Engineering school TELECOM Bretagne
Technopôle Brest-Iroise - CS 83818
29238 Brest Cedex 3 - France
Supervisor: Professor **Frédéric Guilloud**

University Royal Institute of Technology
School of Electrical Engineering - Communication Theory
Osquldas väg 10 - SE-100 44 Stockholm - Sweden
Supervisor: Professor **Lars K. Rasmussen**
Co-Supervisor: Doctor **Vishwambhar Rathi**

Rate-Distortion Bounds for Sparse-Graph Codes

Author: **Grégory Demay**,
Research Engineer at the Royal Institute of Technology

Submitted in candidature for the degree of Master of Engineering from TELECOM Bretagne

Stockholm,
September 1st, 2009 - August 31st, 2010.

Abstract

The aim of this document is to present my work during the last 6 months of my gap year from Mars 2010 to August 2010. They will constitute my final internship before graduating from my master of engineering at TELECOM Bretagne. A part of my work done during the first half of this year is also presented since it is closely related to what have been done after.

This internship takes place in the Swedish university KTH and mainly deals with research on information theory, which can be viewed as a branch of applied mathematics aiming to quantify the information. This research area is a core component in any communication system and as such it is undoubtedly useful to anyone interested in telecommunications.

More technically, from Shannon's classical theory any point-to-point communication can be depicted as a two-step process involving first compression of an information source output, and then transmission of the compressed data to a receiver. In this thesis, we focus on lossy source coding using linear sparse-graph codes. The source considered, introduced by Martinian and Yedidia, is the binary erasure source (BES). It is a discrete memoryless source with ternary output alphabet, and can be viewed as a generalization of the binary symmetric source (BSS). The compression is done using low-density generator matrix (LDGM) codes, and compound codes introduced by Martinian and Wainwright, which are based on LDGM codes and on regular low-density parity-check (LDPC) codes.

The main goal of this thesis is to bound the rate-distortion performance of the aforementioned sparse-graph codes for lossy compression of a BES. As our main contributions, we first derive lower bounds on the rate-distortion performance of LDGM codes for the BES, which are valid for any LDGM code of a given rate and generator node degree distribution and any encoding function. Our approach follows that of Kudekar and Urbanke, where lower bounds were derived for the BSS case. They introduced two methods for deriving lower bounds, namely the counting method and the test channel method. Based on numerical results they observed that the two methods lead to the same bound. We generalize these two methods for the BES and prove that indeed both methods lead to identical rate-distortion bounds for the BES and hence, also for the BSS. Secondly, based on the technique introduced by Martinian and Wainwright, we upper bound the rate-distortion performance of the check regular Poisson LDGM (CRP LDGM) ensemble and the compound LDGM-LDPC ensemble for the BES. We also show that there exist compound LDGM-LDPC codes, with degrees independent of the blocklength, which can achieve any given point on the Shannon rate-distortion curve of the BES.

Résumé

Le but de ce document est de présenter mon travail durant les 6 derniers mois de mon année de césure, de mars 2010 à août 2010. Ceux-ci constitueront mon stage de fin d'étude, nécessaire à l'obtention de mon diplôme d'ingénieur de TELECOM Bretagne. Une partie de mon travail qui s'est déroulée durant la première moitié de mon stage est aussi présentée, étant donné qu'elle est étroitement liée à ce qui a été fait durant la seconde.

Ce stage se déroule dans l'université suédoise KTH et a pour principal sujet la recherche en théorie de l'information. Celle-ci peut être vue comme une branche des mathématiques appliquées, dont le but serait de quantifier l'information. Cette théorie est une composante clée de n'importe quel système de communication et est par conséquent sans aucun doute utile pour quiconque intéressé par le domaine des télécommunications.

Plus techniquement, depuis le travail avant-gardiste de Shannon, n'importe quel communication point-à-point peut être représentée par un processus en deux temps impliquant d'une part la compression des données produites par une source d'information, et d'autre part la transmission de ces données compressées. Dans cette thèse, nous portons notre attention sur la compression de source avec perte en utilisant des codes linéaires ayant une représentation graphique de faible densité. La source considérée, introduite par Martinian et Yedidia, est la source binaire à effacement (BES). Celle-ci est une source discrète sans mémoire et peut être vue comme une genéralisation de la source binaire symmétrique (BSS). La compression est faite en utilisant des codes ayant une matrice génératrice de faible densité (LDGM), ainsi que des codes hybrides introduits par Martinian et Wainwright. Ces derniers sont construits à partir de codes LDGM et de codes réguliers ayant une matrice de parité de faible densité (LDPC).

Le but principal de cette thèse est de minorer et majorer les performances de ces codes par rapport au couple rendement/distorsion pour la compression avec perte de la BES. Nos principales contributions sont dans un premier temps d'obtenir des minorants pour les performances des codes LDGM par rapport au couple rendement/distorsion qui soient valides pour n'importe quel code LDGM avec un rendement et une distribution des degrés des noeuds générateurs donnés, ainsi que n'importe quelle fonction utilisée pour l'encodage. Notre approche suit celle de Kudekar et Urbanke, qui dérivèrent des minorants dans le cas de la BSS. Ils introduisirent deux méthodes pour obtenir ces minorants : la méthode par comptage et celle par canal test. Basé sur des résultats numériques, ils observèrent que les deux méthodes, bien que fondamentalement différentes, menaient au même résultat. Nous généralisons ces deux méthodes au cas de la BES et nous prouvons rigoureusement que ces deux méthodes sont bien équivalentes, dans le sens où elles mènent au même minorant. Deuxièmement, basé sur la technique introduite par Martinian et Wainwright, nous majorons, par rapport au couple rendement/distorsion, les performances des codes LDGM ayant une distribution des degrés des bits d'information suivant une loi de Poisson (CRP LDGM) et des codes hybrides LDGM-LDPC pour le cas de la BES. Finalement, nous prouvons qu'il existe des codes hybrides, avec des degrés indépendants de la longueur de bloc de codage utilisée, pouvant atteindre n'importe quel point sur la courbe de Shannon rendement/distorsion de la BES.

Acknowledgments

I would like to take the opportunity to acknowledge all those who have supported me during this year internship at KTH.

First and foremost, I am indebted to all my supervisors: Prof. Lars K. Rasmussen, Dr. Vishwambhar Rathi, and Prof. Frédéric Guilloud for their wise guidance and thorough proofreading. I would like to especially thank Lars for trusting me and giving me a total freedom in my research. I am also extremely grateful to Vish, who despite a clumsy lock always tried to keep an open door for my questions and needs of enlightenment. More generally I want to thank the Communication Theory Laboratory, which hosted me during this year, for their welcome and generosity.

In addition, I would like to thank Dr. Ingmar Land and the Australian Institute for Telecommunications Research for having given me my first research experience and introduced me to Lars.

More practically, since at the end of the day you still need a roof to sleep under, notably with the Swedish cloudy weather, I would like to particularly thank Prof. Jan Hillert and Dr. Gabrielle Åhlberg for providing me so much more than just the cheapest accommodation in the most beautiful house I ever lived in.

Last but not least, none of this would have been possible without the love and support of my dear parents.

CONTENTS

LIST OF FIGURES

Notations

$\mathscr{A}, \mathscr{B}, \cdots, \mathscr{X}, \mathscr{Y}, \mathscr{Z}$	sets (generally finite).
$A, B, \cdots, X, Y, Z$	random variables.
$a, b, \cdots, x, y, z$	realizations of the random variables $A, B, \cdots, X, Y, Z$ respectively.
$X^n = (X_1, X_2, \cdots, X_n)$	n instances of the random variable X.
$\{S_i\}_{i=1}^{\infty}$	source (generally discrete and memoryless).
S	a source letter.
$\mathscr{S}$	source alphabet.
$\hat{S}$	a reconstruction letter.
$\hat{\mathscr{S}}$	reconstruction alphabet.
$\mathrm{BES}(\varepsilon)$	binary erasure source with erasure probability ε.
$H(X)$	entropy of the random variable X.
$h(p)$	binary entropy function, that is, $-p \log_2 p - (1-p) \log_2 (1-p)$.
$H(X \mid Y)$	conditional entropy of the random variable X given the random variable Y.
$I(X; Y)$	mutual information between the random variables X and Y.
$\log$	natural logarithm.
$\log_2$	logarithm to base 2.
$\mathbb{P}\{A\}$	probability the event A happens.
p_X	probability mass function of the random variable X.
$\mathbb{N}$	set of natural numbers.
$\star$	erasure symbol.

Acronyms

AEP	Asymptotic Equipartition Property.
BEC	binary erasure channel.
BEEC	binary error/erasure channel.
BES	binary erasure source.
BSC	binary symmetric channel.
BSS	binary symmetric source.
CRP LDGM	check regular Poisson LDGM.
DCC	Data Compression Conference.
i.e.	id est.
i.i.d.	independent and identically distributed.
ISITA	International Symposium on Information Theory and its Applications.
LDGM	low-density generator matrix.
LDPC	low-density parity-check.
LHS	left hand side.
RHS	right hand side.

I
Introduction

I.1 Background

Communication can be broadly defined as a process of transferring information between several entities. The problem of reliable communications is far from being new, and the way of solving it always used a certain form of redundancy: by repeating what was just said, by sending many carrier pigeons, by using several post riders, etc.

The invention of electrical telegraph in the beginning of the 19^{th} century revolutionized our way of communicating. The use of an electrical signal as a mean to convey information has two huge benefits in comparison to the pre-telegraphic solutions: it is lightning fast and scalable. The main drawback being the increased complexity of this communication system, since an additional pair of entities, an encoder and a decoder, is now needed in order to transform information into an electrical signal and vice-versa. This invention created a new paradigm in communication, the messages exchanged are now electrical signals, that is to say mathematical functions with power constraints.

This mathematical abstraction is required in order to develop a general theory of communication which should answer two main questions, what is the best possible performance of this system and how to achieve it. A basis of this theory was given by Nyquist and Hartley in [1] and [2] respectively, but the complete and rigorous point-to-point communication theory was not developed until 1948, in Shannon's groundbreaking paper [3]. Roughly, the best possible performance is limited by a nonnegative number called the channel capacity, and it can be achieved using coding. This latter can be broken down into two successive tasks, source coding and channel coding. From Shannon's non-constructive random coding argument, we know that capacity achieving codes exist, and since then, the goal of the coding community has been to find such codes.

A milestone is reached in 1993, when Berrou, Glavieux, and Thitmajshima discovered in [4] the turbo codes, which were the first known capacity approaching codes. A few years later, MacKay proved in [5] that low-density parity-check (LDPC) codes are also capacity achieving under optimal decoding. These codes as well as their decoding algorithm known as the message-passing algorithm, were first created by Gallager in his thesis [6] in 1963, and forgotten due to their overwhelming complexity as compared to computing capabilities of this time. In [7], Tanner generalized the message-passing algorithm using bipartite graphs, a structure well suited for LDPC codes, which led to a new paradigm in the field of coding. Codes can now be described by a sparse graphical model, and the complete task of encoding and decoding can be decomposed by a series of operations done at node level.

Following the remarkable success of sparse-graph codes for the channel coding problem, a natural progression is to explore the capabilities of such codes for the source coding problem. One of the first contributions in this direction was made in [8], where Martinian and Yedidia introduced the binary erasure source (BES), and showed that duals of good sparse-graph channel codes for the binary erasure channel (BEC) are good sparse-graph compression codes for the BES.

Ciliberti, Mezard, and Zecchina used the statistical-physics-based replica method to show in [9] that a low-density generator matrix (LDGM) code with a Poisson generator degree distribution can achieve the Shannon rate-distortion function of the binary symmetric source (BSS) as the average degree increases. Based on this method, they also designed a message-passing encoding algorithm termed Survey Propagation (SP). It was later shown by Wainwright and Maneva [10], and independently by Braunstein and Zecchina [11], that in the context of sparse-graph code compression using decimation over LDGM codes the SP algorithm can be interpreted as a special case of the Belief Propagation (BP) algorithm. More recently Filler and Friedrich proposed a decimation-based BP algorithm, termed bias propagation, that can also perform close to Shannon's rate distortion bound using optimized degree distributions for LDGM codes [12].

Another interesting approach to code construction for lossy compression is based on polar codes introduced by Arikan [13]. Polar codes are based on a deterministic code construction that achieves the channel capacity. It was subsequently shown by Korada and Urbanke that polar codes are also optimal for various lossy compression problems including those for the BES and the BSS [14]. In terms of implementation, however, the encoding and decoding complexities of polar codes are higher than the corresponding complexities for sparse-graph codes with iterative message-passing. A compound sparse-graph code construction was proposed by Martinian and Wainwright in [15, 16, 17, 18], where desirable features of LDPC codes and LDGM codes were combined. They further showed that a randomly chosen code from such ensemble under optimal encoding and decoding achieves the rate-distortion bound with high probability.

The first performance bounds for LDGM-based lossy compression of a BSS were derived by Dimakis, Wainwright, and Ramchandran in [19] for ensembles of codes. In contrast Kudekar and Urbanke derived lower bounds on the rate-distortion function of individual LDGM codes for the BSS [20]. For the BES, it was shown in [8] that sparse-graph codes can achieve the optimal rate only for zero distortion. Furthermore, so far the analysis for lossy compression using sparse-graph codes was mainly focus on the BSS case and there are no known bounds for the BES case. The use of a more general source such as the BES would allow to gain fundamental insight into the behavior of sparse-graph codes used as lossy compressors. As our main contributions in this thesis, we studied the asymptotic behavior of some sparse graphical structures used for lossy compression of a BES. More precisely, we derived lower and upper bounds for the rate-distortion performance of a BES using LDGM codes and the compound construction [21, 22]. Finally, we proved the optimality of the compound construction for lossy compression of a BES.

I.2 Outline and Contributions

In this section, an outline of the thesis is presented along with a summary of contributions.

Chapter II

This chapter aims at giving a tutorial background to the unfamiliar reader in the field of communication theory, and introducing rate-distortion theory using sparse-graph codes. We start by presenting the general digital communication system model, and then shift our focus to discrete memoryless systems. We go through the fundamentals of information theory for discrete random variables, then we discuss linear block codes. Finally, we present the basics of rate-distortion theory by defining distortion measures and rate-distortion codes. As an example, we calculate the rate-distortion function of a BES.

Chapter III

We first detail lossy compression using LDGM codes. After explaining some necessary simplifications used through this chapter, we derive a lower bound on the achievable distortion for lossy compression of a BES using LDGM codes. As our contributions, we derive lower bounds on the rate-distortion performance of LDGM codes for lossy compression of a BES, which are valid for any LDGM code with a given rate and generator node degree distribution, and any encoding function. To do so, we generalize the counting and test-channel method in [20] to the BES case, and formally prove the equivalence of both methods which was numerically observed by Kudekar and Urbanke.

The results of this work have been published in the proceedings of the Data Compression Conference (DCC) [21]. Note that although this work was done during the first half of my gap year, it is highly relevant to include it in this thesis. First and foremost, it is closely related to what have been done after, and secondly the presentation of our results at DCC at the end of March was done by myself, thanks to the generosity of the Communication Theory Laboratory.

Chapter IV

We focus on the check-regular Poisson distributed LDGM ensemble, and on the LDGM-LDPC compound construction which will be introduced. Considering these codes to do lossy compression of a BES, we derive upper bounds on their rate-distortion performance by generalizing the second moment method exposed in [18] to the BES case. We also prove the source coding optimality of the compound construction for the BES. These results have been accepted to the International Symposium on Information Theory and its Applications (ISITA) as [22].

Appendix A

The goal of this appendix is to briefly describe the Swedish university KTH and the Communication Theory Laboratory which hosted me during my year internship.

II
Communication Theory

To provide context for the remainder of the thesis, some technical background in the field of communication is required, which is exposed from Section II.1 to Section II.4. The purpose of these sections is to assist the unfamiliar reader in understanding communication theory, as well as the key concepts of rate-distortion theory using sparse-graph codes, which is the main subject of this document. The reader is expected to have some general knowledge about probability theory, stochastic processes, and linear algebra.

This chapter is organized as follows. In Section II.1, we state the basics of a digital communication system for point-to-point communication and detail discrete memoryless systems. In Section II.2, we define some fundamental concepts of information theory useful to understand rate-distortion theory within the area of source coding. In Section II.3, we detail block coding for source and channel coding problems. We emphasize the importance of linear block codes and define sparse-graph codes. Finally in Section II.4, we present the fundamentals of rate-distortion theory .

II.1 Digital Communication System Model

The fundamental problem in *point-to-point* communication is the reliable transmission of information through an imperfect medium between a source and a sink. The transmission medium, called a *channel*, is imperfect in the sense that a channel output might be different from the channel input in an unpredictable way. This randomness is usually due to physical considerations, grouped under the generic term *noise*. The goal is to retrieve the channel input from the corrupted output of the channel. The rigorous mathematical framework required to study this problem was introduced and developed by Shannon in his groundbreaking paper [3]. We will briefly explain here the key concepts. A more thorough explanation can be found in classical literature, such as [23].

Usually a source can produce a variety of messages, which can be analog or digital, and not necessarily adapted for transmission over the channel considered. For this reason, an intermediate entity is required in order to guarantee the suitability of the signal to be transmitted. Similarly, the channel output might be not directly understandable by the destination, and requires another intermediate entity. Thus, a typical communication system requires 5 different entities:

- a *source*, which produces either analog or digital information messages to be communicated to the sink;
- a *transmitter*, which modifies the source output to produce a signal suitable for transmission over the channel;
- a channel, which is the transmission medium;
- a *receiver*, which works on the channel output to reconstruct the initial message for the destination;
- a *destination*, which is the entity for which the message is intended.

A block diagram of a general communication system is depicted in Figure II.1.

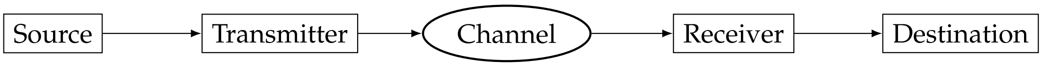


Figure II.1: *A typical point-to-point communication system.*

So far, reliability in point-to-point communications is enforced by a two steps process. Firstly, given the imperfection of the channel, the number of channel uses is minimized. Secondly, a certain amount of *redundancy* is added to the initial message in a clever manner, such that it will help the receiver to recover the source message from the noisy output of the channel. Both steps and their inverse operations are done at the transmitter and at the receiver respectively.

Consequently, we can refine the model for the transmitter and the receiver. Indeed, producing a signal suitable for transmission from the source message can be decomposed into 3 parts:

- a *source encoder*, which will compress the digital or analog message from the source into a minimal representation in some finite field (usually a binary sequence) in order to minimize the use of an unreliable channel;
- a *channel encoder*, which will add a certain amount of redundancy to the source encoder output in order to increase the reliability of the received data;
- a *modulator*, which will transform the digital message into an analog signal, since channel inputs and outputs are generally analog waveforms.

The details of a general transmitter are shown in Figure II.2.

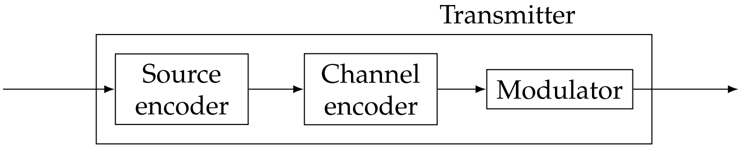


Figure II.2: *Details of a transmitter.*

In the same manner, we can detail the operations done at the receiver. Producing a message suitable for the destination can also be divided into 3 parts:

- a *demodulator*, which will digitize the corrupted signal output by the channel ;
- a *channel decoder*, which will hopefully be able to retrieve the output of the source encoder from the output of the demodulator using the added redundancy at the transmitter;
- a *source decoder*, which will determine the initial source message from the output of the channel decoder.

