoding is used. Later, the authors of [32], prove that an LDGM-based code construction, termed multilevel LDGM code, can achieve the rate distortion bound for general finite alphabet memoryless sources with general distortion measure under optimum encoding as well. A condition for this is that the number of 1's in each row of the generator matrix is unbounded as the source block length n grows, but it scales sub-linearly with n.

All established optimality results for LDGM-based source codes hinge on optimal encoding, in other words selecting the reconstruction sequence which minimizes the distortion. Unfortunately, optimal encoding is infeasible in practice. On the other hand, the simple belief propagation (BP) algorithm, which is successful in channel decoding, is not satisfactory in LDGM-based source coding. To circumvent this problem, the more sophisticated survey propagation algorithm (SP), originally introduced for solving random satisfiability problems, was adopted in [33, 34], showing excellent practical performance. Interestingly, it was shown in [35,36] that SP can be regarded as a form of BP with decimation, where the underlying probability distribution is defined on a larger space of so-called generalized codewords. Alternatively, in [37], the authors verify empirically that the simpler BP with decimation, where the underlying probability distribution is defined only on the set of normal codewords, has very good practical performance.

1.3 Multiple Description Coding From Information Theory Perspective

In this section we provide a brief background on the MD problem from information theoretical point of view. First we review the fundamental result on rate distortion trade-off in single description source coding. Then we present the MD problem formulation and the relevant achievable rate distortion regions for the case of two descriptions, and for the L-description case with individual and central distortion constraints.

1.3.1 Single Description Source Coding

Consider a sequence $X^n = (X(1), X(2), \cdots, X(n))$ drawn from an i.i.d. source X with generic distribution p_X. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be a distortion measure, where $\mathcal{X}$ and $\hat{\mathcal{X}}$ are the source alphabet and the reconstruction alphabet, respectively. Assume that the alphabets $\mathcal{X}$ and $\hat{\mathcal{X}}$ are finite. The rate distortion pair (R, D) is said to be achievable, if for any $\epsilon > 0$ and any n sufficiently large, there exists an encoding mapping

$$f^{(n)} : \mathcal{X}^n \to \{1, 2, \cdots, \lfloor 2^{nR+\epsilon} \rfloor\}$$

and a decoding function

$$g^{(n)} : \{1, 2, \cdots, \lfloor 2^{nR+\epsilon} \rfloor\} \to \hat{\mathcal{X}}^n$$

such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{l=1}^{n} d(X(l), \hat{X}(l))\right] \leq D + \epsilon$$

where $\hat{X}^n = g^{(n)}(f^{(n)}(X^n))$.

The rate distortion region is the closure of the set of all achievable (R, D) pairs. The rate distortion function $R(D)$ is the infimum of rates R such that (R, D) is in the rate distortion region, for a given distortion D. The source coding theorem [38] states that

$$R(D) = \min_{p(\hat{x}|x)} I(X; \hat{X}),$$

where the minimum is taken over all test channels $p(\hat{x}|x)$ which satisfy the distortion constraint $\mathbb{E}d(X, \hat{X}) = p(x)p(\hat{x}|x)d(x, \hat{x}) \leq D$.

1.3.2 Multiple Description Problem

The multiple description (MD) source coding problem was first posed by Wolf, Wyner, Ziv, Ozarow and Witsenhausen at the 1979 IEEE Information Theory Workshop. Some of the early results on this problem appear in [39–43]. Since then various information theoretical approaches have been exploited in an attempt to characterize the MD rate-distortion region, but the complete characterization has not been found yet, not even for the case of two descriptions. In this thesis we restrict our attention to the two description case, and L-description case with individual and central distortion constrains. Note that the former is a special case of the latter, therefore we present the MD problem for the latter case.

The formal definition of the L-description problem with individual and central distortion constraints is given as follows. The rate-distortion tuple $(R_1, \cdots, R_L, d_1, \cdots, d_L, d_{\{1,\cdots,L\}})$ is said to be achievable, if for any $\epsilon > 0$ and all sufficiently large n, there exist encoding function $f_\ell^{(n)} : \mathcal{X}^n \to \{1, \cdots, \lfloor 2^{n(R_\ell+\epsilon)} \rfloor\}$, $\ell = 1, \cdots, L$, and decoding functions $g_{\{\ell\}}^{(n)} : \{1, \cdots, \lfloor 2^{n(R_\ell+\epsilon)} \rfloor\} \to \hat{\mathcal{X}}^n$, $\ell = 1, \cdots, L$, and $g_{\{1,\cdots,L\}}^{(n)} : \prod_{\ell=1}^{L} \{1, \cdots, \lfloor 2^{n(R_\ell+\epsilon)} \rfloor\} \to \hat{\mathcal{X}}^n$ such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d(X(t), \hat{X}_{\{\ell\}}(t))\right] \leq d_{\{\ell\}} + \epsilon, \quad \ell = 1, \cdots, L, \tag{1.1}$$

$$\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d(X(t), \hat{X}_{\{1,\cdots,L\}}(t))\right] \leq d_{\{1,\cdots,L\}} + \epsilon, \tag{1.2}$$

where $\hat{X}_{\{\ell\}}^n = g_{\{\ell\}}^{(n)}(f_\ell^{(n)}(X^n))$, $\ell = 1, \cdots, L$, and $\hat{X}_{\{1,\cdots,L\}}^n = g_{\{1,\cdots,L\}}^{(n)}(f_1^{(n)}(X^n), \cdots, f_L^{(n)}(X^n))$. Finally, the rate-distortion region is defined as the closure of the set of all achievable tuples $(R_1, \cdots, R_L, d_1, \cdots, d_L, d_{\{1,\cdots,L\}})$.

1.3.3 Inner Bounds for the Two Description Rate-Distortion Region

Let us denote by $\mathcal{RD}$ the rate-distortion region in the two description case. The first inner bound of $\mathcal{RD}$, referred to as the EGC region, was provided by El Gamal and Cover in [41]. This was shown to be tight for the no excess sum-rate case by Ahlswede [43]. Later Zhang and Berger [42] proposed a different inner bound, termed the ZB region, and showed that it contains points not included in the EGC region. Recent work of Wang *et al.* [44] proves that the ZB region includes the EGC region and establishes the ZB region as the best inner bound known to date.

The EGC region, denoted by $\mathcal{RD}_{\mathrm{EGC}}$, is defined as the convex closure of the set of quintuples $(R_1, R_2, d_1, d_2, d_0)$, where $d_0 = d_{\{1,2\}}$, for which there exist auxiliary random variables X_t, $t = 0, 1, 2$, jointly distributed with X such that

$$
\begin{aligned}
&R_t \geq I(X; X_t), \quad t = 1, 2, \\
&R_1 + R_2 \geq I(X; X_0, X_1, X_2) + I(X_1; X_2), \\
&\mathbb{E}[d(X, X_t)] \leq d_t, \quad t = 0, 1, 2.
\end{aligned}
$$

It was proved in [41] that $\mathcal{RD}_{\mathrm{EGC}} \subseteq \mathcal{RD}$. Moreover, it was shown by Ahlswede [43] that the previous relation holds with equality in the no excess sum-rate case.

The ZB region, denoted by $\mathcal{RD}_{\mathrm{ZB}}$, is defined as the set of quintuples $(R_1, R_2, d_1, d_2, d_0)$ such that there exist random variables X_c, X_1, X_2 jointly distributed with X and functions ψ_t, $t = 0, 1, 2$, satisfying

$$
\begin{aligned}
&R_t \geq I(X; X_c, X_t), \quad t = 1, 2, \qquad (1.3)\\
&R_1 + R_2 \geq 2I(X; X_c) + I(X; X_1, X_2 | X_c) + I(X_1; X_2 | X_c), \\
&\mathbb{E}d(X, \psi_t(X_c, X_t)) \leq d_t, \quad t = 1, 2, \\
&\mathbb{E}d(X, \psi_0(X_c, X_1, X_2)) \leq d_0.
\end{aligned}
$$

Let $\mathcal{RD}_{\mathrm{EGC}^*}$ denote the convex closure of the set of quintuples $(R_1, R_2, d_1, d_2, d_0)$ for which there exist random variables X_1 and X_2 jointly distributed with X and functions $\psi_t(\cdot)$, $t = 0, 1, 2$, such that

$$
\begin{aligned}
&R_t \geq I(X; X_t), \quad t = 1, 2, \qquad (1.4)\\
&R_1 + R_2 \geq I(X; X_1, X_2) + I(X_1; X_2),\\
&\mathbb{E}[d(X, \psi_t(X_t))] \leq d_t, \quad t = 1, 2,\\
&\mathbb{E}[d(X, \psi_0(X_1, X_2))] \leq d_0.
\end{aligned}
$$

$\mathcal{RD}_{\mathrm{EGC}^*}$ is referred to as the EGC* region and is attributed to El Gamal and Cover as well. Although $\mathcal{RD}_{\mathrm{EGC}^*}$ appears to be weaker than $\mathcal{R}D_{\mathrm{EGC}}$, it was shown in [44] that the two regions are equivalent and both of them are contained in $\mathcal{RD}_{\mathrm{ZB}}$, i.e., $\mathcal{RD}_{\mathrm{EGC}^*} = \mathcal{RD}_{\mathrm{EGC}} \subseteq \mathcal{RD}_{\mathrm{ZB}}$.

1.3.4 EGC Region for L-Description Case

Recall that the EGC region is a general inner bound of the two description rate-distortion region. Ozarow [39] proved that the EGC region is tight in the quadratic Gaussian case. In fact, it has been shown, by refining and generalizing Ozarow's proof technique, that a natural extension of the EGC region to the L-description case is tight for Gaussian multiple description coding with individual and central distortion constraints [45–47].

Define the EGC region for the L-description case with individual and central distortion constraints, as the convex closure of the set of tuples $(R_1, \cdots, R_L, d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$ such that there exist L auxiliary random variables $U_{\{1\}}, \cdots, U_{\{L\}}$ jointly distributed with the generic source variable X, and functions $g_{\{\ell\}}$, $\ell = 1, \cdots, L$, and $g_{\{1,\cdots,L\}}$, satisfying

$$\sum_{\ell \in \mathcal{A}} R_\ell \geq \sum_{\ell \in \mathcal{A}} H(U_{\{\ell\}}) - H(U_{\{\ell\}}, \ell \in \mathcal{A}|X), \quad \emptyset \subset \mathcal{A} \subseteq \{1, \cdots, L\}.$$

and

$$\mathbb{E}[d(X, g_{\{\ell\}}(U_{\{\ell\}}))] \leq d_{\{\ell\}}, \quad \ell = 1, \cdots, L,$$
$$\mathbb{E}[d(X, g_{\{1,\cdots,L\}}(U_{\{1\}}, \cdots, U_{\{L\}}))] \leq d_{\{1,\cdots,L\}}$$

Notice that technically, the above definition corresponds to the EGC* region, however, in view of [44], the two regions coincide.

1.4 Contribution

In this thesis, we present two practical coding schemes based on LDGM codes, for the MD problem for two important cases. The first coding scheme is for the case of two descriptions, finite-alphabet sources and Hamming distortion measure. It consists of three sequential encoders, one for the common part of the two descriptions, and the other two for the remaining component of the first, respectively second description. This scheme is applicable to corner points of a certain rate region derived from the ZB rate-distortion region of two description, but this is not a limitation since any other rate pair can be obtained through timesharing of two such corner points. At each encoding stage a multilevel LDGM code is used to generate each codebook, whose construction needs the knowledge of the auxiliary variables involved in the ZB region. Finally, message passing algorithm on the associated factor graph is employed to perform the encoding at each stage. We discuss in detail the application of this coding scheme to the following cases of the two description problem with Hamming distortion measure.

1. No excess sum-rate case for binary sources. In this case the coding scheme is simplified consisting only of two sequential encoders. The characterization of the MD region is available, as the EGC region, therefore the rate-distortion limits are computable. However, their computation is not an easy task, and analytical expressions are only partly available for the uniform binary source. In order to aid the code design we provide the exact expression of the distortion region and of the auxiliary variables necessary to achieve its boundary. This result is another important contribution of this thesis.

2. No excess marginal rate for the uniform binary source. For this case Zhang and Berger proposed in [48] a method to compute an upper bound for the central distortion given the values of the side distortions. We improve Zhang and Berger's result by providing the analytical expression of this upper bound.

3. Successive refinement (SR) for general finite-alphabet sources. This is another special case of the MD problem, where the fundamental limits are known [49–51]. In this case the second description is of no interest alone, but only in conjunction with the first description. Thus, the first description constitutes the base layer, while the second description is the refinement layer. However, practical SR coding for the discrete source with Hamming distortion has not been attempted so far. In this thesis, we also show how the proposed LDGM-based coding scheme can be adapted to the SR problem.

The second coding scheme is for the case of L descriptions with individual and central distortion constraints, for Gaussian sources with squared error distortion. This scheme can be regarded as an alternative implementation of the Gaussian two description code proposed in [52] with lattice codes replaced by multilevel LDGM codes, as well as the extension to the general L-description case. Although we follow the general strategy of [52, 53] by reducing the Gaussian multiple description problem to a sequence of Gaussian quantization problems via Gram-Schmidt orthogonalization, there are several noteworthy conceptual differences.

1. We exploit the special structure of the covariance matrix associated with the sum-rate optimal EGC region, which not only simplifies the calculation of the coeffficients, but also leads to an efficient implementation of Gram-Schmidt orthogonalization.

2. We give a new interpretation of the quantization splitting method developed in [52, 53] by eliminating the use of conditional codebooks.

Finally, the effectiveness of all proposed schemes is verified through extensive simulation experiments.

1.5 Organization of the Thesis

The thesis is organized as follows. In Chapter 2 we describe the proposed coding scheme for the ZB region for the two description case. First the theoretical sequential coding scheme based on random codebooks is presented. Then the details of the multilevel LDGM codes used at each stage are addressed, together with the associated factor graph and message passing encoding algorithm. Chapter 3 presents the application of the proposed coding system for three cases of the two description problem with Hamming distortion measure, including the no excess sum-rate case for general binary sources, no excess marginal rate case for the uniform binary source and successive refinement for finite-alphabet sources. A notable result presented in this chapter is the analytical expression of the distortion region in the no excess sum-rate case for binary sources with Hamming distortion. Chapter 4 describes the proposed LDGM-based successive coding scheme for Gaussian quadratic L description coding with individual and central distortion constraints. Experimental results for all proposed codes are presented in Chapter 5. Finally, Chapter 6 concludes this thesis.

Chapter 2

Practical LDGM-based Coding Scheme for the ZB Region

Although the question whether $\mathcal{RD}_{\text{ZB}}$ is tight is an open problem, $\mathcal{RD}_{\text{ZB}}$ is the best inner bound known so far for $\mathcal{RD}$. Therefore, in this chapter we propose a practical coding scheme based on LDGM codes, tailored for the ZB region. The chapter first presents a theoretical sequential coding system using random codebooks. Next the use of multilevel LDGM codes is proposed to generate the codebooks. The previous result of [32] implies the asymptotic optimality of this scheme when used in conjunction with the strong typicality encoding rule. However, such an encoder is impractical, therefore an efficient suboptimal encoding algorithm is proposed. For this the encoder problem is first formulated as an unconstrained optimization problem similar in spirit to the problem in the single description case. Then, the belief propagation algorithm with decimation is presented as a heuristic solution. Finally, an upper bound on the performance gap of the algorithm is established, in terms of the error introduced in the computation of marginal probabilities. The chapter ends with a discussion of the efficient implementation of multilevel LDGM codes needed in the scheme.

2.1 Sequential Coding System

Given the auxiliary random variables X_c, X_1, X_2 jointly distributed with X, specified by the ZB region, i.e. obeying relations (1.3), the distortion triple $(d_{\{1\}}, d_{\{2\}}, d_{\{1,2\}})$ can be achieved by the set of rate pairs (R_1, R_2) satisfying

$$
\begin{aligned}
&R_1 \geq I(X; X_c, X_1), \\
&R_2 \geq I(X; X_c, X_2), \\
&R_1 + R_2 \geq 2I(X; X_c) + I(X; X_1, X_2|X_c) + I(X_1; X_2|X_c).
\end{aligned}
$$

Let us denote by $\mathcal{R}(p_{X_cX_1X_2|X})$ this rate region, and by $\mathcal{F}(p_{X_cX_1X_2|X})$ its dominant face, which is defined as the set of rate pairs $(R_1, R_2) \in \mathcal{R}(p_{X_cX_1X_2|X})$ for which $R_1 + R_2 = 2I(X; X_c) + I(X; X_1, X_2|X_c) + I(X_1; X_2|X_c)$. Without loss of generality, we shall focus on $\mathcal{F}(p_{X_cX_1X_2|X})$ since every point in $\mathcal{R}(p_{X_cX_1X_2|X})$ is dominated, in a componentwise sense, by some point $(R_1, R_2) \in \mathcal{F}(p_{X_cX_1X_2|X})$. Moreover, the two corner points E_1 and E_2 of $\mathcal{R}(p_{X_cX_1X_2|X})$ (i.e., the end points of $\mathcal{F}(p_{X_cX_1X_2|X})$) are of particular importance. Specifically, one can achieve an arbitrary point on the dominant face by timesharing E_1 and E_2. Precisely, the coordinates of E_1 are

$$
\begin{aligned}
R_1 &= I(X; X_c) + I(X; X_1|X_c) \\
R_2 &= I(X; X_c) + I(X; X_2|X_c, X_1) + I(X_1; X_2|X_c) \\
&= I(X; X_c) + I(X, X_1; X_2|X_c).
\end{aligned}
$$

The coordinates of E_2 are obtained by swapping indices 1 and 2 in the above relations.

Next we present a sequential coding scheme for the corner point E_1. This system consists of three encoders, labeled $0, 1$ and 2. We mention that a similar sequential coder, consisting of only two encoders, was described in [52] for the corner rate points related to the EGC region.

Codebook generation. Encoder 0 randomly generates a codebook $\mathcal{C}_0 = \{x_{c,i}^n\}_{i=1}^{2^{n(I(X;X_c)+\epsilon_1)}}$ according to the distribution $\prod_{l=1}^n p_{X_c}(\cdot)$, where ϵ_1 depends on ϵ and $\epsilon_1 \to 0$ as $\epsilon \to 0$. For each index $i, 1 \leq i \leq 2^{n(I(X;X_c)+\epsilon_1)}$ in $\mathcal{C}_0$, encoder 1 randomly generates a conditional codebook $\mathcal{C}_{1,i} = \{x_{1,i,j}^n\}_{j=1}^{2^{n(I(X;X_1|X_c)+\epsilon_1)}}$ according to the conditional distribution $\prod_{l=1}^n p_{X_1|X_c}(\cdot|x_{c,i}(l))$. Further, for each index $i, 1 \leq i \leq 2^{n(I(X;X_c)+\epsilon_1)}$ in $\mathcal{C}_0$, encoder 2 randomly generates a conditional codebook $\mathcal{C}_{2,i} = \{x_{2,i,k}^n\}_{k=1}^{2^{n(I(X,X_1;X_2|X_c)+\epsilon_1)}}$ according to the conditional distribution $\prod_{l=1}^n p_{X_2|X_c}(\cdot|x_{c,i}(l))$.

Encoding. Given the source sequence x^n, encoder 0 finds a codeword $x_{c,i}^n$ which is jointly strongly typical with x^n, if such a codeword exists.

Encoder 1 has knowledge of index i picked by encoder 0, and it chooses an index j such that the codeword $x_{1,i,j}^n$ is jointly strongly typical with $(x^n, x_{c,i}^n)$, if possible.

Encoder 2 has knowledge of both indices i and j picked by encoders 0 and 1, and it chooses an index k such that the codeword $x_{2,i,k}^n$ is jointly strongly typical with $x^n, x_{c,i}^n$ and $x_{1,i,j}^n$, if such a codeword exists.

The pair of indices (i, j) forms description 1, and the pair (i, k) forms description 2.

Decoding. Decoder 1 receives the index pair (i, j) and takes $g_1^{(n)}(i, j) = (\psi_1(x_{1,i,j}^n(l)))_{l=1}^n$ as the reconstruction of x^n.

Decoder 2 receives index pair (i, k) and takes $g_2^{(n)}(i, k) = (\psi_2(x_{2,i,k}^n(l)))_{l=1}^n$ as the reconstruction of the source sequence x^n.

The central decoder obtains the reconstruction of x^n according to decoding function $g_0^{(n)}(i,j,k) = (\psi_0(x_{c,i}^n(l), x_{1,i,j}^n(l), x_{2,i,k}^n(l)))_{l=1}^n$.

Since the probability that the encoder finds a triple of indices (i,j,k) such that $x_{c,i}^n$, $x_{1,i,j}^n$, $x_{2,i,k}^n$ are jointly strongly typical with x^n, approaches 1 as $n \to \infty$, it follows that relations (1.1) and (1.2) are satisfied for sufficiently large n.

Following the above theoretical coding system we propose a practical coding scheme for the corner rate points E_1 and E_2. At each encoding stage a multilevel LDGM code is used to generate the codebook, and a message passing algorithm is used to perform the encoding. Multilevel LDGM codes were used in [32] to generate codebooks with codewords of non-uniform empirical distribution and it was shown that they are able to achieve the rate-distortion bound for single description coding when used with the strong typicality encoding rule. Next we describe the multilevel LDGM code employed at various encoding stages, by considering a general encoding scenario which covers all three encoders $0, 1, 2$.

2.2 Multilevel LDGM Code for General Encoding Scenario

Let Y and Z be jointly distributed random variables over the alphabets $\mathcal{Y}$ and $\mathcal{Z}$, respectively. Consider the probability distributions $p_1(\cdot), \cdots, p_n(\cdot)$ over the alphabet $\mathcal{Z}$. Given the input sequence y^n, the goal of the encoder is to select a binary sequence v^m such that $z^n(v^m)$ is jointly strongly typical with y^n. The first key step of encoding is generating a codebook $\mathcal{C} = \{z^n(v^m) | v^m \in \{0,1\}^m\}$, where $m = Rn$, and $R > 0$, such that the marginal empirical distribution of the l-th symbol z_l approximates the distribution $p_l(\cdot)$, for all $1 \leq l \leq n$.

To see that this general setting covers all three encoders $0, 1, 2$, note that for encoder 0, the variables Y and Z are X and X_c, respectively, and the distributions $p_1(\cdot), \cdots, p_n(\cdot)$ are identical to $p_{X_c}(\cdot)$. The input sequence is the source sequence, i.e., $y^n = x^n$, and $R = I(X; X_c) + \epsilon_1$. For encoder 1, we use different LDGM codes for different indices i selected by encoder 0. In each case, the two inputs x^n and $x^n_{c,i}$ are modeled as a single input sequence $y^n \in \{\mathcal{X} \times \mathcal{X}_c\}^n$, where $y(l) = (x(l), x_{c,i}(l))$ for $1 \leq l \leq n$, and $\mathcal{X}_c$ is the alphabet of X_c. Consequently, $Y = (X, X_c)$ and $Z = X_1$. Moreover, $p_l(\cdot) = p_{X_1|X_c}(\cdot|x_{c,i}(l))$, $1 \leq l \leq n$, and $R = I(X; X_1|X_c) + \epsilon_1)$. For encoder 2, again we use a different LDGM code for each index i selected by encoder 0. In each case, we have $y^n \in \{\mathcal{X} \times \mathcal{X}_c \times \mathcal{X}_1\}^n$, with $y(l) = (x(l), x_{c,i}(l), x_{1,i,j}(l))$, $1 \leq l \leq n$, where j is the index output by encoder 1. Consequently, $Y = (X, X_c, X_1)$, $Z = X_2$, $p_l(\cdot) = p_{X_2|X_c}(\cdot|x_{c,i}(l))$ for all l and $R = I(X, X_1; X_2|X_c) + \epsilon_1$.

Let us go back to the general setting. To generate the codebook $\mathcal{C}$ we use a multi-level LDGM code constructed as follows. Consider a low-density generator matrix A of dimension $n\omega \times m$, with elements in the binary field $GF(2)$, where ω is a positive integer. The choice of the integer ω will be clarified shortly. The matrix A is used to generate $n\omega$-length bit sequences which are further mapped into n-length sequences over the alphabet $\mathcal{Z}$ via a mapping $\Phi : \{0,1\}^{n\omega} \rightarrow \mathcal{Z}^n$. Φ is defined based on some mappings $\phi_l : \{0,1\}^{\omega} \rightarrow \mathcal{Z}$, $1 \leq l \leq n$, as follows. For every $c^{n\omega} \in \{0,1\}^{n\omega}$, the l-th symbol of the sequence $\Phi(c^{n\omega})$ equals $\phi_l(c(l), c(l+n), c(l+2n), \cdots, c(l+(\omega-1)n))$. Finally, the mappings $\phi_l(\cdot)$ are constructed to ensure that the marginal distribution of the l-th symbol of the codewords $z^n \in \mathcal{C}$ approximates the distribution $p_l(\cdot)$. To this end the positive integer ω is chosen such that $2^\omega p_l(b)$ is approximately an integer for all $1 \le l \le n$ and $b \in \mathcal{Z}$. Moreover, the positive integers $q_{l,b}$ are selected such that $p_l(b) \approx \frac{q_{l,b}}{2^\omega}$ and $\sum_{b\in\mathcal{Z}} q_{l,b} = 2^\omega$, for all $1 \le l \le n$ and $b \in \mathcal{Z}$. Further, for each l, the function $\phi_l(\cdot)$ is designed to map exactly $q_{l,b}$ ω-length bit sequences to every symbol $b \in \mathcal{Z}$. Finally, the codebook $\mathcal{C}$ is defined as

$$\mathcal{C} = \{\Phi(Av^m)|v^m \in \{0,1\}^m\},$$

where Av^m is the product between the matrix A and the column vector v^m over the binary field $GF(2)$.

Based on the result of [32, Theorems 1 and 2], for a code ensemble using random generator matrices at each encoding stage, with entries selected independently according to $Ber(p_n)$, where $p_n \to 0$ and $np_n \to \infty$ as $n \to \infty$, the expected distortion at each decoder satisfies the constrains (1.1) and (1.2) asymptotically as $n \to \infty$ and $\omega \to \infty$, when the strong typicality encoder is used.

2.3 Encoder Optimization Problem

In previous work on LDGM-based source coding [32,33,37], the purpose of the encoder is, given the input sequence y^n, to select the information sequence v^m such that the distortion between the corresponding codeword $z^n(v^m)$ and the input sequence to be minimized. In other words, a cost is assigned to each v^m and the encoder's task is to solve the unconstrained optimization problem of minimizing this cost. Since the optimal solution is intractable, this problem is solved heuristically using various message passing algorithms on the associated factor graph [32–34, 37]. Specifically, the authors of [37] assign a probability to each information sequence v^m such that sequences with smaller cost have higher probability, and use belief propagation with decimation in a attempt to find the sequence with highest probability.