The details of a general receiver are shown in Figure II.3.

From this partitioning of our initial communication system (Figure II.1), we have 3 pairs of entities which work in a transparent manner to each other:

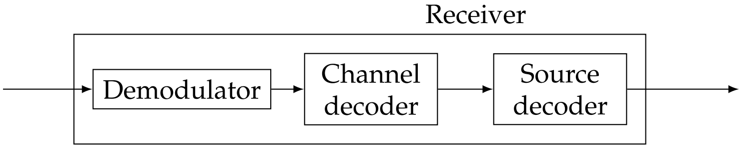


Figure II.3: *Details of a receiver.*

- source encoder/source decoder;
- channel encoder/channel decoder;
- modulator/demodulator.

Our primary focus is on source coding. Thus we will assume that the problem of efficient modulation / demodulation is solved and will not be mentioned any further. As a consequence, the complete point-to-point communication system model we will consider is represented in Figure II.4, where modulation / demodulation is now considered as part of the channel.

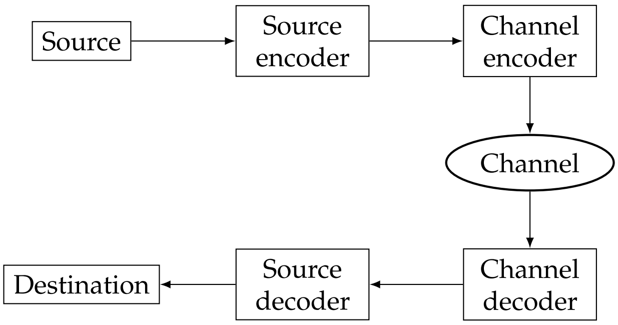


Figure II.4: *Complete point-to-point communication problem.*

Thus, we now have two main problems:

- for a given source, how do we do an efficient source encoder and decoder ? This is the *source coding* problem;
- for a given channel, how do we do an efficient channel encoder and decoder ? This is the *channel coding* problem.

This key idea of splitting a communication problem into source and channel coding problems is known as the Shannon's *source-channel separation* theorem and allows a great flexibility. Indeed, a good source coding solution can be used for a variety of channels, while a good channel coding solution can be used for different sources. Although our main interest is in the source coding problem, it is closely related to the channel coding one (notion of duality), and that is the reason why some notions of channel coding will be explained.

But before going into more details on these problems, we need to specify the models used for sources and channels. Basically, a source will be considered as a sequence of random variables, whereas a channel will be viewed as a probabilistic mapping. We will focus on discrete memoryless sources and discrete memoryless channels.

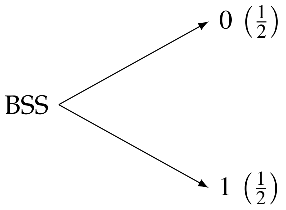
II.1.1 Discrete Memoryless Sources

In this subsection we will define a discrete memoryless source and consider two examples of interest for our case. The concept of a discrete source is important, since in a practical system any message from the source will be discretized due to the finite storage capacity of the system considered. The property of being memoryless is usually a simplification from the real system. In a more mathematical way, a discrete memoryless source can be defined as follows.

Definition II.1.1 - *Discrete memoryless source [24]*:
A *discrete source* is a sequence of random variables $\{S_i\}_{i=1}^{\infty}$ taking values in a finite set $\mathscr{S}$, called the *source alphabet*. If the S_i's are independent and identically distributed (i.i.d.), we speak of a *discrete memoryless source*.

Thus, a discrete memoryless source is characterized by a source alphabet and a probability distribution. In the sequel of this subsection we give two examples of discrete memoryless sources. First we consider the simplest one which is the binary symmetric source (BSS), and then the binary erasure source (BES), which is the source we will mainly focus on.

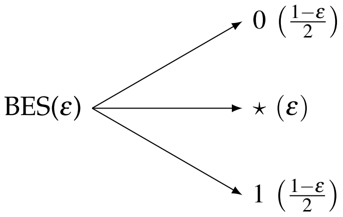
Example II.1.1 - *Binary symmetric source*
The most well-known example of a discrete memoryless source is the BSS. This source has equiprobable binary output, and is shown in Figure II.5.



$$\mathscr{S} = \{0,1\}$$
$$\mathbb{P}\{S_i = 0\} = \mathbb{P}\{S_i = 1\} = \frac{1}{2}$$

Figure II.5: *The binary symmetric source.*

Example II.1.2 - *Binary erasure source*
The BES was introduced in [8] for the binary erasure quantization problem. It is a discrete memoryless source, with a ternary alphabet $\mathscr{S} \triangleq \{0,1,\star\}$, where $\star$ is the erasure symbol. Generally, the BES models the situation where some of the bits output by a BSS are considered to be irrelevant, lost, or corrupted by noise, and thus are uniformly represented by erasures. In particular, this source could be a good model for some network applications, where bits could be lost during transmission, or to represent the output of a BEC (see Example II.1.4). A BES whose source symbol can take on the value $\{\star\}$ with probability ε, or the values $\{0,1\}$ with equal probabilities is denoted by BES(ε). A BES(ε) is shown in Figure II.6.


0 $\left(\frac{1-\varepsilon}{2}\right)$
BES(ε)
$\star$ (ε)
1 $\left(\frac{1-\varepsilon}{2}\right)$

$$\mathscr{S} = \{0,1,\star\}$$
$$\mathbb{P}\{S_i = \star\} = \varepsilon$$
$$\mathbb{P}\{S_i = 0\} = \mathbb{P}\{S_i = 1\} = \frac{1-\varepsilon}{2}$$

Figure II.6: *The binary erasure source.*

Note that a BES can be viewed as a generalization of a BSS, since a BES(ε) with zero erasure probability ($\varepsilon = 0$) is a BSS.

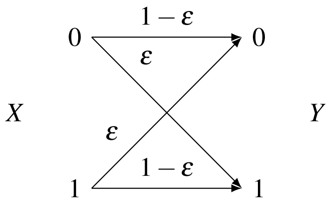
II.1.2 Discrete Memoryless Channels

The concept of a discrete channel is of interest since the whole chain "modulator–channel–demodulator" can be seen as a discrete channel, irrespective of the transmission medium. Note that we will always consider channels without any feedback, and thus in any definition concerning channels this assumption is implicitly made.

Definition II.1.2 - *Discrete memoryless channel [24, 25]*:
A *discrete channel* is a stochastic process characterized by a finite input set $\mathscr{X}$, a finite output set $\mathscr{Y}$ and a transition matrix $W : \mathscr{X} \to \mathscr{Y}$, where $W(y|x)$ denotes the probability of observing y, given that x was the channel input, $\forall(x,y) \in \mathscr{X} \times \mathscr{Y}$. The channel is said to be *memoryless* if the probability distribution of the output depends only on the input at that time, and is conditionally independent of past inputs and outputs.

We now give two simple examples of discrete memoryless channels. First we consider the binary symmetric channel (BSC), and then the binary erasure channel (BEC).

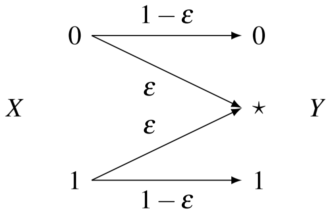
EXAMPLE II.1.3 - *Binary symmetric channel*
The BSC is a discrete memoryless channel with binary input and output. This latter is characterized by its *crossover probability*, which is the probability that a bit is flipped during the transmission. A BSC with crossover probability ε is shown in Figure II.7.



$$\mathscr{X} = \mathscr{Y} = \{0,1\}$$
$$\mathbb{P}\{Y=0 \,|\, X=0\} = \mathbb{P}\{Y=1 \,|\, X=1\} = 1-\varepsilon$$
$$\mathbb{P}\{Y=0 \,|\, X=1\} = \mathbb{P}\{Y=1 \,|\, X=0\} = \varepsilon$$

Figure II.7: *The binary symmetric channel.*

EXAMPLE II.1.4 - *Binary erasure channel*
The BEC is also a discrete memoryless channel but with binary input alphabet and ternary output alphabet $\{0,1,\star\}$, where $\star$ stands for an erasure. A BEC is represented in Figure II.8.



$$\mathscr{X} = \{0,1\}$$
$$\mathscr{Y} = \{0,1,\star\}$$
$$\mathbb{P}\{Y=\star \,|\, X=0\} = \mathbb{P}\{Y=\star \,|\, X=1\} = \varepsilon$$
$$\mathbb{P}\{Y=0 \,|\, X=0\} = \mathbb{P}\{Y=1 \,|\, X=1\} = 1-\varepsilon$$

Figure II.8: *The binary erasure channel.*

II.2 NOTIONS OF INFORMATION THEORY

We will present here some basic notions of information theory for discrete systems, implying that all the random variables considered are defined over a discrete set. For more details, a good starting point is [25], or [24] for more advanced notions.

One of the main goals of information theory is to provide a mathematical framework which defines rigorously the concept of ”information” and its related notions. A key step toward the definition of information is to realize that the less is known about an event, the more information its realization will provide.

Entropy is the core information measure in information theory and characterizes the average uncertainty of a given random variable.

Definition II.2.1 - *Entropy [25]*:
Consider a discrete random variable X defined over a finite set $\mathscr{X}$ with probability mass function p_X. The *entropy* of X is denoted by $H(X)$ and is defined as

$$H(X) \triangleq \sum_{x\in\mathscr{X}} p_X(x) \log_2 \frac{1}{p_X(x)}, \tag{II.1}$$

where $H(X)$ will be measured in *bits*, since we use logarithms to base 2. In other words, the entropy corresponds to the average number of bits needed to describe the random variable considered.

An important special case is when the random variable considered is binary.

$$X \sim \text{Ber}(p) \implies H(X) = h(p),$$

where $X \sim \text{Ber}(p)$ means that X is Bernoulli distributed with parameter p, and $h(\cdot)$ is the *binary entropy function* id est (i.e.), $h(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$. We can now define the joint entropy and the conditional entropy.

Definition II.2.2 - *Joint entropy [25]*:
Consider two discrete random variables (X,Y) defined over the finite set $\mathscr{X} \times \mathscr{Y}$ with joint probability mass function p_{XY}. The *joint entropy* of the random variables X and Y is denoted by $H(X,Y)$ and is defined as

$$H(X,Y) \triangleq \sum_{(x,y)\in\mathscr{X}\times\mathscr{Y}} p_{XY}(x,y) \log_2 \frac{1}{p_{XY}(x,y)}. \tag{II.2}$$

Definition II.2.3 - *Conditional entropy [25]*:
Consider two discrete random variables (X,Y) defined over the finite set $\mathscr{X} \times \mathscr{Y}$ with joint probability mass function p_{XY}. The *conditional entropy* of the random variable Y given the random variable X is denoted by $H(Y \mid X)$ and is defined as

$$H(Y \mid X) \triangleq \sum_{(x,y)\in\mathscr{X}\times\mathscr{Y}} p_{XY}(x,y) \log_2 \frac{1}{p_{Y|X}(y|x)}. \tag{II.3}$$

A natural property of the conditional entropy is that *conditioning reduces entropy*, i.e.,

$$H(Y \mid X) \leq H(Y).$$

Note that we have the following relationship between entropy, joint entropy, and conditional entropy

$$H(Y \mid X) = H(X,Y) - H(X). \tag{II.4}$$

Another important concept of information theory is the mutual information between two random variables, which quantifies how much information one provides about the other.

Definition II.2.4 - *Mutual information [25]*:
Consider two discrete random variables (X,Y) defined over the finite set $\mathscr{X} \times \mathscr{Y}$ with joint probability mass function p_{XY}, and marginals p_X, p_Y. The *mutual information* between the random variables X and Y is denoted by $I(X;Y)$ and is defined as

$$I(X;Y) \triangleq \sum_{(x,y)\in\mathscr{X}\times\mathscr{Y}} p_{XY}(x,y)\log_2 \frac{p_{XY}(x,y)}{p_X(x)\,p_Y(y)}. \tag{II.5}$$

Note that the mutual information is symmetric $I(X;Y) = I(Y;X)$. A practical way of computing the mutual information is to use its close relationship with entropy,

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X). \tag{II.6}$$

For a given channel, a key parameter called the *channel capacity* quantifies the maximum amount of information possible to transmit reliably per channel use.

Definition II.2.5 - *Capacity [25]*:
Consider two discrete random variables X,Y defined over the finite sets $\mathscr{X}$, and $\mathscr{Y}$ with probability mass function p_X, and p_Y respectively. Let $\mathscr{Q}$ be the set of probability mass functions p_X defined on $\mathscr{X}$. Consider a discrete memoryless channel with input random variable X, and output random variable Y. Then the *capacity* C for this channel is defined by

$$C = \max_{p_X\in\mathscr{Q}} I(X;Y), \tag{II.7}$$

where C is expressed in bits/channel use (assuming $\log_2$ was used to compute the mutual information).

II.3 Codes on Finite Fields

Source and channel coding problems can be seen as a mapping from a set of messages to another set of messages, called a code. Since we consider only discrete systems, it is natural to use finite fields. Moreover, we will consider here only block codes, meaning that the length of the input and output sequences at the encoder are fixed. In the sequel, we shall say simply code for a block code, and in the remaining of this section k,n,m, and M will denote natural numbers. More details about various coding techniques can be found in [26].

A block code over a finite field $\mathbb{F}$ can be defined as follows.

Definition II.3.1 - *Block code [26]*:
A *block code* $\mathscr{C}$ of length n and cardinality M over a finite field $\mathbb{F}$ is a collection of M elements from $\mathbb{F}^n$, i.e.,

$$\mathscr{C}(n,M) = \{c_1^n, \cdots, c_M^n\}, c_i^n \in \mathbb{F}^n, 1 \le i \le M. \tag{II.8}$$

The elements of a code $\mathscr{C}(n,M)$ are called *codewords* and the parameter n is called the *blocklength*.

Note that all operations will be done in $\mathbb{F}$. The most used finite field for coding is the binary field $\mathbb{F}_2 \triangleq \{0,1\}$. Usually, we will only consider binary codes, meaning that every codeword is a binary sequence.

For a given code $\mathscr{C}(n,M)$, an important measure called the rate is the proportion of initial information contained in the code. In the general case, let $|\mathbb{F}|$ denote the cardinality of $\mathbb{F}$. Note that $\log_{|\mathbb{F}|} M$ corresponds to the number of source symbols taken as inputs at the encoder, and will also be denoted by k in the rest of the chapter. Then, we have

Definition II.3.2 - *Rate [26]*:
The *rate* R of a block code $\mathscr{C}(n,M)$ defined over $\mathbb{F}$ is

$$R = \frac{1}{n}\log_{|\mathbb{F}|} M, \tag{II.9}$$

R is expressed in information symbols per transmitted symbol.

Generally, the optimal source encoder or optimal channel decoder for a block coding strategy have exponential complexity with respect to the blocklength. Indeed, they both try to find the closest codeword from a given vector, requiring a search through $|\mathbb{F}|^{nR}$ codewords.

II.3.1 Linear Block Codes

In order to simplify the coding and decoding complexity, some algebraic structure has to be introduced in the definition of codes. One of the most important classes of block codes is the class of linear block codes.

Definition II.3.3 - *Linear block codes*:
A *block code* $\mathscr{C}(n,M)$ over the finite field $\mathbb{F}$ is said to be *linear* if the codewords of the code span a *linear subspace* of $\mathbb{F}^n$.

Since a linear code $\mathscr{C}(n,M)$ is a linear subspace of $\mathbb{F}^n$, there exists some integer k, $0 \le k \le n$ such that $\mathscr{C}$ has *dimension* k. Consequently, there exists a *generator matrix* $\mathbf{G}$ (generally not unique) of dimension $k \times n$ which generates the set of codewords. $\mathscr{C}(\mathbf{G})$ will denote a code generated by the matrix $\mathbf{G}$,

$$\mathscr{C}(\mathbf{G}) = \left\{ c^n \in \mathbb{F}^n : \; c^n = u^k\mathbf{G}, \; u^k \in \mathbb{F}^k \right\}. \tag{II.10}$$

Note that both notations $\mathscr{C}(n,M)$ and $\mathscr{C}(\mathbf{G})$ are equivalent, since they produce the same set of codewords.

Each code $\mathscr{C}$ can also be associated with a dual code, denoted by $\mathscr{C}^{\perp}$, which is a set of elements in $\mathbb{F}^n$ orthogonal to any codeword in $\mathscr{C}$.

Definition II.3.4 - *Dual code [26]*:
Consider a linear block code $\mathscr{C}(\mathbf{G})$, where $\mathbf{G} \in \mathbb{F}^{k\times n}$. The *dual code* $\mathscr{C}^{\perp}$ associated to $\mathscr{C}(\mathbf{G})$ is

$$\mathscr{C}^{\perp} = \left\{ v^n \in \mathbb{F}^n : \; c^n (v^n)^T = 0, \; \forall c^n \in \mathscr{C}(\mathbf{G}) \right\} = \left\{ v^n \in \mathbb{F}^n : \; \mathbf{G}(v^n)^T = 0 \right\}, \tag{II.11}$$

where $(v^n)^T$ denotes the transpose of the vector v^n of length n.

The generator matrix of a dual code is called a *parity-check matrix* and is denoted by $\mathbf{H}$, $\mathbf{H} \in \mathbb{F}^{(n-k)\times n}$.

A code $\mathscr{C}$ defined by a generator matrix $\mathbf{G}$, can also be defined by its parity-check matrix $\mathbf{H}$

$$\mathscr{C}(\mathbf{H}) = \left\{ v^n \in \mathbb{F}^n : \; \mathbf{H}(v^n)^T = 0 \right\}. \tag{II.12}$$

Thus, a code can be defined by a generator matrix or a parity-check matrix and both definitions are equivalent in the sense that they produce the same set of codewords. A row in the parity-check matrix will be called a *parity-check equation*, since it corresponds to an equation that the bits of a codeword must satisfy.

In a less mathematical way, the generator matrix will expand the encoder input by applying a set of linear combinations, whereas the parity-check matrix will restrict the encoder output by applying a set of linear constraints.

Tanner Graphs

Tanner Graphs were invented by R. Michael Tanner in [7]. Although they have a much broader use, we will use them only to represent linear block codes in a more convenient and visual way. Any generator matrix or parity-check matrix (basically any matrix) has a Tanner graph representation which is a *bipartite graph*. In one set of vertices, a node corresponds to a specific row in the matrix, whereas in the other set, a node represents a column in the matrix. There is an edge in the graph if and only if the corresponding coefficient in the matrix considered is not null (in non binary case the edge also has a label corresponding to the value of this coefficient). The terminology varies if we consider a Tanner graph associated to a parity-check matrix or a Tanner graph associated to a generator matrix. In Figure II.9 we give an example of a Tanner graph for a LDPC code, while an example of a Tanner graph for a LDGM code is given in Figure II.10.

II.3.2 Linear Sparse-Graph Codes

Linear sparse-graph codes are linear block codes which have at least one *sparse* Tanner graph, meaning that the number of edges in the code grows linearly with the blocklength instead of a square dependency. In the following we will describe LDPC and LDGM codes, which are two different types of sparse-graph codes. In Section IV.1, we will introduce another type of sparse-graph codes, which are based on a compound construction using both LDPC and LDGM codes.

Low-Density Parity-Check Codes

LDPC codes are a class of linear sparse-graph codes introduced by Gallager in [6] and rediscovered by MacKay in [5]. They are characterized by *sparse* random parity-check matrices. This sparseness allows for a linear complexity with the blocklength within the iterative decoding algorithm. Thus, a LDPC code is nothing more than a classical linear block code, where its parity-check matrix is mostly full of zeroes.