We will follow this idea, and therefore we need first to define a cost function meaningful for our case. Recall that the objective of the encoder is to select a codeword $z^n = z^n(v^m)$ which is jointly strongly typical with the input sequence y^n. Therefore, let us first review the definition of joint typicality. For each pair of symbols $(a, b) \in \mathcal{Y} \times \mathcal{Z}$ consider the indicator function $I_{a,b} : \mathcal{Y} \times \mathcal{Z} \to \{0, 1\}$, defined as $I_{a,b}(y, z) = 1$ if and only if $y = a$ and $z = b$. Extend this function to pairs of sequences as follows $I_{a,b}(y^n, z^n) = \sum_{l=1}^{n} I_{a,b}(y_l, z_l)$. Let us assume that $p_{YZ}(a, b) > 0$ for all $(a, b) \in \mathcal{Y} \times \mathcal{Z}$. Then, according to [38], sequences y^n and z^n are said to be jointly ϵ-strongly typical if

$$-\epsilon/\nu < I_{a,b}(y^n, z^n)/n - p_{YZ}(a, b) < \epsilon/\nu \text{ for all } (a, b) \in \mathcal{Y} \times \mathcal{Z}, \tag{2.1}$$

where ν denotes the size of $\mathcal{Y} \times \mathcal{Z}$. Consider now the set $\mathcal{S} \subseteq \mathcal{Z}^n$ of sequences z^n satisfying the following relations

$$I_{a,b}(y^n, z^n)/n - p_{YZ}(a, b) \leq \epsilon/(\nu^2) \text{ for all } (a, b) \in \mathcal{Y} \times \mathcal{Z}. \tag{2.2}$$

We will show that for any $z^n \in \mathcal{S}$, the inequalities in (2.1) are satisfied. Indeed, the second inequality in (2.1) holds obviousily. In order to prove that the first inequality is valid too, assume that there is a pair of symbols $(a_0, b_0) \in \mathcal{Y} \times \mathcal{Z}$ such that

$-\epsilon/\nu \geq I_{a_0,b_0}(y^n, z^n)/n - p_{YZ}(a_0, b_0)$. Then we obtain that

$$\begin{aligned}\epsilon/\nu &\leq 1 - I_{a_0,b_0}(y^n, z^n)/n - (1 - p_{YZ}(a_0, b_0)) \\ &= \sum_{(a,b)\in\mathcal{Y}\times\mathcal{Z}\setminus\{(a_0,b_0)\}} (I_{a,b}(y^n, z^n)/n - p_{YZ}(a, b)) \\ &\leq \frac{(\nu-1)\epsilon}{\nu^2} < \epsilon/\nu,\end{aligned}$$

which is a contradiction.

According to the above discussion, in order to ensure that sequences y^n and z^n are jointly ϵ-strongly typical, it is enough to require that inequalities (2.2) be satisfied. Intuitively, such a sequence z^n can be found by minimizing a weighted sum of the quantities $I_{a,b}(y^n, z^n)/n$, with higher weights assigned to pairs (a, b) with smaller $p_{YZ}(a, b)$, in other words, by solving the problem

$$\min_{z^n\in\mathcal{C}} \textstyle\sum_{(a,b)\in\mathcal{Y}\times\mathcal{Z}} \rho'(a, b) I_{a,b}(y^n, z^n), \tag{2.3}$$

for some nonnegative values $\rho'(a, b) \geq 0$, $(a, b) \in \mathcal{Y} \times \mathcal{Z}$. The above formulation also has an interpretation using Lagrangian multipliers as follows. Consider the problem of minimizing $I_{a_0,b_0}(y^n, z^n)/n$ for some fixed pair (a_0, b_0), subject to the constraints (2.2) . Assume that this problem has a solution z_0^n corresponding to a point on the boundary of the convex hull of the set $\mathcal{P}$, where $\mathcal{P} \triangleq \{(I_{a,b}(y^n, z^n))_{(a,b)\in\mathcal{Y}\times\mathcal{Z}} : z^n \in \mathcal{C}\}$. Then, according to [54], [55], there are nonnegative values $\rho'(a, b) \geq 0$, $(a, b) \in \mathcal{Y} \times \mathcal{Z} \setminus \{(a_0, b_0)\}$, such that z_0^n is also a solution to problem (2.3), where $\rho'(a_0, b_0) = 1$.

Guided by the previous discussion we will adopt the formulation (2.3) for the optimization problem at the encoder. Notice that a further simplification of the cost function can be obtained based on the observation that for each $a \in \mathcal{Y}$, the summation $\sum_{b\in\mathcal{Z}} I_{a,b}(y^n, z^n)$ equals $I_a(y^n)$, which is fixed. Thus, by defining $b(a)$ for each $a \in \mathcal{Y}$, to be an element of $\mathcal{Z}$ for which

$$\rho'(a, b(a)) \leq \rho'(a, b) \text{ for all } b \in \mathcal{Z},$$

and by denoting $\rho(a, b) = \rho'(a, b) - \rho'(a, b(a))$, for all $a \in \mathcal{Y}$, and $b \in \mathcal{Z} \setminus \{b(a)\}$. problem (2.3) becomes equivalent to

$$\min_{z^n\in\mathcal{C}} \sum_{a\in\mathcal{Y}} \sum_{b\in\mathcal{Z}\setminus\{b(a)\}} \rho(a, b) I_{a,b}(y^n, z^n). \tag{2.4}$$

Note that all $\rho(a, b)$ in (2.4) are nonnegative. It is easy to see that the problem of distortion minimization for single description coding is a special case of (2.4). Precisely, in such a case, the symbol $b(a)$ is such that $d(a, b(a)) = 0$, and $\lambda(a, b)$ coincides with $d(a, b)$.

Although the above argumentation is intended to support the proposed problem formulation, we were not able to prove, that such values $\rho(a, b)$ exist which to ensure that the solution to (2.4) satisfies the constraints (2.2), in other words to prove that the lower boundary of the convex hull of set $\mathcal{P}$ contains points corresponding to codewords z^n satisfying the strong typicality condition. However, this approach proved itself successful in our experiments. In order to find appropriate values of the parameters $\rho(a, b)$, we use an input training sequence, and tried different parameter values until the solution z^n obtained via the message passing algorithm, satisfied the constraints (2.1) for some small enough ϵ.

On the other hand, the strong typicality condition (2.1) imposed at the encoder may be too strong in some cases. For example, encoder 2 only needs to choose a codeword such that the distortion of the second description and of the central distortion to satisfy certain constraints. This problem can be formulated as the minimization of a weighted sum of the two distortions, which also has the form of (2.4), but with a smaller number of unknown parameters (only one parameter corresponding to the ratio of the two weights). Such a formulation simplifies our task because the number of parameters which need to be appropriately set is smaller. Whenever such simplification in the problem is possible we will take advantage of it. On the other hand, the simplified problem still fits into the general form of (2.4).

In conclusion, after having set the parameters $\rho(a, b)$, we formulate the encoder task as the task to solve problem (2.4). Next we present the factor graph and the message passing algorithm as a suboptimal solution to the problem (2.4).

2.4 Factor Graph Representation

The multilevel LDGM code is associated with a factor graph (Figure 2.1). A factor graph consists of factor (or function) nodes, variable nodes, and edges connecting factor nodes to variable nodes [56]. In our case there are three types of factor nodes: 1) $n\omega$ check nodes $C_1, \cdots, C_{n\omega}$, 2) n network nodes $N_1, \cdots, N_n$, and 3) n source nodes $S_1, \cdots, S_n$. We also distinguish between three types of variable nodes: 1) m information variable nodes $V_1, \cdots, V_m$, 2) $n\omega$ check variable nodes $CV_1, \cdots, CV_{n\omega}$ and 3) n network variable nodes $NV_1, \cdots, NV_n$. For convenience, the notations introduced in this section are summarized in Table (2.1).

Each source node S_l represents the l-th symbol $y(l)$ in the input sequence y^n. Note

Symbol	Explanation	Page
y^n	input sequence	25
v^m	information sequence(index of codeword)	26
z^n	output sequence(codeword on network nodes)	26
$c^{n\omega}$	the sequence of check nodes values	26
n	length of input sequence	25
m	number of variable nodes, $m = nR$	26
ω	number of check nodes connecting with one network node	25
A	low density generator matrix of dimension $n\omega \times m$	26
S_l	source node	25
N_l	network node	25
C_ℓ	check node	25
V_k	information variable node	25
CV_ℓ	check variable node	25
NV_l	network variable node	25
$\mathcal{CV}(\ell)$	the set of indices ℓ such that CV_ℓ is connected to N_l	26
$\mathcal{V}(\ell)$	the set of indices k such that V_k is connected to C_ℓ	26
$\mathcal{C}(\ell)$	the set of indices ℓ such that C_ℓ is adjacent to N_l	26
$f_{S_l}(.)$	the compatibility function at source node S_l	26
$f_{C_\ell}(.)$	the compatibility function at check node C_ℓ	27
$f_{N_l}(.)$	the compatibility function at network node N_l	27
$\lambda_{(.,.)}$	parameter used in the message passing algorithm	26
γ	parameter used in the message passing algorithm	26
$\rho_{(.,.)}$	parameter used in the encoding optimization problem	26
$\phi_l(.)$	mapping from adjacent check nodes to network node N_l	27
$I_{(.,.)}$	indicator function	28

Table 2.1: Table of notations used in Section 2.4 and page number where they were introduced.

that $y(l)$ is not a variable. Each information variable node V_k is associated with the variable $v(k)$ which represents the k-th bit in the information sequence v^m. Thus, the variable $v(k)$ can take values in the set $\{0,1\}$. The node V_k is connected by an edge to every check node C_ℓ such that the entry of generator matrix $A(\ell,k)=1$. Each check variable node CV_ℓ is associated with the variable $c(\ell)$ which represents the ℓ-th bit in the sequence $c^{n\omega}=Av^m$. Consequently, $c(\ell)$ takes values in $\{0,1\}$, too. Node CV_ℓ is connected by an edge to check node C_ℓ and to network node N_l where $l=((\ell-1) \bmod n)+1$. Each network variable node NV_l is associated with the variable $z(l)$ which represents the l-th symbol of the codeword z^n. Thus, the variable $z(l)$ takes values in $\mathcal{Z}$. Node NV_l is connected to the source node S_l and to the network node N_l. Note that, according to this description, each node N_l is connected to variable nodes NV_l and $CV_l, CV_{l+n}, CV_{l+2n}, \cdots, CV_{l+(\omega-1)n}$.

For each network node N_l, denote by $\mathcal{CV}(l)$ the set of indices ℓ such that CV_ℓ is connected to N_l, i.e., $\mathcal{CV}_\ell=\{l, l+n, l+2n, \cdots, l+(\omega-1)n\}$. Further, for each check node C_ℓ, denote by $\mathcal{V}(\ell)$ the set of indices k such that V_k is connected to C_ℓ. Finally, for each variable node V_k let $\mathcal{C}(k)$ denote the set of indices ℓ such that C_ℓ is adjacent to V_k.

Each factor node is assigned a function of the variables associated with the adjacent variable nodes. This function is called compatibility function. The compatibility function at source node S_l is

$$f_{S_l}(z(l)) = \begin{cases} 1, & \text{if } z(l)=b(y(l)) \\ \exp(-\lambda(y(l),z(l))), & \text{otherwise} \end{cases}, \tag{2.5}$$

for all $z(l)\in\mathcal{Z}$, $1\le l\le n$, where $\lambda(a,b)=\gamma\rho(a,b)$ for all $a\in\mathcal{Y}$ and $b\in\mathcal{Z}\setminus\{b(a)\}$,

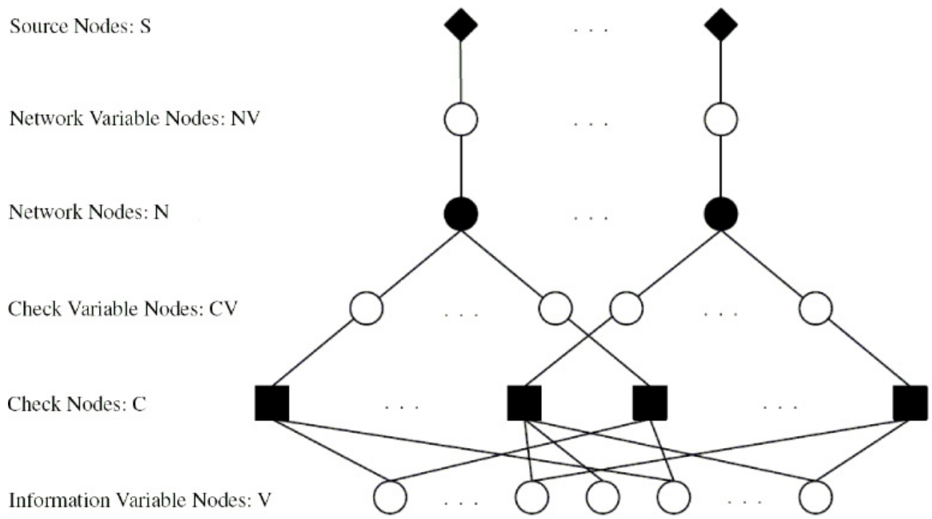

Figure 2.1: Factor graph of a multilevel LDGM code

for some positive value γ. The significance of the parameter γ will be revealed shortly.

The compatibility function at the check node C_ℓ is

$$f_{C_\ell}(c(\ell),(v(k))_{k\in\mathcal{V}(\ell)}) = \begin{cases} 1, & \text{if } c(\ell) = \oplus_{k\in\mathcal{V}(\ell)} v(k) \\ 0, & \text{otherwise} \end{cases}, \tag{2.6}$$

for all $c(\ell) \in \{0,1\}$, and $(v(k))_{k\in\mathcal{V}(\ell)} \in \{0,1\}^{\mathcal{V}(\ell)}$, where $\oplus$ denotes modulo 2 addition.

Finally, the compatibility function at each network node N_l is

$$f_{N_l}(z(l),(c(\ell))_{\ell\in\mathcal{CV}(l)}) = \begin{cases} 1, & \text{if } z(l) = \phi_l((c(\ell))_{\ell\in\mathcal{CV}(l)}) \\ 0, & \text{otherwise} \end{cases}, \tag{2.7}$$

for all $z(l) \in \mathcal{Z}$, and $(c(\ell))_{\ell\in\mathcal{CV}(l)} \in \{0,1\}^{\mathcal{CV}(l)}$.

By assigning a value to the variable associated with each variable node, from its corresponding space of values, a triple of sequences $(v^m, c^{n\omega}, z^n) \in \{0,1\}^m \times \{0,1\}^{n\omega} \times$

$\mathcal{Z}^n$, is obtained. Such a triple is said to be compatible if and only if the compatibility function at each check and network node takes the value 1. It is easy to see that a triple $(v^m, c^{n\omega}, z^n)$ is compatible if and only if $c^{n\omega} = Av^m$ and $z^n = \Phi(c^n)$. Thus, a sequence z^n is a codeword if and only if there are v^m and $c^{n\omega}$ such that the triple $(v^m, c^{n\omega}, z^n)$ is compatible. Let us denote by $\mathcal{T}$ the space of compatible triples $(v^m, c^{n\omega}, z^n)$ and let us define a conditional probability distribution over $\mathcal{T}$, conditioned on y^n, as follows

$$p_{\mathcal{T}}(v^m, c^{n\omega}, z^n|y^n) = \frac{\exp(-\gamma F(z^n))}{Z} \tag{2.8}$$

where

$$F(z^n) = \sum_{a\in\mathcal{Y}} \sum_{b\in\mathcal{Z}\setminus\{b(a)\}} \rho(a,b) I_{a,b}(y^n, z^n)) \tag{2.9}$$

and Z is a normalization constant. Then any solution the following problem

$$\max\nolimits_{(v^m,c^{n\omega},z^n)\in\mathcal{T}} p_{\mathcal{T}}(v^m, c^{n\omega}, z^n|y^n) \tag{2.10}$$

provides a solution to problem (2.4) as well.

Further notice that equations (2.5), (2.8) and (2.9) lead to the factorization of $p_{\mathcal{T}}(v^m, c^{n\omega}, z^n|y^n)$ as follows

$$p_{\mathcal{T}}(v^m, c^{n\omega}, z^n|y^n) = \frac{\prod_{l=1}^{n} f_{S_l}(z(l))}{Z}. \tag{2.11}$$

The next step is to extend the probability distribution $p_{\mathcal{T}}$ to a probability distribution

p over the set of all triples $(v^m, c^{n\omega}, z^n) \in \{0,1\}^m \times \{0,1\}^{n\omega} \times \mathcal{Z}^n$ by letting

$$p(v^m, c^{n\omega}, z^n|y^n) = \begin{cases} p_{\mathcal{T}}(v^m, c^{n\omega}, z^n|y^n), & \text{if } (v^m, c^{n\omega}, z^n) \in \mathcal{T} \\ 0, & \text{otherwise} \end{cases}. \tag{2.12}$$

With this notation the optimization problem (2.10) becomes

$$\max_{(v^m, c^{n\omega}, z^n) \in \{0,1\}^m \times \{0,1\}^{n\omega} \times \mathcal{Z}^n} p(v^m, c^{n\omega}, z^n|y^n). \tag{2.13}$$

Based on (2.11),(2.12), and using the compatibility functions (2.6),(2.7) associated to the factor nodes, the conditional probability $p(\cdot|y^n)$ can be factorized as follows

$$p(v^m, c^{n\omega}, z^n|y^n) = \frac{\prod_{l=1}^{n} f_{S_l}(z(l)) \prod_{\ell=1}^{n\omega} f_{C_\ell}(c(\ell), (v(k))_{k\in\mathcal{V}(\ell)}) \prod_{l=1}^{n} f_{N_l}(z(l), (c(\ell))_{\ell\in\mathcal{CV}(l)})}{Z}.$$

Consequently, the factor graph represents the factorization of the conditional probability distribution $p(\cdot|y^n)$. This factorization is crucial in designing the message passing algorithm as a heuristic solution to problem (2.4).

2.5 Message Passing Algorithm

As in [37] we employ belief propagation with decimation as our message passing algorithm on the associated factor graph. The algorithm proceeds in a series of rounds. Each round consists of two phases: 1) the message passing phase and 2) the decimation phase. During the message passing phase, messages are transmitted between graph nodes in a series of iterations. At the end of this phase the marginal probabilities $p_{V_k}(\cdot)$ at information variable nodes are computed. During the decimation phase, the information variable nodes V_k for which the bias of $p_{V_k}(\cdot)$ is greater than a certain threshold η are fixed and removed from the graph.

For the message passing phase we use the sum-product algorithm of [56], which is equivalent to belief propagation. Consequently, assuming that the graph has large cycles, the sum-product algorithm can be used to approximate the marginal distributions of variables. During each iteration, messages are passed from each variable node to each adjacent factor node, then from each factor node to each adjacent variable node. The message passed from some variable node U to some factor node F, denoted by $M_{U\to F}(u)$, is a function of the variable u associated to node U. Likewise, the message from F to U, denoted by $M_{F\to U}(u)$, is also a function of the variable u. Thus, because the variables associated to the information variable nodes and to check variable nodes are binary, the messages passed from and to these nodes can be regarded as vectors consisting of two values. On the other hand, the messages passed to and from the network variable nodes are vectors of $|\mathcal{Z}|$ values.

The message to be passed by a node along some edges is computed based on the latest messages received by that node along the other edges, according to the sum-product update rules in [56, Eq. (5,6)]. Applying these rules to our factor graph yields the message equations in Figure 2.2. Note that network variable nodes NV_l and check variable nodes CV_ℓ only forward messages. Precisely, each NV_l forwards to N_l the message received from S_l and vice versa. Thus, to expedite the algorithm we can skip the intermediate node NV_l and directly pass messages between S_l and N_l. Further, we can simplify the graph by removing the node NV_l and its adjacent edges and replacing them by an edge between S_l and N_l. Likewise, since the node CV_ℓ only forwards the message received from C_ℓ to node N_l, and vice versa, where

$$
\begin{aligned}
M_{S_l \to NV_l}(z(l)) =& f_{S_l}(z(l)), \\
M_{NV_l \to N_l}(z(l)) =& M_{S_l \to NV_l}(z(l)), \\
& \text{for all } z(l) \in \mathcal{Z}, 1 \leq l \leq n, \\
M_{N_l \to CV_\ell}(c(\ell)) =& \sum_{z(l) \in \mathcal{Y}} M_{NV_l \to N_l}(z(l)) \sum_{(c(\kappa))_{\kappa \in \mathcal{CV}(l) \setminus \{\ell\}}} f_{N_l}(z(l), (c(\kappa))_{\kappa \in \mathcal{CV}(l)}) \prod_{\kappa \in \mathcal{CV}(l) \setminus \{\ell\}} M_{CV_\kappa \to N_l}(c(\kappa)), \\
M_{CV_\ell \to N_l}(c(\ell)) =& M_{C_\ell \to CV_\ell}(c(\ell)), \\
& \text{for all } c(\ell) \in \{0,1\}, 1 \leq l \leq n, 1 \leq \ell \leq n\omega, \text{ with } l = ((\ell - 1) \mod n) + 1, \\
M_{CV_\ell \to C_\ell}(c(\ell)) =& M_{N_l \to CV_\ell}(c(\ell)), \text{ where } l = ((\ell - 1) \mod n) + 1, \\
M_{C_\ell \to CV_\ell}(c(\ell)) =& f_{C_\ell}(c(\ell), (v(k))_{k \in \mathcal{V}(\ell)}) \prod_{k \in \mathcal{V}(\ell)} M_{V_k \to C_\ell}(v(k)) \\
& \text{for all } c(\ell) \in \{0,1\}, 1 \leq \ell \leq n\omega, \\
M_{C_\ell \to V_k}(v(k)) =& \sum_{c(\ell)} M_{CV_\ell \to C_\ell}(c(\ell)) \sum_{(v(j))_{j \in \mathcal{V}(\ell) \setminus \{k\}}} f_{C_\ell}(c(\ell), (v(j))_{j \in \mathcal{V}(\ell)}) \prod_{j \in \mathcal{V}(\ell) \setminus \{k\}} M_{V_j \to C_\ell}(v(j)), \\
M_{V_k \to C_\ell}(v(k)) =& \prod_{\kappa \in \mathcal{C}(k) \setminus \{\ell\}} M_{C_\kappa \to V_k}(v(k)), \\
& \text{for all } v(k) \in \{0,1\}, 1 \leq \ell \leq n\omega, 1 \leq k \leq m, \text{ such that } A(\ell, k) = 1.
\end{aligned}
$$

Figure 2.2: Equations of messages passed in the factor graph.

$l = \ell \mod n$, we can remove CV_ℓ and its adjacent edges and replace them by an edge between C_ℓ to node N_l. The simplified graph is illustrated in Figure 2.3. The messages passed between adjacent nodes in the simplified graph are computed according to the formulae in Figure 2.4. At the initialization phase, the source nodes transmit messages computed as in Figure 2.4, and the check nodes transmit the constant vector $(0.5, 0.5)$. After that, the message passing schedule at each iteration is the following. First the network nodes send messages to check nodes, then the information variable nodes transmit messages to the check nodes. Finally, the check nodes and source nodes

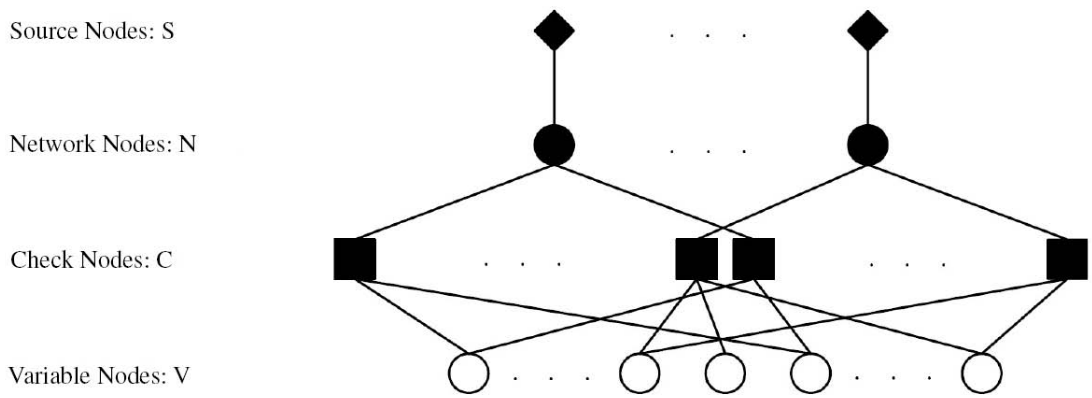

Figure 2.3: Simplified graph associated to the multilevle LDGM code.

send messages to the network nodes.

The message passing phase ends when all the messages $M_{V_k \to C_\ell}(0)$ converge or a maximum number of iterations, typically 100, is reached. Then the marginal probability distributions $p_{V_k}(\cdot)$ are computed according to the following equation

$$p_{V_k}(v(k)) = \frac{1}{T_k} \prod_{\ell \in \mathcal{C}(k)} M_{C_\ell \to V_k}(v(k)), \tag{2.14}$$

where T_k is a normalization factor which ensures that $p_{V_k}(0) + p_{V_k}(1) = 1$. Further, during the decimation phase we fix the information variable nodes whose bias is greater than a certain threshold η (i.e. $|p_{V_k}(0) - p_{V_k}(1)| > \eta$). If no such variable node exists then we fix the one with highest bias. After that we remove the fixed variables from the graph. The message passing during the next round proceeds on the modified graph. Any check node C_ℓ whose adjacent variable nodes are all fixed (hence removed from the graph) computes the message to be sent to the adjacent

$$
\begin{aligned}
M_{S_l \to N_l}(z(l)) =& f_{S_l}(z(l)), \\
& \text{for all } z(l) \in \mathcal{Z}, 1 \le l \le n, \\
M_{N_l \to C_\ell}(c(\ell)) =& \sum_{z(l)\in\mathcal{Y}} M_{S_l \to N_l}(z(l)) \sum_{(c(\kappa))_{\kappa\in\mathcal{CV}(l)\setminus\{\ell\}}} f_{N_l}(z(l),(c(\kappa))_{\kappa\in\mathcal{CV}(l)}) \prod_{\kappa\in\mathcal{CV}(l)\setminus\{\ell\}} M_{C_\kappa \to N_l}(c(\kappa)), \\
& \text{where } l = ((\ell-1) \mod n)+1 \\
M_{C_\ell \to N_l}(c(\ell)) =& f_{C_\ell}(c(\ell),(v(k))_{k\in\mathcal{V}(\ell)}) \prod_{k\in\mathcal{V}(\ell)} M_{V_k \to C_\ell}(v(k)), \\
& \text{for all } c(\ell) \in \{0,1\}, 1 \le l \le n, 1 \le \ell \le n\omega, \text{ with } l = ((\ell-1) \mod n)+1, \\
M_{C_\ell \to V_k}(v(k)) =& \sum_{c(\ell)} M_{CV_\ell \to C_\ell}(c(\ell)) \sum_{(v(j))_{j\in\mathcal{V}(\ell)\setminus\{k\}}} f_{C_\ell}(c(\ell),(v(j))_{j\in\mathcal{V}(\ell)}) \prod_{j\in\mathcal{V}(\ell)\setminus\{k\}} M_{V_j \to C_\ell}(v(j)), \\
M_{V_k \to C_\ell}(v(k)) =& \prod_{\kappa\in\mathcal{C}(k)\setminus\{\ell\}} M_{C_\kappa \to V_k}(v(k)), \\
& \text{for all } v(k) \in \{0,1\}, 1 \le \ell \le n\omega, 1 \le k \le m, \text{ such that } A(\ell,k)=1.
\end{aligned}
$$

Figure 2.4: Equations of messages passed in the simplified graph.

network node N_l as follows

$$M_{C_\ell \to N_l}(0) = (1-c)\exp(\delta) + c\exp(-\delta), \tag{2.15}$$

$$M_{C_\ell \to N_l}(1) = c\exp(\delta) + (1-c)\exp(-\delta), \tag{2.16}$$

where c is the result of modulo-2 addition of the values of all adjacent variable nodes, and $\delta > 0$. The encoding process ends till all variable nodes are fixed, at that time, the information sequence v^m is obtained.

Notice that in belief propagation, after using the message equations in Figure 2.2, the components of each message are normalized. As pointed out in [56] this normalization is not necessary. Precisely, the same result is obtained in (2.14) with or without normalization of messages in Figure 2.4, provided that normalization is applied in (2.14). This observation implies that we can multiply the source node compatibility function given in (2.5) by $\exp(\mu(y(l)))$, for any function $\mu : \mathcal{Y} \to \mathbb{R}$, without changing the output of the algorithm. With such a change the messages passed by the source nodes acquire a form similar to that used in prior work [33], [37], [32].

The message passing algorithm is presented in Figure 2.5 in pseudo code form. In order to evaluate the time complexity of this algorithm, we assume that each message is computed in constant time. Now, recall that at each iteration in the message passing phase, one message is passed along each edge in each direction. Therefore the total time complexity is $O(E \cdot MI \cdot MR)$, where E is the number of edges in the factor graph, MI is the maximum number of iterations and MR is the number of rounds. Furthermore, MI can be considered a constant, while MR can be upper bounded by $m = R \times n$, this bound being achieved when only one variable node is decimated at each round. Notice that E equals ωn plus the number of ones in the generator matrix of the LDGM code, which is bounded by its dimension $m \times \omega n$. However, this bound is loose, because the matrix is sparse. Finally, according to the above discussion and by considering R a constant as well, it follows that a loose upper bound on the time complexity of the message passing algorithm is $O(\omega n^3)$.

2.6 Bound on Algorithm Performance Gap

In this section we derive an upper bound of the performance gap of the proposed message passing algorithm, in terms of the parameters γ and the maximum absolute error δ in the computation of the marginal probabilities at the information variable nodes, under the assumption that $\delta < \frac{1}{2}$ and a single variable is decimated at a time. An immediate consequence of this result is that, if there exists some $\delta_0 > 0$ such that $\delta < \frac{1}{2} - \delta_0$ for all $\gamma > 0$, then the algorithm is optimal for γ large enough (since the space of feasible solutions is finite).

Proposition. 2.1 Assume that only one variable is decimated at every decimation phase and that the marginal probabilities of the information variable nodes computed by the algorithm after every message passing phase, are within a value δ, $0 \le \delta < 1/2$, from the true probabilities. Let F_{opt} denote the optimal cost of problem (2.4) and let F^* denote the cost corresponding to the solution output by the algorithm. Then the following relation holds

$$F^* \le F_{opt} + \frac{m}{\gamma}((m+1)/2\ln 2 - \ln(1-2\delta)), \tag{2.17}$$

where m is the number of information variable nodes.

Proof. For each $v^m \in \{0,1\}^m$ let $F(v^m) = F(z^n(v^m))$, where $z^n(v^m)$ is the codeword corresponding to v^m. For each $0 \le t \le m-1$, let $\mathcal{U}_t$ denote the set of information sequences remaining in the search range for a solution, after the t-th decimation phase. In other words all sequences in $\mathcal{U}_t$ have the bit values on the t positions which were decimated during the first t decimation phases, equal to the bit values already decided for these positions. Further, let F_t denote the smallest cost $F(v^m)$ over all $v^m \in \mathcal{U}_t$. We will first show that

$$F_t \le F_{t-1} + \frac{(m-t+1)\ln 2 - \ln(1-2\delta)}{\gamma}, \tag{2.18}$$

for every $t, 0 \le t \le m-1$. Let us fix some t and assume by contradiction that (2.18) does not hold. This implies that

$$F(v^m) > F_{t-1} + \xi, \text{for all } v^m \in \mathcal{U}_t, \tag{2.19}$$

where $\xi = \frac{(m-t+1)\ln 2 - \ln(1-2\delta)}{\gamma}$.

Assume now that at the t-th decimation phase, variable $v(k)$ is chosen to be decimated and its assigned value is b. Let $p_{V_k}(v(k))$ denote the marginal probability computed after the message passing phase, and let $P_{V_k}(v(k))$ denote the true marginal probability. Then $p_{V_k}(b) \geq 1/2$, and, by the hypothesis, one has $|p_{V_k}(b) - P_{V_k}(b)| \leq \delta$. These inequalities imply that

$$P_{V_k}(b) \geq 1/2 - \delta. \tag{2.20}$$

Further, one has

$$\begin{aligned} P_{V_k}(b) &=^{(a)} \sum_{v^m \in \mathcal{U}_t} \frac{1}{Z_t} \exp(-\gamma F(v^m)) \\ &<^{(b)} \frac{2^{m-t}}{Z_t} \exp(-\gamma(F_{t-1} + \xi)) \\ &<^{(c)} \frac{2^{m-t}}{\exp(-\gamma F_{t-1})} \exp(-\gamma(F_{t-1} + \xi)) \\ &= 2^{m-t} \exp(-\gamma \xi), \end{aligned} \tag{2.21}$$

where $Z_t = \sum_{v^m \in \mathcal{U}_{t-1}} \exp(-\gamma F(v^m))$. Equality (a) follows from the definition of the marginal probability, (b) follows from (2.19) and (c) from $Z_t > \exp(-\gamma F_{t-1})$. Finally, (2.20) and (2.21) lead to

$$2^{m-t} \exp(-\gamma \xi) \geq 1/2 - \delta, \tag{2.22}$$

which implies $\xi < \frac{(m-t+1)\ln 2 - \ln(1-2\delta)}{\gamma}$, and thus contradicts the definition of ξ.

Finally, we will show that

$$F^* \leq F_{m-1} + \frac{1}{\gamma} \ln \frac{2}{1-2\delta}. \tag{2.23}$$

If $F^* = F_{m-1}$ then (2.23) is trivially satisfied. Assume that $F^* > F_{m-1}$, and let $v(\kappa)$ be the non decimated variable after the first $m-1$ decimation phases. Further, let b' denote the bit value assigned to $v(\kappa)$ after the m-th decimation phase. Since $\mathcal{U}_{m-1}$ has only two elements, then $P_{V_\kappa}(b') = \frac{\exp(-\gamma F^*)}{\exp(-\gamma F^*)+\exp(-\gamma F_{m-1})} < \frac{\exp(-\gamma F^*)}{\exp(-\gamma F_{m-1})}$. By the hypothesis, we have $P_{V_\kappa}(b') \geq 1/2 - \delta$, and further, combining with the previous relation, we obtain $\exp(-\gamma(F^* - F_{m-1})) > 1/2 - \delta$, which, in turn, implies (2.23).

To conclude the proof, note that $F_{opt} = F_0$, and relation (2.17) follows by combining (2.18) for all $t, 0 \leq t \leq m-1$, and (2.23). $\square$

2.7 Decoding Process

The final step of single source coding is the decoding process. This is very easily performed on the simplified graph. Given the information sequence v^m, the bit value $v(k)$ is assigned to each information variable node V_k. Further, each value $c(\ell)$ is computed at the check node C_ℓ as the modulo-2 sum of the bit values at adjacent information variable nodes. At last, each codeword symbol $z(l)$ is computed at the network node N_l based on the bit values at adjacent check nodes, via the mapping $\phi_l(\cdot)$.

2.8 Efficient Implementation of Multiple LDGM-based Codes at Encoders 1 and 2

As emphasized in previous subsection, for each $t = 1, 2$, encoder t needs $2^{n(I(X;X_c)+\epsilon_1)}$ different multilevel LDGM-based codes. This fact apparently raises space complexity concerns. To address these concerns we choose a common value of the integer ω and a common low density generator matrix A for all these codes. It is clear then that the associated graphs are identical, consequently a single graph needs to be stored. What differs from one code to another are only the compatibility functions at network nodes. However, their storage need is not an issue either, since these functions are chosen from a small pool containing only $|\mathcal{X}_c|$ functions.

Another way of simplifying encoders 1 and 2 is by replacing the multiple codebooks by a single codebook of the same size, via a variable substitution as follows. According to the Variable Substitution Lemma [44, Lemma 1], there exist random variables U_t, $t = 1, 2$, taking values in the finite sets $\mathcal{U}_t$, respectively, with $|\mathcal{U}_t| \leq |\mathcal{X}_c|(|\mathcal{X}_t| - 1) + 1$, $t = 1, 2$, and the functions $\pi_t : \mathcal{X}_c \times \mathcal{U}_t \to \mathcal{X}_t$, such that

C1) U_t is independent of X_c;

C2) $X_t = \pi_t(X_c, U_t)$;

C3) $X - (X_c, X_1) - U_1$, and $(X, X_1) - (X_c, X_2) - U_2$ form Markov chains.

The above relations imply [44] that

$$I(X, X_c; U_1) = I(X_c; U_1) + I(X; U_1|X_c) = I(X; X_1|X_c),$$

$$I(X, X_c, X_1; U_2) = I(X_c; U_2) + I(X, X_1; U_2|X_c) = I(X, X_1; X_2|X_c).$$

Then encoders 1 and 2 can be modified as follows. Encoder 1 randomly generates codebook $\mathcal{C}_1 = \{u_{1,j}^n\}_{j=1}^{n(I(X,X_c;U_1)+\epsilon_1)}$, according to the distribution $\prod_{l=1}^n p_{U_1}(\cdot)$. Encoder 2 randomly generates codebook $\mathcal{C}_2 = \{u_{2,k}^n\}_{k=1}^{n(I(X,X_c,X_1;U_2)+\epsilon_1)}$, according to the distribution $\prod_{l=1}^n p_{U_2}(\cdot)$. Given the input sequence x^n and index i selected by encoder 0, encoder 1 finds an index j such that $u_{1,j}^n$ is jointly typical with x^n and $x_{c,i}^n$. Further, encoder 2 finds an index k such that $u_{2,k}^n$ is jointly typical with x^n, $x_{c,i}^n$ and $u_{1,j}^n$. On the decoder side, decoder 1 reconstructs the codeword $x_{1,i,j}^n$ based on $u_{1,j}^n$ and $x_{c,i}^n$ by setting $x_{1,i,j}(l) = \pi_1(x_{c,i}(l), u_{1,j}(l))$, $1 \le l \le n$. Decoder 2 reconstructs the codeword $x_{2,i,k}^n$ based on $u_{2,k}^n$ and $x_{c,i}^n$ by setting $x_{2,i,k}(l) = \pi_2(x_{c,i}(l), u_{2,k}(l))$, $1 \le l \le n$.

The two methods described in this subsection for simplifying encoders 1 and 2 are conceptually equivalent. The slight difference between them is not significant theoretically, but it could raise practical concerns. Next we analyze the connection between these two methods.

Let W denote a uniform variable over the alphabet $\mathcal{W} = \{0,1\}^\omega$. Let $\mathcal{C}_{t,lin} = \{c^{n\omega} = A_t v^m | v^m \in \{0,1\}^m\}$ be the linear codebook used in the construction of method 1, where A_t is a low density generator matrix. By identifying each bit sequence $c^{n\omega}$ with an n-length sequence w^n over the alphabet $\mathcal{W}$ as follows $w(l) = (c(l), c(l+n) \cdots, c(l+(\omega-1)n))$, $\mathcal{C}_{t,lin}$ can be regarded as a codebook representing the variable W. Then each conditional codebook $\mathcal{C}_{t,i}$ is actually obtained from $\mathcal{C}_{t,lin}$ via a component-wise mapping function applied to the codewords w^n.

In method 2 the conditional codebooks are constructed based on the common codebook $\mathcal{C}_t$ via componentwise mapping functions. By implementing the common codebook with a multilevel LDGM code, $\mathcal{C}_t$ is derived from $\mathcal{C}_{t,lin}$ using a component-wise mapping function, thus the conditional codebooks are also derived from $\mathcal{C}_{t,lin}$

based on component-wise mapping functions. It is clear now that the conditional codebooks in method 2 can be obtained via method 1. On the other hand, notice that the linear codebook $\mathcal{C}_{t,lin}$ in method 1 can be regarded as the common codebook $\mathcal{C}_t$ for some choice of the variable U_t satisfying conditions C1-C3 (but not necessarily the cardinality condition on $\mathcal{U}_t$).

Thus, both methods ensure equivalent construction of the conditional codebooks (if we remove the cardinality constraint on $\mathcal{U}_t$ in method 2). The only difference between them consists in the encoding procedure. In method 2, the encoder searches for a codeword u^n in the common codebook, while in method 1 the encoder looks for a codeword in the conditional codebook directly . This difference is not significant theoretically, but it could be in practice. Precisely, if the size of alphabet $\mathcal{U}_t$ is larger than the size of $\mathcal{X}_t$, then the message passing algorithm for method 2 is more complex since the messages passed by the source nodes have more components. Moreover, additional complexity arises at the design phase since more parameters need to be selected to define the source messages. On the other hand, if $\mathcal{U}_t$ and $\mathcal{X}_t$ have equal sizes, then the message passing algorithms have equal complexity. In our simulations we chose to implement method 2 whenever the latter case arises.

$max_{iter} \leftarrow 100$, $\eta \leftarrow 0.9$
$\mathcal{C}(k) \leftarrow \{\ell | A(\ell, k) = 1\}$, $\mathcal{V}(\ell) \leftarrow \{k | A(\ell, k) = 1\}$, for all $k = 1, \cdots, m$, and $\ell = 1, \cdots, n\omega$
$\mathcal{NV} \leftarrow \{1, \cdots, m\}$
for $l = 1$ to n **do**
  Compute $M_{S_l \to N_l}$ as in Fig. 2.4
  for $s = 0$ to $\omega - 1$ **do**
    $M_{C_{l+sn} \to N_l} \leftarrow (\frac{1}{2}, \frac{1}{2})$
    for $k \in \mathcal{V}(l + sn)$ **do**
      $M_{C_{l+sn} \to V_k} \leftarrow (\frac{1}{2}, \frac{1}{2})$
    end for
  end for
end for
while $\mathcal{NV} \neq \emptyset$ **do**
  for $i = 1$ to max_{iter} **do**
    Compute $M_{N_l \to C_{l+sn}}$, $M_{V_k \to C_{l+sn}}$ as in Fig. 2.4
    for all $l = 1, \cdots, n$, $s = 0, \cdots, \omega - 1$, $k \in \mathcal{V}(l + sn) \cap \mathcal{NV}$.
    for $l = 1$ to n **do**
      for $s = 0$ to $\omega - 1$ **do**
        if $\mathcal{V}(l + sn) \cap \mathcal{NV} \neq \emptyset$ **then**
          Compute $M_{C_{l+sn} \to N_l}$ as in Fig. 2.4
          for $k \in \mathcal{V}(l + sn) \cap \mathcal{NV}$ **do**
            Compute $M_{C_{l+sn} \to V_k}$ as in Fig. 2.4
          end for
        else
          Compute $M_{C_{l+sn} \to N_l}$ as in Eq. (2.16)
        end if
      end for
    end for
    if $M_{C_\ell \to V_k}$ converge for all $\ell = 1, \cdots, n\omega, k \in \mathcal{V}(\ell) \cap \mathcal{NV}$ **then**
      Break
    end if
  end for
  for $k \in \mathcal{NV}$ **do**
    Compute $p_{V_k}(v(k))$ as in Eq.(2.14)
    if $|p_{V_k}(0) - p_{V_k}(1)| > \eta$ **then**
      $\mathcal{NV} \leftarrow \mathcal{NV} - \{k\}$
      $\hat{b} \leftarrow \mathrm{argmax}_{b \in \{0,1\}} p_{V_k}(b)$, $v(k) \leftarrow \hat{b}$
      $M_{V_k \to C_\ell}(\hat{b}) \leftarrow 1$, $M_{V_k \to C_\ell}(1 - \hat{b}) \leftarrow 0$, for all $q \in \mathcal{C}(k)$
    end if
  end for
end while

Figure 2.5: Message passing algorithm. $\mathcal{NV}$ denotes the set of indices k of currently non-decimated variable nodes V_k.

Chapter 3

Applications of the Proposed Coding Scheme

In this chapter we discuss the application of the proposed practical coding scheme to approach the theoretical rate-distortion limits or known bounds for three cases of the MD problem with Hamming distortion measure:

1) No excess sum-rate case for general binary sources. For this case the characterization of the MD rate-distortion region is known as the EGC region. However, analytical expressions are partly available only for the uniform source. To aid the code design, we derive the exact expression of the distortion region and of the auxiliary variables necessary to achieve its boundary.

2) No excess marginal rate case for the uniform binary source. We improve an analytical expression of the upper bound of central distortion.

3) Successive refinement for finite-alphabet sources.

3.1 No Excess Sum-rate Case for Binary Sources

The term no excess sum-rate for the two descriptions problem refers to the case when $R_1 + R_2 = R(d_0)$, where R_t is the rate of description t, for $t = 1, 2$, and $R(\cdot)$ denotes the rate-distortion function. Since in this case the MD rate-distortion coincides with the EGC* region, and the EGC* region is a particular case of the ZB region, corresponding to X_c constant, we can readily apply the coding scheme proposed in Chapter 2 to approach its boundary provided that the auxiliary variables X_1 and X_2 are known. However, the computation of the boundary points and of the variables X_1 and X_2, achieving them is not an easy task, and analytical expressions are available only in part for the uniform binary source with Hamming distortion. Our next result represents some progress in this direction by providing the expression of the distortion region for general binary sources.

First we introduce some notations. Let $\mathcal{D}$ denote the distortion region for the no excess sum-rate case. Precisely, $\mathcal{D}$ is defined as

$$\begin{aligned}\mathcal{D} = \{(d_1, d_2, d_0) : \exists (R_1, R_2) \text{ such that } R_1 + R_2 = R(d_0) \text{ and}\\ (R_1, R_2, d_1, d_2, d_0) \in \mathcal{RD}\}.\end{aligned}$$

For $d_0 \geq 0$ let us additionally define

$$\mathcal{D}(d_0) = \{(d_1, d_2) : (d_1, d_2, d_0) \in \mathcal{D}\}.$$

Clearly, $\mathcal{D} = \cup_{d_0 : d_0 \geq 0} \{(d_1, d_2, d_0) : (d_1, d_2) \in \mathcal{D}(d_0)\}$. Since $\mathcal{D}(d_0)$ is convex, it suffices to characterize its supporting lines. Two of the supporting lines are already known, namely the lines of equations $d_1 = d_0$ and $d_2 = d_0$. Therefore, to complete the task, it is enough to solve the following optimization problem

$$\min_{(d_1,d_2)\in\mathcal{D}(d_0)} \alpha d_1 + d_2 \tag{3.1}$$

for all $\alpha > 0$.

As shown by Ahlswede in [43] the EGC region is tight for the no excess rate case. Specifically, we have

$$\begin{aligned}&\{(R_1, R_2, d_1, d_2, d_0) \in \mathcal{RD} : R_1 + R_2 = R(d_0)\}\\&= \{(R_1, R_2, d_1, d_2, d_0) \in \mathcal{Q} : R_1 + R_2 = R(d_0)\},\end{aligned} \tag{3.2}$$

where $\mathcal{Q}$ denotes the convex closure of the set of quintuples $(R_1, R_2, d_1, d_2, d_0)$ for which there exist auxiliary random variables X_t, $t = 0, 1, 2$, jointly distributed with X such that

$$\begin{aligned}&I(X_1; X_2) = 0,\\&R_t \geq I(X; X_t), \quad t = 1, 2,\\&R_1 + R_2 \geq I(X; X_0, X_1, X_2),\\&d_t \geq \mathbb{E}[d(X, X_t)], \quad t = 0, 1, 2.\end{aligned}$$

Now we proceed to characterize $\mathcal{D}(d_0)$ for binary sources with Hamming distortion measure (i.e., $d(x, \hat{x}) = d_H(x, \hat{x}) = 0$ if $x = \hat{x}$, and 1 otherwise). Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$ and $p_X(0) = \delta$, where $0 < \delta \leq 1/2$. With no loss of generality, we shall assume

$0 \le d_0 < \delta$. In this case $R(d_0)$ is a strictly convex function of d_0. Furthermore, if

$$R(d_0) = R_1 + R_2 \ge I(X; X_0, X_1, X_2),$$
$$d_0 \ge \mathbb{E}[d(X, X_0)],$$

then one must have $X = X_0 \oplus Z$, where $\oplus$ denotes the modulo-2 addition operation, and $Z \sim \text{Ber}(d_0)$ is independent of (X_0, X_1, X_2). Therefore, it suffices to specify $p_{X_1X_2|X_0}$ in order to determine $p_{XX_0X_1X_2}$ due to the fact that p_{XX_0} is completely determined once d_0 is given and that $X - X_0 - (X_1, X_2)$ form a Markov chain.

Now in view of (3.2), it can be readily shown that

$$\begin{aligned}&\min_{(d_1,d_2)\in\mathcal{D}(d_0)} \alpha d_1 + d_2\\ &\ge \min_{p_{X_1X_2|X_0}:I(X_1;X_2)=0} \alpha\mathbb{E}[d(X, X_1)] + \mathbb{E}[d(X, X_2)].\end{aligned}$$

It turns out that the inequality can be replaced by an equality. To see this, for any $p_{X_1X_2|X_0}$ such that the induced X_1 and X_2 are independent, let

$$R_1 = I(X; X_1),$$
$$R_2 = I(X, X_1; X_2) + I(X; X_0|X_1, X_2),$$
$$d_t = \mathbb{E}[d(X, X_t)], \quad t = 0, 1, 2.$$

It can be verified that $(R_1, R_2, d_1, d_2, d_0) \in \mathcal{RD}_{\text{EGC}}$ and $R_1 + R_2 = R(d_0)$. Therefore, (3.1) is equivalent to

$$\min_{p_{X_1X_2|X_0}:I(X_1;X_2)=0} \alpha\mathbb{E}[d(X,X_1)] + \mathbb{E}[d(X,X_2)]. \tag{3.3}$$

The next result is proved in Appendix A.

Proposition 3.1 The following relation holds

$$\begin{aligned}\mathcal{D}(d_0) = \{(d_1,d_2) : &(d_1+1-2d_0-\delta)(d_2+1-2d_0-\delta)\\ &\geq (1-d_0-\delta)(1-2d_0), d_1 \geq d_0, d_2 \geq d_0\}.\end{aligned}$$

Moreover, for each $\alpha > 0$, the solution (d_1, d_2) to (3.1) is unique, and a corresponding solution to (3.3) is specified by

$$\begin{aligned}
p_{XX_1X_2}(0,1,1) &= \frac{d_0}{1-2d_0}(1-\delta-d_0), \\
p_{XX_1X_2}(1,0,0) &= \frac{d_0}{1-2d_0}(\delta+d_0-d_1-d_2), \\
p_{XX_1X_2}(1,1,0) &= \frac{d_0}{1-2d_0}(d_1-d_0), \\
p_{XX_1X_2}(1,0,1) &= \frac{d_0}{1-2d_0}(d_2-d_0), \\
p_{XX_1X_2}(0,1,0) &= \frac{1-d_0}{1-2d_0}(d_1-d_0), \\
p_{XX_1X_2}(0,0,1) &= \frac{1-d_0}{1-2d_0}(d_2-d_0), \\
p_{XX_1X_2}(0,0,0) &= \frac{1-d_0}{1-2d_0}(\delta+d_0-d_1-d_2), \\
p_{XX_1X_2}(1,1,1) &= \frac{1-d_0}{1-2d_0}(1-\delta-d_0), \\
X_0 = \psi_0(X_1,X_2) &= \begin{cases} 1, & X_1=X_2=1 \\ 0, & \text{otherwise} \end{cases};
\end{aligned} \tag{3.4}$$

furthermore, when $\delta < 1/2$, the solution to (3.3) is also unique, while for $\delta = 1/2$ and $d_1 \neq d_0, d_2 \neq d_0$, there is only one alternative solution, which is obtained by swapping 0 and 1 in (3.4).

Note that for the case of uniform binary source part of the result of Proposition 3.1 can be inferred from [42, Section IV]. Figure 3.1 depicts the distortion region $\mathcal{D}(d_0)$ for a binary source with $p_X(0) = 1/4$, $d_0 = 0$ and $d_0 = 0.013$. Consider now a distortion pair (d_1, d_2) on the lower boundary[1] of $\mathcal{D}(d_0)$ and let $p_{X_0X_1X_2|X}$ be the joint conditional distribution specified by (3.4). Since X_0 is a function of X_1 and X_2, it follows that X_1 and X_2 are the auxiliary variables in (1.4) achieving the distortion

[1]The lower boundary of $\mathcal{D}(d_0)$ is defined as the set of solutions to (3.1) for $\alpha > 0$.