Definition II.3.5 - *LDPC code*:
Consider a *sparse* parity-check matrix $\mathbf{H}$, where $\mathbf{H} \in \mathbb{F}^{(n-k)\times n}$. Then, a LDPC code denoted by $\mathbb{M}(\mathbf{H})$, is a linear block code of rate $R_H \geq \frac{k}{n}$ induced by $\mathbf{H}$, i.e.,

$$\mathbb{M}(\mathbf{H}) = \left\{c^n \in \mathbb{F}^n : \ \mathbf{H}(c^n)^T = 0\right\}. \tag{II.13}$$

An important subclass of LDPC codes are the ones which are said to be *regular*.

Definition II.3.6 - *Regular LDPC code*:
A (n, p, q)-regular LDPC code is a linear block code of length n characterized by a parity-check matrix $\mathbf{H}$, $\mathbf{H} \in \mathbb{F}^{(n-k)\times n}$, where $\mathbf{H}$ has exactly p ones per column and q ones per row.

Example II.3.1 - $(10,2,4)$-*regular LDPC code*
The parity-check matrix of a $(10,2,4)$-regular LDPC code is given in (II.14) and its corresponding Tanner graph is shown in Figure II.9.

$$H = \begin{array}{cccccccccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} & \\ \multirow{5}{*}{} & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & c_1 \\ & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & c_2 \\ & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & c_3 \\ & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & c_4 \\ & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & c_5 \end{array} \tag{II.14}$$

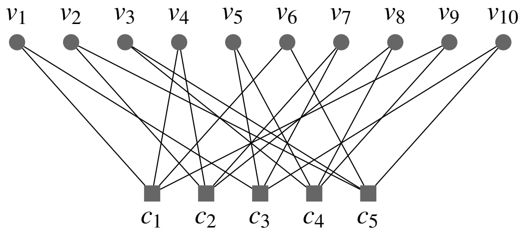


Figure II.9: *Tanner graph of a* $(10,2,4)$*- regular LDPC code.*

A parity-check equation (a row in the parity-check matrix) corresponds to a *parity-check node* represented by a square node (■), and a column in the parity-check matrix corresponds to a *variable node* represented by a circle node (●). The edges in the Tanner graph corresponds to the ones in the parity-check matrix.

Note that the performance of LDPC codes can be greatly improved for channel coding if we consider irregular codes [27], meaning that the number of ones per column and/or per row in the parity-check matrix is not constant anymore.

Low-Density Generator-Matrix codes

LDGM codes have a representation in term of a sparse generator matrix.

Definition II.3.7 - *LDGM code*:
Consider a *sparse* generator matrix $\mathbf{G}$, where $\mathbf{G} \in \mathbb{F}^{k \times n}$. Then, a LDGM code denoted by $\mathbb{L}(\mathbf{G})$, is a linear block code of rate $R_G \leq \frac{k}{n}$ defined by $\mathbf{G}$, i.e.,

$$\mathbb{L}(\mathbf{G}) = \left\{ c^n \in \mathbb{F}^n : \exists w^k \in \mathbb{F}^k \text{ s.t. } c^n = w^k \mathbf{G} \right\}. \tag{II.15}$$

Analogously to regular LDPC codes, we can define regular LDGM codes.

Definition II.3.8 - *Regular LDGM code*:
A (n,p,q)-regular LDGM code is a linear block code of length n characterized by a generator matrix $\mathbf{G}$, $\mathbf{G} \in \mathbb{F}^{k \times n}$, where $\mathbf{G}$ has exactly p ones per column and q ones per row.

Example II.3.2 - $(12,3,4)$-*regular LDGM code*
The generator matrix of a $(12,3,4)$-regular LDGM code is given in (II.16) and its corresponding Tanner graph is shown in Figure II.10.

$$G=\begin{array}{c|cccccccccccc|c} & c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 & c_{10} & c_{11} & c_{12} & \\ \hline & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_1 \\ & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & w_2 \\ & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & w_3 \\ & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & w_4 \\ & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & w_5 \\ & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & w_6 \\ & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & w_7 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & w_8 \\ & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & w_9 \end{array} \tag{II.16}$$

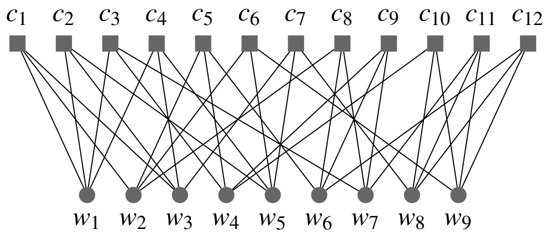

c_1 c_2 c_3 c_4 c_5 c_6 c_7 c_8 c_9 c_{10} c_{11} c_{12}
w_1 w_2 w_3 w_4 w_5 w_6 w_7 w_8 w_9

Figure II.10: *Tanner graph of a* $(12,3,4)$- *regular LDGM code.*

Similarly to LDPC codes, we have a one-to-one correspondence between each node of the graph and a row or a column of the generator matrix. Each single row is represented by a circle node (●), and is called a *generator node*. Each column is represented by a square node (■). The edges in the Tanner graph corresponds to the ones in the generator matrix.

Like for LDPC codes, if the number of ones per row and/or column is not constant, we speak of *irregular LDGM codes*. For these codes, an important characteristic is the generator node degree distribution, denoted by $L(x)$.

Definition II.3.9 - *Generator node degree distribution*:
For a given Tanner graph of a LDGM code, let L_i be the proportion of generator nodes having degree i. Then, the generator node degree distribution $L(x)$ is

$$L(x) \triangleq \sum_i L_i x^i. \tag{II.17}$$

Thus, for Example II.3.2, $L(x) = x^4$.

II.4 Rate-Distortion Theory

We will now detail source coding problems. In source coding, one can distinguish lossless source coding and lossy source coding. From classical theory [25], we perfectly know how to deal with lossless source coding, the minimum achievable rate being the entropy of the source and several algorithms are known to asymptotically achieve it. However, what is the minimal rate we could achieve if we allow some loss between the original source sequence and the reconstructed sequence output by the source decoder ? This problem leads to rate-distortion theory that we will briefly describe.

We first need to define precisely the distortion between the source and reconstructed sequences. Consider a discrete memoryless source $\{S_i\}_{i=1}^{\infty}$ and its finite alphabet $\mathscr{S}$ as defined in Definition II.1.1. We consider source sequences of length n, denoted by S^n, and where $n \in \mathbb{N}$.

$$S^n = (S_1, S_2, \cdots, S_n)\,,\ S_i \in \mathscr{S}, 1 \le i \le n. \tag{II.18}$$

Let the source decoder output be $\hat{S}^n$, where

$$\hat{S}^n = \left(\hat{S}_1, \hat{S}_2, \cdots, \hat{S}_n\right)\,,\ \hat{S}_i \in \hat{\mathscr{S}}, 1 \le i \le n, \tag{II.19}$$

and $\hat{\mathscr{S}}$ denotes the reconstruction alphabet, which is a finite set, and not necessarily the same as the source alphabet $\mathscr{S}$.

II.4.1 Distortion Measures

We are now able to define the idea of distortion between a source sequence and its reconstructed version. To do so, we first define what a single-letter distortion measure is, and then extend it naturally to sequences of n symbols.

Definition II.4.1 - *Single-letter distortion measure*:
A single-letter distortion measure d is a mapping between the source alphabet $\mathscr{S}$, the reconstruction alphabet $\hat{\mathscr{S}}$, and the set of non negative real numbers $\mathbb{R}_+$

$$\begin{array}{rccc} d: & \mathscr{S} \times \hat{\mathscr{S}} & \rightarrow & \mathbb{R}_+ \\ & (s, \hat{s}) & \mapsto & d(s, \hat{s}). \end{array} \tag{II.20}$$

A natural extension of this definition to source sequences is as following

Definition II.4.2 - *Multi-letter distortion measure*:
A multi-letter distortion measure between a source sequence $s^n \in \mathscr{S}^n$ and a reconstructed sequence $\hat{s}^n \in \hat{\mathscr{S}}^n$, induced by a single-letter distortion measure d is

$$d^{(n)}(s^n, \hat{s}^n) = \frac{1}{n} \sum_{i=1}^{n} d(s_i, \hat{s}_i). \tag{II.21}$$

In the sequel we will denote a single or a multi-letter distortion measure by d and simply talk about distortion measure since there is no ambiguity between those two, the meaning being clear from the context.

An important example of a distortion measure for finite alphabets, and especially for binary sequences is the Hamming distortion measure.

Example II.4.1 - *Hamming distortion measure*
Assume for this example that $\mathscr{S} = \hat{\mathscr{S}}$ (with $\mathscr{S}$ finite). Then, the Hamming distortion measure is defined as follow

$$\begin{array}{rccl} d: & \mathscr{S} \times \hat{\mathscr{S}} & \rightarrow & \{0,1\} \\ & (s, \hat{s}) & \mapsto & \left\{ \begin{array}{ll} 0, & \text{if } s = \hat{s}, \\ 1, & \text{if } s \neq \hat{s}. \end{array} \right. \end{array} \tag{II.22}$$

For the BES, we will use a slightly generalized version of the Hamming distortion measure. Note that when a source letter is an erasure it would seem logical to be able to encode it by either a zero or a one without incurring any distortion penalty, since having an erasure implies that we already lost the information about this specific source symbol.

Example II.4.2 - *Generalized Hamming distortion measure for the BES*
Consider a BES and a binary reconstruction alphabet $\hat{\mathscr{S}} = \{0,1\}$. A generalized Hamming distortion measure for the BES can be defined as follows

$$\begin{array}{rccl} d: & \{0,1,\star\} \times \{0,1\} & \rightarrow & \{0,1\} \\ & (s, \hat{s}) & \mapsto & \left\{ \begin{array}{ll} 0, & \text{if } s = \star \text{ or } s = \hat{s}, \\ 1, & \text{otherwise.} \end{array} \right. \end{array} \tag{II.23}$$

This distortion measure will be used when lossy compression of a BES is considered.

II.4.2 Rate-Distortion Codes

For this subsection, assume that a source S, a source alphabet $\mathscr{S}$, a reproduction alphabet $\hat{\mathscr{S}}$, and a distortion measure d defined over $\mathscr{S} \times \hat{\mathscr{S}}$ are given.

Definition II.4.3 - *(n,M) rate-distortion code*:
A (n,M) rate-distortion code is constituted by an encoding function f, and a decoding function g, where

$$f: \ \mathscr{S}^n \rightarrow \{1,2,\cdots,M\}, \tag{II.24}$$
$$g: \ \{1,2,\cdots,M\} \rightarrow \hat{\mathscr{S}}^n. \tag{II.25}$$

The encoding function maps the source sequences to some index from the set $\{1,2,\cdots,M\}$, while the decoder maps the index to a codeword. Consequently, the reconstruction set $\hat{\mathscr{S}}^n$ corresponds to a codebook.

Note that to each rate-distortion code and distortion measure d there is an associated distortion D, $D \in [0,1]$, which is the expected distortion over the set of source sequences

$$D = \mathbb{E}\left[d\left(S^n, g\left(f\left(S^n\right)\right)\right)\right] = \sum_{s^n \in \mathscr{S}^n} \mathbb{P}\left\{S^n = s^n\right\} d\left(s^n, g\left(f\left(s^n\right)\right)\right). \tag{II.26}$$

Definition II.4.4 - *Achievable rate-distortion pair*:
A rate-distortion pair (R,D) is said to be achievable if for any $\varepsilon > 0$ there exists for a sufficiently large blocklength n_ε, a $(n_\varepsilon, M_\varepsilon)$ rate-distortion code defined over $\hat{\mathscr{S}}$ such that

$$\frac{1}{n}\log_{|\hat{\mathscr{S}}|} M_\varepsilon \leq R + \varepsilon, \tag{II.27}$$
$$\mathbb{E}\left[d\left(S^n, g\left(f\left(S^n\right)\right)\right)\right] \leq D + \varepsilon. \tag{II.28}$$

Definition II.4.5 - *Rate-distortion region*:
The rate-distortion region is the closure of the set of all achievable rate-distortion pairs (R,D).

Definition II.4.6 - *Rate-distortion function*:
The rate-distortion function, denoted by $R(D)$, is the minimum of all rates R for a given distortion D, such that the rate-distortion pair (R,D) is achievable.

The rate-distortion function could also be defined as the infimum of rates R such that (R,D) is in the rate-distortion region for a given distortion D [25]. The rate-distortion function is a central quantity in rate-distortion theory since it characterizes the theoretical limits of lossy compression for a given source. In a nutshell, the quantity $R(D)$ corresponds to the minimum achievable rate at distortion D.

A key theorem due to Shannon in [28] gives us a way to compute the rate-distortion function.

Theorem II. 4.1 - *The Rate-Distortion Theorem*:
Consider a discrete memoryless source S with source alphabet $\mathscr{S}$, distribution $p_S(s)$, and a distortion measure d defined over $\mathscr{S} \times \hat{\mathscr{S}}$. Then the rate-distortion function $R(D)$ is given by

$$R(D) = \min_{p_{\hat{S}|S}(\hat{s}|s) \in \mathscr{P}} \left\{ I(S;\hat{S}) \right\}, \text{ where} \tag{II.29}$$

$$\mathscr{P} = \left\{ p_{\hat{S}|S}\left(\hat{s} \mid s\right) : \sum_{(s,\hat{s}) \in \mathscr{S} \times \hat{\mathscr{S}}} p_S\left(s\right) p_{\hat{S}|S}\left(\hat{s} \mid s\right) d\left(s,\hat{s}\right) \leq D \right\}.$$

We will now compute the rate-distortion function for the BES using the distortion measure defined in Example II.4.2.

EXAMPLE II.4.3 - *Shannon rate-distortion function for the BES*
Consider a BES(ε) shown in Example II.1.2 and the distortion measure d introduced in Example II.4.2. Then, the Shannon rate-distortion function for the BES(ε) is denoted by $R_{\varepsilon}^{\text{sh}}(D)$, defined as (II.30), and plotted in Figure II.11.

$$R_{\varepsilon}^{\text{sh}}(D) \triangleq \begin{cases} (1-\varepsilon)\left[1-h\left(\frac{D}{1-\varepsilon}\right)\right], & \text{if } D < \frac{1-\varepsilon}{2}, \\ 0, & \text{if } D \geq \frac{1-\varepsilon}{2}. \end{cases} \tag{II.30}$$

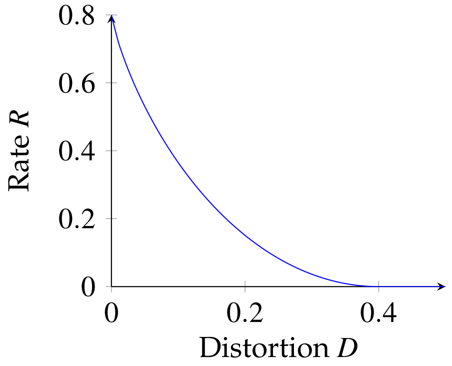


Figure II.11: *Shannon rate-distortion function $R_{\varepsilon}^{sh}(D)$ for a BES(ε) with $\varepsilon = 0.2$.*

Proof. We will show this result in a classical way. First we show that the mutual information between S and $\hat{S}$ is lower bounded by (II.30) and then we show that this value is achievable [25].

Consider a source letter $S \in \mathscr{S} = \{0, 1, \star\}$ and a reconstruction letter $\hat{S} \in \hat{\mathscr{S}} = \{0, 1\}$. For this example, define the $\oplus$ operator as follows

$$S \oplus \hat{S} = \begin{cases} S + \hat{S} \mod 2 & \text{if } S \neq \star \\ \star & \text{if } S = \star. \end{cases}$$

Consequently, $\left(S \oplus \hat{S}\right) \oplus \hat{S} = S$. Then,

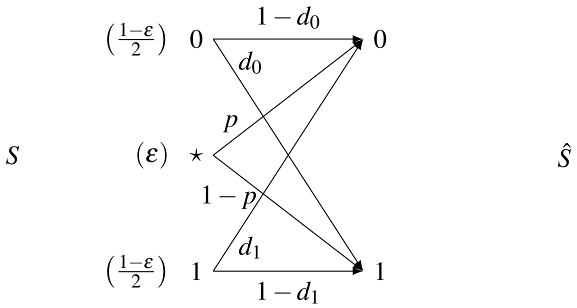
$$\begin{aligned} I(S;\hat{S}) &= H(S) - H(S \mid \hat{S}) \\ &= H(S) - H(S \oplus \hat{S} \mid \hat{S}) \\ &\geq H(S) - H(S \oplus \hat{S}). \end{aligned} \tag{II.31}$$

Looking back at Figure II.6, the entropy of the source is $H(S) = h(\varepsilon) + (1-\varepsilon)$. Let $d \triangleq \mathbb{P}\left\{S \oplus \hat{S} = 1\right\}$, which corresponds to the probability that the reconstruction letter is different from the source letter when it is not an erasure. The probability of having an erasure in $S \oplus \hat{S}$ is the same as the probability of having an erasure in the source letter, i.e., ε. The probability of having a 1 in $S \oplus \hat{S}$ is $d(1-\varepsilon)$, while the probability of having a 0 is $(1-d)(1-\varepsilon)$. Thus, $H(S \oplus \hat{S}) = h(\varepsilon) + (1-\varepsilon)h(d)$, and

$$I(S;\hat{S}) \geq (1-\varepsilon)(1-h(d)). \tag{II.32}$$

We got our final bound by observing that $\mathbb{E}[d(S,\hat{S})] = (1-\varepsilon)d \leq D$, and thus $I(S;\hat{S}) \geq R_{\varepsilon}^{\text{sh}}(D)$, since the binary entropy function is an increasing function on $[0, 1/2]$.

We will now show that this lower bound is indeed the rate-distortion function by finding an input distribution which achieves it, and which satisfies the distortion constraint. Consider the test-channel shown in Figure II.12 with $\varepsilon, d_0, p, d_1 \in [0,1]$.



$\left(\frac{1-\varepsilon}{2}\right)$ 0
$1-d_0$
0
d_0
p
S
(ε) ⋆
$\hat{S}$
$1-p$
d_1
$\left(\frac{1-\varepsilon}{2}\right)$ 1
1
$1-d_1$

Figure II.12: *Test channel for the BES.*

Note that the test channel shown in Figure II.12 can be fully characterized by the tuple $(\varepsilon, d_0, p, d_1)$. Consider now 3 channels which produce an average mutual information $I_i(S;\hat{S})$, $i \in \{1,2,3\}$ respectively. Furthermore, assume that the first two channels are identical with parameters $(\varepsilon, d_0, p, d_1)$, while the third one has $(\varepsilon, \frac{d_0+d_1}{2}, p, \frac{d_0+d_1}{2})$ for parameters. Since the average mutual information is a convex function of the conditional distribution [25], by applying Jensen's inequality we have

$$I_1(S;\hat{S}) = \frac{I_1(S;\hat{S}) + I_2(S;\hat{S})}{2} \geq I_3(S;\hat{S}).$$

Thus, if $d_0 \neq d_1$, it is possible to achieve the rate-distortion function by considering the same channel but with same crossover probabilities $\frac{d_0+d_1}{2}$.