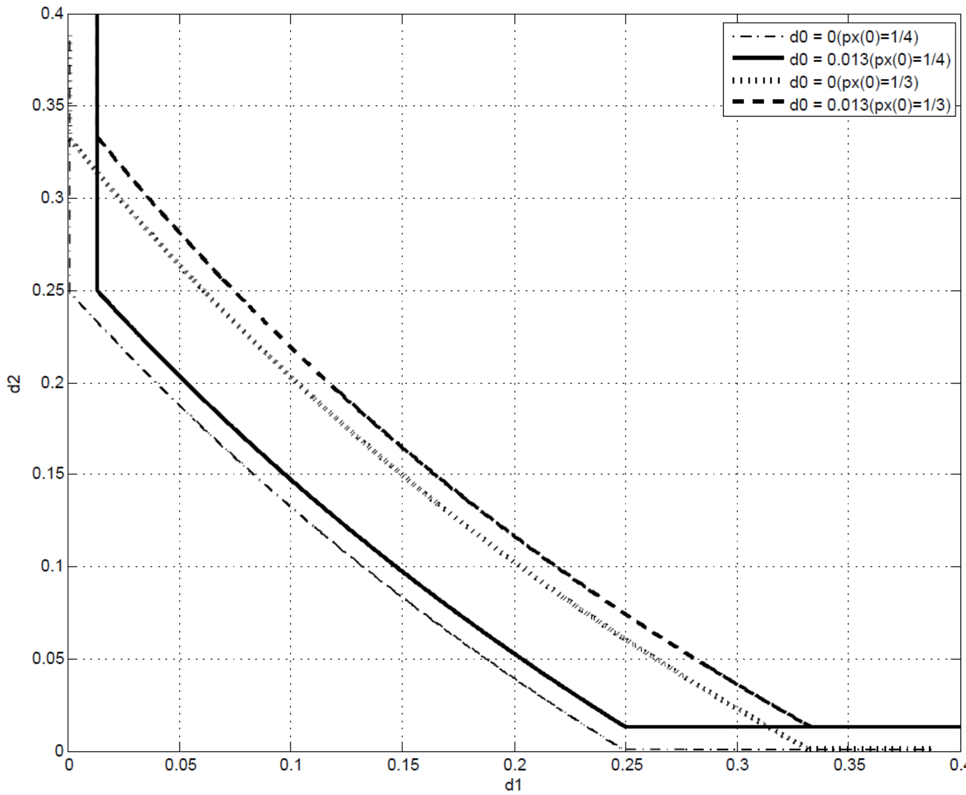

Figure 3.1: Distortion region $\mathcal{D}(d_0)$ for a binary source with $p_X(0) = 1/4$, $d_0 = 0$ and $d_0 = 0.013$.

triple (d_1, d_2, d_0), with ψ_t being the identity function, for $t = 1, 2$, and ψ_0 defined as in Proposition 3.1. Define the rate region

$$\mathcal{R}(d_1, d_2, d_0) \triangleq \{(R_1, R_2) : R_1 + R_2 = R(d_0), \text{ and } (R_1, R_2, d_1, d_2, d_0) \in \mathcal{RD}\}.$$

It can be shown by leveraging Proposition 3.1 that $\mathcal{R}(d_1, d_2, d_0)$ coincides with the set of rate pairs satisfying

$$R_1 \geq I(X; X_1)$$
$$R_2 \geq I(X; X_2)$$
$$R_1 + R_2 = R(d_0).$$

Thus, $\mathcal{R}(d_1, d_2, d_0)$ coincides with the dominant face $\mathcal{F}(p_{X_c X_1 X_2|X})$ defined in Section 2.1, where X_c is a constant. Thus, the two corner points E_1 and E_2 are the extremities of $\mathcal{R}(d_1, d_2, d_0)$.

Let us analyze now the application of the coding scheme of Section 2 for the corner point E_1. Since X_c is a constant, clearly the scheme does not need the encoder 0.

Encoder 1. This encoder needs only one non conditional codebook $\mathcal{C}_1 = \{x_{1,j}^n\}_{j=1}^{n(I(X;X_1)+\epsilon_1)}$ randomly generated according to $\prod_{l=1}^n p_{X_1}(\cdot)$. The multilevel LDGM code for this codebook is designed as described in Subsection 2.2, for $(Y, Z) = (X, X_1)$. The distributions $p_l(\cdot)$, $1 \leq l \leq n$, are all identical to $p_{X_1}(\cdot)$. The input sequence is $y^n = x^n$ and the output sequence is $z^n = x_1^n$. The encoder needs to select the codeword $x_1^n \in \mathcal{C}_1$ jointly typical with x^n according to the distribution p_{XX_1}. We set $b(0) = 0$ and $b(1) = 1$. Then the optimization problem (2.4) has the form

$$\min_{x_1^n \in \mathcal{C}} \lambda(0,1) I_{0,1}(x^n, x_1^n) + \lambda(1,0) I_{1,0}(x^n, x_1^n).$$

It can be easily verified that $p_{X_1|X}(1|0) > p_{X_1|X}(0|1)$. Therefore, guided by the intuition that $\lambda(a, b)$ should be inverse proportional to $p_{X_1|X}(b|a)$ we set these parameters such that $\lambda(0, 1) < \lambda(1, 0)$.

Encoder 2. This encoder also needs a single codebook $\mathcal{C}_2 = \{x_{2,k}^n\}_{k=1}^{n(I(X,X_1;X_2)+\epsilon_1)}$

randomly generated according to $\prod_{l=1}^{n} p_{X_2}(\cdot)$. The multilevel LDGM code for this codebook is constructed as described in Section 2.2, for $Y = (X, X_1)$ and $Z = X_2$. Note that $\mathcal{Y} = \{0,1\}^2$, and that the input sequence is y^n with $y(l) = (x(l), x_1(l))$, $1 \le l \le n$, where x_1^n is the codeword selected by encoder 1. The strong typicality condition is too strong for this encoder since all we are interested in is to obtain low central and side 2 distortions. Thus, the optimization problem can be formulated as

$$\min_{x_2^n \in \mathcal{C}_2} \beta_0 d_H^{(n)}(x^n, x_0^n) + \beta_2 d_H^{(n)}(x^n, x_2^n), \tag{3.5}$$

for some non-negative weights β_0 and β_1, where x_0^n is the central reconstruction, hence $x_0(l) = 1$ if $x_1(l) = x_2(l) = 1$, and 0 otherwise, $1 \le l \le n$, and $d_H^{(n)}(x^n, y^n) \triangleq \sum_{l=1}^{n} d_H(x(l), y(l))$. It follows that

$$\begin{aligned}
d_H^{(n)}(x^n, x_0^n) &= I_{(0,1),1}(y^n, x_2^n) + I_{(1,1),0}(y^n, x_2^n) + I_{(1,0),0}(y^n, x_2^n) + I_{(1,0),1}(y^n, x_2^n), \\
&= I_{(0,1),1}(y^n, x_2^n) + I_{(1,1),0}(y^n, x_2^n) + I_{(1,0)}(y^n), \\
d_H^{(n)}(x^n, x_2^n) &= I_{(0,0),1}(y^n, x_2^n) + I_{(0,1),1}(y^n, x_2^n) + I_{(1,0),0}(y^n, x_2^n) + I_{(1,1),0}(y^n, x_2^n).
\end{aligned}$$

Since $I_{(1,0)}(y^n)$ depends only on the input y^n, consequently is a constant for the optimization problem, problem (3.5) becomes

$$\min_{x_2^n \in \mathcal{C}_2} (\beta_0 + \beta_2)(I_{(0,1),1}(y^n, x_2^n) + I_{(1,1),0}(y^n, x_2^n)) + \beta_2 (I_{(0,0),1}(y^n, x_2^n) + I_{(1,0),0}(y^n, x_2^n)).$$

This cost function fits under the framework of (2.4) by setting $b(0,1) = 0$, $b(1,1) = 1$, $b(0,0) = 0$, $b(1,0) = 1$, $\lambda((0,1),1) = \lambda((1,1),0) = \beta_0 + \beta_2$ and $\lambda((0,0),1) = \lambda((1,0),0) = \beta_2$.

3.2 No Excess Marginal Rate for the Uniform Binary Source

The term no excess marginal rate refers to the case when $R_1 = R(d_1)$ and $R_2 = R(d_2)$. An interesting question in such a case is, given the distortion pair (d_1, d_2), what is the minimum distortion d_0? Precisely, let us define $d_0(d_1, d_2)$ as follows

$$d_0(d_1, d_2) \triangleq \min\{d_0 : (R(d_1), R(d_2), d_1, d_2, d_0) \in \mathcal{RD}\}.$$

Zhang and Berger proposed in [48] an upper bound for $d_0(d_1, d_2)$ by choosing the auxiliary random variables X_c, X_1, X_2 specified by the ZB region such that $X - X_1 - X_c$, $X - X_2 - X_c$, and $X_1 - (X, X_c) - X_2$ form Markov chains, $I(X; X_t) = R(d_t)$, $\mathbb{E}[d(X, X_t)] \leq d_t$ and $\psi(X_t, X_c) = X_t$, for $t = 1, 2$.

Further they showed that for the uniform binary source with Hamming distortion this upper bound is strictly better than the upper bound derived from the EGC region, namely than $\min(d_1, d_2)$. Precisely, let

$$\begin{aligned}
&\mathbb{P}(X_t = X|X) = 1 - d_t, \quad t = 1, 2,\\
&\mathbb{P}(X_t \neq X|X) = d_t, \quad t = 1, 2,\\
&\mathbb{P}(X_c = X|X) = 1 - s,\\
&\mathbb{P}(X_c \neq X|X) = s,\\
&\mathbb{P}(X_c = X_t|X) = 1 - s_t, \quad t = 1, 2,\\
&\mathbb{P}(X_c \neq X_t|X) = s_t, \quad t = 1, 2,
\end{aligned}$$

for some s such that $\max(d_1, d_2) \leq s \leq 1/2$, and $s_t = \frac{s-d_t}{1-2d_t}$, $t = 1, 2$. Then the upper bound is computed in [48] as

$$UB(d_1, d_2) = \inf_{s:\max(d_1,d_2)\leq s\leq 1/2} \mathbb{P}(X \neq \psi_0(X_c, X_1, X_2)). \tag{3.6}$$

The mapping ψ_0 used in [48] is defined as follows

$$\psi_0(x_c, x_1, x_2) = \begin{cases} x_1, & x_1 = x_2 \\ x_c, & \text{otherwise} \end{cases} .$$

We argue that this mapping is not optimal and provide the optimal mapping ψ_0 which minimizes the central distortion, along with the optimal value of s_0 which achieves the infimmum in (3.6). Let us assume without restricting the generality that $d_1 \leq d_2$. Define $\alpha(d_1, d_2) \triangleq \frac{d_1(1-d_2)}{d_1+d_2-2d_1d_2)}$. Then relations $d_2 \leq \alpha(d_1, d_2) \leq 1/2$ hold. In Appendix B we show that the mapping ψ_0 which minimizes the central distortion is the following

$$\psi_0(x_c, x_1, x_2) = \begin{cases} x_1, & x_1 = x_2 \\ 1 - x_c, \quad \text{if } x_1 \neq x_2, & d_2 \leq s \leq \alpha(d_1, d_2) \\ x_1, \quad \text{if } x_1 \neq x_2, & \alpha(d_1, d_2) < s \leq 1/2 \end{cases} . \tag{3.7}$$

Then the upper bound $UB(d_1, d_2)$ (3.6) becomes

$$\min_{s:d_2\leq s\leq\alpha(d_1,d_2)} \big(1 - s - \frac{(1-d_1)(1-d_2)(1-s-d_1)(1-s-d_2)}{(1-s)(1-2d_1)(1-2d_2)} + \frac{d_1d_2(1-s-d_1)(1-s-d_2)}{s(1-2d_1)(1-2d_2)}\big). \tag{3.8}$$

Define $\beta(d_1, d_2) \triangleq \frac{d_1+d_2-2d_1d_2}{d_1d_2(1-d_1)(1-d_2)}$. As proved in Appendix B, the value s_0 achieving the above minimum is

$$s_0 = \frac{1}{2} - \frac{1}{2}\sqrt{1 - 4\frac{\sqrt{1+\beta(d_1,d_2)}-1}{\beta(d_1,d_2)}}. \tag{3.9}$$

Then the value of $UB(d_1, d_2)$ can be recovered by replacing s by s_0 in the expression to be minimized in (3.8).

In order to approach in practice the upper bound $UB(d_1, d_2)$ we will use the coding scheme proposed in Chapter 2. At each encoder we formulate the optimization problem as minimization of distortion between the source sequence x^n and the output sequence of that encoder.

Notably, in this case, the variables U_t, $t = 1, 2$ introduced in Subsection 2.8 for the purpose of simplifying encoders 1 and 2, take values in the binary alphabet. Precisely, we have $U_t = X_c \oplus X_t$ and $X_t = \pi_t(X_c, U_t) = X_c \oplus U_t$, $t = 1, 2$. Since $d_H(x, x_t) = d_H(x \oplus x_c, x_t \oplus x_c)$, the optimization problem at the simplified encoder t can be equivalently formulated as minimizing $d_H^{(n)}(\Delta x^n, u_t^n)$ over all $u_t^n \in \mathcal{C}_t$, where $\Delta x(l) \triangleq x_{c,i}(l) \oplus x(l)$, for $1 \leq l \leq n$, i being the index output by the base layer encoder. In conclusion, at encoder 0, $Y = X$ holds, while at encoders $t, t = 1, 2$, $Y = \Delta X$ holds, where $\Delta X \triangleq X \oplus X_c$. Thus, we have $\mathcal{Y} = \mathcal{Z} = \{0, 1\}$ at each encoder $t = 0, 1, 2$, and we set $b(a) = a$ for $a = 0, 1$, and $\lambda(0, 1) = \lambda(1, 0) > 0$.

3.3 Successive Refinement

Successive refinement (SR) can be regarded as a special form of multiple description coding in which the distortion constraint on the second description (i.e., d_2) is not imposed. In this scenario it is common to refer to the first description as the base layer and the second description as the refinement layer.

The successive refinement coding rate-distortion region $\mathcal{RD}_{\mathrm{SR}}$ is given by

$$\mathcal{RD}_{\mathrm{SR}} = \{(R_1, R_2, d_1, d_0) : (R_1, R_2, d_1, \infty, d_0) \in \mathcal{RD}\}.$$

As shown in [51], $\mathcal{RD}_{\mathrm{SR}}$ is equal to the set of quadruples (R_1, R_2, d_1, d_0) for which there exist auxiliary random variables X_0 and X_1, jointly distributed with the generic source variable X, such that

$$\begin{aligned} &R_1 \geq I(X; X_1), \\ &R_1 + R_2 \geq I(X; X_0, X_1), \\ &\mathbb{E}[d(X, X_t)] \leq d_t, \quad t = 0, 1. \end{aligned}$$

It is easy to see that $\mathcal{RD}_{\mathrm{SR}}$ is equivalent to $\mathcal{RD}_{\mathrm{EGC}}$ with X_2 set to be a constant.

An important case is when $(R(d_1), R(d_0) - R(d_1), d_1, d_0) \in \mathcal{RD}_{\mathrm{SR}}$. The sources for which this relation holds for all $d_1 > d_0 \geq 0$, are called in [50] successively refinable sources. Equitz and Cover showed in [50] that $(R(d_1), R(d_0) - R(d_1), d_1, d_0) \in \mathcal{RD}_{\mathrm{SR}}$ if and only if there are random variables X_0 and X_1, jointly distributed with the generic source variable X, such that

$$\begin{aligned} &R(d_1) = I(X; X_1), \\ &R(d_0) = I(X; X_0, X_1), \\ &\mathbb{E}[d(X, X_t)] \leq d_t, \quad t = 0, 1, \end{aligned}$$

and $X - X_0 - X_1$ form a Markov chain. They further proved, based on results of [57], that finite-alphabet sources with Hamming distortion are successively refinable and the joint distribution $p_{XX_1X_0}$ is specified by

$$p_{X_t}(x) = \frac{(p_X(x) - \lambda_t)^+}{\sum_{x' \in \mathcal{X}} (p_X(x') - \lambda_t)^+}, \quad x \in \mathcal{X}, t = 0, 1,$$

$$p_{X|X_0}(x|x_0) = \begin{cases} 1 - d_0, & x = x_0 \in \mathcal{X}_0^+ \\ \lambda_0, & x \neq x_0, x \in \mathcal{X}_0^+, x_0 \in \mathcal{X}_0^+ \\ p_X(x) & x \notin \mathcal{X}_0^+, x_0 \in \mathcal{X}_0^+ \end{cases},$$

$$p_{X_0|X_1}(x_0|x_1) = \begin{cases} \frac{1-d_1-\lambda_0}{1-d_0-\lambda_0}, & x_0 = x_1 \in \mathcal{X}_1^+ \\ \frac{\lambda_1-\lambda_0}{1-d_0-\lambda_0}, & x_0 \neq x_1, x_0 \in \mathcal{X}_1^+, x_1 \in \mathcal{X}_1^+ \\ \frac{p_X(x_0)-\lambda_0}{1-d_0-\lambda_0} & x_0 \in \mathcal{X}_0^+, x_0 \notin \mathcal{X}_1^+, x_1 \in \mathcal{X}_1^+ \end{cases},$$

where $\lambda_0 \in [0, \lambda_1]$ and λ_1 are determined by

$$\sum_{x_t \in \mathcal{X}_t^+} p_{X_t}(x_t) p_{X|X_t}(x|x_t) = p_X(x), \quad x \in \mathcal{X}, t = 0, 1,$$

and $\mathcal{X}_t^+ = \{x \in \mathcal{X} : p_X(x) - \lambda_t > 0\}$, $t = 0, 1$.

The following theoretical coding scheme for SR was suggested in [50].

Codebook generation. For the base layer, the codebook $\mathcal{C}_1 = \{x_{1,i}^n\}_{i=1}^{2^{nR_1}}$ where $R_1 = I(X; X_1) + \epsilon_1$, is randomly generated according to $\prod_{l=1}^n p_{X_1}(\cdot)$. For the refinement layer, for each index $i, 1 \leq i \leq 2^{nR_1}$, a codebook $\mathcal{C}_{0,i} = \{x_{0,i,j}^n\}_{j=1}^{2^{nR_2}}$, where $R_2 = I(X; X_0|X_1) + \epsilon_2$, is randomly generated according to the conditional distribution $\prod_{l=1}^n p_{X_0|X_1}(\cdot|x_{1,i}(l))$.

Encoder for the base layer. Given the source sequence x^n, the encoder of the base layer finds the index $i, 1 \leq i \leq 2^{nR_1}$, which minimizes $d_H^{(n)}(x^n, x_{1,i}^n)$.

Encoder for the refinement layer. This encoder has knowledge of the index

i of the base layer, and it chooses the index $j, 1 \leq j \leq 2^{nR_2}$, which minimizes $d_H^{(n)}(x^n, x_{0,i,j}^n)$.

Decoder. The decoder of the base layer receives index i and takes $x_{1,i}^n$ as the reconstruction of x^n. When additionally the refinement index j is received, the decoder takes $x_{0,i,j}^n$ as the reconstruction of x^n.

As justified in [50], this theoretical coding scheme is able to achieve the distortion pair $(d_1 + \epsilon, d_0 + \epsilon)$, as n grows to infinity.

Note that this encoding scheme resembles the succession of encoders 0 and 1 for the ZB region. The main difference resides in the fact that in the successive refinement case, the output of each encoder forms a separate description.

Our practical coding scheme for the SR problem uses multilevel LDGM codes for the encoder of each layer as described in Subsection 2.2. Precisely, for the base layer we have $(Y, Z) = (X, X_1)$, mR_1, and the distributions $p_l(\cdot)$ are identical to $p_{X_1}(\cdot)$. At the encoder for the refinement layer we have $(Y, Z) = (X, X_0)$, mR_2 and $p_l(\cdot) = p_{X_0|X_1}(\cdot|x_{1,i}(l))$, $1 \leq l \leq n$. The optimization problem at each encoder is formulated as minimization of the Hamming distortion between the input and output sequences, in other words, we set $b(a) = a$ for any $a \in \mathcal{X}$ for both encoders and let $\lambda(a, b) = \delta_1 > 0$ for first encoder and $\lambda(a, b) = \delta_2 > 0$, for the second encoder, for all $a, b \in \mathcal{X}$ with $a \neq b$.

Moreover, the second encoder can be simplified as described in Subsection 2.8 by substituting variable X_0 by a variable U_0 over the same alphabet $\mathcal{X}_0^+$, satisfying the following requirements

1) U_0 is independent of X_1;

2) $X_0 = \pi(X_1, U_0)$ for some function $\pi : \mathcal{X}_1^+ \times \mathcal{X}_0^+ \to \mathcal{X}_0^+$;

3) $X - (X_1, X_0) - U_0$ forms a Markov chain.

Next we construct the variable U_0 and the mapping π. Assume without restricting the generality that $\mathcal{X}_1^+ = \{0, 1, \cdots, \tau - 1\}$. Consider the function $\sigma : \mathcal{X}_1^+ \times \mathcal{X} \to \mathcal{X}$, defined as follows

$$\sigma(x_1, x) \triangleq \begin{cases} x \oplus_\tau (\tau - x_1), & \text{if } x \in \mathcal{X}_1^+ \\ x, & \text{if } x \in \mathcal{X} \setminus \mathcal{X}_1^+ \end{cases} .$$

where $\oplus_\tau$ denotes modulo-τ addition. Define

$$\pi(x_1, x_0) \triangleq \begin{cases} x_0 \oplus_\tau x_1, & \text{if } x_0 \in \mathcal{X}_1^+ \\ x_0, & \text{otherwise} \end{cases} .$$

Further define $\pi_1 : \mathcal{X}_1^+ \times \mathcal{X}_0^+ \to \mathcal{X}_0^+$ as the restriction of σ over the set $\mathcal{X}_1^+ \times \mathcal{X}_0^+$. Let $U_0 \triangleq \pi_1(X_1, X_0)$. It can be easily verified that the all the aforementioned requirements for U_0 are satisfied. Further define $\Delta X = \sigma(X_1, X)$. According to the argument in Subsection 2.8, the encoder of the refinement layer can be converted into an encoder of the variable ΔX, with input sequence Δx^n, where $\Delta x(l) = \sigma(x_{1,i}(l), x(l))$, $1 \le l \le n$, and i is the index selected by the encoder of the base layer. Encoder 2 needs to randomly generate only one codebook $\mathcal{C}_0 = \{u_{0,j}^n\}_{j=1}^{nR_2}$ according to the distribution $\prod_{l=1}^n p_{U_0}(\cdot)$. The equality $d_H(x, x_0) = d_H(\sigma(x_1, x), \sigma(x_1, x_0))$ can be easily derived. Therefore, the optimization problem at encoder 2 can be equivalently formulated as the minimization of $d_H^{(n)}(\Delta x^n, u_0^n)$ over all $u_0^n \in \mathcal{C}_0$.

Chapter 4

Practical LDGM-based Coding Scheme For Quadratic Gaussian L-Description Coding

In this chapter we propose a practical LDGM-based coding scheme for quadratic Gaussian L-description coding, with individual and central distortion constraints. In the first section, it is first pointed out that the coding problem for an arbitrary point on the dominant face of an L-description El Gamal-Cover (EGC) region can be converted to that for a vertex of a K-description EGC region for some $K \leq 2L-1$, where the latter problem can be solved via successive coding. Then it is shown how, for the quadratic Gaussian case, each step in the successive coding can be reduced to a Gaussian quantization operation via Gram-Schmidt orthogonalization. The special structure of the covariance matrix associated with the sum-rate optimal EGC region, is exploited in order to simplify the calculation of the related coefficients and to derive an efficient implementation of the Gram-Schmidt orthogonalization. Furthermore, a new interpretation of the quantization splitting method developed in [52, 53] is given, by eliminating the use of conditional codebooks. In the second section the practical coding scheme is introduced, which consists in implementing the Gaussian quantization operation involved at each step using a multilevel LDGM code with the message passing algorithm described in section 2.5.

4.1 Successive Coding Scheme

For any L auxiliary random variables $U_{\{1\}}, \cdots, U_{\{L\}}$ jointly distributed with the generic source variable X, we define $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ as the set of rate tuples $(R_1, \cdots, R_L)$ satisfying

$$\sum_{\ell\in\mathcal{A}} R_\ell \geq \sum_{\ell\in\mathcal{A}} H(U_{\{\ell\}}) - H(U_{\{\ell\}}, \ell \in \mathcal{A}|X), \quad \emptyset \subset \mathcal{A} \subseteq \{1, \cdots, L\}.$$

Let $\mathcal{P}(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$ be the set of conditional distributions $p_{U_{\{1\}},\cdots,U_{\{L\}}|X}$ such that

$$\mathbb{E}[d(X, g_{\{\ell\}}(U_{\{\ell\}}))] \leq d_{\{\ell\}}, \quad \ell = 1, \cdots, L, \tag{4.1}$$

$$\mathbb{E}[d(X, g_{\{1,\cdots,L\}}(U_{\{1\}}, \cdots, U_{\{L\}}))] \leq d_{\{1,\cdots,L\}} \tag{4.2}$$

for some functions $g_{\{\ell\}}$, $\ell = 1, \cdots, L$, and $g_{\{1,\cdots,L\}}$. Then, for given distortion tuple $(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$, the set of rate tuples $(R_1, \cdots, R_L)$ such that $(R_1, \cdots, R_L, d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$ is in the EGC region defined in section 1.3.4, coincides with the convex closure of

$$\bigcup_{p_{U_{\{1\}},\cdots,U_{\{L\}}|X}\in\mathcal{P}(d_{\{1\}},\cdots,d_{\{L\}},d_{\{1,\cdots,L\}})}\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X}).$$

It is known [45] that

$$\mathcal{R}_{EGC}(d_{\{1\}},\cdots,d_{\{L\}},d_{\{1,\cdots,L\}})\subseteq\mathcal{R}(d_{\{1\}},\cdots,d_{\{L\}},d_{\{1,\cdots,L\}}),$$

where $\mathcal{R}(d_{\{1\}},\cdots,d_{\{L\}},d_{\{1,\cdots,L\}})$ denotes the set rate tuples $(R_1,\cdots,R_L)$ such that $(R_1,\cdots,R_L,d_{\{1\}},\cdots,d_{\{L\}},d_{\{1,\cdots,L\}})$ is in the L-description region with individual and central constraints, defined in section 1.3.2.

Roughly speaking, one may view $U_{\{1\}},\cdots,U_{\{L\}}$ as L descriptions of source X; moreover, $g_{\{1\}}(U_{\{1\}}),\cdots,g_{\{L\}}(U_{\{L\}})$ can be interpreted as the reconstructions based on individual descriptions while $g_{\{1,\cdots,L\}}(U_{\{1\}},\cdots,U_{\{L\}})$ can be interpreted as the reconstruction based on the complete set of descriptions.

Therefore, we shall primarily focus on $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ and simply refer to it as the EGC region when no confusion can arise. As observed in [46,47], $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ is a contra-polymatroid and its vertices can be easily characterized. Specifically, $(R_1(\pi),\cdots,R_L(\pi))$ is a vertex of $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ for every permutation π on $\{1,\cdots,L\}$, where

$$R_{\pi(1)}(\pi)=I(X;U_{\{\pi(1)\}}),\tag{4.3}$$

$$R_{\pi(\ell)}(\pi)=I(X,U_{\{\pi(1)\}},\cdots,U_{\{\pi(\ell-1)\}};U_{\{\pi(\ell)\}}),\quad \ell=2,\cdots,L.\tag{4.4}$$

The dominant face of $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$, denoted as $\mathcal{D}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$, is the set of rate tuples satisfying

$$\sum_{\ell=1}^{L} R_\ell = \sum_{\ell=1}^{L} H(U_{\{\ell\}}) - H(U_{\{1\}}, \cdots, U_{\{L\}}|X).$$

It can be readily verified by leveraging (4.3) and (4.4) that all the vertices are on the dominant face.

The vertices of $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ are of particular importance since every rate tuple in $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ is dominated in a component-wise manner by some rate tuple in $\mathcal{D}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ and the latter can be expressed as a convex combination of no more than L vertices. As pointed out in [52,53], the expression of the vertices (see (4.3) and (4.4)) suggests a successive coding scheme which can be roughly described as follows[1]: for vertex $(R_1(\pi), \cdots, R_L(\pi))$, one first uses X to produce $U_{\{\pi(1)\}}$, then successively from $\ell = 2$ to L, uses $(X, U_{\{\pi(1)\}}, \cdots, U_{\{\pi(\ell-1)\}})$ to produce $U_{\{\pi(\ell)\}}$. Furthermore, every rate tuple in $\mathcal{D}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ is achievable via suitable timesharing of such successive coding schemes.

4.1.1 Efficient Version for Quadratic Gaussian Source

Now we proceed to propose an efficient implementation of the aforementioned successive coding scheme in the quadratic Gaussian case, where p_X is a Gaussian distribution with mean zero and variance σ_X^2, and $d(\cdot, \cdot)$ is the standard squared error distortion measure. In this setting it is known [47] that $\mathcal{R}_{EGC}(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}}) = \mathcal{R}(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$; moreover, it suffices to assume that $U_{\{1\}}, \cdots, U_{\{L\}}$ are zero-mean and jointly Gaussian with the generic source variable X. By exploiting

[1]For simplicity, here we describe the scheme in the form of single-letter operation. However, it should be noted that to approach the information-theoretic limits, one has to implement such a scheme over long blocks.

the properties of the Gaussian distribution, the following simplified version of the successive coding scheme, referred to as the successive quantization scheme, was developed in [52,53]. Without loss of generality, we shall assume $\pi(\ell)=\ell$, $\ell=1,\cdots,L$. Using the Gram-Schmidt orthogonalization procedure, we can write

$$U_{\{\ell\}}=\hat{U}_{\{\ell\}}+\Delta_\ell, \quad \ell=1,\cdots,L,$$

where

$$\begin{aligned}
&\hat{U}_{\{1\}}=\mathbb{E}[U_{\{1\}}|X],\\
&\hat{U}_{\{\ell\}}=\mathbb{E}[U_{\{\ell\}}|X,U_{\{1\}},\cdots,U_{\{\ell-1\}}], \quad \ell=2,\cdots,L,
\end{aligned}$$

and $X,\Delta_1,\cdots,\Delta_L$ are jointly independent and Gaussian. Now (4.3) and (4.4) can be rewritten as

$$R_\ell(\pi)=I(\hat{U}_{\{\ell\}};U_{\{\ell\}})=I(\hat{U}_{\{\ell\}};\hat{U}_{\{\ell\}}+\Delta_\ell), \quad \ell=1,\cdots,L. \tag{4.5}$$

As observed in [52, 53], one can readily obtain an L-step successive quantization scheme by interpreting $\hat{U}_{\{\ell\}}$, $U_{\{\ell\}}$, and Δ_ℓ in (4.5) respectively as the quantization input, the quantization output, and the quantization error at step ℓ. Note that the explicit expressions of $\hat{U}_{\{1\}},\cdots,\hat{U}_{\{L\}}$ in terms of $(X,U_{\{1\}},\cdots,U_{\{L\}})$ depend on the covariance matrix of $(X,U_{\{1\}},\cdots,U_{\{L\}})$, which in turn depends on distortion constraints $d_{\{1\}},\cdots,d_{\{L\}}$, and $d_{\{1,\cdots,L\}}$. A direct derivation of such expression, though possible for the case $L=2$ [52], appears to be cumbersome for general L. Fortunately, it turns out that the special structure of the optimal covariance matrix of

$(X, U_{\{1\}}, \cdots, U_{\{L\}})$ allows for an efficient implementation of Gram-Schmidt orthogonalization as well as a simple calculation of the relevant coefficients.

Without loss of generality, we shall assume $0 < d_{\{\ell\}} \leq \sigma_X^2$, $\ell = 1, \cdots, L$, and $0 < d_{\{1,\cdots,L\}} \leq \sigma_X^2$. Define

$$R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}}) = \min \Big\{ \sum_{\ell=1}^{L} R_\ell : (R_1, \cdots, R_L) \in \mathcal{R}(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}}) \Big\}.$$

It is known [46, 47] that

$$R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}}) = \max_{a \in [0, \sigma_X^2]} \frac{1}{2} \log \left(\frac{\sigma_X^{4L-2}(\sigma_X^2 d_{\{1,\cdots,L\}} - a d_{\{1,\cdots,L\}} + a\sigma_X^2)}{d_{\{1,\cdots,L\}} \prod_{\ell=1}^{L} (\sigma_X^2 d_{\{\ell\}} - a d_{\{\ell\}} + a\sigma_X^2)} \right).$$

In particular,

$$R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}}) = \begin{cases} \frac{1}{2} \sum_{\ell=1}^{L} \log(\frac{\sigma_X^2}{d_{\{\ell\}}}), & d_{\{1,\cdots,L\}} \geq \overline{d}_{\{1,\cdots,L\}} \\ \frac{1}{2} \log(\frac{\sigma_X^2}{d_{\{1,\cdots,L\}}}), & d_{\{1,\cdots,L\}} \leq \underline{d}_{\{1,\cdots,L\}} \end{cases},$$

where

$$\overline{d}_{\{1,\cdots,L\}} = \Big(\sum_{\ell=1}^{L} \frac{1}{d_{\{\ell\}}} - \frac{L-1}{\sigma_X^2} \Big)^{-1}; \qquad \underline{d}_{\{1,\cdots,L\}} = \sum_{\ell=1}^{L} d_\ell - (L-1)\sigma_X^2.$$

Therefore, if $d_{\{1,\cdots,L\}} \geq \overline{d}_{\{1,\cdots,L\}}$, then one can decrease $d_{\{1,\cdots,L\}}$ until $d_{\{1,\cdots,L\}} = \overline{d}_{\{1,\cdots,L\}}$ without affecting $R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$; similarly, if $d_{\{1,\cdots,L\}} \leq \underline{d}_{\{1,\cdots,L\}}$, then one can decrease one of $d_{\{\ell\}}$, $\ell = 1, \cdots, L$, until $d_{\{1,\cdots,L\}} = \underline{d}_{\{1,\cdots,L\}}$ without affecting $R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$. As a consequence, there is no loss of generality in assuming $\underline{d}_{\{1,\cdots,L\}} \leq d_{\{1,\cdots,L\}} \leq \overline{d}_{\{1,\cdots,L\}}$. In this case we have

$$R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}}) = \frac{1}{2} \log \Big(\frac{\sigma_X^{4L-2}(\sigma_X^2 d_{\{1,\cdots,L\}} - \hat{a} d_{\{1,\cdots,L\}} + \hat{a}\sigma_X^2)}{d_{\{1,\cdots,L\}} \prod_{\ell=1}^L (\sigma_X^2 d_{\{\ell\}} - \hat{a} d_{\{\ell\}} + \hat{a}\sigma_X^2)} \Big),$$

where $\hat{a} \in [0, \sigma_X^2]$ is the solution to the following equation

$$\Big(\frac{\sigma_X^2 d_{\{1,\cdots,L\}}}{\sigma_X^2 - d_{\{1,\cdots,L\}}} + a \Big)^{-1} = \sum_{\ell=1}^{L} \Big(\frac{\sigma_X^2 d_{\{\ell\}}}{\sigma_X^2 - d_{\{\ell\}}} + a \Big)^{-1}.$$

In particular, when $d_{\{1\}} = \cdots = d_{\{L\}} = d$, we have

$$\hat{a} = \frac{\sigma_X^4 d - L\sigma_X^4 d_{\{1,\cdots,L\}} - \sigma_X^2 d d_{\{1,\cdots,L\}} + L\sigma_X^2 d d_{\{1,\cdots,L\}}}{(L-1)(\sigma_X^2 - d)(\sigma_X^2 - d_{\{1,\cdots,L\}})}$$

and

$$R_\Sigma(d, \cdots, d, d_{\{1,\cdots,L\}}) = \frac{1}{2} \log \Big(\frac{(L-1)^{L-1}\sigma_X^2(\sigma_X^2 - d_{\{1,\cdots,L\}})^L}{L^L d_{\{1,\cdots,L\}}(\sigma_X^2 - d)(d - d_{\{1,\cdots,L\}})^{L-1}} \Big).$$

Note that in the current setting (4.1) and (4.2) can be rewritten as

$$\mathbb{E}[(X - \mathbb{E}[X|U_{\{\ell\}}])^2] \leq d_{\{\ell\}}, \quad \ell = 1, \cdots, L, \tag{4.6}$$

$$\mathbb{E}[(X - \mathbb{E}[X|U_{\{1\}}, \cdots, U_{\{L\}}])^2] \leq d_{\{1,\cdots,L\}}. \tag{4.7}$$

It is known [46, 47] that if $U_{\{1\}}, \cdots, U_{\{L\}}$ are zero-mean and jointly Gaussian with X

such that

$$\mathbb{E}[XU_{\{\ell\}}] = \sigma_X^2, \quad \ell = 1, \cdots, L, \tag{4.8}$$

$$\mathbb{E}[U_{\{\ell\}}U_{\{\ell'\}}] = \begin{cases} \sigma_X^2 + \sigma_{\{\ell\}}^2, & \ell = \ell' \\ \sigma_X^2 - \hat{a}, & \ell \neq \ell' \end{cases}, \tag{4.9}$$

where $\sigma_{\{\ell\}}^2 = \frac{\sigma_X^2 d_{\{\ell\}}}{\sigma_X^2 - d_{\{\ell\}}}$, $\ell = 1, \cdots, L$, then rate tuples in $\mathcal{D}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ achieve the minimum sum rate $R_\Sigma(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$ and distortion constraints (4.6) and (4.7) are satisfied; the corresponding $\mathcal{R}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$ will be referred to as the sum-rate optimal Gaussian EGC region. Now we proceed to give an explicit construction of such $(U_{\{1\}}, \cdots, U_{\{L\}})$. Let

$$\sigma_{\{1,\cdots,\ell\}}^2 = \left(\sum_{i=1}^{\ell} \left(\sigma_{\{i\}}^2 + \hat{a}\right)^{-1}\right)^{-1} - \hat{a}, \quad \ell = 2, \cdots, L.$$

Note that

$$\hat{a} = \sqrt{(\sigma_{\{1,\cdots,\ell-1\}}^2 - \sigma_{\{1,\cdots,\ell\}}^2)(\sigma_{\{\ell\}}^2 - \sigma_{\{1,\cdots,\ell\}}^2)} - \sigma_{\{1,\cdots,\ell\}}^2, \quad \ell = 2, \cdots, L.$$

Let $N_{\{1,\cdots,L\}}$ and $N'_{\{1,\cdots,\ell\}}$, $\ell = 2, \cdots, L$, be L zero-mean Gaussian random variables, where the variance of $N_{\{1,\cdots,L\}}$ is $\sigma_{\{1,\cdots,L\}}^2$ and the rest have unit variance. We assume that X, $N_{\{1,\cdots,L\}}$, and $N'_{\{1,\cdots,\ell\}}$, $\ell = 2, \cdots, L$, are jointly independent. One can successively construct

$$N_{\{1,\cdots,\ell\}} = N_{\{1,\cdots,\ell+1\}} + \sqrt{\sigma_{\{1,\cdots,\ell\}}^2 - \sigma_{\{1,\cdots,\ell+1\}}^2} N'_{\{1,\cdots,\ell+1\}}$$

from $\ell = L-1$ to 1. Now let

$$U_{\{1,\cdots,\ell\}} = X + N_{\{1,\cdots,\ell\}}, \quad \ell = 1,\cdots,L.$$
$$U_{\{\ell\}} = U_{\{1,\cdots,\ell\}} - \sqrt{\sigma^2_{\{\ell\}} - \sigma^2_{\{1,\cdots,\ell\}}} N'_{\{1,\cdots,\ell\}}, \quad \ell = 2,\cdots,L.$$

It can be verified that the constructed $(U_{\{1\}},\cdots,U_{\{L\}})$ satisfies (4.8) and (4.9). Note that $(U_1,\cdots,U_{\{\ell\}}) - U_{\{1,\cdots,\ell\}} - (X, U_{\{\ell+1\}},\cdots,U_{\{L\}})$ form a Markov chain, $\ell = 2,\cdots,L$. Therefore, we have

$$\hat{U}_{\{1\}} = X,$$
$$\hat{U}_{\{\ell\}} = \mathbb{E}[U_{\{\ell\}}|X, U_{\{1,\cdots,\ell-1\}}] = \gamma_{\ell-1}X + \beta_{\ell-1}U_{\{1,\cdots,\ell-1\}}, \quad \ell = 2,\cdots,L, \tag{4.10}$$

where $\gamma_{\ell-1} = 1 + \frac{\hat{a}}{\sigma^2_{\{1,\cdots,\ell-1\}}}$ and $\beta_{\ell-1} = -\frac{\hat{a}}{\sigma^2_{\{1,\cdots,\ell-1\}}}$. It is easy to see that

$$U_{\{1,\cdots,\ell\}} = \eta_{\ell-1}U_{\{1,\cdots,\ell-1\}} + \bar{\eta}_{\ell-1}U_{\{\ell\}}, \quad \ell = 2,\cdots,L, \tag{4.11}$$

where $\eta_{\ell-1} = 1 - \bar{\eta}_{\ell-1} = \frac{\sqrt{\sigma^2_{\{\ell\}} - \sigma^2_{\{1,\cdots,\ell\}}}}{\sqrt{\sigma^2_{\{1,\cdots,\ell-1\}} - \sigma^2_{\{1,\cdots,\ell\}}} + \sqrt{\sigma^2_{\{\ell\}} - \sigma^2_{\{1,\cdots,\ell\}}}}$. This recurrence relation leads to an efficient implementation of Gram-Schmidt orthogonalization (see Figure 4.1). It can also be verified that

$$\mathbb{E}[X|U_{\{\ell\}}] = \alpha_{\{\ell\}}U_{\{\ell\}}, \quad \ell = 1,\cdots,L, \tag{4.12}$$
$$\mathbb{E}[X|U_{\{1\}},\cdots,U_{\{\ell\}}] = \mathbb{E}[X|U_{\{1,\cdots,\ell\}}] = \alpha_{\{1,\cdots,\ell\}}U_{\{1,\cdots,\ell\}}, \quad \ell = 2,\cdots,L, \tag{4.13}$$
$$\mathbb{E}[\Delta_1^2] = \sigma^2_{\{1\}},$$
$$\mathbb{E}[\Delta_\ell^2] = \sigma^2_{\{\ell\}} - \frac{\hat{a}^2}{\sigma^2_{\{1,\cdots,\ell-1\}}}, \quad \ell = 2,\cdots,L.$$

where $\alpha_{\{\ell\}} = \frac{\sigma_X^2}{\sigma_X^2+\sigma_{\{\ell\}}^2}$ and $\alpha_{\{1,\cdots,\ell\}} = \frac{\sigma_X^2}{\sigma_X^2+\sigma_{\{1,\cdots,\ell\}}^2}$. In particular, when $d_{\{1\}} = \cdots = d_{\{\ell\}} = d$, we have

$$\begin{aligned}
&\sigma_{\{1\}}^2 = \cdots = \sigma_{\{L\}}^2 = \frac{\sigma_X^2 d}{\sigma_X^2 - d} \triangleq \sigma^2,\\
&\sigma_{\{1,\cdots,\ell\}}^2 = \frac{1}{\ell}\sigma^2 - \frac{\ell-1}{\ell}\hat{a} = \frac{\sigma_X^2 d}{\ell(\sigma_X^2 - d)} - \frac{\ell-1}{\ell}\hat{a}, \quad \ell = 2, \cdots, L,\\
&U_{\{1,\cdots,\ell\}} = \frac{\ell-1}{\ell}U_{\{1,\cdots,\ell-1\}} + \frac{1}{\ell}U_{\{\ell\}} = \frac{1}{\ell}\sum_{i=1}^{\ell} U_{\{i\}}, \quad \ell = 2, \cdots, L,\\
&\mathbb{E}[\Delta_1^2] = \sigma^2,\\
&\mathbb{E}[\Delta_\ell^2] = \sigma^2 - \frac{(\ell-1)\hat{a}^2}{\sigma^2 - (\ell-2)\hat{a}}, \quad \ell = 2, \cdots, L.
\end{aligned}$$

It should be pointed out that although we have mainly focused on the minimum sum rate, one can obtain similar results in a more general setting. Indeed, it can be shown by leveraging the construction in [47] that the Gram-Schmidt orthogonalization procedure can be simplified in essentially the same manner for a vertex that achieves a general minimum weighted sum rate.

4.1.2 Quantization Splitting

As mentioned earlier, every rate tuple on the dominant face of an EGC region is achievable via timesharing of the successive coding schemes for vertices. Alternatively, one can use the splitting method developed in [52, 53].

Given $(X, U_{\{1\}}, \cdots, U_{\{L\}})$, we say $(U_{\{1\},1}, \cdots, U_{\{L\},1})$ is split from $(U_{\{1\}}, \cdots, U_{\{L\}})$ if $U_{\{\ell\},1} - U_{\{\ell\}} - (X, U_{\{\ell'\},1}, U_{\{\ell'\}}, \ell' \neq \ell)$ form a Markov chain for all ℓ. Let $\mathcal{U} = \{U_{\{1\},1}, U_{\{1\}}, \cdots, U_{\{L\},1}, U_{\{L\}}\}$. We say μ is a well-ordered permutation on $\mathcal{U}$ if $\mathcal{U}_{\{\ell\},1}$

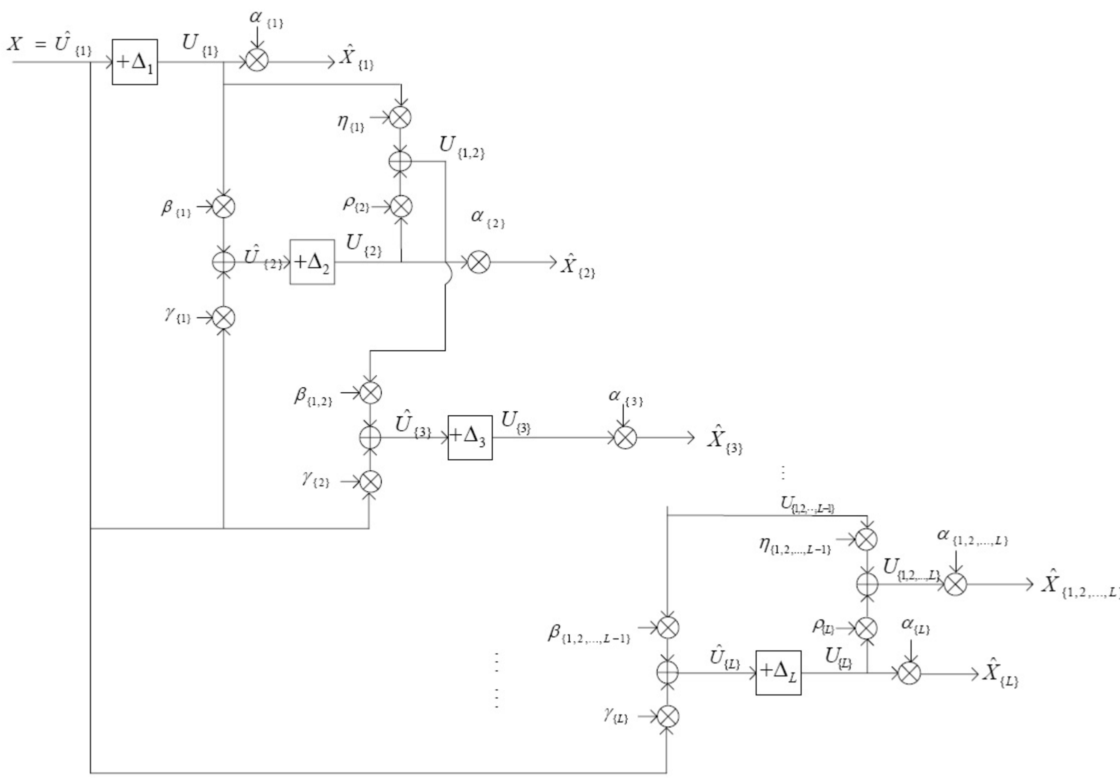

Figure 4.1: Successive quantization scheme for a vertex of the sum-rate optimal Gaussian EGC region.

is placed before $\mathcal{U}_{\{\ell\}}$ for all ℓ. For any $U \in \mathcal{U}$, let $\{U\}_\mu^-$ denote the set of random variables placed before U in μ.

It is known [52, 53] that for any $(R_1, \cdots, R_L) \in \mathcal{D}(p_{U_{\{1\}},\cdots,U_{\{L\}}|X})$, one can find $(U_{\{1\},1}, \cdots, U_{\{L\},1})$ split from $(U_{\{1\}}, \cdots, U_{\{L\}})$ and a well-ordered permutation μ such that

$$R_\ell = R_{\ell,1} + R_{\ell,2}, \quad \ell = 1, \cdots, L,$$

where

$$R_{\ell,1} = I(X, \{U_{\{\ell\},1}\}_\mu^-; U_{\{\ell\},1}),$$
$$R_{\ell,2} = I(X, \{U_{\{\ell\}}\}_\mu^-; U_{\{\ell\}}|U_{\{\ell\},1});$$

moreover, at least one $U_{\{\ell\},1}$ can be set to zero[2] and removed from μ.

Note that $R_{\ell,2}$ is expressed as a conditional mutual information. This is why in [52, 53] conditional codebooks are used in the random coding argument for the splitting method. In fact, an inspection of the random coding argument in [52, 53] reveals that the resulting scheme requires one to construct and store $2^{nR_{\ell,1}}$ conditional codebooks, each of size $2^{nR_{\ell,2}}$, for the ℓ-th description. Here we shall give a new interpretation of the splitting method by converting the expression of $R_{\ell,2}$ from a conditional form to an unconditional form and consequently eliminating the use of conditional codebooks[3]. By applying Lemma 1 in [58] for every $\ell = 1, \cdots, L$, it follows that there exist random variables $U_{\{1\},2}, \cdots, U_{\{L\},2}$ jointly distributed with $(X, U_{\{1\},1}, U_{\{1\}}, \cdots, U_{\{L\},1}, U_{\{L\}})$ such that the following properties are satisfied for all ℓ:

P1) $U_{\{\ell\},2}$ is independent of $U_{\{\ell\},1}$;

P2) $U_{\{\ell\}}$ is a deterministic function of $U_{\{\ell\},1}$ and $U_{\{\ell\},2}$;

P3) $U_{\{\ell\},2} - (U_{\{\ell\},1}, U_{\{\ell\}}) - (X, U_{\{\ell'\},1}, U_{\{\ell'\},2}, U_{\{\ell'\}}, \ell' \neq \ell)$ form a Markov chain.

Let $\mathcal{U}' = \{U_{\{1\},1}, U_{\{1\},2}, \cdots, U_{\{L\},1}, U_{\{L\},2}\}$ and μ' be a permutation on $\mathcal{U}'$ induced by μ with $U_{\{\ell\}}$ replaced by $U_{\{\ell\},2}$ at the corresponding positions. For any $U \in \mathcal{U}'$, let

[2] In this case we have $R_{\ell,1} = 0$ and $R_\ell = R_{\ell,2} = I(X, \{U_{\{\ell\}}\}_\mu^-; U_{\{\ell\}})$.

[3] More precisely, the new interpretation allows one to replace those $2^{nR_{\ell,1}}$ conditional codebooks with a single codebook of size $2^{nR_{\ell,2}}$.

$\{U\}_{\mu'}^-$ denote the set of random variables placed before U in μ'. By P2) and P3), we can rewrite $R_{\ell,1}$ as

$$R_{\ell,1} = I(X, \{U_{\{\ell\},1}\}_{\mu'}^-; U_{\{\ell\},1}).$$

Moreover, it follows by P1), P2), and P3) that

$$\begin{aligned} I(X, \{U_{\{\ell\},2}\}_{\mu'}^-; U_{\{\ell\},2}) &= I(X, \{U_{\{\ell\},2}\}_{\mu'}^-; U_{\{\ell\},2}|U_{\{\ell\},1}) + I(U_{\{\ell\},1}; U_{\{\ell\},2}) \\ &= I(X, \{U_{\{\ell\},2}\}_{\mu'}^-; U_{\{\ell\},2}|U_{\{\ell\},1}) = I(X, \{U_{\{\ell\}}\}_{\mu}^-; U_{\{\ell\}}|U_{\{\ell\},1}). \end{aligned}$$

Therefore, we have

$$R_{\ell,2} = I(X, \{U_{\{\ell\},2}\}_{\mu'}^-; U_{\{\ell\},2}).$$

Now by ordering $R_{1,1}, R_{1,2}, \cdots, R_{L,1}, R_{L,2}$ according to μ', one can readily see that the coding problem for an arbitrary point on the dominant face of an L-description EGC region can be converted to that for a vertex of a K-description EGC region for some $K \leq 2L-1$ (due to the fact that at least one $U_{\{\ell\},1}$ can be set to zero and removed from μ and μ'), where the latter problem can be solved via successive coding. Note that we essentially split each description into two coarse descriptions; moreover, according to P2), the original description can be recovered from the two coarse descriptions.

More concrete results can be obtained in the quadratic Gaussian case. In this setting there is no loss of generality in assuming that $U_{\{1\},1}, U_{\{1\},2}, \cdots, U_{\{L\},1}, U_{\{L\},2}$

are zero-mean and jointly Gaussian with $(X, U_{\{1\}}, \cdots, U_{\{L\}})$. Specifically, we can let

$$U_{\{\ell\},1} = U_{\{\ell\}} + Z_\ell, \quad \ell = 1, \cdots, L,$$
$$U_{\{\ell\},2} = U_{\{\ell\}} - b_\ell Z_\ell, \quad \ell = 1, \cdots, L,$$

where Z_ℓ is a Gaussian random variable with mean zero and variance $\sigma^2_{Z_\ell}$, and $b_\ell = \frac{\mathbb{E}[U^2_{\{\ell\}}]}{\sigma^2_{Z_\ell}}$; moreover, $Z_1, \cdots, Z_L$, and $(X, U_{\{1\}}, \cdots, U_{\{L\}})$ are jointly independent. The values of $\sigma^2_{Z_1}, \cdots, \sigma^2_{Z_L}$ are determined by $(R_1, \cdots, R_L)$. Note that in the extreme case when $\sigma^2_{Z_\ell} = \infty$, we let $U_{\{\ell\},1} = 0$ and $U_{\{\ell\},2} = U_{\{\ell\}}$; similarly, when $\sigma^2_{Z_\ell} = 0$, we let $U_{\{\ell\},1} = U_{\{\ell\}}$ and $U_{\{\ell\},2} = 0$. It is easy to verify that P1), P2), and P3) are satisfied; in particular, we have

$$U_{\{\ell\}} = \tau_\ell U_{\{\ell\},1} + \bar{\tau}_\ell U_{\{\ell\},2}, \quad \ell = 1, \cdots, L.$$

where $\tau_\ell = 1 - \bar{\tau}_\ell = \frac{b_\ell}{b_\ell + 1}$. To obtain a successive quantization scheme, one can apply the Gram-Schmidt orthogonalization procedure to $(U_{\{1\},1}, U_{\{1\},2}, \cdots, U_{\{L\},1}, U_{\{L\},2})$ with the projection order specified by μ'.