Consequently it is sufficient to consider the test-channel described in Figure II.12 with parameters (ε, d, p, d), $d \in [0,1]$. For this channel, the average mutual information is

$$I(S;\hat{S}) = \underbrace{h\left(\varepsilon\left(\frac{1}{2} - p\right) + \frac{1}{2}\right)}_{H(\hat{S})} - \underbrace{[(1-\varepsilon)h(d) + \varepsilon h(p)]}_{H(\hat{S}|S)}. \tag{II.33}$$

Minimizing (II.33) over p, we find that the minimum is achieved at $p = \frac{1}{2}$ and that the minimal value of mutual information is $I(S;\hat{S}) \mid_{p=\frac{1}{2}} = (1-\varepsilon)[1-h(d)]$. We get (II.30) by observing that in this setup the average distortion is $\mathbb{E}\left[d(S,\hat{S})\right] = (1-\varepsilon)d$. □

III

Lower Bounds on the Rate-Distortion Performance of LDGM Codes

In this chapter, we focus on lossy compression of a BES using LDGM codes. Using two different arguments, we derive two lower bounds on their rate-distortion performance, and then show that both bounds are equal. More specifically, in Section III.1, we describe lossy source coding of a BES using LDGM codes, and then expose some key ideas behind the derivation of our bounds. Thereafter, we derive two lower bounds on the rate-distortion performance of binary LDGM codes for the BES. The first one is obtained in Section III.2 using a counting method, while the second one is obtained in Section III.3 using a test-channel method, which uses a probabilistic argument. Both methods have been introduced by Kudekar and Urbanke in [20] for the BSS, and we generalize them to the BES case. In the same paper, the authors observed for several numerical examples that both lower bounds seem to be equal for the BSS case. In Section III.4, we formally prove the equivalence of both methods for the BES case, and as a consequence also for the BSS. Note that these lower bounds are valid for any LDGM code with a given rate and generator node degree distribution, and any encoding function. Finally, an open question is addressed in Section III.5.

III.1 Preliminaries

For this chapter, we consider the BES(ε) introduced in Example II.1.2, with source alphabet $\mathscr{S} = \{0, 1, \star\}$. We focus on source sequences of length n, and we denote the set of source sequences of length n by $\mathscr{S}^n$, where $n \in \mathbb{N}$. Let $S^n = \{S_1, \cdots, S_n\}$, $S^n \in \mathscr{S}^n$ be a random source string. In a similar way, we define $\hat{\mathscr{S}}^n$ to be the set of reconstruction sequences of length n, and $\hat{S}^n$ is a reconstruction sequence.

III.1.1 LDGM Codes as Lossy Compressors

Consider a binary LDGM code of rate R, $R \in [0, 1]$, defined by a generator matrix $\mathbf{G}$. The process of lossy compression of a source sequence $s^n \in \mathscr{S}^n$ is explained below and illustrated in Example III.1.1.

1. The encoder f maps the source sequence s^n to one of the 2^{nR} index words $w^{nR} \in \mathscr{W}^{nR} \triangleq \mathbb{F}_2^{nR}$,

where

$$f: \begin{array}{ccc} \mathscr{S}^n & \rightarrow & \mathscr{W}^{nR} \\ s^n & \mapsto & w^{nR}. \end{array} \tag{III.1}$$

2. The index word w^{nR} is transmitted to the source decoder without any noise.
3. The decoder g maps the index word w^{nR} to a reconstruction sequence $\hat{s}^n \in \hat{\mathscr{S}}^n$, where

$$g: \begin{array}{ccc} \mathscr{W}^{nR} & \rightarrow & \hat{\mathscr{S}}^n \\ w^{nR} & \mapsto & \hat{s}^n \triangleq w^{nR}\mathbf{G}. \end{array} \tag{III.2}$$

Since we consider only binary codes, the index alphabet $\mathscr{W}$ and the reconstruction alphabet $\hat{\mathscr{S}}$ are both binary. Moreover, the set of reconstruction sequences of length n, corresponds to the codebook $\left\{w^{nR}\mathbf{G}: \forall w^{nR} \in \mathscr{W}^{nR}\right\}$.

The distortion $d(s, \hat{s})$ between the source letter s and the reconstructed letter $\hat{s}$ is given by the generalized Hamming distortion measure for the BES defined by (II.23). The distortion between a source sequence s^n and a reconstruction sequence $\hat{s}^n$ is given by the average of the single-letter distortion between the letters of both sequences, i.e.,

$$d(s^n, \hat{s}^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} d(s_i, \hat{s}_i). \tag{III.3}$$

This distortion measure will also be called a normalized distortion measure, since it is normalized with respect to the blocklength n. The average normalized distortion is $\mathbb{E}[d(S^n, g(f(S^n)))]$, where the average is over all the source sequence realizations

$$\mathbb{E}[d(S^n, g(f(S^n)))] \triangleq \sum_{s^n \in \mathscr{S}^n} \mathbb{P}\{S^n = s^n\}\, d(s^n, g(f(s^n))). \tag{III.4}$$

Example III.1.1 - *Lossy compression of a source sequence $s^n \in \mathscr{S}^n$*

For this example, assume $n = 6$, and the sequence output by the source is $s^6 = (0{\star}0111)$. Furthermore consider a naive encoder f, which maps a source sequence of length six to its first four non erased digits (if there are not enough non erased bits, the remaining positions are filled with zeroes). Thus, using this simple encoder we have

$$f((0{\star}0111)) = (0011)$$

Now consider a LDGM code with rate $R = \frac{2}{3}$ and generator matrix $\mathbf{G}$, where

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Then the reconstructed sequence $\hat{s}^6$ output by the decoder g is

$$\hat{s}^6 = g((0011)) = f((0{\star}0111))\mathbf{G} = (011011)$$

For this specific source sequence we achieve a distortion of $\frac{1}{3}$.

In the next subsection, we will explain the main ideas behind the counting method and the test-channel method.

III.1.2 Main ideas

The basic idea of the counting or the test-channel method is to bound the number of source sequences which have at least one codeword within a normalized distortion less or equal to D, $D \in [0, \frac{1-\varepsilon}{2}]$.

To do so, we need to introduce a few notations which will be useful for this chapter. For a source sequence s^n, let $H_E\left(s^n\right)$ be the number of erasures in s^n

$$\begin{array}{rccc} H_E : & \mathscr{S}^n & \rightarrow & \{0, 1, \cdots, n\} \\ & s^n & \mapsto & \sum_{i=1}^{n} \delta_\star\left(s_i\right), \end{array} \tag{III.5}$$

where $\delta_\star\left(s_i\right) = 1$ if and only if $s_i = \star$, and 0 otherwise.

In the rest of this chapter, b will denote the number of erasures in a source sequence. Hence, $b \in \{0, 1, \cdots, n\}$. We now do a partition of our set of source sequences of length n according to their number of erasures. Let $\mathscr{S}_b^n$ be the set of source sequences which have b erasures

$$\forall b \in \{0, 1, \cdots, n\}, \; \mathscr{S}_b^n \triangleq \{s^n \in \mathscr{S}^n : \; H_E\left(s^n\right) = b\}. \tag{III.6}$$

For each codeword $\hat{s}^n \in \hat{\mathscr{S}}^n$, we can consider the ball $\mathscr{B}_b^n(\hat{s}^n, D)$ centered around $\hat{s}^n$ and of radius D, which contains all the source sequences of length n with b erasures and which are within a normalized distortion less or equal to D

$$\forall \hat{s}^n \in \hat{\mathscr{S}}^n, \; \mathscr{B}_b^n(\hat{s}^n, D) \triangleq \{s^n \in \mathscr{S}_b^n : \; d\left(s^n, \hat{s}^n\right) \leq D\}. \tag{III.7}$$

Let $\mathscr{C}_b^n(D)$ be the set corresponding to the union of all these balls $\mathscr{B}_b^n(\hat{s}^n, D)$ over the set of codewords, i.e.,

$$\mathscr{C}_b^n(D) \triangleq \bigcup_{\hat{s}^n \in \hat{\mathscr{S}}^n} \mathscr{B}_b^n(\hat{s}^n, D). \tag{III.8}$$

$\mathscr{C}_b^n(D)$ contains the set of source sequences with b erasures and which have at least one codeword within normalized distortion D.

The main idea behind the counting and test-channel method is to upper bound $\left|\mathscr{C}_b^n(D)\right|$, the cardinality of $\mathscr{C}_b^n(D)$, for the most probable case which is $b = n\varepsilon$. To do so, the counting method uses a counting argument, while the test-channel method uses a probabilistic argument. The reason why the cardinality of this set is so important will be clear in Lemma III.1.2, presented in the next subsection.

Before going any further, some simplifications are necessary. As explained in [20], it is sufficient to prove our results

- for regular LDGM codes only;
- in the limit of infinite blocklength.

Indeed, we will assume that the LDGM code considered has generator node degree distribution $L(x) = x^l, l \in \mathbb{N}$. For the general case, it suffices to replace l by the average generator node degree L' of the code.

Another important simplification, is that it is sufficient to prove our results in the limit of infinite blocklength. This is because, from any code with a given blocklength n, we are able to construct another code with blocklength tending to infinity, and which has exactly the same characteristics in terms of rate, achievable distortion, and generator node degree distribution. A

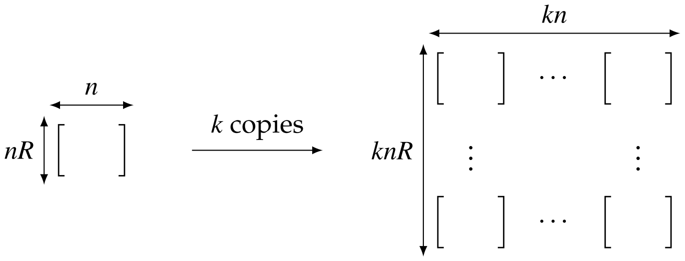


Figure III.1: *Construction of an arbitrary large code with same rate, achievable distortion, and generator node degree distribution.*

way of doing this, is to take k copies of the original code in a two dimensional manner, as shown in the schematic Figure III.1.

It is noteworthy that due to the Asymptotic Equipartition Property (AEP) (see [25, Chapter 3]) we have, $\lim_{n\to\infty} \mathbb{P}\left\{\frac{H_E(S^n)}{n} = \varepsilon\right\} = 1$. As a consequence, and with a slight abuse of notation, we will treat $n\varepsilon$ as an integer, although this is true only in the limit of blocklength tending to infinity.

III.1.3 Lower Bound on the Average Distortion

The main idea behind the counting or the test-channel method is to find an upper bound on $|\mathscr{C}_b^n(D)|$, the cardinality of $\mathscr{C}_b^n(D)$. We show that we are only interested in the cardinality of $\mathscr{C}_b^n(D)$ for $b = n\varepsilon$. Before doing that, we give an asymptotic approximation for binomial coefficient in the next lemma.

Lemma III.1.1 - *Asymptotic approximation for binomial coefficients*:
For $\varepsilon \in [0,1]$ and large n, the binomial coefficient $\binom{n}{\varepsilon n}$ can be approximated as

$$\binom{n}{\varepsilon n} = \frac{2^{nh(\varepsilon)}}{\sqrt{2\pi n\varepsilon(1-\varepsilon)}}(1+o(1)).$$

Further, for $\Delta \in \left[-\sqrt{n}(\log n)^{\frac{1}{3}}, \sqrt{n}(\log n)^{\frac{1}{3}}\right]$

$$\binom{n}{\varepsilon n+\Delta} = \binom{n}{\varepsilon n}\left(\frac{\varepsilon}{1-\varepsilon}\right)^{-\Delta} e^{-\frac{\Delta^2}{2n\varepsilon(1-\varepsilon)}}(1+o(1)).$$

Proof. The proof is based on the saddle point approximation of coefficients of powers of polynomials. A detailed explanation of which can be found in [26, Appendix D]. The approximations for binomial coefficients can be found in [26, p. 513]. □

In the next lemma, we derive a lower bound on the average distortion in terms of $|\mathscr{C}_{\varepsilon n}^n(D)|$. This is a key lemma for both the counting and test-channel methods, since it provides a link between the cardinality of $\mathscr{C}_{\varepsilon n}^n(D)$ and the expected normalized distortion.

Lemma III.1.2 - *Lower bound on the average distortion*:
Consider lossy compression of the BES(ε) using a LDGM code **G** of rate R with encoding map

$f : s^n \mapsto w^{nR}$, where $s^n \in \mathscr{S}^n$ and w^{nR} is an index word of length nR. Let $\hat{s}^n$ be the reconstruction associated to the word w^{nR} with the decoding map g given by $\hat{s}^n = w^{nR}\mathbf{G}$. If

$$\lim_{n\to\infty} \frac{1}{n} \log \left(\varepsilon^{\varepsilon n} \left(\frac{1-\varepsilon}{2} \right)^{n-\varepsilon n} |\mathscr{C}^n_{\varepsilon n}(D)| \right) < 0, \tag{III.9}$$

then

$$\mathbb{E}\left[d(S^n, g(f(S^n)))\right] \geq D(1+o(1)). \tag{III.10}$$

Proof. We observe that, $\forall s^n \in \mathscr{S}^n_b$,

$$d\left(s^n, g\left(f\left(s^n\right)\right)\right) \geq \begin{cases} 0, & \text{if } s^n \in \mathscr{C}^n_b(D), \\ D, & \text{if } s^n \notin \mathscr{C}^n_b(D). \end{cases} \tag{III.11}$$

and $|\mathscr{S}^n_b| = \binom{n}{b} 2^{n-b}$. Defining $\delta(n) = \sqrt{n}(\log n)^{1/3}$ and using the above arguments, we obtain

$$\begin{aligned}
\mathbb{E}\left[d(S^n, g(f(S^n)))\right] &= \sum_{b=0}^{n} \mathbb{P}\left\{S^n \in \mathscr{S}^n_b\right\} \mathbb{E}\left[d(S^n, g(f(S^n))) | S^n \in \mathscr{S}^n_b\right], \\
&\geq \sum_{b=\varepsilon n - \delta(n)}^{\varepsilon n + \delta(n)} \mathbb{P}\left\{S^n \in \mathscr{S}^n_b\right\} \mathbb{E}\left[d(S^n, g(f(S^n))) | S^n \in \mathscr{S}^n_b\right], \\
&\geq \sum_{b=\varepsilon n - \delta(n)}^{\varepsilon n + \delta(n)} \mathbb{P}\left\{S^n \in \mathscr{S}^n_b\right\} \sum_{s^n \in \mathscr{S}^n_b} \mathbb{P}\left\{S^n = s^n | S^n \in \mathscr{S}^n_b\right\} d\left(s^n, g(f(s^n))\right), \\
&\geq \sum_{b=\varepsilon n - \delta(n)}^{\varepsilon n + \delta(n)} \binom{n}{b} \varepsilon^b (1-\varepsilon)^{n-b} \sum_{s^n \in \mathscr{S}^n_b \setminus \mathscr{C}^n_b(D)} \frac{D}{\binom{n}{b} 2^{n-b}}, \\
&\geq \underbrace{D \sum_{b=\varepsilon n - \delta(n)}^{\varepsilon n + \delta(n)} \binom{n}{b} \varepsilon^b (1-\varepsilon)^{n-b}}_{A} \\
&\quad \underbrace{-D \sum_{b=\varepsilon n - \delta(n)}^{\varepsilon n + \delta(n)} \varepsilon^b \left(\frac{1-\varepsilon}{2}\right)^{n-b} |\mathscr{C}^n_b(D)|}_{B}.
\end{aligned} \tag{III.12}$$

As $|\mathscr{C}^n_b(D)| = e^{o(n)} \mathscr{C}^n_{\varepsilon n}(D)$ for $b \in [\varepsilon n - a(n), \varepsilon n + a(n)]$, summation B is of the order of $o(1)$ due to $(III.9)$. We use Stirling's formula and Laplace's method of summation [29, p. 755] to show that

$A = 1 + o(1)$.

$$
\begin{aligned}
A &= \sum_{b=\varepsilon n-\delta(n)}^{\varepsilon n+\delta(n)} \binom{n}{b} \varepsilon^{b}(1-\varepsilon)^{n-b}, \\
&\overset{(i)}{=} \sum_{\Delta=-\delta(n)}^{\Delta=\delta(n)} \binom{n}{\varepsilon n+\Delta} \varepsilon^{\varepsilon n+\Delta}(1-\varepsilon)^{n-\varepsilon n-\Delta}, \\
&\overset{(ii)}{=} \sum_{\Delta=-\delta(n)}^{\Delta=\delta(n)} \frac{e^{-\frac{\Delta^2}{2n\varepsilon(1-\varepsilon)}}}{\sqrt{2\pi n\varepsilon(1-\varepsilon)}}(1+o(1)), \\
&\overset{(iii)}{=} \frac{1}{\sqrt{2\pi n\varepsilon(1-\varepsilon)}} \int_{x=-\infty}^{\infty} e^{-\frac{x^2}{2n\varepsilon(1-\varepsilon)}}(1+o(1)), \\
&= 1+o(1).
\end{aligned}
$$

In step (i), we do a change of variable $b = \varepsilon n + \Delta$. Step (ii) follows from Lemma III.1.1, where we use the approximation of $\binom{n}{n\varepsilon}$ and the approximation of $\binom{n}{n\varepsilon+\Delta}$ in terms of $\binom{n}{n\varepsilon}$. In step (iii), we replace the sum by an integral, as the "variance" of summation terms, which is equal to $n\varepsilon(1-\varepsilon)$, tends to infinity as n tends to infinity [29, p. 761]. This completes the proof of the lemma. □

III.2 Bound Using the Counting Method

The counting method, which uses a counting argument to derive a lower bound on the rate-distortion performance of LDGM codes is extended to the BES case. In the next theorem, we state the rate-distortion bound using counting.

Theorem III. 2.1 - *Bound via Counting*:
Consider lossy compression of a BES(ε) using a LDGM code of blocklength n with generator node degree distribution $L(x)$. Let $\hat{\mathscr{S}}^n$ be the set of codewords of the LDGM code. Further let

$$f(x) = \prod_{i=0}^{d}(1+x^i)^{L_i}, \quad a(x) = \sum_{i=0}^{d} iL_i \frac{x^i}{1+x^i}, \tag{III.13}$$

$$R(x) = (1-\varepsilon)\frac{\left(1-h\left(\frac{x}{1+x}\right)\right)}{1-\log_2\left(\frac{f(x)}{x^{a(x)}}\right)}, \tag{III.14}$$

$$D(x) = \frac{x(1-\varepsilon)}{1+x} - R(x)a(x). \tag{III.15}$$

Then, for any blocklength n, the achievable rate-distortion performance of a LDGM code of rate R and generator node degree distribution $L(x)$ is lower bounded by the parametric curve $(D(x), R(x))$, $x \in [0,1]$.