Now we proceed to give a detailed treatment of the case $L = 2$. It is known [39, 46, 47] that there is no loss of generality in assuming $\underline{d}_{\{1,2\}} \leq d_{\{1,2\}} \leq \overline{d}_{\{1,2\}}$; moreover, in this setting

$$\mathcal{R}(d_{\{1\}}, d_{\{2\}}, d_{\{1,2\}}) = \mathcal{R}(p_{U_{\{1\}},U_{\{2\}}|X}),$$

where $p_{U_{\{1\}},U_{\{2\}}|X}$ is the conditional Gaussian distribution specified by (4.8) and (4.9) with

$$\sigma^2_{\{\ell\}} = \frac{\sigma_X^2 d_{\{\ell\}}}{\sigma_X^2 - d_{\{\ell\}}}, \quad \ell = 1, 2,$$
$$\hat{a} = \sqrt{\Big(\frac{\sigma_X^2 d_{\{1\}}}{\sigma_X^2 - d_{\{1\}}} - \frac{\sigma_X^2 d_{\{1,2\}}}{\sigma_X^2 - d_{\{1,2\}}}\Big)\Big(\frac{\sigma_X^2 d_{\{2\}}}{\sigma_X^2 - d_{\{2\}}} - \frac{\sigma_X^2 d_{\{1,2\}}}{\sigma_X^2 - d_{\{1,2\}}}\Big)} - \frac{\sigma_X^2 d_{\{1,2\}}}{\sigma_X^2 - d_{\{1,2\}}}.$$

Note that for any $(R_1, R_2) \in \mathcal{D}(p_{U_{\{1\}},U_{\{2\}}|X})$, we can set $U_{\{2\},1} = 0$, $U_{\{2\},2} = U_{\{2\}}$, and write

$$\begin{aligned} R_1 &= R_{1,1} + R_{1,2}, \\ R_2 &= I(X, U_{\{1\},1}; U_{\{2\}}), \end{aligned} \tag{4.14}$$

where

$$R_{1,1} = I(X; U_{\{1\},1}), \tag{4.15}$$

$$R_{1,2} = I(X, U_{\{1\},1}, U_{\{2\}}; U_{\{1\},2}). \tag{4.16}$$

The Gram-Schmidt orthogonalization procedure yields

$$\begin{aligned} U_{\{1\},1} &= \mathbb{E}[U_{\{1\},1}|X] + \tilde{\Delta}_1, \\ U_{\{2\}} &= \mathbb{E}[U_{\{2\}}|X, U_{\{1\},1}] + \tilde{\Delta}_2, \\ U_{\{1\},2} &= \mathbb{E}[U_{\{1\},2}|X, U_{\{1\},1}, U_{\{2\}}] + \tilde{\Delta}_3, \end{aligned}$$

where $X, \tilde{\Delta}_1, \tilde{\Delta}_2, \tilde{\Delta}_3$ are jointly independent. Therefore, we can rewrite $R_{1,1}$, R_2, and

$R_{1,2}$ as

$$
\begin{aligned}
R_{1,1} &= I(\mathbb{E}[U_{\{1\},1}|X]; U_{\{1\},1}) = I(\mathbb{E}[U_{\{1\},1}|X]; \mathbb{E}[U_{\{1\},1}|X] + \tilde{\Delta}_1),\\
R_2 &= I(\mathbb{E}[U_{\{2\}}|X, U_{\{1\},1}]; U_{\{2\}}) = I(\mathbb{E}[U_{\{2\}}|X, U_{\{1\},1}]; \mathbb{E}[U_{\{2\}}|X, U_{\{1\},1}] + \tilde{\Delta}_2),\\
R_{1,2} &= I(\mathbb{E}[U_{\{1\},2}|X, U_{\{1\},1}, U_{\{2\}}]; U_{\{1\},2}) = I(\mathbb{E}[U_{\{1\},2}|X, U_{\{1\},1}, U_{\{2\}}]; \mathbb{E}[U_{\{1\},2}|X, U_{\{1\},1}, U_{\{2\}}] + \tilde{\Delta}_3).
\end{aligned}
$$

It can be verified that

$$
\begin{aligned}
&\mathbb{E}[U_{\{1\},1}|X] = X,\\
&\mathbb{E}[U_{\{2\}}|X, U_{\{1\},1}] = X + \mathbb{E}[U_{\{2\}} - X|U_{\{1\},1} - X]\\
&= X - \frac{\hat{a}}{\sigma^2_{\{1\}} + \sigma^2_{Z_1}}(U_{\{1\},1} - X) = \nu_1 X + \nu_2 U_{\{1\},1}, \qquad (4.17)\\
&\mathbb{E}[U_{\{1\},2}|X, U_{\{1\},1}, U_{\{2\}}] = X + \mathbb{E}[U_{\{1\},2} - X|U_{\{1\},1} - X, U_{\{2\}} - X]\\
&= X - \frac{\sigma^2_X\sigma^2_{\{2\}} + \hat{a}^2}{(\sigma^2_{\{1\}} + \sigma^2_{Z_1})\sigma^2_{\{2\}} - \hat{a}^2}(U_{\{1\},1} - X) - \frac{\sigma^2_X\hat{a} + (\sigma^2_{\{1\}} + \sigma^2_{Z_1})\hat{a}}{(\sigma^2_{\{1\}} + \sigma^2_{Z_1})\sigma^2_{\{2\}} - \hat{a}^2}(U_{\{2\}} - X)\\
&= \nu_3 X + \nu_4 U_{\{1\},1} + \nu_5 U_{\{2\}}, \qquad (4.18)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}[\tilde{\Delta}^2_1] &= \sigma^2_{\{1\}} + \sigma^2_{Z_1},\\
\mathbb{E}[\tilde{\Delta}^2_2] &= \sigma^2_{\{2\}} - \frac{\hat{a}^2}{\sigma^2_{\{1\}} + \sigma^2_{Z_1}},\\
\mathbb{E}[\tilde{\Delta}^2_3] &= \sigma^2_{\{1\}} + \frac{(\sigma^2_X + \sigma^2_{\{1\}})^2}{\sigma^2_{Z_1}} - \frac{\sigma^4_X\sigma^2_{\{2\}} + 2\sigma^2_X\hat{a}^2 + (\sigma^2_{\{1\}} + \sigma^2_{Z_1})\hat{a}^2}{(\sigma^2_{\{1\}} + \sigma^2_{Z_1})\sigma^2_{\{2\}} - \hat{a}^2},
\end{aligned}
$$

where $\nu_1 = \frac{\sigma^2_{\{1\}} + \sigma^2_{Z_1} + \hat{a}}{\sigma^2_{\{1\}} + \sigma^2_{Z_1}}$, $\nu_2 = -\frac{\hat{a}}{\sigma^2_{\{1\}} + \sigma^2_{Z_1}}$, $\nu_3 = \frac{(\sigma^2_X + \sigma^2_{\{1\}} + \sigma^2_{Z_1})(\sigma^2_{\{2\}} + \hat{a})}{(\sigma^2_{\{1\}} + \sigma^2_{Z_1})\sigma^2_{\{2\}} - \hat{a}^2}$, $\nu_4 = -\frac{\sigma^2_X\sigma^2_{\{2\}} + \hat{a}^2}{(\sigma^2_{\{1\}} + \sigma^2_{Z_1})\sigma^2_{\{2\}} - \hat{a}^2}$, and $\nu_5 = -\frac{\sigma_X^2 \hat{a} + (\sigma_{\{1\}}^2 + \sigma_{Z_1}^2)\hat{a}}{(\sigma_{\{1\}}^2 + \sigma_{Z_1}^2)\sigma_{\{2\}}^2 - \hat{a}^2}$. Note that

$$R_2 = h(U_{\{2\}}) - h(\tilde{\Delta}_2) = \frac{1}{2}\log\left(\frac{\sigma_X^2 + \sigma_{\{2\}}^2}{\sigma_{\{2\}}^2 - \frac{\hat{a}^2}{\sigma_{\{1\}}^2 + \sigma_{Z_1}^2}}\right),$$

which implies

$$\sigma_{Z_1}^2 = \frac{\hat{a}^2}{\sigma_{\{2\}}^2 - 2^{-2R_2}(\sigma_X^2 + \sigma_{\{2\}}^2)} - \sigma_{\{1\}}^2.$$

In particular, when $d_{\{1\}} = d_{\{2\}} = d$ and $R_1 = R_2 = R$, we have

$$R = \frac{1}{4}\log\left(\frac{(\sigma_X^2 + \sigma^2)^2}{\sigma^4 - \hat{a}^2}\right),$$

and consequently,

$$\sigma_{Z_1}^2 = \frac{\hat{a}^2}{\sigma^2 - 2^{-2R}(\sigma_X^2 + \sigma^2)} - \sigma^2 = \sqrt{\sigma^4 - \hat{a}^2},$$

where $\sigma^2 = \frac{\sigma_X^2 d}{\sigma_X^2 - d}$ and $\hat{a} = \frac{\sigma_X^2 d}{\sigma_X^2 - d} - \frac{2\sigma_X^2 d_{\{1,2\}}}{\sigma_X^2 - d_{\{1,2\}}}$.

The quantization splitting system for the 2-description case is depicted in Figure 4.2.

4.2 Practical Scheme Based On LDGM Codes

As discussed previously, each stage ℓ in the successive quantization scheme reduces to a Gaussian quantization operation interpreted as the forward channel $U_{\{\ell\}} = \hat{U}_{\{\ell\}} + \Delta_\ell$, which we implement based on LDGM codes. Therefore, let us first describe the

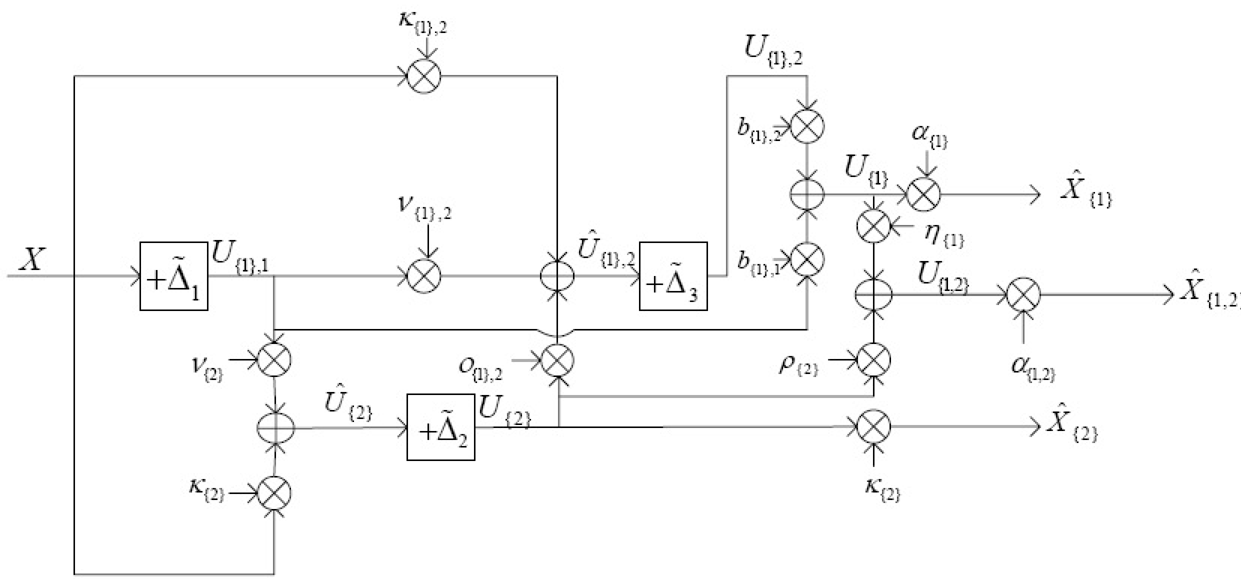

Figure 4.2: Quantization splitting scheme for the Gaussian 2-description case.

proposed Gaussian quantization scheme in a general setting.

4.2.1 Gaussian Quantization with LDGM Codes

Consider an i.i.d. Gaussian source $\hat{U} \sim \mathcal{N}(0, \sigma^2_{\hat{U}})$ and an additive Gaussian noise channel $U = \hat{U} + \Delta$, where $\Delta \sim \mathcal{N}(0, \sigma^2)$ denotes the noise. Our goal is to construct an n-block quantizer of rate $R = I(\hat{U}; U) + \epsilon$ and codebook $\mathcal{C} = \{u_i^n : 1 \leq i \leq 2^{nR}\}$, to approximate this Gaussian channel. To this end we need first to construct a finite random variable $\tilde{U}$ over some finite alphabet $\tilde{\mathcal{U}}$, to approximate U. An additional requirement on $\tilde{U}$, is that a positive integer ω exists such that $2^\omega p_{\tilde{U}}(\tilde{u})$ is a positive integer for every $\tilde{u} \in \tilde{\mathcal{U}}$. Next we apply a multilevel LDGM code described in Section

2.2 to generate the codebook $\mathcal{C} \subseteq \tilde{\mathcal{U}}^n$, such that the marginal distribution of the codewords approximates $p_{\tilde{U}}$, and thus approximates p_U as well; finally, given an input sequence $\hat{u}^n = \hat{u}(1)\cdots\hat{u}(n)$, the quantizer output $u^n = u(1)\cdots u(n)$ is selected from $\mathcal{C}$ using the message passing algorithm described in Section 2.5.

Notice that the MMSE estimator of $\hat{U}$ given the variable U is $\mathbb{E}[\hat{U}|U] = \alpha U$, where $\alpha = \frac{\sigma^2_{\hat{U}}}{\sigma^2_{\hat{U}}+\sigma^2}$. Thus, if the quantizer output is u^n then the optimal source reconstruction is αu^n. Therefore, given the quantizer input sequence $\hat{u}^n$, we formulate the quantizer encoder problem as the problem of selecting the output sequence $u^n \in \mathcal{C}$ such that the mean squared error between $\hat{u}^n$ and αu^n to be minimized.

The performance of this scheme depends on how well the message passing algorithm solves the encoder problem, but also on the choice of random variable $\tilde{U}$. The choice of $\tilde{U}$ is constrained by the requirement that a positive integer ω exists such that $2^\omega p_{\tilde{U}}(\tilde{u})$ is a positive integer for every $\tilde{u} \in \tilde{\mathcal{U}}$. Notice that this condition also constrains the size of the alphabet $\tilde{\mathcal{U}}$ to be at most 2^ω. Therefore it is interesting to consider the problem of optimizing $\tilde{U}$ subject to fixed ω. In order to formulate this problem we will disregard the dependence on the particular behavior of the message passing algorithm. Moreover, we will assume that the LDGM encoder approximates a theoretical random coding scheme[4]. Then the problem of optimizing $\tilde{U}$ is equivalent to optimizing the alphabet $\tilde{\mathcal{U}}$ and the conditional probability $p_{\tilde{U}|\hat{U}}$, formulated

[4]Such an assumption is supported by the results of [32]. Although the argument in [32] is for discrete-valued sources and bounded distortion measures, it can be extended to cover the quadratic Gaussian case using standard techniques.

as follows

$$\begin{aligned} \min_{\tilde{\mathcal{U}}, p_{\tilde{U}|\hat{U}}} \quad & \mathbb{E}[(\hat{U} - \alpha\tilde{U})^2] \\ \text{subject to} \quad & I(\hat{U};\tilde{U}) = I(\hat{U};U) \\ & \tilde{\mathcal{U}} \subseteq \mathbb{R} \\ & 2^{\omega} p_{\tilde{U}}(\tilde{u}) \in \mathbb{N}, \qquad \forall \tilde{u} \in \tilde{\mathcal{U}} \end{aligned} \tag{4.19}$$

The above optimization problem has similar flavor to the problems considered in the context of alphabet constrained rate-distortion theory for continuous-valued sources in [59, 60], but appears to be more difficult due to the additional integer constraint. On the other hand, as a practical solution to LDGM code design, one can modify the requirements to fit the problems solved in [59, 60]. Specifically, one can replace first $\hat{U}$ by the output $\breve{U}$ of a fine scalar quantizer; then drop the last constraint in (4.19), impose instead the condition that $p_{\tilde{U}}(\tilde{u})$ are equal to some fixed values, and use the algorithm of [60] to determine the optimal alphabet $\tilde{\mathcal{U}}$. Alternatively, upon replacing $\hat{U}$ by $\breve{U}$, one can fix only the size of the alphabet $\tilde{\mathcal{U}}$ and determine the probabilities $p_{\tilde{U}}(\tilde{u})$ via the algorithm of [59]; then choose an integer ω such that $2^{\omega} p_{\tilde{U}}(\tilde{u})$ are close to some integer values.

We leave the quest for a solution algorithm to problem (4.19) and/or the investigation of the performance of the aforementioned strategies for future work. In our experiments we confine ourselves to a simple heuristic for the selection of $\tilde{U}$, inspired by the central limit theorem and by the intuition that $\tilde{U}$ has to be a good approximation of variable U. Let $W_1, \cdots, W_\omega$ be ω independent random variables, uniformly distributed over the alphabet $\{-1, 1\}$. Define $\bar{W}_\omega = \frac{\sum_{j=1}^{\omega} W_i \sqrt{\sigma_{\hat{U}}^2 + \sigma^2}}{\sqrt{\omega}}$. According to the central limit theorem, the sequence of random variables $\bar{W}_\omega$ converges to $\mathcal{N}(0, \sigma^2_{\hat{U}} + \sigma^2)$ as $\omega \to \infty$. Therefore, we choose $\tilde{U} = \bar{W}_\omega$.

4.2.2 Successive Quantization with LDGM Code

As discussed in previous section, the successive quantization scheme for vertices of the sum-rate optimal Gaussian EGC region follows the block diagram in Figure 4.1. The operation $U_{\{\ell\}} = \hat{U}_{\{\ell\}} + \Delta_\ell$, at the ℓ-th step, $\ell = 1, \cdots, L$, is implemented using an n-block multilevel LDGM code as described in Section 2.2, for $U = U_{\{\ell\}}$, $\hat{U} = \hat{U}_{\{\ell\}}$, and $\Delta = \Delta_\ell$. The input sequence $\hat{u}^n_{\{\ell\}} = \hat{u}_{\{\ell\}}(1), \cdots, \hat{u}_{\{\ell\}}(n)$ coincides with the source sequence $x^n = x(1) \cdots x(n)$ for $\ell = 1$, and for $\ell > 1$, it is computed based on the sequences $u^n_{\{\kappa\}}$ output at all previous stages $1 \leq \kappa \leq \ell - 1$, according to the recursive equations (4.10) and (4.11), applied symbol by symbol. The sequence $u^n_{\{\ell\}}$ output by the quantizer at stage ℓ, is found using the belief propagation algorithm described in Subsection 2.5. The index i_ℓ formed out of the nR_ℓ information bits corresponding to the selected output is transmitted as the ℓ-th description.

The decoder corresponding to the ℓ-th description receives index i_ℓ and constructs the corresponding codeword $u^n_{\{\ell\}}$ using the factor graph for the ℓ-th stage LDGM code. The source reconstruction $\hat{x}^n_{\{\ell\}}$ is formed by $\hat{x}_{\{\ell\}}(l) = \alpha_{\{\ell\}} u_{\{\ell\}}(l)$, $1 \leq l \leq n$, according to (4.12).

Finally, the central decoder receives all indices $i_1, \cdots, i_L$, recovers $u^n_{\{1\}}, \cdots, u^n_{\{L\}}$, and based on them constructs the sequence $u^n_{\{1,\cdots,L\}}$ using (4.11) recursively. Then the source reconstruction is generated according to (4.13).

4.2.3 Quantization Splitting with LDGM Code

The procedure of quantization splitting for the 2-description case is illustrated in the block diagram of Figure 4.2. It consists of three successive n-block quantizers implemented using multilevel LDGM codes, as described in previous subsection. Specifically, the first quantizer models the forward channel $U_{\{1\},1} = X + \tilde{\Delta}_1$. Its input is the source sequence x^n and its output is denoted by $u^n_{\{1\},1}$. The second quantizer approximates the channel $U_{\{2\}} = \hat{U}_{\{2\}} + \tilde{\Delta}_2$. Its n-block input sequence $\hat{u}^n_{\{2\}}$ is constructed based on x^n and $u^n_{\{1\},1}$ according to (4.17) as shown in the block diagram. Its output is denoted by $u^n_{\{2\}}$. Finally, the third quantizer models the channel $U_{\{1\},2} = \hat{U}_{\{1\},2} + \tilde{\Delta}_3$, with input $\hat{u}^n_{\{1\},2}$ generated from x^n, $u^n_{\{1\},1}$, and $u^n_{\{2\}}$ using (4.18). Its output is denoted by $u^n_{\{1\},2}$.

Let i_1, i_2, i_3 denote the information bit sequences corresponding to the outputs of the three quantizers, respectively. Then indices i_1 and i_3 form the first description, while i_2 constitutes the second description. The decoder of the first description receives i_1 and i_3, recovers $u^n_{\{1\},1}$ and $u^n_{\{1\},2}$, based on which it generates $u^n_{\{1\}}$, and further $\hat{x}^n_{\{1\}}$ as the source reconstruction, using the operations described in Figure 4.2. The decoder of the second description receives index i_2, recovers $u^n_{\{2\}}$, and generates the source reconstruction $\hat{x}^n_{\{2\}}$ according to Figure 4.2. When both descriptions are received at the decoder, the sequence $u^n_{\{1,2\}}$ is generated from $u^n_{\{1\}}$ and $u^n_{\{2\}}$, which is used to further generate the source reconstruction $\hat{x}^n_{\{1,2\}}$ as in Figure 4.2.

Chapter 5

Experimental Results

In this chapter, we present test results using the coding schemes for the MD problem, proposed in the previous chapters. The degree distributions of the LDGM codes are from the website (http://lthcwww.epfl.ch.research/ldpcopt) or obtained by implementing the algorithm in [61]. We have used damping as in [32,37] in our message passing algorithm, if the messages do not converge after 30 iterations.

5.1 Discrete Source with Hamming Distortion Measure

We have tested the proposed coding scheme in each of the three cases of the MD problem for discrete source with Hamming distortion measure. The length n of the input sequence is $10,000$. The value of the threshold η is 0.9. Next we present the experimental results for each case. The values of the empirical distortions are averaged over 100 runs.

5.1.1 No Excess Sum-Rate Case for Binary Sources

We have applied the LDGM-based coding scheme to approach several points (d_1, d_2) on the lower boundary of $\mathcal{D}(d_0)$ for a uniform binary source and for a non-uniform binary source with $p_X(0) = 1/4$. In both cases we have considered $d_0 = 0$ and $d_0 = 0.013$. In our tests the sum-rate $R_1 + R_2$ equals $R(d_0)$, in each case.

The values $\lambda(a, b)$ used in our simulations are as follows:

- Uniform source, encoder 1: $\lambda(0, 1) = 1.0$, $\lambda(1, 0) = 3.2 - 3.6$.
- Non-uniform source, encoder 1: $\lambda(0, 1) = 1.0 - 2.0$, $\lambda(1, 0) = 2.8$.
- Uniform source, encoder 2: $\lambda((0, 1), 1) = \lambda((1, 1), 0) = 3.2$, $\lambda((0, 0), 1) = \lambda((1, 0), 0) = 0.0 - 0.6$,
- Non-uniform source, encoder 2: $\lambda((0, 1), 1) = \lambda((1, 1), 0) = 3.2$, $\lambda((0, 0), 1) = \lambda((1, 0), 0) = 0.0$,

The value of the parameter δ used to set the messages passed by check nodes whose all information variable nodes are fixed is 1.8 for the uniform source and 1.6 for the non-uniform source.

Tables 5.1 and 5.2 present the results for the uniform source, respectively the non-uniform source. In each table the first column contains the pair of rates (R_1, R_2) used in the encoding scheme. The second column indicates whether the pair of rates is an E_1 corner point, E_2 corner point, or a point obtained by timesharing the corner points. We use the symbol T to indicate the latter situation. The third column contains the target distortion triple (d_1, d_2, d_0). The remaining three columns present the empirical values of the three distortions, respectively, averaged over 100 runs.

(R_1, R_2)		(d_1, d_2, d_0)	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_0$
$(0.500, 0.500)$	E_1	$(0.147, 0.272, 0)$	0.154	0.279	0.008
$(0.500, 0.500)$	E_1	$(0.156, 0.266, 0)$	0.156	0.269	0.008
$(0.383, 0.617)$	E_1	$(0.207, 0.207, 0)$	0.209	0.212	0.007
$(0.617, 0.383)$	E_2	$(0.207, 0.207, 0)$	0.212	0.209	0.007
$(0.500, 0.500)$	T	$(0.207, 0.207, 0)$	0.211	0.211	0.007
$(0.445, 0.555)$	E_1	$(0.174, 0.242, 0)$	0.176	0.247	0.009
$(0.676, 0.324)$	E_2	$(0.174, 0.242, 0)$	0.175	0.246	0.009
$(0.500, 0.500)$	T	$(0.174, 0.242, 0)$	0.176	0.247	0.009
$(0.349, 0.551)$	E_1	$(0.215, 0.215, 0.013)$	0.219	0.216	0.021
$(0.551, 0.349)$	E_2	$(0.215, 0.215, 0.013)$	0.216	0.219	0.021
$(0.450, 0.450)$	T	$(0.215, 0.215, 0.013)$	0.218	0.218	0.021
$(0.422, 0.478)$	E_1	$(0.174, 0.258, 0.013)$	0.181	0.258	0.022
$(0.619, 0.281)$	E_2	$(0.174, 0.258, 0.013)$	0.177	0.259	0.021
$(0.450, 0.450)$	T	$(0.174, 0.258, 0.013)$	0.180	0.258	0.022

Table 5.1: Test results in the no excess sum-rate case for the uniform binary source: (d_1, d_2, d_0) is a target distortion triple; (R_1, R_2) is the pair of rates used by the code; $\hat{d}_1$, $\hat{d}_2$, $\hat{d}_0$ are the empirical distortions.

As observed from Tables 5.1 and 5.2, the distortions are very close to the theoretical lower bounds.