Proof. As was argued in [20], we only need to prove our bound for the limit of blocklength tending to infinity. We use Lemma III.1.2 to prove the theorem. Thus, we need to bound the cardinality of $\mathscr{C}^n_{\varepsilon n}(D)$. Pick $w \in [0,1]$ such that $wn \in \mathbb{N}$, and $D + w \leq \frac{1-\varepsilon}{2}$. In this section, $A_n(nw)$

denotes the number of codewords of Hamming weight at most nw. In order to obtain the next equation consider two reconstruction sequences $\hat{s}_1^n$, $\hat{s}_2^n \in \hat{\mathscr{S}}^n$. Then, remark that a ball $\mathscr{B}_{\varepsilon n}^n(\hat{s}_1^n, D+w)$ contains every smaller ball $\mathscr{B}_{\varepsilon n}^n(\hat{s}_2^n, D)$ if and only if $d(\hat{s}_1^n, \hat{s}_2^n) \leq w$. By linearity of the code, it follows that each ball $\mathscr{B}_{\varepsilon n}^n(\hat{s}_2^n, D)$ is in exactly $A_n(nw)$ bigger balls $\mathscr{B}_{\varepsilon n}^n(\hat{s}_1^n, D+w)$ [20]. Thus,

$$|\mathscr{C}_{\varepsilon n}^n(D)| = |\bigcup_{\hat{s}^n \in \hat{\mathscr{S}}^n} \mathscr{B}_{\varepsilon n}^n(\hat{s}^n, D)| \leq \frac{1}{A_n(nw)} \sum_{\hat{s}^n \in \hat{\mathscr{S}}^n} |\mathscr{B}_{\varepsilon n}^n(\hat{s}^n, D+w)|.$$

Then by using

$$|\mathscr{B}_{\varepsilon n}^n(\hat{s}^n, D+w)| = \binom{n}{\varepsilon n} \sum_{i=0}^{n(D+w)} \binom{n-\varepsilon n}{n-\varepsilon n-i}, \tag{III.16}$$

Stirling's approximation, and the observation that the entropy function $h(x)$ is increasing for $x \in [0,1/2]$ and decreasing for $x \in [1/2,1]$, we obtain

$$|\mathscr{B}_{\varepsilon n}^n(\hat{s}^n, D+w)| \leq 2^{nh(\varepsilon)} 2^{(n-\varepsilon n)h\left(\frac{D+w}{1-\varepsilon}\right)+o(n-\varepsilon n)}.$$

As was shown in [20], $A_n(nw) \geq \sum_{i=0}^{nw} \text{coef}(f(x)^{nR}, x^i)$, where $\text{coef}(f(x)^{nR}, x^i)$ denotes the coefficient in front of x^i in the expansion of $f(x)^{nR}$. From Theorem 1 in [30],

$$\text{coef}(f(x)^{nR}, x^w) \leq \inf_{x>0} \frac{f(x)^{nR}}{x^{nw}} = \frac{f(x_\omega)^{nR}}{x_\omega^{nw}} \leq A_n(nw),$$

where x_ω is the unique positive solution to the equation $a(x) = \omega$ and we define $\omega = w/R$. As $D+w \leq \frac{1-\varepsilon}{2}$, the maximum of the summation term in (III.16) occurs at $i = n(D+w)$. Thus, we obtain

$$|\mathscr{C}_{\varepsilon n}^n(D)| \leq 2^{n\left[-R\log_2 \frac{f(x_\omega)}{x_\omega^{a(x_\omega)}} + R + h(\varepsilon) + (1-\varepsilon)h\left(\frac{D+Ra(x_\omega)}{1-\varepsilon}\right)\right] + o(n-\varepsilon n)}, \tag{III.17}$$

where the relation $a(x_\omega) = \omega$ was used. From Lemma III.1.2 we know that if $\varepsilon^{\varepsilon n}\left(\frac{1-\varepsilon}{2}\right)^{n-\varepsilon n} |\mathscr{C}_{\varepsilon n}^n(D)|$ is exponentially small in comparison to n, then the average distortion is at least D. We now upper bound the growth rate $\varepsilon^{\varepsilon n}\left(\frac{1-\varepsilon}{2}\right)^{n-\varepsilon n} |\mathscr{C}_{\varepsilon n}^n(D)|$ using (III.17), and we obtain

$$\lim_{n\to\infty} \left[\frac{1}{n} \log_2 \left(\varepsilon^{n\varepsilon} \left(\frac{1-\varepsilon}{2}\right)^{n(1-\varepsilon)} |\mathscr{C}_{\varepsilon n}^n(D)|\right)\right] \leq g(D,R),$$

where

$$g(D,R) = \inf_{\substack{D+a(x)R \leq \frac{1-\varepsilon}{2} \\ x \geq 0}} h_1(x), \tag{III.18}$$

with

$$h_1(x) = -R\log_2 \frac{f(x)}{x^{a(x)}} + R + (1-\varepsilon)\left(h\left(\frac{D+Ra(x)}{1-\varepsilon}\right) - 1\right).$$

If $g(D,R) < 0$ the average normalized distortion is lower bounded by D, and thus, we obtain the rate-distortion bound by considering the condition

$$g(D,R) = 0.$$

In order to compute the infimum in (III.18), we take the derivative of $h_1(x)$ and obtain

$$\frac{dh_1(\beta, x)}{dx} = Ra'(x)\log_2\left[x\left(\frac{1}{\alpha}-1\right)\right],$$

where $\alpha = \frac{D+Ra(x)}{1-\varepsilon}$. The vanishing derivative conditions imply that

$$\alpha = \frac{x}{1+x}. \tag{III.19}$$

From (III.19), we see that if $x \leq 1$, then $D + Ra(x) \leq (1-\varepsilon)/2$. By using (III.19), the condition $g(D,R) = 0$, and varying x over the interval $x \in [0,1]$, we obtain the desired parametric lower bound on the rate-distortion performance. □

III.2.1 Bound for Low Rates

Based on the arguments of [20], the parametric bound can be improved for low rates using the straight-line bound. Indeed, since there are nR generator nodes of degree l (replace l by L' for the general case), there are nRl edges in the corresponding Tanner graph. Thus, the number of parity-check nodes n must be less than nRl to avoid unconnected nodes. If $R \leq \frac{1}{l}$, then there are $n(1-Rl)$ nodes which are not connected in the graph and which achieve a distortion of $\frac{1-\varepsilon}{2}$, while the other connected nodes can only achieve the same best possible distortion for any rate below $\frac{1}{l}$ [20]. More precisely, let $L' = L'(1)$ and if $R \in \left[0, \frac{1}{L'(1)}\right]$, then

$$D = \frac{1}{2}\left(1 - RL'\left(1 - 2\left(\frac{x\left(\frac{1}{L'}\right)(1-\varepsilon)}{1+x\left(\frac{1}{L'}\right)} - \frac{a\left(x\left(\frac{1}{L'}\right)\right)}{L'}\right)\right)\right), \tag{III.20}$$

where $x\left(\frac{1}{L'}\right)$ is the unique solution of $R(x) = \frac{1}{L'}$, and $R(x)$ is defined in (III.14).

The complete lower bound is plotted in Figure III.2 for regular LDGM codes with different generator nodes degree.

As we can see on Figure III.2, the bound is really close to the Shannon rate distortion function for high rates but then moves away from it for lower rates. The existence of this gap shows that LDGM codes with finite degrees are not optimal for lossy compression of a BES. Moreover, there might exist a tighter bound for higher rates.

In the next section, we derive the rate-distortion bound by using the test-channel-based probabilistic arguments developed in [20].

III.3 Bound Using the Test Channel Method

The test channel method, which uses a probabilistic argument to derive an other lower bound on the rate-distortion performance of LDGM codes is extended to the BES case.

In this section we use a probabilistic argument to bound the cardinality of $\mathscr{C}_b^n(D)$. We consider the test channel model represented in Figure III.3 with the binary error/erasure channel (BEEC) represented in Figure III.4. More precisely, choose an index word $w^{nR} \in \mathscr{W}^{nR}$ uniformly at random. This generates the corresponding codeword $\hat{S}^n$, which is then sent component wise through the BEEC defined in Figure III.4.

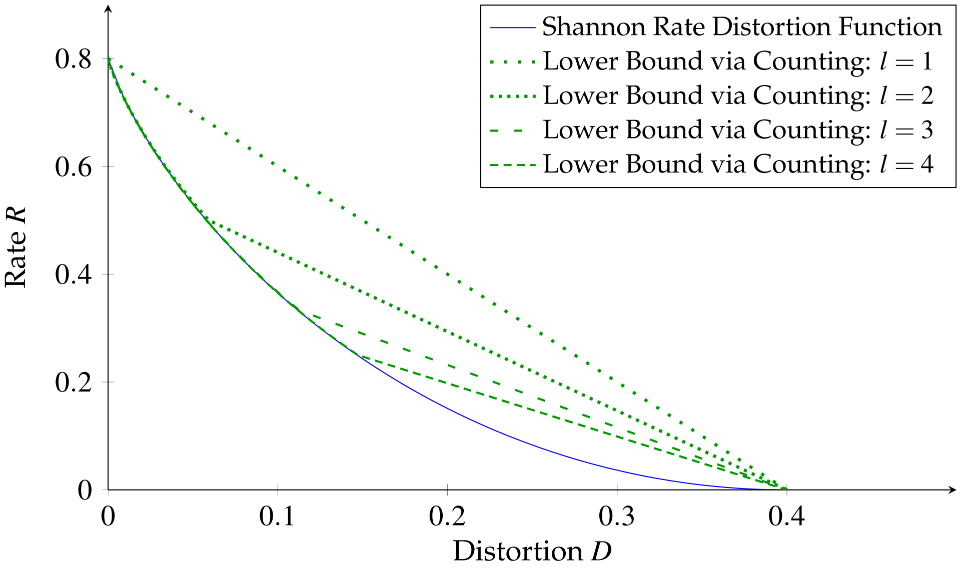


Figure III.2: *Rate-distortion performance of generator regular LDGM codes for lossy compression of a BES(ε) with $\varepsilon = 0.2$.*

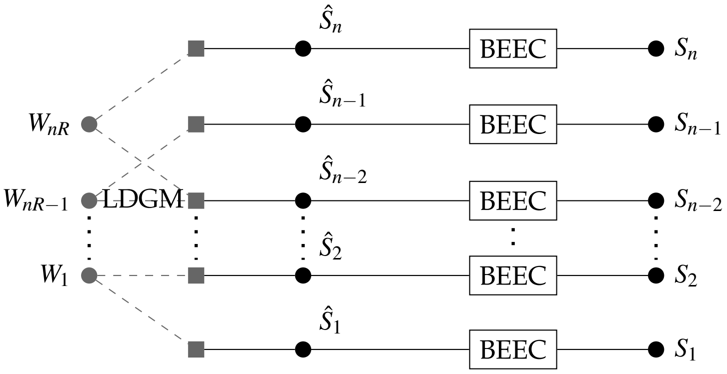


Figure III.3: *Backward test channel. The index word W^{nR} is chosen uniformly at random, then multiplied by the corresponding generator matrix of the LDGM code considered, generating the reconstruction sequence $\hat{S}^n$. Finally each component $\hat{S}_i$ is sent through a binary error/erasure channel (BEEC), $i \in \{1, \cdots, n\}$.*

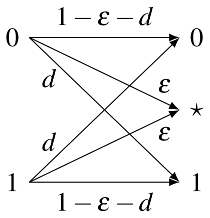


Figure III.4: *The binary error/erasure channel.*

We prove the rate-distortion bound in the following theorem. We only consider the regular case for the sake of simplicity of exposition. As was argued in [20], it suffices to prove the bound only for the limit of blocklength tending to infinity. Also, the obtained bound can be strengthened for low rates using the straight-line bound defined in (III.20).

Theorem III. 3.1 - *Bound via Test Channel*:
Consider lossy compression of a BES(ε) using a LDGM code with codeword set $\hat{\mathscr{S}}^n$, regular generator node degree distribution $L(x) = x^l$, blocklength n, and rate R. Then for an average normalized distortion D, the rate R is lower bounded by,

$$R \geq \sup_{D \leq d \leq \frac{1-\varepsilon}{2}} \frac{(1-\varepsilon)(1-\log_2(1-\varepsilon)) + (1-D-\varepsilon)\log_2(1-\varepsilon-d) + D\log_2(d)}{1-\log_2\left(1+\left(\frac{d}{1-\varepsilon-d}\right)^l\right)}. \tag{III.21}$$

Proof. We will use Lemma III.1.2, so we will focus on source words with a number of erasures equal to εn. In this section, $A_n(nw)$ denotes the number of codewords of Hamming weight exactly nw. Consider a particular source sequence s_0^n which lies in $\mathscr{C}_{\varepsilon n}^n(D)$, $s_0^n \in \mathscr{C}_{\varepsilon n}^n(D)$. This implies that $\exists \hat{s}_0^n \in \hat{\mathscr{S}}^n$ such that $d(s_0^n, \hat{s}_0^n) \leq D$. Let $c = \frac{d}{1-\varepsilon-d}$, and $D \leq d \leq (1-\varepsilon)/2$. Then,

$$\begin{aligned}
\mathbb{P}\{S^n = s_0^n\} &= \sum_{\hat{s}^n \in \hat{\mathscr{S}}^n} \mathbb{P}\left\{S^n = s_0^n, \hat{S}^n = \hat{s}^n\right\}, \\
&= \sum_{\substack{w \in [0,1]: \\ nw \in \mathbb{N}}} \sum_{\substack{\hat{s}^n \in \hat{\mathscr{S}}^n: \\ d(\hat{s}_0^n, \hat{s}^n) = w}} \mathbb{P}\left\{S^n = s_0^n, \hat{S}^n = \hat{s}^n\right\}, \\
&= 2^{-nR} \varepsilon^{\varepsilon n} (1-\varepsilon-d)^{n-\varepsilon n} \sum_{\substack{w \in [0,1]: \\ nw \in \mathbb{N}}} \sum_{\substack{\hat{s}^n \in \hat{\mathscr{S}}^n: \\ d(\hat{s}_0^n, \hat{s}^n) = w}} c^{nd(s_0^n, \hat{s}^n)}, \\
&\overset{\text{(i)}}{\geq} 2^{-nR} \varepsilon^{\varepsilon n} (1-\varepsilon-d)^{n-\varepsilon n} \sum_{\substack{w \in [0,1]: \\ nw \in \mathbb{N}}} \sum_{\substack{\hat{s}^n \in \hat{\mathscr{S}}^n: \\ d(\hat{s}_0^n, \hat{s}^n) = w}} c^{n(d(s_0^n, \hat{s}_0^n) + d(\hat{s}_0^n, \hat{s}^n))}, \\
&\overset{\text{(ii)}}{\geq} 2^{-nR} \varepsilon^{\varepsilon n} (1-\varepsilon-d)^{n-\varepsilon n} \sum_{\substack{w \in [0,1]: \\ nw \in \mathbb{N}}} \sum_{\substack{\hat{s}^n \in \hat{\mathscr{S}}^n: \\ d(\hat{s}_0^n, \hat{s}^n) = w}} c^{n(D+w)}, \\
&\geq 2^{-nR} e^{\varepsilon n} (1-\varepsilon-d)^{n-\varepsilon n} \sum_{\substack{w \in [0,1]: \\ nw \in \mathbb{N}}} A_n(nw) c^{n(D+w)},
\end{aligned}$$

where step (i) follows since $d \leq \frac{1-\varepsilon}{2}$, and step (ii) follows since $d(s_0^n, \hat{s}_0^n) \leq D$ and $d(\hat{s}_0^n, \hat{s}^n) = w$.

Assuming $\frac{Rc^l}{1+c^l} < \frac{1}{l}$, we obtain the following inequality (see [20] for proof),

$$\sum_{\substack{w \in [0,1]: \\ nw \in \mathbb{N}}} A_n(nw) c^{nw} \geq \frac{1}{n}\left(1+c^l\right)^{nR}. \tag{III.22}$$

Thus for $s^n \in \mathscr{C}_{\varepsilon n}^n(D)$,

$$\mathbb{P}\{S^n = s^n\} \geq \frac{1}{n} 2^{-nR} \left(1+c^l\right)^{nR} \varepsilon^{\varepsilon n} (1-\varepsilon-d)^{n(1-D)-\varepsilon n} d^{Dn}. \tag{III.23}$$

Also,

$$\binom{n}{\varepsilon n} \varepsilon^{\varepsilon n} (1-\varepsilon)^{n-\varepsilon n} = \sum_{s^n \in \mathscr{S}_{\varepsilon n}^n} \mathbb{P}\{S^n = s^n\} \geq \sum_{s^n \in \mathscr{C}_{\varepsilon n}^n(D)} \mathbb{P}\{S^n = s^n\}. \tag{III.24}$$

We obtain the following upper bound on $|\mathscr{C}^n_{\varepsilon n}(D)|$ by combining (III.23) and (III.24)

$$|\mathscr{C}^n_{\varepsilon n}(D)| \leq n2^{nR}(1+c^l)^{-nR}(1-\varepsilon-d)^{-n(1-D)}(1-\varepsilon)^n d^{-Dn}\binom{n}{\varepsilon n}\left(\frac{1-\varepsilon-d}{1-\varepsilon}\right)^{\varepsilon n}. \tag{III.25}$$

By using Lemma III.1.2, we know that

$$\mathbb{E}\left[d(S,g(f(S)))\right] \geq D(1+o(1)),$$

if (III.9) is true, i.e.,

$$\lim_{n\to\infty}\frac{1}{n}\log\left(\varepsilon^{\varepsilon n}\left(\frac{1-\varepsilon}{2}\right)^{n-\varepsilon n}|\mathscr{C}^n_{\varepsilon n}(D)|\right) < 0.$$

We use the bound on $\mathscr{C}^n_{\varepsilon n}(D)$ from (III.25), plug it into the above condition, and obtain that if

$$R - R\log_2(1+c^l) + (1-\varepsilon)(-1+\log_2(1-\varepsilon)) - (1-D-\varepsilon)\log_2(1-\varepsilon-d) - D\log_2(d) < 0, \tag{III.26}$$

then the distortion is at least D. This proves (III.21).

We still need to prove that $\frac{Rc^l}{1+c^l} < \frac{1}{l}$, which is trivially fulfilled for $R < \frac{1}{l}$ since $c < 1$. Assuming $R > \frac{1}{l}$, we need to prove that $d < \frac{1-\varepsilon}{1+(Rl-1)^{\frac{1}{l}}}$. By taking the derivative of the left hand side (LHS) of (III.26) with respect to d, we find the expression for our optimal d to be

$$d = \frac{1-\varepsilon}{1+\left[\frac{Rl}{d-D}-1\right]^{\frac{1}{l}}}.$$

As $D \leq d \leq \frac{1-\varepsilon}{2}$, it implies $d - D < 1$. This in turn implies that the optimal d is less than $\frac{1-\varepsilon}{1+(Rl-1)^{\frac{1}{l}}}$. This gives us the desired result. Thus the inequality $\frac{Rc^l}{1+c^l} < \frac{1}{l}$ always holds and we get (III.22). □

Remark: Note that the lower bound on the rate in (III.21) is lower bounded by the Shannon rate distortion function. In order to see this, we put $d = D$ on the LHS of (III.21). This give,

$$(1-\varepsilon)(1-\log_2(1-\varepsilon)) + (1-D-\varepsilon)\log_2(1-\varepsilon-D) + D\log_2(D).$$

In the above expression, we multiply and divide D by $(1-\varepsilon)$. After simplification we obtain

$$(1-\varepsilon)\left(1+\left(1-\frac{D}{1-\varepsilon}\right)\log_2\left(1-\frac{D}{1-\varepsilon}\right)+\frac{D}{1-\varepsilon}\log_2\left(\frac{D}{1-\varepsilon}\right)\right) = (1-\varepsilon)\left(1-h\left(\frac{D}{1-\varepsilon}\right)\right),$$

which is the Shannon's rate distortion function for BES(ε).

III.4 Equivalence of Both Methods

For the BSS it was numerically observed in [20] that the bounds obtained by the counting method and the test channel method are identical. In the following theorem we prove equality of the two bounds for a BES(ε). As the BES(0) is a BSS, we also implicitly prove the equality of the two bounds for a BSS. Due to simplicity of exposition, we prove our result for regular LDGM codes.

Theorem III. 4.1 - *The Bounds Defined in Theorems III.2.1 and III.3.1 are Equal*:
Consider lossy compression of a BES(ε) using a regular LDGM code with codeword set $\hat{S}^n$, and generator node degree distribution $L(x) = x^l$. Then the lower bounds on the rate-distortion of $\hat{S}^n$ obtained in Theorem III.2.1 and Theorem III.3.1 are identical.

Proof. Considering (III.21), we define the function $v(d)$

$$v(d) = \frac{(1-\varepsilon)\,(1-\log_2(1-\varepsilon)) + (1-D-\varepsilon)\log_2(1-\varepsilon-d) + D\log_2(d)}{1-\log_2\left(1+\left(\frac{d}{1-\varepsilon-d}\right)^l\right)}.$$

By determining the maximum of $v(d)$ in parametric form, we show that the resulting expression is the same as the bound in Theorem III.2.1.

We first do a change of variable in $v(d)$ as $x = \frac{d}{1-\varepsilon-d}$ and write $v(d)$ as $v(x)$,

$$v(x) = \frac{(1-\varepsilon)\,(1-\log_2(1-\varepsilon)) + (1-D-\varepsilon)\log_2\left(\frac{1-\varepsilon}{1+x}\right) + D\log_2\left((1-\varepsilon)\frac{x}{1+x}\right)}{1-\log_2\left(1+x^l\right)}. \tag{III.27}$$

In order to compute the maximum, we take the derivative of $v(x)$ which is given by,

$$\frac{dv(x)}{dx} = \frac{1}{1-\log_2(1+x^l)} \times \left(\frac{D}{x} - \frac{1-\varepsilon}{1+x} + \frac{lx^{l-1}}{(1+x^l)(1-\log_2(1+x^l))}\left((1-\varepsilon)\,(1-\log_2(1+x)) + D\log_2(x)\right)\right).$$

Equating the derivative to zero and solving for D results in

$$D = (1-\varepsilon)\frac{x}{1+x} - a(x)(1-\varepsilon)\frac{\frac{x}{1+x}\log_2(x) + 1 - \log_2(1+x)}{1-\log_2\left(\frac{f(x)}{x^{a(x)}}\right)},$$

where

$$f(x) = 1+x^l, a(x) = \frac{lx^l}{1+x^l}.$$

Note that

$$1-h\left(\frac{x}{1+x}\right) = 1-\log_2(1+x) + \frac{x}{1+x}\log_2(x).$$

This results in

$$D = (1-\varepsilon)\frac{x}{1+x} - (1-\varepsilon)a(x)\frac{1-h\left(\frac{x}{1+x}\right)}{1-\log_2\left(\frac{f(x)}{x^{a(x)}}\right)} = (1-\varepsilon)\frac{x}{1+x} - a(x)R(x), \tag{III.28}$$

where $R(x)$ is the expression for the rate given in Theorem III.2.1. This gives the same expression for D as in Theorem III.2.1. Substituting the expression for D from (III.28) into (III.27), we obtain the expression for R which is the same as that of Theorem III.2.1. □

III.5 Open Question

We derived lower bounds on the rate-distortion performance of LDGM codes over the BES using two different methods and then showed that these latter lead to the same result. Although we considered regular LDGM codes for the sake of simplicity, the generalization to the irregular case is straightforward and the two methods we used are valid for any sparse-graph code.

An open question was asked in [20]: "What is the relationship between the test channel model and the rate-distortion problem?" It was further conjectured that a (R,D) pair is only achievable for the BSS if the entropy $H(s^n) = n$ in the test channel model. The equivalent conjecture for BES(ε) will be that an (R,D) pair is only achievable if $H(s^n) = n(1-\varepsilon) + nh(\varepsilon)$ in the test channel model. Based on this conjecture and Gallager's bound, it was conjectured that for the BSS a stronger rate-distortion bound is valid. The extension of this conjecture to BES(ε) would result in the following bound

$$R \geq \frac{1-\varepsilon+h(\varepsilon)-\left(\varepsilon\log_2\frac{1}{\varepsilon}+D\log_2\frac{1}{D}+(1-\varepsilon-D)\log_2\frac{1}{1-\varepsilon-D}\right)}{1-\sum_{f=0}^{l}\binom{l}{f}\varepsilon^f\sum_{w=0}^{l-f}\binom{l-f}{w}D^w(1-\varepsilon-D)^{l-f-w}h\left(\frac{1}{1+\left(\frac{D}{1-\varepsilon-D}\right)^{l-f-2w}}\right)}. \tag{III.29}$$

Since the details are a bit cumbersome, the way this bound is derived in appendix III.A.

III.A Details of the Conjecture

In this section, we expose how to get the lower bound (III.29).

First observe that

$$H(W^{nR}) = H(S^n) - H(S^n \mid W^{nR}) + H(W^{nR} \mid S^n) = nR. \tag{III.30}$$

In the following lemmas, we will calculate each term in (III.30) using the test-channel model defined in Figures III.3, III.4, and a generator regular LDGM code with generator node degree distribution $L(x) = x^l$, $l \in \mathbb{N}$.

Lemma III.1.1 - $H(S^n)$:

$$H(S^n) = n(1-\varepsilon) + nh(\varepsilon). \tag{III.31}$$

Proof.

$$\begin{aligned}
H(S^n) &= -\sum_{b=0}^{n} \sum_{s^n \in \mathscr{S}_b^n} \left(\frac{1-\varepsilon}{2}\right)^{n-b} \varepsilon^b \log_2 \left[\left(\frac{1-\varepsilon}{2}\right)^{n-b} \varepsilon^b\right], \\
&= -\sum_{b=0}^{n} \binom{n}{b} 2^{n-b} \left(\frac{1-\varepsilon}{2}\right)^{n-b} \varepsilon^b \log_2 \left[\left(\frac{1-\varepsilon}{2}\right)^{n-b} \varepsilon^b\right], \\
&= \sum_{b=0}^{n} \binom{n}{b} (1-\varepsilon)^{n-b} \varepsilon^b \left[n\left(1-\log_2(1-\varepsilon)\right) - b\left(1-\log_2(1-\varepsilon)+\log_2 \varepsilon\right)\right], \\
&= n\left(1-\log_2(1-\varepsilon)\right) - \left(1+\log_2 \frac{\varepsilon}{1-\varepsilon}\right) \sum_{b=0}^{n} \binom{n}{b} (1-\varepsilon)^{n-b} \varepsilon^b b, \\
&= n(1-\varepsilon) + nh(\varepsilon),
\end{aligned}$$

where the last step follows since $\sum_{b=0}^{n} \binom{n}{b} (1-\varepsilon)^{n-b} \varepsilon^b b = n\varepsilon$. □

Lemma III.1.2 - $H(S^n \mid W^{nR})$:

$$H(S^n \mid W^{nR}) = n\left[\varepsilon \log_2 \frac{1}{\varepsilon} + d \log_2 \frac{1}{d} + (1-\varepsilon-d) \log_2 \frac{1}{1-\varepsilon-d}\right]. \tag{III.32}$$

Proof. Since $\hat{s}^n = w^{nR}\mathbf{G}$, we have

$$\begin{aligned}
H(S^n \mid W^{nR}) &= H(S^n \mid \hat{S}^n), \\
&= -\sum_{\hat{s}^n \in \hat{\mathscr{S}}^n} 2^{-nR} \sum_{b=0}^{n} \sum_{s^n \in \mathscr{S}_b^n} \mathbb{P}\left\{S^n = s^n \mid \hat{S}^n = \hat{s}^n, s^n \in \mathscr{S}_b^n\right\} \times \\
&\qquad \log_2 \mathbb{P}\left\{S^n = s^n \mid \hat{S}^n = \hat{s}^n, s^n \in \mathscr{S}_b^n\right\}.
\end{aligned}$$

Since the BEEC is used, we have

$$\mathbb{P}\left\{S^n = s^n \mid \hat{S}^n = \hat{s}^n, s^n \in \mathscr{S}_b^n\right\} = \varepsilon^b d^{nd(s^n,\hat{s}^n)} (1-\varepsilon-d)^{n-b-nd(s^n,\hat{s}^n)}.$$

Let f be $nd(s^n, \hat{s}^n)$, $f \in \{0, 1, \cdots, n-b\}$. Note that $\left| s^n \in \mathscr{S}_b^n \text{ s.t. } nd(s^n, \hat{s}^n) = f \right| = \binom{n}{b}\binom{n-b}{f}$. Thus,

$$\begin{aligned} H(S^n \mid W^{nR}) &= \sum_{\hat{s}^n \in \hat{\mathscr{S}}^n} 2^{-nR} \sum_{b=0}^{n} \sum_{f=0}^{n-b} \sum_{\substack{s^n \in \mathscr{S}_b^n, \\ nd(s^n, \hat{s}^n) = f}} \varepsilon^b d^f (1-\varepsilon-d)^{n-b-f} \log_2 \frac{1}{\varepsilon^b d^f (1-\varepsilon-d)^{n-b-f}}, \\ &= \sum_{b=0}^{n} \binom{n}{b} \varepsilon^b \left(b \log_2 \frac{1}{\varepsilon} + (n-b) \log_2 \frac{1}{1-\varepsilon-d} \right) (1-\varepsilon)^{n-b} + \\ &\quad \log_2 \frac{1-\varepsilon-d}{d} \sum_{b=0}^{n} \binom{n}{b} \varepsilon^b \sum_{f=0}^{n-b} \binom{n-b}{f} d^f (1-\varepsilon-d)^{n-b-f} f, \\ &= n \left[\varepsilon \log_2 \frac{1}{\varepsilon} + d \log_2 \frac{1}{d} + (1-\varepsilon-d) \log_2 \frac{1}{1-\varepsilon-d} \right], \end{aligned} \tag{III.33}$$

where the last step follows since $\sum_{f=0}^{n-b} \binom{n-b}{f} d^f (1-\varepsilon-d)^{n-b-f} f = d(n-b)(1-\varepsilon)^{n-b-1}$. □

LEMMA III.1.3 - *Lower bound on $H(W^{nR} \mid S^n)$:*

$$H(W^{nR} \mid S^n) \geq nR \sum_{f=0}^{l} \binom{l}{f} \varepsilon^f \sum_{w=0}^{l-f} \binom{l-f}{w} d^w (1-\varepsilon-d)^{l-f-w} h\left(\frac{1}{1 + \frac{D}{1-\varepsilon-D}^{l-f-2w}} \right). \tag{III.34}$$

Proof. Recall that $W^{nR} = \{W_1, \cdots, W_{nR}\}$. Let $W_{\sim g}^{nR}$ denotes the set of generator nodes without the g^{th} one, i.e., $W_{\sim g}^{nR} \triangleq \{W_1, \cdots, W_{g-1}, W_{g+1}, \cdots, W_{nR}\}$. The subset of source bits connected to the g^{th} generator node W_g in the test-channel model (see Figure III.3) is denoted by $\mathbf{S}_g$, while the subset of reconstructed bits connected to W_g is denoted by $\hat{\mathbf{S}}_g$. Then,

$$H(W^{nR} \mid S^n) = \sum_{g=1}^{nR} H(W_g | S^n, W_1, \cdots, W_{g-1}) \geq \sum_{g=1}^{nR} H(W_g | S^n, W_{\sim g}^{nR}) = \sum_{g=1}^{nR} H(W_g | \mathbf{S}_g, W_{\sim g}^{nR}). \tag{III.35}$$

W_g is a binary random variable (chosen uniformly at random in our test channel model), thus $\mathbb{P}\left\{ W_g = 0 \mid \mathbf{S}_g, W_{\sim g}^{nR} \right\} = 1 - \mathbb{P}\left\{ W_g = 1 \mid \mathbf{S}_g, W_{\sim g}^{nR} \right\}$ and

$$H(W_g | \mathbf{S}_g, W_{\sim g}^{nR}) = \sum_{\mathbf{S}_g, W_{\sim g}^{nR}} \mathbb{P}\left\{ \mathbf{S}_g, W_{\sim g}^{nR} \right\} h\left(\mathbb{P}\left\{ W_g = 0 \mid \mathbf{S}_g, W_{\sim g}^{nR} \right\} \right). \tag{III.36}$$

We can calculate $\mathbb{P}\left\{ W_g = 0 \mid \mathbf{S}_g, W_{\sim g}^{nR} \right\}$ and $\mathbb{P}\left\{ \mathbf{S}_g, W_{\sim g}^{nR} \right\}$ using our test channel model. To do so, let $\hat{\mathbf{s}}_g^0$ be the values of the subset of reconstructed bits connected to W_g, when $W_g = 0$ and $W_{\sim g}^{nR} = w_{\sim g}^{nR}$. In the same manner we define $\hat{\mathbf{s}}_g^1$, when $W_g = 1$ and $W_{\sim g}^{nR} = w_{\sim g}^{nR}$. In addition, let n_g be the number of elements in $\mathbf{S}_g$ (or equivalently in $\hat{\mathbf{S}}_g$). Thus,

$$\begin{aligned}
\mathbb{P}\left\{W_g = 0 \mid \mathbf{S}_g, W_{\sim g}^{nR}\right\} &= \frac{\mathbb{P}\left\{W_g = 0, \mathbf{S}_g, W_{\sim g}^{nR}\right\}}{\sum_{i\in\{0,1\}} \mathbb{P}\left\{W_g = i, \mathbf{S}_g, W_{\sim g}^{nR}\right\}} = \frac{\mathbb{P}\left\{\mathbf{S}_g, \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^0\right\}}{\mathbb{P}\left\{\mathbf{S}_g, \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^0\right\} + \mathbb{P}\left\{\mathbf{S}_g, \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^1\right\}}, \\
&= \frac{\mathbb{P}\left\{\mathbf{S}_g \mid \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^0\right\}}{\mathbb{P}\left\{\mathbf{S}_g \mid \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^0\right\} + \mathbb{P}\left\{\mathbf{S}_g | \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^1\right\}}, \text{ since } \mathbb{P}\left\{W_g = 0\right\} = \mathbb{P}\left\{W_g = 1\right\}, \\
&= \frac{\varepsilon^{H_E(\mathbf{S}_g)} d^{n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^0)} (1-\varepsilon-d)^{l - H_E(\mathbf{S}_g) - n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^0)}}{\sum_{i\in\{0,1\}} \varepsilon^{H_E(\mathbf{S}_g)} d^{n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^i)} (1-\varepsilon-d)^{l - H_E(\mathbf{S}_g) - n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^i)}}, \\
&= \frac{1}{1 + \left(\frac{d}{1-\varepsilon-d}\right)^{l - H_E(\mathbf{S}_g) - 2 n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^0)}},
\end{aligned}$$

the last equality follows since $n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^1) = l - H_E(\mathbf{S}_g) - n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^0)$.

Let $\xi_{\varepsilon,d,l}\left(H_E(\mathbf{S}_g), n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^0)\right) \triangleq \frac{1}{1+\left(\frac{d}{1-\varepsilon-d}\right)^{l - H_E(\mathbf{S}_g) - 2 n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^0)}}$.

In the same way for $\mathbb{P}\left\{\mathbf{S}_g, W_{\sim g}^{nR}\right\}$ we have

$$\begin{aligned}
\mathbb{P}\left\{\mathbf{S}_g, W_{\sim g}^{nR}\right\} &= \mathbb{P}\left\{\mathbf{S}_g \mid W_g = 0, W_{\sim g}^{nR}\right\} \mathbb{P}\left\{W_g = 0, W_{\sim g}^{nR}\right\} + \mathbb{P}\left\{\mathbf{S}_g \mid W_g = 1, W_{\sim g}^{nR}\right\} \mathbb{P}\left\{W_g = 1, W_{\sim g}^{nR}\right\}, \\
&= 2^{-nR}\left(\mathbb{P}\left\{\mathbf{S}_g \mid \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^0\right\} + \mathbb{P}\left\{\mathbf{S}_g \mid \hat{\mathbf{S}}_g = \hat{\mathbf{s}}_g^1\right\}\right), \\
&= 2^{-nR}\left(\sum_{i\in\{0,1\}} \varepsilon^{H_E(\mathbf{S}_g)} d^{n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g^i)} (1-\varepsilon-d)^{l - H_E(\mathbf{S}_g) - n_g d(\mathbf{S}_g, \hat{\mathbf{s}}_g^i)}\right).
\end{aligned}$$

Let f be $H_E(\mathbf{S}_g)$, with $f \in \{0, 1, \cdots, l\}$, and let w be $n_g d(\mathbf{S}_g, \hat{\mathbf{s}}_g^0)$, with $w \in \{0, 1, \cdots, l-f\}$. Then looking back at (III.36) we have

$$\begin{aligned}
H(W_g \mid \mathbf{S}_g, W_{\sim g}^{nR}) &= 2^{-nR} \sum_{W_{\sim g}^{nR}} \sum_{f=0}^{l} \sum_{w=0}^{l-f} \sum_{\substack{\mathbf{S}_g : H_E(\mathbf{S}_g) = f \\ n_g d(\mathbf{s}_g, \hat{\mathbf{s}}_g) = w}} \left(\varepsilon^f d^w (1-\varepsilon-d)^{l-f-w} + \varepsilon^f d^{l-f-w} (1-\varepsilon-d)^w\right) \times \\
&\quad h\left(\xi_{\varepsilon,d,l}(f,w)\right), \\
&= \frac{1}{2} \sum_{f=0}^{l} \binom{l}{f} \varepsilon^f \sum_{w=0}^{l-f} \binom{l-f}{w} \left(d^w (1-\varepsilon-d)^{l-f-w} + d^{l-f-w} (1-\varepsilon-d)^w\right) \times \\
&\quad h\left(\xi_{\varepsilon,d,l}(f,w)\right), \\
&= \sum_{f=0}^{l} \binom{l}{f} \varepsilon^f \sum_{w=0}^{l-f} \binom{l-f}{w} d^w (1-\varepsilon-d)^{l-f-w} h\left(\xi_{\varepsilon,d,l}(f,w)\right).
\end{aligned}$$

Finally, using (III.35)

$$H(W^{nR} \mid S^n) \geq nR \sum_{f=0}^{l} \binom{l}{f} \varepsilon^f \sum_{w=0}^{l-f} \binom{l-f}{w} d^w (1-\varepsilon-d)^{l-f-w} h\left(\frac{1}{1+\left(\frac{d}{1-\varepsilon-d}\right)^{l-f-2w}}\right).$$

□

Isolating R in the previous equation, and using (III.30) as well as the results shown in the previous lemmas, we obtain the lower bound on the rate stated in (III.29).

IV

Upper Bounds on the Rate-Distortion Performance of the CRP LDGM and Compound LDGM-LDPC ensembles

In this chapter, we are interested in lossy compression of a BES using linear block codes from the Check Regular Poisson (CRP) LDGM ensemble and from the compound LDGM-LDPC ensemble. We first derive upper bounds on their rate-distortion performance, and then show that there exist compound LDGM-LDPC codes, with degrees independent of the blocklength, which can achieve any point on the Shannon rate-distortion curve of the BES. More precisely, in Section IV.1 we define the CRP LDGM ensemble and the compound construction introduced by Martinian and Wainwright in [15]. In Section IV.2, we generalize the techniques in [15, 16, 17], which are based on the second moment method, to the case of a BES. Then, in Section IV.3, we derive upper bounds on the rate-distortion performance for the BES using codes from the CRP LDGM ensemble and from the compound LDGM-LDPC ensemble. Finally, we prove the optimality of the compound construction for lossy compression of a BES in Section IV.4.

IV.1 Linear Block Codes Considered

For this chapter, we consider the BES(ε) introduced in Example II.1.2, with source alphabet $\mathscr{S} \triangleq \{0,1,\star\}$. We focus on source sequences of length n, and we denote the set of source sequences of length n by $\mathscr{S}^n$, where $n \in \mathbb{N}$. Let $S^n = \{S_1, \cdots, S_n\}$, $S^n \in \mathscr{S}^n$ be a random source string.

Moreover, we focus on binary LDGM and LDPC codes introduced in Subsection II.3.2. These latter are binary linear block codes defined by a sparse generator matrix and a sparse parity-check matrix respectively. Recall from Definition II.3.7, that a binary LDGM code of rate $R_G \leq \frac{m}{n}$ maps binary sequences $w^m \in \{0,1\}^m$ into a codeword $c^n \in \{0,1\}^n$, $m \leq n$. More precisely, let $\mathbb{L}(\mathbf{G})$ be a binary LDGM code of rate $0 \leq R_G \leq 1$ and blocklength n, generated by a sparse binary generator matrix $\mathbf{G} \in \{0,1\}^{m \times n}$

$$\mathbb{L}(\mathbf{G}) = \{c^n \in \{0,1\}^n : \exists w^m \in \{0,1\}^m \text{ s.t. } c^n = w^m \mathbf{G}\}.$$

In the same manner, let $\mathbb{M}(\mathbf{H})$ be a LDPC code of rate $R_H \geq \frac{k}{n}$ and blocklength n, defined by a sparse parity-check matrix $\mathbf{H} \in \{0,1\}^{(n-k) \times n}$

$$\mathbb{M}(\mathbf{H}) = \left\{c^n \in \{0,1\}^n : \ \mathbf{H}(c^n)^T = 0\right\}.$$

We are interested in specific random ensembles of sparse-graph codes based on ensemble of LDGM and LDPC codes, which will be detailed in the next subsections. For the remaining of this chapter d_c, d_v, d'_c, m, n, and k denote natural integers.