5.1.2 No Excess Marginal Rate for Uniform Binary Source

We have used the proposed coding scheme to approach the upper bound $UB(d_1, d_2)$ of equation (3.6) for the central distortion, in the case of no excess marginal rate for the uniform binary source. We have considered three target distortion pairs (d_1, d_2): $d_1 = d_2 = 0.1, 0.2, 0.3$. In our experiments we have set $\lambda(0,1) = \lambda(1,0)$ at each encoder, and the value used for this parameter is 0.5 for encoder 0 and 1.2 for both encoders 1 and 2. The value of δ is 0.5 at encoder 0 and 0.8 at encoders 1 and 2. In all the cases $R_1 = R(d_1) = R_2 = R(d_2)$. The experimental results are summarized in

(R_1, R_2)		(d_1, d_2, d_0)	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_0$
$(0.492, 0.319)$	E_1	$(0.116, 0.116, 0)$	0.120	0.120	0.014
$(0.319, 0.492)$	E_2	$(0.116, 0.116, 0)$	0.120	0.120	0.014
$(0.424, 0.387)$	T	$(0.116, 0.116, 0)$	0.120	0.120	0.014
$(0.502, 0.309)$	E_1	$(0.060, 0.176, 0)$	0.065	0.181	0.014
$(0.162, 0.649)$	E_2	$(0.176, 0.060, 0)$	0.184	0.065	0.016
$(0.373, 0.439)$	T	$(0.104, 0.132, 0)$	0.110	0.137	0.015
$(0.291, 0.423)$	E_1	$(0.123, 0.123, 0.013)$	0.130	0.128	0.025
$(0.423, 0.291)$	E_2	$(0.123, 0.123, 0.013)$	0.128	0.130	0.025
$(0.369, 0.344)$	T	$(0.123, 0.123, 0.013)$	0.129	0.129	0.025
$(0.492, 0.222)$	E_1	$(0.060, 0.191, 0.013)$	0.064	0.197	0.027
$(0.122, 0.591)$	E_2	$(0.191, 0.060, 0.013)$	0.198	0.066	0.027
$(0.377, 0.337)$	T	$(0.101, 0.151, 0.013)$	0.105	0.156	0.027

Table 5.2: Test results in the no excess sum-rate case for a binary source with $p_X(0) = 1/4$: (d_1, d_2, d_0) is a target distortion triple; (R_1, R_2) is the pair of rates used by the code; $\hat{d}_1$, $\hat{d}_2$, $\hat{d}_0$ are the empirical distortions.

Table 5.3. The first column contains the common value of the target side distortions $d_1 = d_2$. The second column contains the target value $UB(d_1, d_2)$ for the central distortion. The remaining three columns present the empirical distortions averaged over 100 runs.

$d_1 = d_2$	$UB(d_1, d_2)$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_0$
0.1	0.0626	0.1159	0.1136	0.0783
0.2	0.1574	0.2161	0.2170	0.1720
0.3	0.2658	0.3173	0.3134	0.2823

Table 5.3: Test results in the no excess marginal rate for the uniform binary source: $d_1 = d_2$ are target distortion values for the side descriptions; $UB(d_1, d_2)$ is the target value for the central distortion; $\hat{d}_1$, $\hat{d}_2$, $\hat{d}_0$ are the empirical distortions.

5.1.3 Successive Refinement

We have tested the proposed SR coding scheme for two binary and one ternary sources. The binary sources are 1) uniform binary source; 2) non-uniform binary source with $p_X(0) = 0.25$. In both cases we have used the following parameters: $\gamma_1 = \gamma_2 = 2.8$, and $\delta = 1.6$.

For the uniform ternary source the parameter values are: $\gamma_1 = \gamma_2 = 1.91 - 2.31$ and $\delta = 1.1$. The results of our experiments are presented in Figures 5.1, 5.2 and 5.3. The empirical distortions $\hat{d}_1$ and $\hat{d}_0$ correspond with R_1 and $R_1 + R_2$ respectively. In each figure the point $(\hat{d}_1, R_1)$ corresponds to the first stage coding results and the points $(\hat{d}_0, R_1 + R_2)$ correspond to the second stage coding results. As presented in Figure 5.1, 5.2 and 5.3 the empirical distortions are very close to the theoretical lower bounds.

5.2 Gaussian Source with Squared Error Distortion

We have tested the proposed successive quantization and quantization splitting scheme for an i.i.d. zero-mean unit-variance Gaussian source. We have considered input sequences of various lengths $n = 100, 1000$ and 10000. In all our tests we set $\eta = 0.9$, and $\omega = 4$.

Tables 5.4 and 5.5 present the simulation results of the LDGM-based successive quantization scheme for a 2-description symmetric and asymmetric distortion tuples, respectively. Tables 5.6 and 5.7 exhibit the results of the proposed scheme for an L-description symmetric distortion tuple with $L = 3$ and $L = 4$, respectively. In each

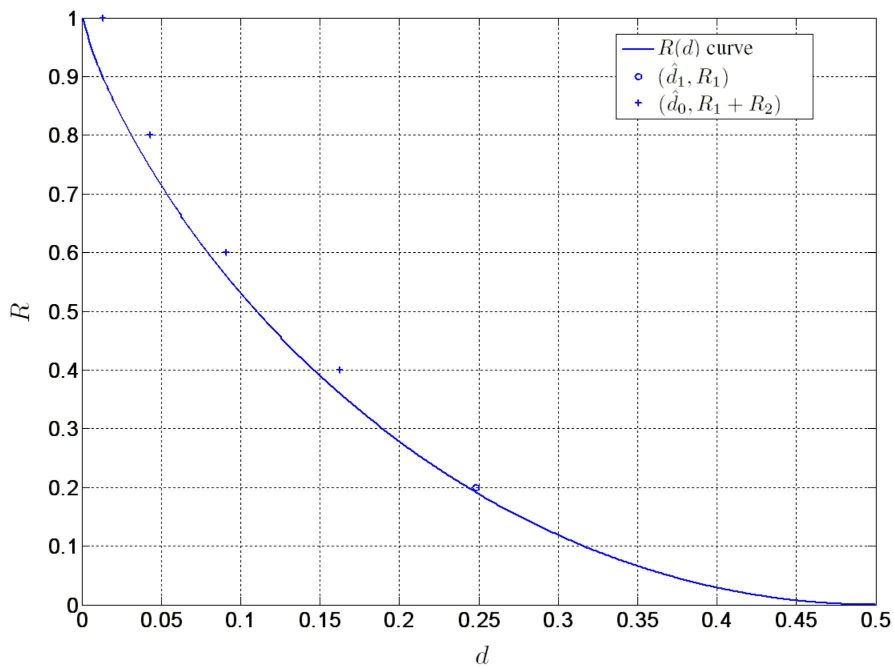

Figure 5.1: Simulation results for the uniform binary source at $R_1 = 0.2$ and various R_2 values.

table $(d_{\{1\}}, \cdots, d_{\{L\}}, d_{\{1,\cdots,L\}})$ denotes the target distortion tuple, while $\hat{d}_{\{1\}}, \cdots, \hat{d}_{\{L\}}$, and $\hat{d}_{\{1,\cdots,L\}}$ denote the empirical distortions; $(R_1, \cdots, R_L)$ denotes the rate pair used in the experiments, which corresponds to a vertex of the sum-rate optimal Gaussian EGC region, defined by (4.5); λ_i and δ_i are for the parameters λ and δ, respectively, used in the message passing algorithm at the encoding stage i, $i = 1, \cdots, L$. Next we list the values of the parameters for $L = 4$, which did not fit in Table 5.7. For $n = 10000$, we have $\lambda_1 = \cdots = \lambda_4 = 1.7$ and $\delta_1 = \cdots = \delta_4 = 1.9$, while for $n = 1000$ and $n = 100$, we have $\lambda_1 = \lambda_2 = 1.6$, $\lambda_3 = \lambda_4 = 1.7$, $\delta_1 = \delta_2 = 1.8$ and $\delta_3 = \delta_4 = 1.9$.

Table 5.8 presents the result obtained using the proposed quantization splitting

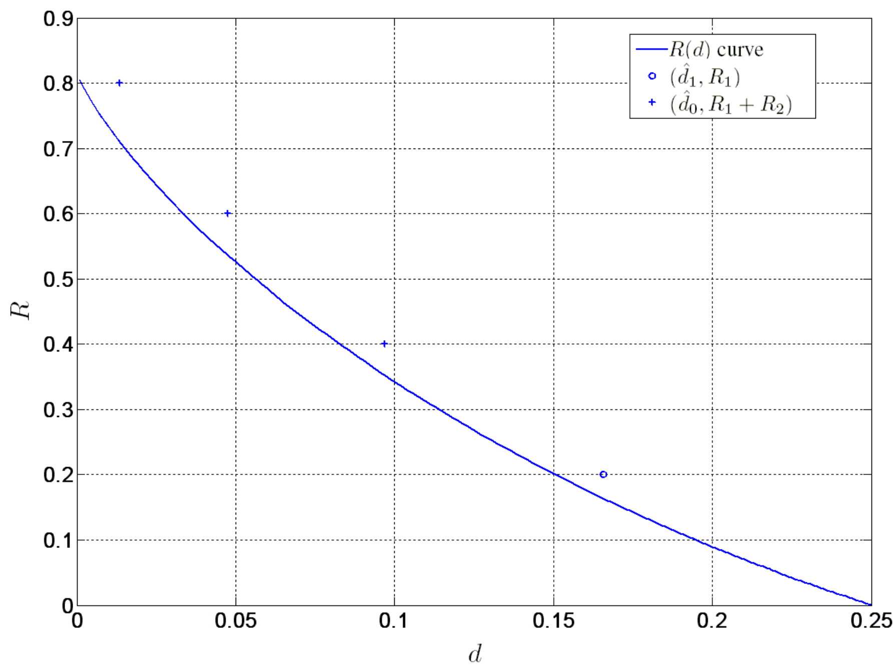

Figure 5.2: Simulation results for a non-uniform binary source with $p_X(0) = 0.25$ at $R_1 = 0.2$ and various R_2 values.

scheme for a 2-description problem with symmetric rates and symmetric distortions: $(d_{\{1\}}, d_{\{2\}}, d_{\{1,2\}})$ denotes the target distortion triple and $\hat{d}_{\{1\}}$, $\hat{d}_{\{2\}}$, $\hat{d}_{\{1,2\}}$ are the empirical distortions; $R_{1,1}, R_{1,2}, R_2$ represent the rates defined in (4.15), (4.16), and (4.14); $\lambda_{1,1}$, $\delta_{1,1}$ and $\lambda_{1,2}$, $\delta_{1,2}$ are the parameters of the LDGM codes used for the first description, in other words for encoder of stage 1 and encoder of stage 3, respectively; λ_2 and δ_2 are the parameters of the LDGM code used for the second description, i.e., for encoder of stage 2.

From the results showed in the tables, we can observe that the empirical distortions are very close to the theoretical distortion bounds.

n	(R_1, R_2)	$(d_{\{1\}}, d_{\{2\}}, d_{\{1,2\}})$	$\hat{d}_{\{1\}}$	$\hat{d}_{\{2\}}$	$\hat{d}_{\{1,2\}}$	λ_1	λ_2	δ_1	δ_2
10000	$(1, 1.015)$	$(0.25, 0.25, 0.125)$	0.267	0.262	0.135	1.7	1.7	1.9	1.9
1000	$(1, 1.015)$	$(0.25, 0.25, 0.125)$	0.267	0.264	0.136	1.6	1.6	1.8	1.8
100	$(1, 1.015)$	$(0.25, 0.25, 0.125)$	0.270	0.272	0.148	1.6	1.6	1.8	1.8

Table 5.4: Parameters and results for a 2-description case with symmetric distortions.

n	(R_1, R_2)	$(d_{\{1\}}, d_{\{2\}}, d_{\{1,2\}})$	$\hat{d}_{\{1\}}$	$\hat{d}_{\{2\}}$	$\hat{d}_{\{1,2\}}$	λ_1	λ_2	δ_1	δ_2
10000	$(1.161, 0.914)$	$(0.2, 0.25, 0.1)$	0.211	0.266	0.111	2.6	1.7	2.8	1.9
10000	$(1, 1.075)$	$(0.25, 0.2, 0.1)$	0.262	0.219	0.111	1.7	2.5	1.9	2.7
1000	$(1.161, 0.914)$	$(0.2, 0.25, 0.1)$	0.212	0.269	0.113	2.8	1.7	2.8	1.9
1000	$(1, 1.075)$	$(0.25, 0.2, 0.1)$	0.265	0.220	0.114	1.8	2.6	1.9	2.8
100	$(1.161, 0.914)$	$(0.2, 0.25, 0.1)$	0.219	0.277	0.121	2.8	1.7	2.8	1.8
100	$(1, 1.075)$	$(0.25, 0.2, 0.1)$	0.271	0.225	0.122	1.8	2.6	1.9	2.8

Table 5.5: Parameters and results for a 2-description case with asymmetric distortions.

n	(R_1, R_2, R_3)	$(d_{\{1\}}, d_{\{2\}}, d_{\{3\}}, d_{\{1,2,3\}})$	$\hat{d}_{\{1\}}$	$\hat{d}_{\{2\}}$	$\hat{d}_{\{3\}}$	$\hat{d}_{\{1,2,3\}}$	λ_1	λ_2	λ_3	δ_1	δ_2	δ_3
10000	$(1.161, 1.165, 1.169)$	$(0.2, 0.2, 0.2, 0.067)$	0.210	0.210	0.212	0.076	1.7	1.7	1.7	1.9	1.9	1.9
1000	$(1.161, 1.165, 1.169)$	$(0.2, 0.2, 0.2, 0.067)$	0.212	0.214	0.217	0.079	1.6	1.6	1.7	1.8	1.9	1.9
100	$(1.161, 1.165, 1.169)$	$(0.2, 0.2, 0.2, 0.067)$	0.219	0.222	0.226	0.088	1.6	1.6	1.7	1.8	1.9	1.9

Table 5.6: Parameters and results for a 3-description case with symmetric distortions.

n	(R_1, R_2, R_3)	$(d_{\{1\}}, d_{\{2\}}, d_{\{3\}}, d_{\{4\}}, d_{\{1,2,3,4\}})$	$\hat{d}_{\{1\}}$	$\hat{d}_{\{2\}}$	$\hat{d}_{\{3\}}$	$\hat{d}_{\{4\}}$	$\hat{d}_{\{1,2,3,4\}}$
10000	$(1.161, 1.163, 1.165, 1.168)$	$(0.2, 0.2, 0.2, 0.2, 0.05)$	0.209	0.209	0.211	0.213	0.059
1000	$(1.161, 1.163, 1.165, 1.168)$	$(0.2, 0.2, 0.2, 0.2, 0.05)$	0.210	0.211	0.213	0.214	0.061
100	$(1.161, 1.163, 1.165, 1.168)$	$(0.2, 0.2, 0.2, 0.2, 0.05)$	0.218	0.219	0.222	0.227	0.075

Table 5.7: Results for a 4-description case with symmetric distortions.

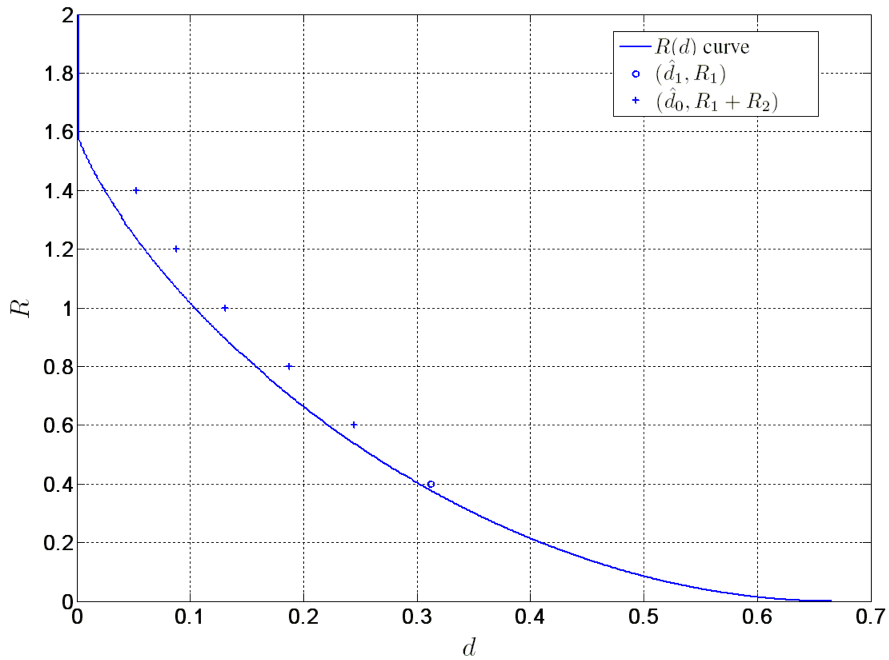

Figure 5.3: Simulation results for the uniform ternary source at $R_1 = 0.4$ and various R_2 values.

n	$(R_{1,1}, R_{1,2}, R_2)$	$(d_{\{1\}}, d_{\{2\}}, d_{\{1,2\}})$	$\hat{d}_{\{1\}}$	$\hat{d}_{\{2\}}$	$\hat{d}_{\{1,2\}}$	$\lambda_{1,1}$	$\lambda_{1,2}$	λ_2	$\delta_{1,1}$	$\delta_{1,2}$	δ_2
10000	$(0.661, 0.346, 1.007)$	$(0.25, 0.25, 0.125)$	0.268	0.269	0.138	1.0	0.3	1.5	1.2	0.6	1.7
1000	$(0.661, 0.346, 1.007)$	$(0.25, 0.25, 0.125)$	0.272	0.271	0.141	1.0	0.3	1.5	1.2	0.6	1.7
100	$(0.661, 0.346, 1.007)$	$(0.25, 0.25, 0.125)$	0.278	0.276	0.147	1.0	0.3	1.5	1.2	0.6	1.7

Table 5.8: Parameters and results for a 2-description case with symmetric rates and symmetric distortions, using quantization splitting.

Chapter 6

Conclusions

In this thesis, we present two practical coding schemes based on low density generator matrix (LDGM) codes for two cases of the multiple description problem. The first one is for the case of two descriptions, for finite-alphabet sources with Hamming distortion measure. The scheme is devised for corner points of a rate region corresponding to Zhang-Berger rate-distortion region, and is sequential in nature. The proposed practical code consists in replacing the random codebooks in the theoretical sequential scheme by multilevel LDGM codebooks, in conjunction with a message passing algorithm to perform the encoding at each stage. In order to derive the latter algorithm, an unconstrained optimization problem formulation with undetermined coefficients, is proposed instead of the strong typicality condition. Details of the implementation of the scheme in several cases of the MD problem are discussed, specifically, the no excess sum-rate case for binary sources, the case of successive refinement, and the no excess marginal case for the uniform binary source. For all the aforementioned cases, extensive tests were performed, which verify the effectiveness of the code. One of the remaining open problems, which is currently under investigation, is to determine the coefficients involved in the encoder optimization problem, which is proposed as an alternative formulation to the strong typicality encoding, and to establish the equivalence between the two approaches. Moreover, for future work, it is of interest to address the optimization of the involved LDGM codes using, for instance, techniques similar to the density evolution developed for LDPC codes.

The second proposed scheme is for the quadratic Gaussian source, in the L-description case with individual and central distortion constraints. It is shown first that the coding problem for an arbitrary point on the dominant face of the El Gamal-Cover (EGC) rate region, can be converted to that for a vertex of a K-description EGC rate region for some $K \leq 2L-1$. The latter problem can be solved via successive coding, and in the quadratic Gaussian case, each successive coding step can be reduced to a Gaussian quantization operation via Gram-Schmidt orthogonalization. Finally, each quantization step is implemented using a multilevel LDGM code with a message passing algorithm. Our tests show very good performance of the scheme in practice. The approach used in our experiments for selecting the finite output alphabet needed in the multilevel LDGM code construction, is inspired from the central limit theorem. An interesting problem which remains open is to design the optimal output alphabet and the corresponding output distribution, given some fixed parameter of the code which limits the message passing algorithm complexity.

Appendix A

Proof of Proposition 3.1

In this appendix we present the proof of Proposition 1.

Proof of Proposition 1. Since $\mathcal{D}(d_0)$ is symmetric, it suffices to solve (3.3) for $\alpha \geq 1$. Note that

$$\begin{aligned}
&\mathbb{E}[d(X, X_t)] \\
&= \mathbb{E}[d(X_0 \oplus Z, X_t)] \\
&= \mathbb{E}[\mathbb{E}[d(X_0 \oplus Z, X_t)|Z]] \\
&= (1 - d_0)\mathbb{E}[d(X_0, X_t)] + d_0(1 - \mathbb{E}[d(X_0, X_t)]) \\
&= (1 - 2d_0)\mathbb{E}[d(X_0, X_t)] + d_0, \quad t = 1, 2.
\end{aligned} \tag{A.1}$$

By plugging (A.1) into (3.3), dividing the cost function by $(1-2d_0)$, which is positive because $d_0 < 1/2$, and eliminating the additive constant, (3.3) becomes equivalent to

$$\min_{p_{X_1X_2|X_0}:I(X_1;X_2)=0} \alpha\mathbb{E}[d(X_0, X_1)] + \mathbb{E}[d(X_0, X_2)]. \tag{A.2}$$

Note that the above optimization problem has a linear cost function, but the constraint $I(X_1; X_2) = 0$ is non-linear. On the other hand, if the marginal distribution of X_1 is fixed, the constraint $I(X_1; X_2) = 0$ becomes linear. Therefore, our approach to solve the problem is to fix the marginal distribution of X_1, in other words, let $p_{X_1}(0) = \epsilon$, for some $\epsilon, 0 < \epsilon < 1$, and solve the parameterized problem $P(\epsilon)$; then optimize over all possible values of ϵ. In order to proceed let us adopt the following notation

$$\begin{aligned} y_1 &= p_{X_0X_1X_2}(0,1,1), \quad y_2 = p_{X_0X_1X_2}(1,0,0), \\ y_3 &= p_{X_0X_1X_2}(1,1,0), \quad y_4 = p_{X_0X_1X_2}(1,0,1), \\ y_5 &= p_{X_0X_1X_2}(0,1,0), \quad y_6 = p_{X_0X_1X_2}(0,0,1), \\ y_7 &= p_{X_0X_1X_2}(0,0,0), \quad y_8 = p_{X_0X_1X_2}(1,1,1). \end{aligned} \tag{A.3}$$

Then

$$\begin{aligned} d_1^* &\triangleq \mathbb{E}[d(X_0, X_1)] = y_1 + y_2 + y_4 + y_5, \\ d_2^* &\triangleq \mathbb{E}[d(X_0, X_2)] = y_1 + y_2 + y_3 + y_6. \end{aligned} \tag{A.4}$$

Let $\mathbf{y} = (y_1, \cdots, y_8)$. Note that

$$\delta^* \triangleq p_{X_0}(0) = \frac{\delta - d_0}{1 - 2d_0}.$$

Clearly, we have $\delta^* \leq 1/2$. Thus, problem (A.2) is equivalent to

$$\begin{aligned}\min_{0<\epsilon<1}\min_{\mathbf{y}} C(\epsilon,\mathbf{y}) &= (\alpha+1)y_1 + (\alpha+1)y_2 + \\ &\quad +y_3 + \alpha y_4 + \alpha y_5 + y_6 \end{aligned} \tag{A.5}$$

subject to the following constraints

$$\begin{aligned} &\sum_{i=1}^{8} y_i = 1, \\ &y_1 + y_5 + y_6 + y_7 = \delta^*, \\ &y_2 + y_4 + y_6 + y_7 = \epsilon, \\ &y_2 + y_7 = \frac{\epsilon}{1-\epsilon}(y_3 + y_5), \\ &y_i \geq 0, \quad 1 \leq i \leq 8. \end{aligned} \tag{A.6}$$

The second constraint in (A.6) is due to the fact that $p_{X_0}(0) = \delta^*$, while the third follows from $p_{X_1}(0) = \epsilon$. The fourth is implied by the independence of X_1 and X_2. Precisely, it is derived from the relation $p_{X_1}(0)p_{X_2}(0) = p_{X_1X_2}(0,0)$. It is easy to see that the equalities in (A.6) are linearly independent equations. Let $P(\epsilon)$ denote the linear optimization problem of minimizing $C(\epsilon,\mathbf{y})$ over all vector variables $\mathbf{y}$ satisfying constraints (A.6).

First we solve problem $P(\epsilon)$ for the case when $\epsilon \leq 1/2$. In this case we eliminate the variables y_5, y_6, y_7, y_8 from the four equations in (A.6) and obtain

$$\begin{aligned}
y_5 &= -y_1 + y_2 + y_4 + \delta^* - \epsilon, \\
y_6 &= \frac{\epsilon}{1-\epsilon} y_1 - \frac{\epsilon}{1-\epsilon} y_2 - \frac{\epsilon}{1-\epsilon} y_3 - \frac{1}{1-\epsilon} y_4 + \frac{\epsilon(1-\delta^*)}{1-\epsilon}, \\
y_7 &= -\frac{\epsilon}{1-\epsilon} y_1 + \left(\frac{\epsilon}{1-\epsilon} - 1\right) y_2 + \frac{\epsilon}{1-\epsilon} y_3 + \frac{\epsilon}{1-\epsilon} y_4 \\
&\quad + \frac{\epsilon(\delta^* - \epsilon)}{1-\epsilon}, \\
y_8 &= -y_2 - y_3 - y_4 + 1 - \delta^*.
\end{aligned} \tag{A.7}$$

Plugging into the cost function, we obtain

$$\begin{aligned}
C(\epsilon, \mathbf{y}) = \alpha(\delta^* - \epsilon) + \frac{\epsilon(1-\delta^*)}{1-\epsilon} \\
+ \left(1 + \frac{\epsilon}{1-\epsilon}\right) y_1 + \left(2\alpha + 1 - \frac{\epsilon}{1-\epsilon}\right) y_2 \\
+ \left(1 - \frac{\epsilon}{1-\epsilon}\right) y_3 + \left(2\alpha - \frac{1}{1-\epsilon}\right) y_4.
\end{aligned} \tag{A.8}$$

Denote $g_1(\epsilon) = \alpha(\delta^* - \epsilon) + \frac{\epsilon(1-\delta^*)}{1-\epsilon}$. It can be easily verified that the coefficients of all variables in (A.8) are non-negative when $\epsilon \leq 1/2$, and they are strictly positive when $\epsilon < 1/2$ (recall that $\alpha \geq 1$). Thus, by setting all variables y_1, y_2, y_3, y_4 to 0 we obtain a lower bound on the cost function as $C(\epsilon, \mathbf{y}) \geq g_1(\epsilon)$. This lower bound can be achieved by a feasible solution $\mathbf{y}$ if the values of y_5, y_6, y_7, y_8 obtained by replacing y_1, y_2, y_3, y_4 by 0 in equalities (A.7) are non-negative. Obviously, the latter condition holds when $\epsilon \leq \delta^*$. Denote by $C^*(\epsilon)$ the value of the cost function of problem $P(\epsilon)$ at optimality. We formulate the conclusion of the above discussion in the following assertion.

Assertion1. For $\alpha \geq 1$ and $0 < \epsilon \leq 1/2$, we have

$$C^*(\epsilon) \geq g_1(\epsilon).$$

Moreover, when $0 < \epsilon \leq \delta^*$ the above relation holds with equality and a solution of $P(\epsilon)$ is $\mathbf{y}(\epsilon)$ defined by

$$\begin{aligned}
&y_1(\epsilon) = y_2(\epsilon) = y_3(\epsilon) = y_4(\epsilon) = 0,\\
&y_5(\epsilon) = \delta^* - \epsilon, \quad y_6(\epsilon) = \frac{\epsilon(1-\delta^*)}{1-\epsilon},\\
&y_7(\epsilon) = \frac{\epsilon(\delta^* - \epsilon)}{1-\epsilon}, \quad y_8(\epsilon) = 1 - \delta^*.
\end{aligned} \tag{A.9}$$

Furthermore, when $0 < \epsilon \leq \delta^*$ and $\epsilon \neq 1/2$, the above solution of $P(\epsilon)$ is unique.