IV.1.1 The CRP LDGM Ensemble

We now introduce the CRP LDGM ensemble which contains LDGM codes constructed in a certain random fashion. This specific randomness allows us to easily characterizes the distribution of a random codeword, as we will see in Lemma IV.1.1. We first define this CRP LDGM ensemble and then show a code belonging to it in Figure IV.1.

Definition IV.1.1 - *The CRP LDGM ensemble*:
The CRP LDGM ensemble denoted by $\mathfrak{L}_P\left(d_c, m, n\right)$ contains every LDGM code $\mathbb{L}\left(\mathbf{G}\right)$, where the generator matrix $\mathbf{G}$ lies in $\{0,1\}^{m\times n}$ and is generated by the following procedure. Each check node is connected to d_c information bits chosen uniformly at random and with replacement.

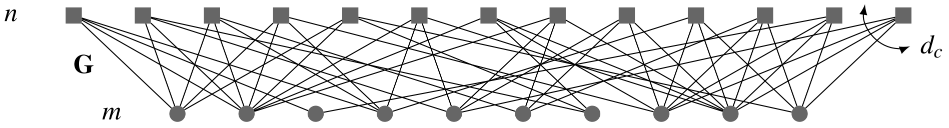


Figure IV.1: *A LDGM code* $\mathbb{L}\left(\mathbf{G}\right)$ *belonging to the CRP LDGM ensemble* $\mathfrak{L}_P\left(4, 10, 13\right)$.

Note that, contrary to Example II.3.2, a code from the CRP LDGM ensemble $\mathfrak{L}_P\left(d_c, m, n\right)$ is usually irregular. Moreover, the degree of a check node is at most d_c, since it is possible, although asymptotically negligible when the blocklength increases, that the same information bit is chosen more than once for a given parity-check node.

A code randomly chosen from this ensemble is known to be suboptimal for both source and channel coding [16, 17], but it has interesting theoretical characteristics, such as the degree distribution of its information bits. Indeed, an information bit of a randomly chosen code $\mathbb{L}\left(\mathbf{G}\right) \in \mathfrak{L}_P\left(d_c, m, n\right)$ has degree d with probability $\binom{n}{d}\left(\frac{1}{m}\right)^d\left(1-\frac{1}{m}\right)^{n-d}$. Thus, for a given rate and in the limit of blocklength tending to infinity, the degree distribution of the information bits tends to a Poisson distribution.

The next lemma characterizes the distribution of a random codeword generated by a code randomly chosen from the CRP LDGM ensemble.

LEMMA IV.1.1 - *Distribution of random codewords generated by CRP LDGM codes*:
Consider the CRP LDGM ensemble $\mathfrak{L}_P\left(d_c, m, n\right)$ and an information sequence $w^m \in \{0,1\}^m$ of weight νm, $\nu \in [0,1]$. Then, the distribution of the random codeword $C^n(\nu)$ generated by $w^m\mathbf{G}$, where $\mathbf{G}$ is the generator matrix of $\mathbb{L}\left(\mathbf{G}\right)$ chosen uniformly at random in $\mathfrak{L}_P\left(d_c, m, n\right)$, is i.i.d. Bernoulli with parameter

$$\delta(\nu, d_c) \triangleq \frac{1}{2}\left[1-(1-2\nu)^{d_c}\right]. \tag{IV.1}$$

Proof. This proof is a more detailed version of the one in [18].

Consider a LDGM code $\mathbb{L}\left(\mathbf{G}\right)$ chosen uniformly at random in $\mathfrak{L}_P\left(d_c, m, n\right)$, and where the information part is fixed by $w^m \in \{0,1\}^m$ with Hamming weight νm, $\nu \in [0,1]$. Consider a specific code bit $C_i(\nu)$, $i \in \{1, \cdots, n\}$. This code bit is connected to d_c information bits, chosen uniformly at random. Let $V_j(i)$ be an indicator variable, which is equal to one if and only if the j^{th} edge starting from $C_i(\nu)$ ends up to an information bit being equal to one, $j \in \{1, \cdots, d_c\}$. Note that we neglect the event of $C_i(\nu)$ having a degree strictly less than d_c, since the probability of such event is asymptotically zero. Thus,

$$C_i(\nu) = V_1(i) \oplus V_2(i) \oplus \cdots \oplus V_{d_c}(i). \tag{IV.2}$$

From (IV.2), $C_i(\nu) = 1$ if and only if the number of $V_j(i)$ being equal to one is odd, whereas $C_i(\nu) = 0$ if and only if the number of $V_j(i)$ being equal to one is even. By construction $\mathbb{P}\left\{V_j(i) = 1\right\} = \nu,\ j \in \{1, \cdots, d_c\}$, and we have

$$\begin{aligned}\mathbb{P}\left\{C_i(\nu) = 1\right\} &= \sum_{\substack{k=0,\\ k\text{ odd}}}^{d_c} \binom{d_c}{k} \nu^k (1-\nu)^{d_c-k},\\ \mathbb{P}\left\{C_i(\nu) = 0\right\} &= \sum_{\substack{k=0,\\ k\text{ even}}}^{d_c} \binom{d_c}{k} \nu^k (1-\nu)^{d_c-k}.\end{aligned}$$

Since $\mathbb{P}\left\{C_i(\nu) = 0\right\} = 1 - \mathbb{P}\left\{C_i(\nu) = 1\right\}$,

$$\mathbb{P}\left\{C_i(\nu) = 1\right\} = \frac{1}{2}\left[1 - \left(\sum_{\substack{k=0,\\ k\text{ even}}}^{d_c} \binom{d_c}{k} \nu^k (1-\nu)^{d_c-k} - \sum_{\substack{k=0,\\ k\text{ odd}}}^{d_c} \binom{d_c}{k} \nu^k (1-\nu)^{d_c-k}\right)\right].$$

We get the claim by observing that

$$((1-\nu)-\nu)^{d_c} = \sum_{\substack{k=0,\\ k\text{ even}}}^{d_c} \binom{d_c}{k} \nu^k (1-\nu)^{d_c-k} - \sum_{\substack{k=0,\\ k\text{ odd}}}^{d_c} \binom{d_c}{k} \nu^k (1-\nu)^{d_c-k}.$$

□

IV.1.2 The Compound LDGM-LDPC Ensemble

We are now able to define the compound LDGM-LDPC construction. The basic idea is to add parity-check constraints to the generator nodes of a LDGM code in order to eliminate its low-weight codewords. This can be done by using a LDGM code compounded with a LDPC code.

Definition IV.1.2 - *Compound LDGM-LDPC code*:
Let $\mathbb{L}(\mathbf{G})$, $\mathbb{M}(\mathbf{H})$ be a LDGM code of rate R_G and a LDPC code of rate R_H respectively, where $\mathbf{G} \in \{0,1\}^{m \times n}$ and $\mathbf{H} \in \{0,1\}^{k \times m}$. Then, $\mathbb{M}(\mathbf{H})\,\mathbf{G}$ defines a new linear block code with blocklength n and rate $R = R_G R_H$, called a compound LDGM-LDPC code and denoted by $\mathbb{C}(\mathbf{G}, \mathbf{H})$,

$$\mathbb{C}(\mathbf{G}, \mathbf{H}) \triangleq \left\{c^n \in \{0,1\}^n :\ \exists w^m \in \{0,1\}^m \text{ s.t. } c^n = w^m\mathbf{G} \text{ and } \mathbf{H}(w^m)^T = \mathbf{0}\right\}. \tag{IV.3}$$

In order to analyze the behavior of such codes for lossy compression of a BES we will focus on a certain ensemble of compound LDGM-LDPC codes, detailed in the following definition.

Definition IV.1.3 - *Compound LDGM-LDPC ensemble*:
The compound LDGM-LDPC ensemble denoted by $\mathfrak{C}(d_c, d_v, d_c', m, n)$, is the set of compound

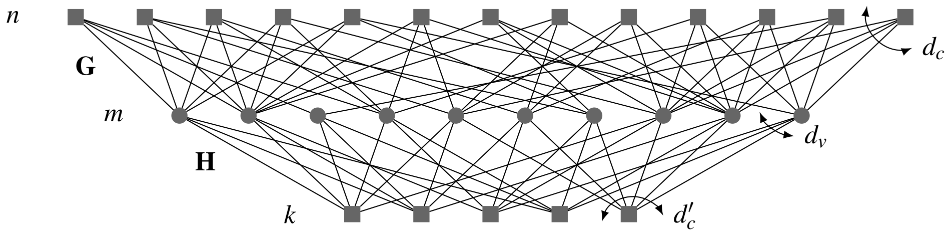


Figure IV.2: *Compound LDGM-LDPC code. The top layer is a (n,m) LDGM code whereas the bottom part is a (m,k) LDPC code. The top layer of nodes are the check bits of the LDGM code, the middle layer are the information bits of the LDGM code, and the bottom layer are the parity-check nodes.*

codes $\mathbb{C}(\mathbf{G},\mathbf{H})$, where $\mathbf{G}$ and $\mathbf{H}$ are both chosen uniformly at random in $\mathfrak{L}_P(d_c,m,n)$ and in the standard (m,d_v,d'_c)-regular LDPC ensemble respectively.

An example of a compound LDGM-LDPC code from $\mathfrak{C}(4,3,6,13,10)$ is given in Figure IV.2.

In the next section, we generalize the techniques in [15, 16, 17] based on the second moment method, to lossy compression of a BES.

IV.2 Second Moment Method

We consider rate-distortion encoding of a BES using the CRP LDGM ensemble and the compound LDGM-LDPC ensemble. For a given code $\mathbb{C}$ with rate R and blocklength n, let $N \triangleq 2^{nR}$ be the total number of codewords. We index the codewords as $\{C_1^n,\cdots,C_N^n\}$, where C_1^n is the codeword generated by the all-zero information word. Let $\mathscr{I}(\cdot)$ be the bijection from $\{C_1^n,\cdots,C_N^n\}$ to $\{1,\cdots,N\}$, and $\mathscr{I}^{-1}(\cdot)$ its inverse.

As in Chapter III,the distortion between a source sequence s^n and a codeword c^n is given by the average of the generalized Hamming single-letter distortion between letters of both sequences, i.e.,

$$d(s^n,c^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} d(s_i,c_i). \tag{IV.4}$$

Furthermore, the source encoder f is assumed to be optimal, that is

$$\begin{aligned} f: \quad \mathscr{S}^n &\rightarrow \{1,\cdots,N\} \\ s^n &\mapsto \mathscr{I}\left(\arg\min_{c^n\in\mathbb{C}}\{d(s^n,c^n)\}\right). \end{aligned} \tag{IV.5}$$

The reconstructed sequence $\hat{S}^n \in \hat{\mathscr{S}}^n$ associated to $S^n \in \mathscr{S}^n$ is then given by

$$\hat{S}^n = \mathscr{I}^{-1}(f(S^n)). \tag{IV.6}$$

The average normalized distortion is $\mathbb{E}\left[d\left(S^n,\hat{S}^n\right)\right]$, where the expectation is taken over all source sequences realizations.

For a source word $S^n \in \mathscr{S}^n$, let $X_i(\mathbb{C},S^n,D)$ be the indicator function which evaluates to one if C_i^n is within normalized distortion D from S^n. We define,

$$Z(\mathbb{C},S^n,D) = \sum_{i=1}^{N} X_i(\mathbb{C},S^n,D). \tag{IV.7}$$

For the sake of notational simplicity, we drop the arguments and write X_i and Z. We want to derive a lower bound on the probability $\mathbb{P}\{Z>0\}$. Let $\mathscr{S}_b^n$ be the set of source sequences with b erasures and let $\mathscr{S}_{\mathscr{E}}^n$ be the set of source sequences whose erasure positions are elements of the set $\mathscr{E}$, $\mathscr{E} \subset \{1,\dots,n\}$. We write

$$\begin{aligned}
\mathbb{P}\{Z>0\} &= \sum_{\mathscr{E}} \mathbb{P}\{S^n \in \mathscr{S}_{\mathscr{E}}^n\}\mathbb{P}\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{E}}^n\},\\
&= \sum_{b=0}^{n} \sum_{\mathscr{E}:|\mathscr{E}|=b} \mathbb{P}\{S^n \in \mathscr{S}_{\mathscr{E}}^n\}\mathbb{P}\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{E}}^n\},\\
&\overset{(a)}{=} \sum_{b=0}^{n} \binom{n}{b} \epsilon^b (1-\epsilon)^{n-b}\mathbb{P}\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\},
\end{aligned} \tag{IV.8}$$

where $\mathscr{B}=\{1,\dots,b\}$. (a) follows because of the i.i.d. behavior of the BES. We are interested in finding a lower bound on the probability $\mathbb{P}\left\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right\}$. Note that we are only interested in the growth rate of $\mathbb{P}\{Z>0\}$ due to the following concentration result in [18].

LEMMA IV.2.1 - *Upper bound on the average distortion*:
Assume that for a given distortion D and a given ensemble of codes, we have

$$\lim_{n\to\infty} \frac{1}{n}\log_2 \mathbb{P}\{Z(\mathbb{C},S^n,D)>0\} \geq 0. \tag{IV.9}$$

Then, $\forall\theta>0$, there exists a code in the ensemble such that for sufficiently large blocklength n the average distortion is less than than $D+\theta$.

We derive a lower bound on $\mathbb{P}\left\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right\}$ by the second moment method,

$$\mathbb{P}\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\} \geq \frac{\mathbb{E}\left(Z \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right)^2}{\mathbb{E}\left(Z^2 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right)}. \tag{IV.10}$$

In the next lemma, we compute the first moment $\mathbb{E}\left(Z \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right)$ for *any* linear code. We also show that if $nD \geq \frac{n-b}{2}$, then the growth rate of $\mathbb{P}\left\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right\}$ is zero.

LEMMA IV.2.2 - *First moment for any linear block code*:
Consider any linear block code with rate R and blocklength n. The first moment $\mathbb{E}\left(Z \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right)$ is given by

$$\mathbb{E}(Z \mid S^n \in \mathscr{S}_{\mathscr{B}}^n) = \begin{cases} o\left(2^{n-b}\right) 2^{nR-n(1-\beta)\left(1-h\left(\frac{D}{1-\beta}\right)\right)}, & \text{if } D \leq \frac{1-\beta}{2},\\ o\left(2^{n-b}\right) 2^{nR}, & \text{otherwise}, \end{cases} \tag{IV.11}$$

where $\beta \triangleq \frac{b}{n}$. If $D \geq \frac{1-\beta}{2}$, then

$$\lim_{n\to\infty} \frac{\log_2\left(\mathbb{P}\left\{Z>0 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\right\}\right)}{n} = 0.$$

Proof. By using the definition of Z in terms of random variables X_i given in (IV.7), we obtain

$$\begin{aligned}
\mathbb{E}(Z \mid S^n \in \mathscr{S}_{\mathscr{B}}^n) &= \sum_{i=1}^{N} \mathbb{E}(X_i \mid S^n \in \mathscr{S}_{\mathscr{B}}^n),\\
&= 2^{nR}\mathbb{P}\{X_1 = 1 \mid S^n \in \mathscr{S}_{\mathscr{B}}^n\}.
\end{aligned}$$

We obtain the expression for $\mathbb{E}\left(Z \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right)$ by noting that

$$\mathbb{P}\left\{X_1 = 1 \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right\} = \sum_{j=0}^{nD} \binom{n-b}{j} \frac{1}{2^{n-b}},$$

and the maximum of the summation term is attained at $j = nD$ if $nD \leq \frac{n-b}{2}$; otherwise it is attained at $\frac{n-b}{2}$.

Thus, when $D \geq \frac{1-\beta}{2}$, we obtain

$$\lim_{n\to\infty} \frac{\log_2\left(\mathbb{P}\left\{X_1 > 0 \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right\}\right)}{n} = (1-\beta)\left(h\left(\frac{1}{2}\right) - 1\right) = 0.$$

The claim of the lemma follows by noting that $\mathbb{P}\left\{X_1 > 0 \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right\} \leq \mathbb{P}\left\{Z > 0 \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right\}$ and the probability of any event is upper bounded by one. □

Remark: From the lemma above, we need to lower bound $\mathbb{P}\left\{Z > 0 \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right\}$ only for $\beta < 1-2D$. In addition, the rate distortion performance analysis is non-trivial only for $D < (1-\varepsilon)/2$.

In the next lemma, we compute the second moment of Z in terms of its expectation. The proof of this lemma is identical to that of Lemma 3 in [16].

Lemma IV.2.3 - *Second moment for any linear block code*:
For any linear code, the second moment satisfies the relation

$$\mathbb{E}\left(Z^2 \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right) = \mathbb{E}\left(Z \mid S^n \in \mathscr{S}^n_{\mathscr{B}}\right)\left(1 + \sum_{j\neq 1} \mathbb{P}\left\{X_j = 1 \mid X_1 = 1, S^n \in \mathscr{S}^n_{\mathscr{B}}\right\}\right). \tag{IV.12}$$

In the next section, we derive upper bounds on the rate distortion performance of CRP LDGM ensemble and compound LDGM-LDPC ensemble for the BES.

IV.3 Upper Bounds on the Rate-Distortion Performance

Consider a randomly chosen code $\mathbb{L}(\mathbf{G}) \in \mathfrak{L}_P(d_c, m, n)$. Let $C^n(\nu)$ be a codeword which is generated by an information word of weight νm, $\nu \in [0,1]$. Then, from Lemma IV.1.1 each component of $C^n(\nu)$ is Bernoulli distributed with parameter $\delta(\nu, d_c)$, defined in (IV.1). Note that to simplify notation, we sometimes drop the arguments of $\delta(\nu, d_c)$ and denote it by δ. Define

$$\mathscr{Q}(\nu,\beta) \triangleq \mathbb{P}\left\{d\left(C^n(\nu), S^n\right) \leq D \mid d\left(C^n_1, S^n\right) \leq D, S^n \in \mathscr{S}^n_{\mathscr{B}}\right\}. \tag{IV.13}$$

The following lemma bounds the exponential behavior of $\mathscr{Q}(\nu,\beta)$.

Lemma IV.3.1 - *Upper bound on the exponential growth rate of $\mathscr{Q}$*:
Let $\beta < 1-2D$. For a randomly chosen code from the CRP LDGM ensemble $\mathfrak{L}_P(d_c, m, n)$ or the compound LDGM-LDPC ensemble $\mathfrak{C}(d_c, d_v, d'_c, m, n)$, the exponential growth rate of the conditional probability defined in (IV.13) is upper bounded as

$$\frac{1}{n}\log_2 \mathscr{Q}(\nu,\beta) \leq F\left(\delta(\nu, d_c), \beta, D\right) + o(1), \tag{IV.14}$$

where

$$F(\gamma,\beta,D)=\inf_{\lambda<0}\max_{\tau\in[0,D]}G(\tau,\lambda,\gamma,\beta,D)=\max_{\tau\in[0,D]}G(\tau,\lambda^\star,\gamma,\beta,D).$$

In order to define $G(\tau,\lambda,\gamma,\beta,D)$, consider $f_1(\gamma,\lambda)\triangleq(1-\gamma)e^\lambda+\gamma$. Then,

$$G(\tau,\lambda,\gamma,\beta,D)\triangleq(1-\beta)\left[h\left(\frac{\tau}{1-\beta}\right)-h\left(\frac{D}{1-\beta}\right)\right]+\tau\log_2\left(f_1(\gamma,\lambda)\right)+$$
$$(1-\beta-\tau)\log_2\left(f_1(1-\gamma,\lambda)\right)-\frac{\lambda D}{\log 2}.$$

The definition of λ^* is based on the following quadratic equation

$$x^2(1-\beta-D)\delta(1-\delta)+x\left(\tau(1-2\delta)+\delta^2(1-\beta)\right)-x\left(D\left(\delta^2+(1-\delta)^2\right)\right)-D\delta(1-\delta)=0. \tag{IV.15}$$

Let ρ^* be the only positive solution of (IV.15). Then $\lambda^*=\min\left(0,\log\left(\rho^*\right)\right)$.