We proceed by computing the minimum value of $g_1(\epsilon)$. The derivative of $g_1(\epsilon)$ with respect to ϵ is $g_1'(\epsilon) = -\alpha + \frac{1-\delta^*}{(1-\epsilon)^2}$. Equation $g_1'(\epsilon) = 0$ has a unique solution in the interval $(0, 1)$, namely

$$\epsilon_1 = 1 - \sqrt{(1-\delta^*)/\alpha}.$$

Clearly, when $\alpha \leq 1/(1-\delta^*)$, we have $\epsilon_1 \leq \delta^*$. Moreover, the function $g_1'(\cdot)$ has a negative sign to the left of ϵ_1, and a positive sign to the right of ϵ_1. The aforementioned discussion validates the following assertion.

Assertion 2. For $1 \leq \alpha \leq 1/(1-\delta^*)$,

$$\min_{0<\epsilon\leq 1/2} g_1(\epsilon) = g_1(\epsilon_1) = 2\sqrt{\alpha(1-\delta^*)} - (\alpha+1)(1-\delta^*),$$

and ϵ_1 is the unique point of minimum.

Next we will analyze the case when $\epsilon > 1/2$. By eliminating the variables y_3, y_4, y_7, y_8 from the four equalities in (A.6), and replacing in the cost function, we obtain

$$\begin{aligned} C(\epsilon, \mathbf{y}) = {} & \alpha(\epsilon - \delta^*) + \frac{\delta^*(1-\epsilon)}{\epsilon} \\ & + \left(2\alpha + 1 - \frac{1-\epsilon}{\epsilon}\right) y_1 + \left(1 + \frac{1-\epsilon}{\epsilon}\right) y_2 \\ & + \left(2\alpha - \frac{1}{\epsilon}\right) y_5 + \left(1 - \frac{1-\epsilon}{\epsilon}\right) y_6. \end{aligned} \tag{A.10}$$

Denote $g_2(\epsilon) = \alpha(\epsilon - \delta^*) + \frac{\delta^*(1-\epsilon)}{\epsilon}$. Because $1/2 < \epsilon < 1$ and $\alpha \geq 1$, the coefficients of variables in (A.10) are non-negative. Thus we obtain a lower bound on the cost function by setting y_1, y_2, y_5, y_6 to 0. Precisely, the following holds

$$C^*(\epsilon) \geq g_2(\epsilon), \quad 1/2 < \epsilon < 1, \alpha \geq 1. \tag{A.11}$$

The derivative of $g_2(\epsilon)$ is $g_2'(\epsilon) = \alpha - \frac{\delta^*}{\epsilon^2}$, and it has a unique zero point in the interval $(0, 1)$, namely $\epsilon_2 = \sqrt{\delta^*/\alpha}$. It can be easily verified that $g_2'(\epsilon)$ is negative to the left of ϵ_2, and it is positive to the right of ϵ_2. Consequently, ϵ_2 is a point of minimum for $g_2(\epsilon)$ on the interval $(0, 1)$. Thus, we conclude that

$$\min_{1/2<\epsilon<1} g_2(\epsilon) \geq g_2(\epsilon_2) = 2\sqrt{\alpha\delta^*} - (\alpha + 1)\delta^*. \tag{A.12}$$

Finally, in order to reach the solution of problem (A.5), we need one more result, which is stated next.

Assertion 3. The following relation holds

$$g_1(\epsilon_1) \leq g_2(\epsilon_2), \tag{A.13}$$

with strict inequality when $\delta < 1/2$, and with equality when $\delta = 1/2$.

Proof of Assertion 3. Inequality (A.13) is equivalent to

$$2\sqrt{\alpha(1-\delta^*)} - (\alpha+1)(1-\delta^*) \leq 2\sqrt{\alpha\delta^*} - (\alpha+1)\delta^*,$$

which is further equivalent to

$$2\sqrt{\alpha}(\sqrt{1-\delta^*} - \sqrt{\delta^*}) \leq (\alpha+1)((1-\delta^*) - \delta^*). \tag{A.14}$$

When $\delta = 1/2$, it follows that $\delta^* = 1/2$ and relation (A.14) holds with equality. When $\delta^* < 1/2$, one has $\sqrt{1-\delta^*} - \sqrt{\delta^*} > 0$ and by multiplying both sides with $1/(\sqrt{1-\delta^*} - \sqrt{\delta^*})$, (A.14) reduces to

$$2\sqrt{\alpha} \leq (\alpha+1)(\sqrt{1-\delta^*} + \sqrt{\delta^*}),$$

which holds with strict inequality due to $0 < 2\sqrt{\alpha} \leq \alpha + 1$ and $1 < \sqrt{1-\delta^*} + \sqrt{\delta^*}$. □

Further, by combining Assertions 1, 2, 3, (A.11) and (A.12), we conclude that when $1 \leq \alpha \leq 1/(1-\delta^*)$, a solution to problem (A.5) is $(\epsilon_1, \mathbf{y}(\epsilon_1))$. Moreover, when $\delta < 1/2$ this solution is unique. This solution yields the following values for d_1, d_2

(via (A.4) and (A.1))

$$d_1 = d_0 + (1 - 2d_0)(\sqrt{1-\delta^*}(1/\sqrt{\alpha} - \sqrt{1-\delta^*})), \tag{A.15}$$
$$d_2 = d_0 + (1 - 2d_0)(\sqrt{1-\delta^*}(\sqrt{\alpha} - \sqrt{1-\delta^*})).$$

Eliminating α from the above relations we obtain

$$(d_1 + 1 - 2d_0 - \delta)(d_2 + 1 - 2d_0 - \delta) =$$
$$(1 - d_0 - \delta)(1 - 2d_0). \tag{A.16}$$

Furthermore, for any pair (d_1, d_2) with $d_0 \leq d_1 \leq d_2$, satisfying (A.16), the value of α satisfying equations (A.15) is

$$\alpha = \frac{(1 - d_0 - \delta)(1 - 2d_0)}{(d_1 + 1 - 2d_0 - \delta)^2}, \tag{A.17}$$

and clearly we have $1 \leq \alpha \leq 1/(1-\delta^*)$. Thus we have shown that the set of solutions to problem (3.1) for all $1 \leq \alpha \leq 1/(1-\delta^*)$, equals the portion of hyperbola (A.16) corresponding to $d_0 \leq d_1 \leq d_2$. The solution pair (d_1, d_2) corresponding to $\alpha = 1/(1-\delta^*)$ is $(d_1, d_2) = (d_0, \delta)$. Since the value of d_1 in the solution to (3.1) is non-increasing as α increases, but on the other hand, d_1 cannot go below d_0, it follows that for any $\alpha > 1/(1-\delta^*)$ the unique solution to problem (3.1) is $(d_1, d_2) = (d_0, \delta)$. Using further the fact that $\mathcal{D}(d_0)$ is symmetric the first conclusion of the proposition follows.

To complete the proof note that, for any pair (d_1, d_2), with $d_1 \leq d_2$, which satisfies equation (A.16), by using (A.17) and computing the solution vector $\mathbf{y}(\epsilon_1)$ via (A.9), we find the joint pdf $p_{X_0X_1X_2}$ specified by

$$\begin{aligned}
p_{X_0X_1X_2}(0,1,1) &= p_{X_0X_1X_2}(1,0,0) = 0, \\
p_{X_0X_1X_2}(1,1,0) &= p_{X_0X_1X_2}(1,0,1) = 0, \\
p_{X_0X_1X_2}(0,1,0) &= \frac{d_1 - d_0}{1 - 2d_0}, \\
p_{X_0X_1X_2}(0,0,1) &= \frac{d_2 - d_0}{1 - 2d_0}, \\
p_{X_0X_1X_2}(0,0,0) &= \frac{\delta + d_0 - d_1 - d_2}{1 - 2d_0}, \\
p_{X_0X_1X_2}(1,1,1) &= \frac{1 - \delta - d_0}{1 - 2d_0}.
\end{aligned}$$

Finally, by computing $p_{XX_1X_2}$ according to

$$p_{XX_1X_2}(x, x_1, x_2) = \sum_{x_0=0,1} p_{X_0X_1X_2}(x_0, x_1, x_2) p_{X|X_0}(x|x_0),$$

which holds because $X - X_0 - (X_1, X_2)$ forms a Markov chain, equalities (3.4) follow. Further, note that relations (3.4) are symmetric in (X_1, d_1) and (X_2, d_2). We conclude that they also hold for $d_1 > d_2$.

We have already established the uniqueness of the conditional probability $p_{X_0X_1X_2|X}$ when $\delta < 1/2$. Let us analyze now the case $\delta = 1/2$. In this case one has $\delta^* = 1/2$ and $1 \leq \alpha \leq 2$. Then $\epsilon_2 = \sqrt{\frac{1}{2\alpha}} = 1\epsilon_1 \geq 1/2$. By Assertion 3, the following equality is valid

$$g_2(\epsilon_2) = \min_{\epsilon: 0 < \epsilon < 1} C^*(\epsilon.)$$

Note that when $\epsilon = \epsilon_2$ and $\alpha < 2$ the coefficients of y_1, y_2, y_5, y_6 in (A.10) are positive for $\epsilon = \epsilon_2$. Therefore, in order to achieve the minimum value $g_2(\epsilon_2)$ in equation (A.10) the variables y_1, y_2, y_5, y_6 must be set to 0. By substituting ϵ by ϵ_2 and y_1, y_2, y_5, y_6 by 0 in (A.6), one obtains another solution to problem (A.5) as $(\epsilon_2, \mathbf{y}')$, where $\mathbf{y}' = (y'_1, \cdots, y'_8)$ defined as

$$\begin{aligned}
&y'_1 = y_2(\epsilon_1), y'_2 = y_1(\epsilon_1), y'_3 = y_6(\epsilon_1), y'_4 = y_5(\epsilon_1),\\
&y'_5 = y_4(\epsilon_1), y'_6 = y_3(\epsilon_1), y'_7 = y_8(\epsilon_1), y'_8 = y_7(\epsilon_1).
\end{aligned}$$

Moreover, we conclude that $(\epsilon_1, \mathbf{y}(\epsilon_1)$ and $(\epsilon_2, \mathbf{y}')$ are the only solutions to problem (A.5). Further, in order to distinguish between the pairs of variables (X_1, X_2) corresponding to the two different solutions, let us denote by (X'_1, X'_2) the pair derived from $(\epsilon_2, \mathbf{y}')$. It can be verified that

$$p_{X_0X'_1X'_2}(x_0, x_1, x_2) = p_{X_0X_1X_2}(x_0 \oplus 1, x_1 \oplus 1, x_2 \oplus 1).$$

Using further the fact that $X - X_0 - (X_1, X_2)$ forms a Markov chain we obtain

$$\begin{aligned}
&p_{XX'_1X'_2}(x_0, x_1, x_2)\\
&= \sum_{x_0=0,1} p_{X_0X'_1X'_2}(x_0, x_1, x_2) p_{X|X_0}(x|x_0)\\
&= \sum_{x_0=0,1} p_{X_0X_1X_2}(x_0 \oplus 1, x_1 \oplus 1, x_2 \oplus 1) p_{X|X_0}(x \oplus 1|x_0 \oplus 1)\\
&= p_{XX_1X_2}(x \oplus 1, x_1 \oplus 1, x_2 \oplus 1).
\end{aligned}$$

With this observation, the proof is complete. $\square$

Appendix B

Proof of Upper Bound Claims of No Excess Marginal Rate Case

In this appendix we prove the claims formulated in Section 3.2.

Clearly, the mapping $\psi_0(\cdot)$ which minimizes the central distortion must satisfy the equality

$$\psi_0(x_c, x_1, x_2) = \arg\max_{x=0,1} p_{XX_cX_1X_2}(x, x_c, x_1, x_2). \tag{B.18}$$

Using the fact that $X - X_1 - X_c$, $X - X_2 - X_c$, and $X_1 - (X, X_c) - X_2$ form Markov chains, we obtain

$$\begin{aligned}
p_{XX_cX_1X_2}(x, x_c, x_1, x_2) &= p_{XX_c}(x, x_c) p_{X_1X_2|XX_c}(x_1, x_2|x, x_c) \\
&= p_{XX_c}(x, x_c) p_{X_1|XX_c}(x_1|x, x_c) p_{X_2|XX_c}(x_2|x, x_c) \\
&= \frac{p_{XX_cX_1}(x, x_c, x_1) p_{XX_cX_2}(x, x_c, x_2)}{p_{XX_c}(x, x_c)} \\
&= \frac{p_{X_1}(x_1) p_{X|X_1}(x|x_1) p_{X_c|X_1}(x_c|x_1) p_{X_2}(x_2) p_{X|X_2}(x|x_2) p_{X_c|X_2}(x_c|x_2)}{p_{X_c}(x_c) p_{X|X_c}(x|x_c)}.
\end{aligned}$$

Thus, relation (B.18) is equivalent to

$$\psi_0(x_c, x_1, x_2) = \arg\max_{x=0,1} \frac{p_{X|X_1}(x|x_1)p_{X|X_2}(x|x_2)}{p_{X|X_c}(x|x_c)}.$$

In order to solve this maximization we will consider different cases for (x_1, x_2, x_c). Let us first define

$$h(b) \triangleq \frac{p_{X|X_1}(b|x_1)p_{X|X_2}(b|x_2)}{p_{X|X_c}(b|x_c)}.$$

Case 1. $x_1 = x_2 = x_c = b$. Then the following relations hold

$$\frac{h(b)}{h(1-b)} = \frac{1-d_1}{d_1}\frac{1-d_2}{d_2}\frac{s}{1-s}.$$

Because the function $\frac{1-x}{x}$ is decreasing for $x > 0$, and $d_2 \leq s$, it follows that $\frac{1-d_2}{d_2}\frac{s}{1-s} \geq 1$. The fact that $d_1 \leq 1/2$ implies $\frac{1-d_1}{d_1} \geq 1$. Combining them we obtain $\frac{h(b)}{h(1-b)} \geq 1$.

Case 2. $x_1 = x_2 = b, x_c = 1-b$. Then we have

$$\frac{h(b)}{h(1-b)} = \frac{1-d_1}{d_1}\frac{1-d_2}{d_2}\frac{1-s}{s} \geq 1,$$

due to $\frac{1-x}{x} \geq 1$ for $0 < x \leq 1/2$.

Case 3. $x_1 = b, x_2 = x_c = 1-b$. In this case, the following is valid

$$\frac{h(b)}{h(1-b)} = \frac{1-d_1}{d_1}\frac{d_2}{1-d_2}\frac{1-s}{s}.$$

Relation $d_1 \leq d_2$ implies that $\frac{1-d_1}{d_1}\frac{d_2}{1-d_2} \geq 1$. Using further the inequality $\frac{1-s}{s} \geq 1$, we obtain $\frac{h(b)}{h(1-b)} \geq 1$.

Case 4. $x_1 = x_c = b, x_2 = 1 - b$. Then

$$\frac{h(1-b)}{h(b)} = \frac{d_1}{1-d_1}\frac{1-d_2}{d_2}\frac{1-s}{s}.$$

We have $\frac{h(1-b)}{h(b)} \geq 1$ if and only if $s \leq \alpha(d_1, d_2) \triangleq \frac{d_1(1-d_2)}{d_1(1-d_2)+d_2(1-d_1)}$. Note that $\alpha(d_1, d_2) \leq 1/2$. Moreover, $d_2 \leq \alpha(d_1, d_2)$, is equivalent to $\frac{1-d_1}{d_1} \leq \frac{(1-d_2)^2}{d_2^2}$, which is true. Consequently, the interval $[d_2, \alpha(d_1, d_2)]$ is nonempty. Thus, the proof of relation (3.7) is complete.

Now let us analyze the central distortion when $s \leq \alpha(d_1, d_2)$. We have

$$\begin{aligned}&\mathbb{P}(X \neq \psi_0(X_c, X_1, X_2)) = \mathbb{P}(X_c = X) - \mathbb{P}(X = X_c = X_1 = X_2) + \mathbb{P}(X \neq X_c, X_c = X_1 = X_2)\\ &= 1 - s - \frac{(1-d_1)(1-d_2)(1-s-d_1)(1-s-d_2)}{(1-s)(1-2d_1)(1-2d_2)} + \frac{d_1 d_2(1-s-d_1)(1-s-d_2)}{s(1-2d_1)(1-2d_2)}.\end{aligned}$$

Let $f(s)$ denote the last expression. By computing the first and second order derivatives with respect to s we obtain

$$\begin{aligned}f'(s) &= \frac{1}{(1-2d_1)(1-2d_2)}[d_1 + d_2 - 2d_1d_2 - d_1d_2(1-d_1)(1-d_2)(\frac{1}{s^2} + \frac{1}{(1-s)^2})],\\ f"(s) &= \frac{2d_1d_2(1-d_1)(1-d_2)}{(1-2d_1)(1-2d_2)}(\frac{1}{s^3} - \frac{1}{(1-s)^3}).\end{aligned}$$

Since $f"(s) \geq 0$ for $0 < s \leq 1/2$, it follows that $f(s)$ is convex. It can be verified that $f(d_2) = f(\alpha(d_1, d_2))$. Consequently, the function $f(\cdot)$ has a point of minimum inside the interval $[d_2, \alpha(d_1, d_2)]$. Since $f(\cdot)$ is strictly convex over $(d_2, \alpha(d_1, d_2))$, it follows that it has a unique point of minimum s_0, which satisfies $f'(s_0) = 0$.

Equation $f'(s) = 0$ is equivalent to

$$\beta(d_1, d_2)s^2(1-s)^2 + 2s(1-s) - 1 = 0,$$

which can be easily solved since it is quadratic in $s(1-s)$. This equation has only two real solutions, of which only solution (3.9) is smaller than 1/2. With this observation, the proof is complete.

Bibliography

[1] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, pp. 74-93, Sep. 2001.

[2] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, pp. 821-834, May 1993.

[3] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple-description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 40, pp. 245-250, Jan. 1994.

[4] V. A. Vaishampayan and J.-C. Batllo, "Asymptotic analysis of multiple description quantizers," *IEEE Trans. Inf. Theory*, vol. 44, pp. 278-284, Jan. 1998.

[5] C. Tian and S. S. Hemami, "Universal multiple description scalar quantization: Analysis and design," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2089-2102, Sep. 2004.

[6] A. Gyorgy, T. Linder, and G. Lugosi, "Tracking the best quantizer," *IEEE Trans. Inf. Theory*, vol. 54,no.4, pp. 1604-1625, April.2008.

[7] S. Dumitrescu and X. Wu, "Optimal two-description scalar quantizer design," *Algorithmica*, vol. 41, no. 4, pp. 269-287, Feb. 2005.

[8] S. Dumitrescu and X. Wu, "Lagrangian optimization of two-description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3990-4012, Nov. 2007.

[9] D. Muresan and M. Effros, "Quantization as histogram segmentation: optimal scalar quantizer design in network systems," *IEEE Trans. Inf. Theory*, vol. 54, pp. 344-366, Jan. 2008.

[10] H. Jafarkhani and V. Tarokh, "Multiple description trellis-coded quantization," *IEEE Trans. Commun.*, vol. 47, pp. 799-803, Jun. 1999.

[11] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple-description vector quantization with lattice codebooks: Design and analysis," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1718-1734, Jul. 2001.

[12] S. N. Diggavi, N. J. A. Sloane, and V. A. Vaishampayan, "Asymmetric multiple description lattice vector quantizers", *IEEE Trans. Inf. Theory*, vol. 48, pp. 174-191, Jan. 2002.

[13] Y. Frank-Dayan and R. Zamir, "Dithered lattice-based quantizers for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 192-204, Jan. 2002.

[14] X. Huang and X. Wu, "Optimal index assignment for multiple description lattice vector quantization," *Proc. of DCC 2006*, pp. 272-281, 2006.

[15] J. Østergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple-description lattice vector quantization," in *IEEE Trans. Inf. Theory*, pp. 1956-1973, May 2006.

[16] S. Moradi, T. Linder, and S. Gazor, "Multiple description coding by successive quantization," *Canadian Journal of Electrical and Computer Engineering* , vol. 33, no. 3/4, pp. 133-138, Dec. 2008.

[17] J. Østergaard and R. Zamir, "Multiple descriptions by dithered Delta-Sigma quantization," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4661-4675, Oct. 2009.

[18] Y. Wang, M. T. Orchard, V. A. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. on Image Processing*, vol. 10, pp. 351-366, Mar. 2001.

[19] V. K. Goyal and J. Kovačević, "Generalized multiple description coding with correlating transforms," *IEEE Trans. on Inf. Theory*, vol. 47, pp. 2199-2224, Sep. 2001.

[20] A. Albanese, J. Blomer, J. Edmonds, M. Luby, M. Sudan "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1737-1744, Oct. 1996.

[21] R. Puri,K. Ramchandran, "Multiple description source coding through forward error correction codes," *Proc. 33rd Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp.342-346, Oct. 1999.

[22] V. Stankovic, R. Hamzaoui, and Z. Xiong, "Efficient channel code rate selection algorithms for forward error correction of packetized multimedia bitstreams in varying channels," *IEEE Trans. Multimedia, Special Issue on Streaming Media*, vol. 6, no.2, pp. 240-248, Apr. 2004

[23] S. Dumitrescu, X. Wu, and Z. Wang, "Globally optimal uneven error-protected packetization of scalable code streams," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp.230-239, Apr. 2004.

[24] S. Dumitrescu, X. Wu, and Z. Wang, "Efficient algorithms for optimal uneven protection packetization of single and multiple scalable code streams against packet erasures," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp.1466-1474, Nov. 2007.

[25] S. Dumitrescu and T. Zheng, "Improved multiple description framework based on successively refinable quantization and uneven erasure protection," *In Proc. IEEE Data Compression Conf.*, pp. 514-514, Mar. 2008.

[26] S. Dumitrescu, G. Rivers, S. Shirani, "Unequal Erasure Protection Technique for Scalable Multi-streams," *IEEE Trans. on Image Processing*, vol. 19, no. 2, pp. 422-434, Feb. 2010.

[27] R. G. Gallager, "Low-density parity-check codes," *Ph.D. dissertation, M.I.T.*, 1963.

[28] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. on Inf. Theory*, vol. 27, no. 5, pp. 533-547, 1981.

[29] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Trans. on Inf. Theory*, vol. 49, no. 9, pp. 2225-2229, Sep. 2003.

[30] E. Martinian and J. Yedidia, “Iterative quantization using codes on graphs,” *in Proc. 35th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, 2003.

[31] E. Martinian, M. J. Wainwright, ”Low-density codes achieve the rate-distortion bound” in *Proc. of Data Compression Conference*, pp. 153-162, Mar, 2006.

[32] Z. Sun, M. Shao, J. Chen, K. M. Wong, and X. Wu, “Achieving the rate-distortion bound with low-density generator matrix codes,” *IEEE Trans. on Commun.*, vol. 58, pp. 1643-1653, Jun. 2010.

[33] M. J. Wainwright and E. Maneva, “Lossy source encoding via message-passing and decimation over generalized codewords of LDGM codes,” in *Proceedings of International Symposium on Information Theory*, pp. 1493-1497, Sept. 4-9, 2005.

[34] S. Ciliberti, K. Mezard, and R. Zecchina, “Message passing algorithms for non-linear nodes and data compression,” *ComPlexUs*, vol. 3, pp. 58-65, Aug. 2006.

[35] E. Maneva, E. Mossel, M. J. Wainwright, ”A New Look at Survey Propagation and Its Generalizations”, *Journal of the ACM*, vol. 54, no. 4, July 2007.

[36] A. Braunstein and R. Zecchina, “Survey propagation as local equilibrium equations,” *Journal of Statistical Mechanics*, Jun. 2004.

[37] T. Filler and J. Friedrich, “Binary quantization using belief propagation with decimation over factor graphs of ldgm codes,” *Proc. Allerton Conf. Comm., Control, and Computing*, Monticello, IL, Sept. 2007.

[38] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd Ed., John Wiley & Sons, New Jersey, 2006.

[39] L. Ozarow, "On a source coding problem with two channels and three receivers", *Bell Syst. Tech. J.*, vol. 59, pp. 1909-1921, 1980.

[40] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions", *Bell Syst. Tech. J.*, vol. 59, pp. 1417-1426, 1980.

[41] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. on Inform. Theory*, vol. IT-28, no. 6, pp. 851-857, Nov. 1982.

[42] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. on Inform. Theory*, vol. IT-33, pp. 502-521, Jul. 1987.

[43] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. on Inform. Theory*, vol. IT-31, pp. 721-726, Nov. 1985.

[44] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, "A random variable substitution lemma with applications to multiple description coding", *IEEE Trans. on Inform. Theory*, submitted for publication.

[45] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2106-2114, Sep. 2003.

[46] H. Wang and P. Viswanath, "Vector gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2133-2153, Jun. 2007.

[47] J. Chen, "Rate region of Gaussian multiple description coding with individual and central distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3991-4005, Sep. 2009.

[48] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. on Inform. Theory*, vol. 41, no. 2, pp. 502-521, March. 1995.

[49] V. N. Koshelev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, vol. 16, pp. 31-49, 1980.

[50] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," in *IEEE Trans. on Inform. Theory*, vol. 37, pp. 269-275, Mar. 1991.

[51] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, pp. 253-259, Jan. 1994.

[52] J. Chen, C. Tian, T. Berger, and S. S. Hemami, "Multiple description quantization via Gram-Schmidt orthogonalization," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5197-5217, Dec. 2006.

[53] X. Zhang, J. Chen, S. B. Wicker, and T. Berger, "Successive coding in multiuser information theory," *IEEE Trans. on Inf. Theory*, vol. 53, pp. 2246-2254, Jun. 2007.

[54] H. Everett III, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Res.*, 11, pp. 399-417, 1963.

[55] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, 1969.

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

[57] J. T. Pinkston, “An application of rate-distortion theory to a converse to the coding theorem,” *IEEE Trans. on Inform. Theory*, vol. IT-15, pp. 269-275, Jan. 1969.

[58] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, “On the role of the refinement layer in multiple description coding and scalable cding,” *IEEE Trans. on Inf. Theory*, vol. 57, pp. 1443-1456, Mar. 2011.

[59] W. A, Finamore, and W. A. Pearlman, ”Optimal encoding of discrete-time continuous-amplitude memoryless sources with finite output alphabets”, *IEEE Trans. Inf. Theory*, vol. 26, pp. 144-155, Mar. 1980.

[60] W. A. Pearlman, and A. Chekima, ”Source coding bounds using quantizer reproduction levels”, *IEEE Trans. Inf. Theory*, vol. 30, pp. 559-567, May 1984.

[61] S.-Y. Chung, G. D. Forney, T. J. Richardson and R. Urbanke, “On the design of low-desity parity check codes within 0.0045 dB of the Shannon limit,” *IEEE Communications Letters*, vol. 5, pp. 58-60, Feb. 2001.