Proof. This proof is an extension of the proof for Lemma 5 in [18]. Since C_1^n is the all-zero codeword, the condition $d\left(C_1^n,S^n\right)\leq D$ is equivalent to $w_H(S^n)\leq nD$, where $w_H(S^n)$ denotes the Hamming weight of S^n (Hamming weight of an erasure is zero). Let T be the random variable corresponding to the Hamming weight of S^n, i.e., $T=w_H(S^n)$, knowing that S^n lies in $\mathscr{S}^n_{\mathscr{B}}$. Then

$$\mathbb{P}\left\{T=t\right\}=\frac{\binom{n-b}{t}}{\sum_{i=0}^{nD}\binom{n-b}{i}}.$$

Let Y be the random variable corresponding to the Hamming distance between $C^n(\nu)$ and S^n, when $S^n\in\mathscr{S}^n_{\mathscr{B}}$. Then

$$Y=\begin{cases}\displaystyle\sum_{j=1}^{T}U_j+\sum_{j=1}^{n-b-T}V_j, & \text{if } 1\leq T\leq nD,\\ \displaystyle\sum_{j=1}^{n-b-T}V_j, & \text{if } T=0,\end{cases}$$

where U_j and V_j are independent Bernoulli random variables with parameters $1-\delta(\nu,d_c)$ and $\delta(\nu,d_c)$ respectively. Then,

$$\mathscr{Q}(\nu,\beta)=\mathbb{P}\left\{Y\leq nD\right\}. \tag{IV.16}$$

To bound this probability we will use the Chernoff bound in the following manner

$$\frac{1}{n}\log_2\mathbb{P}\left\{Y\leq nD\right\}\leq\inf_{\lambda<0}\left(\frac{1}{n}\log_2\mathbb{M}_Y(\lambda)-\frac{\lambda D}{\log 2}\right), \tag{IV.17}$$

where $\mathbb{M}_Y(\lambda)$ denotes the moment generating function of the random variable Y. Then we have

$$\begin{aligned}\mathbb{M}_U(\lambda)&=(1-\delta)e^\lambda+\delta,\\ \mathbb{M}_V(\lambda)&=\delta e^\lambda+1-\delta,\\ \mathbb{M}_Y(\lambda)&=\sum_{t=0}^{nD}\mathbb{P}\left\{T=t\right\}\left[\mathbb{M}_U(\lambda)\right]^t\left[\mathbb{M}_V(\lambda)\right]^{n-b-t}.\end{aligned}$$

Let $\tau = \frac{t}{n}$. Using Stirling's formula, we obtain

$$\frac{1}{n}\log_2 \mathbb{M}_Y(\lambda) = \frac{1}{n}\log_2\left\{\sum_{t=0}^{Dn} 2^{n(1-\beta)\left(h\left(\frac{\tau}{1-\beta}\right)-h\left(\frac{D}{1-\beta}\right)\right)} \times 2^{n(\tau\log_2(\mathbb{M}_U(\lambda))+(1-\beta-\tau)\log_2(\mathbb{M}_V(\lambda)))}\right\}+o(1).$$

Using this, (IV.16), and (IV.17) we have

$$\frac{1}{n}\log_2 \mathscr{Q}(\nu,D) \leq \inf_{\lambda<0}\max_{\tau\in[0,D]} G(\tau,\beta,\lambda,\delta,D)+o(1).$$

Using similar arguments as in the proof of Lemma 5 of [18], it can be shown that the order of infimum with respect to λ and maximum with respect to τ can be interchanged. Equating the partial derivative of $G(\tau,\beta,\lambda,\delta,D)$ with respect to λ to zero results in the quadratic equation (IV.15) in terms of e^{λ}. Solving the quadratic equation gives the desired expression for $F(\gamma,\beta,D)$. This proves the lemma. □

Note that in the previous lemma we derived an upper bound on the growth rate of $\mathscr{Q}(\nu,\beta)$ for any fraction β of erasures such that $\beta < 1-2D$. However, from now on we will only consider $\beta = \varepsilon$. This is because asymptotically the probability of having a source sequence with fraction of erasures equal to ε is almost equal to one. A plot of the function $F(\delta(\nu,d_c),\varepsilon,D)$ defined in the previous lemma is shown in Figure IV.3.

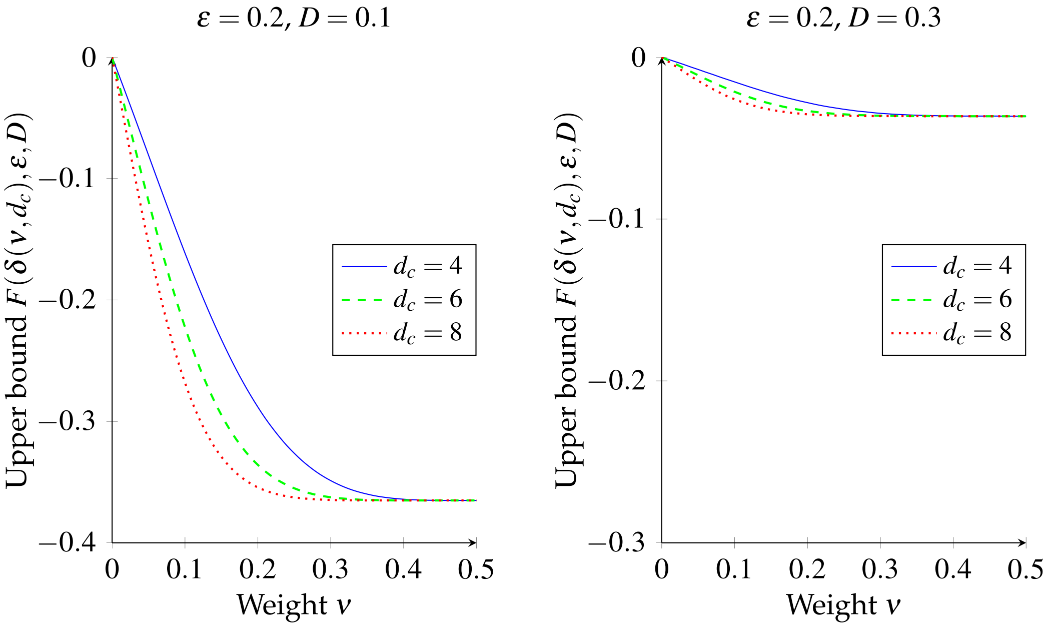


Figure IV.3: *Upper bound on the exponential growth rate of $\mathscr{Q}(\nu,\varepsilon)$ for CRP LDGM codes with top degree $d_c \in \{4,6,8\}$, overall distortion $D \in \{0.1,0.3\}$, and erasure probability $\varepsilon = 0.2$.*

In the following lemma we derive an upper bound on the rate distortion performance of the CRP LDGM ensemble.

Lemma IV.3.2 - *Upper bound on the rate-distortion performance for the CRP LDGM ensemble*: Consider the CRP LDGM ensemble $\mathfrak{L}_P(d_c,m,n)$ with rate $R \leq \frac{m}{n}$. The rate distortion performance of $\mathfrak{L}_P(d_c,m,n)$ for the BES(ε) is upper bounded by

$$R \geq \max_{\nu\in[0,1]} \frac{R_{\varepsilon}^{\text{sh}}(D)+F(\delta(\nu,d_c),\varepsilon,D)}{1-h(\nu)}, \tag{IV.18}$$

where $F(\delta(\nu,d_c),\varepsilon,D)$ is defined in Lemma IV.3.1.

Proof. Combining (IV.10) and (IV.12) we obtain

$$\begin{aligned}\frac{1}{n}\log_2\left(\mathbb{P}\left\{Z>0 \mid S^n\in\mathscr{S}^n_{\mathscr{B}}\right\}\right) \quad \geq \quad & \frac{1}{n}\log_2 E\left(Z \mid S^n\in\mathscr{S}^n_{\mathscr{B}}\right) - \\ & \frac{1}{n}\log_2\left(1+\sum_{j\neq 1}\mathbb{P}\left\{X_j=1 \mid X_1=1, S^n\in\mathscr{S}^n_{\mathscr{B}}\right\}\right).\end{aligned} \tag{IV.19}$$

Assuming $\beta\leq 1-2D$, we can upper bound the last term in (IV.19) by using (IV.14).

$$\begin{aligned}\frac{1}{n}\log_2\left(1+\sum_{j\neq 1}\mathbb{P}\left\{X_j=1 \mid X_1=1, S^n\in\mathscr{S}^n_{\mathscr{B}}\right\}\right) \quad &= \quad \frac{1}{n}\log_2\left(\sum_{\substack{\nu\in[0,1]:\\ \nu m\in\mathbb{N}}}\binom{m}{\nu m}\mathscr{Q}(\nu,\beta)\right) \\ &\leq \quad \max_{\nu\in[0,1]}\left\{Rh(\nu)+F\left(\delta(\nu,d_c),\beta,D\right)\right\}.\end{aligned} \tag{IV.20}$$

Combining the last two equations and (IV.8), and considering only the typical value $\beta=\varepsilon$ we have

$$\frac{1}{n}\log_2\mathbb{P}\left\{Z>0\right\}\geq R-(1-\varepsilon)\left(1-h\left(\frac{D}{1-\varepsilon}\right)\right)-\max_{\nu\in[0,1]}\left\{Rh(\nu)+F(\delta(\nu,d_c),\varepsilon,D)\right\}+o(1). \tag{IV.21}$$

As long as the right hand side (RHS) in (IV.21) stays non-negative, we get the claim from Lemma IV.2.1. □

In the following lemma we derive an upper bound on the rate distortion performance of the compound LDGM-LDPC ensemble.

Lemma IV.3.3 - *Upper bound on the rate distortion performance of the compound LDGM-LDPC ensemble*:
Consider the compound LDGM-LDPC ensemble $\mathfrak{C}(d_c,d_v,d_c',m,n)$ with overall rate R. Let $B(\nu)$ be an upper bound on the growth rate of the weight distribution of the (d_v,d_c')-regular LDPC ensemble. Then the rate distortion performance of $\mathfrak{C}(d_c,d_v,d_c',m,n)$ for the BES(ε) is upper bounded by

$$R \geq \max_{\nu\in[0,1]} \frac{R_{\varepsilon}^{\text{sh}}(D)+F(\delta(\nu,d_c),\varepsilon,D)}{1-\frac{B(\nu)}{R_H}}, \tag{IV.22}$$

where $F(\delta(\nu,d_c),\varepsilon,D)$ is defined in Lemma IV.3.1.

Proof. The proof for this lemma is very similar to the proof of the Lemma IV.3.2. The equivalent of (IV.20) for the compound construction is

$$\sum_{j=1}^{N}\mathbb{P}\left\{X_j=1 \mid X_1=1, S^n\in\mathscr{S}^n_{\mathscr{B}}\right\} = \sum_{\substack{\nu\in[0,1]:\\ \nu m\in\mathbb{N}}}\mathscr{A}_m(\nu)\mathscr{Q}(\nu,\beta),$$

where $\mathscr{A}_m(\nu)$ denotes the number of codewords of weight νm of the LDPC code. We then upper bound the growth rate of $\mathscr{A}_m(\nu)$ by $B(\nu)$ to obtain the desired result. □

To understand the upper bounds in Lemmas IV.3.2 and IV.3.3, the functions inside the maximum of (IV.18) and (IV.22) are plotted in Figure IV.4. From this figure, we can see the suboptimality of CRP LDGM codes. Indeed, contrary to the compound construction, the rate lower bound for the CRP LDGM ensemble is strictly increasing around $\nu = 0$, leading to a rate strictly greater than the minimum achievable rate.

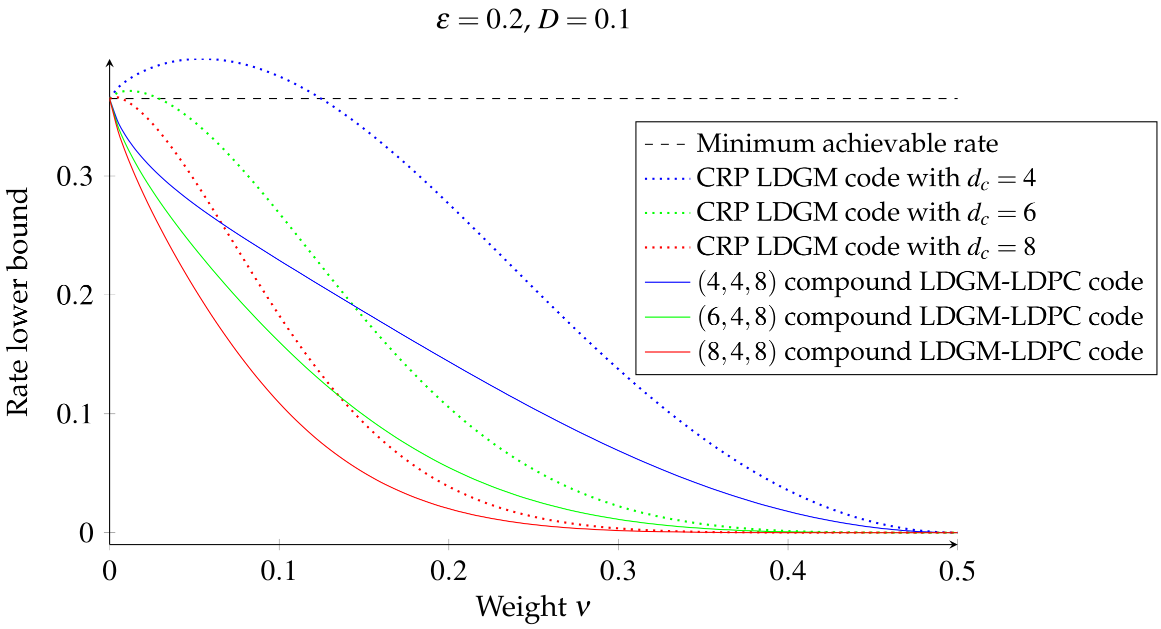


Figure IV.4: *Rate lower bounds for the CRP LDGM and the compound LDGM-LDPC ensembles. Codes from the CRP LDGM ensemble with top degree $d_c \in \{4,6,8\}$ are shown (dotted lines) as well as the compound construction based on this latter and a $(4,8)$-regular LDPC code (solid lines).*

In the next section, we prove the optimality of the compound construction for the BES(ε).

IV.4 Source Coding Optimality of the Compound LDGM-LDPC Ensemble

Before proving the optimality of the compound construction, we recall that $B(\nu)$ is an upper bound on the growth rate of the weight enumerator function for the LDPC code $A_m(\nu)$. Since the dependence of the function B on the degree pair (d_v, d_c') is obvious we will use both notations $B(\nu)$ or $B(\nu, d_v, d_c')$. The following lemma states some properties on the bounding function $B(\nu)$ which are proved in [18].

Lemma IV.4.1 - *Upper bound on the growth rate of $A_m(\nu)$:*
For a LDPC code with degrees (d_v, d_c'), where d_c' is even, an upper bound $B(\nu, d_v, d_c')$ on the growth rate of the weight enumerator function of the code can be defined for any $\nu \in \left[0, \frac{1}{2}\right]$ as

$$B(\nu, d_v, d_c') = (1-d_v)h(\nu) - (1-R_H) + d_v \inf_{\lambda \le 0} \left\{ \frac{1}{d_c'} \log_2 \left(\left(1+2^{\lambda}\right)^{d_c'} + \left(1-2^{\lambda}\right)^{d_c'} \right) - \nu\lambda \right\},$$

and $B(\nu) = B(1-\nu)$ for $\nu \in \left[\frac{1}{2}, 1\right]$. The function $B(\nu)$ we just defined satisfies the following conditions.

1. $B(\nu)$ is symmetric around $\frac{1}{2}$;
2. $B(\nu)$ is twice differentiable on $(0,1)$ with $B'\left(\frac{1}{2}\right)=0$ and $B''\left(\frac{1}{2}\right)<0$;
3. $B(\nu)$ achieves its unique optimum at $\nu=\frac{1}{2}$, where $B\left(\frac{1}{2}\right)=R_H$;
4. $\exists \mu_1>0$ such that $\forall \nu \in (0,\mu_1)$, $B(\nu)<0$.

The upper bound $B(\nu)$ defined in Lemma IV.4.1 is plotted in Figure IV.5 for regular LDPC codes of rate $\frac{1}{2}$.

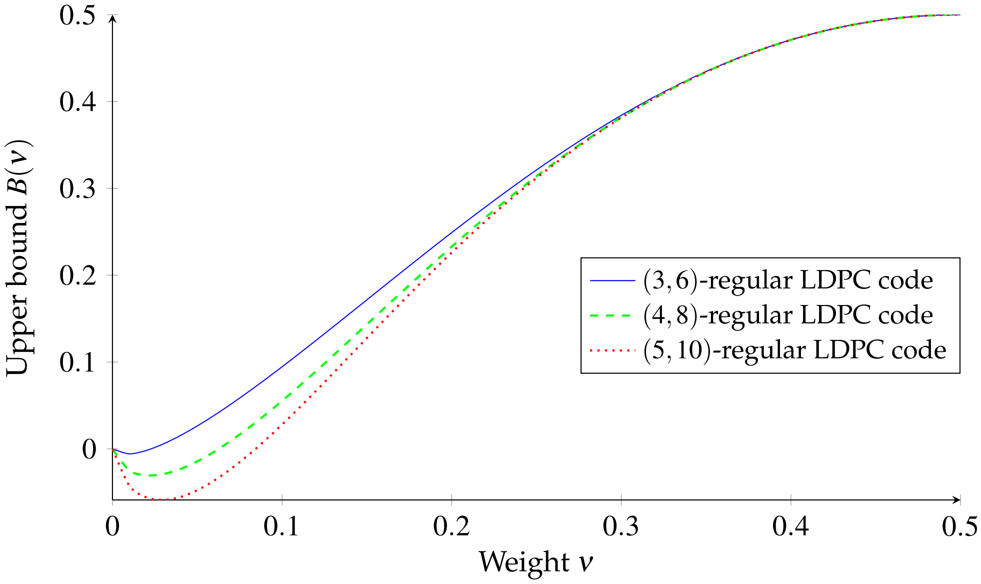


Figure IV.5: *Upper bound on the growth rate of the weight enumerator function for (d_v,d_c')-regular LDPC codes of rate $\frac{1}{2}$ and even d_c'.*

The next lemma derives some properties of the function $F(\delta(\nu,d_c),\varepsilon,D)$ (defined in Lemma IV.3.1) which will be useful in proving the optimality of the compound construction.

Lemma IV.4.2 - *Properties of $F(\delta(\nu,d_c),\varepsilon,D)$*:
For any even degree $d_c \geq 4$, the function $F(\delta(\nu,d_c),\varepsilon,D)$ is differentiable in the neighborhood of $\nu=\frac{1}{2}$ with

$$F\left(\delta\left(\frac{1}{2},d_c\right),\varepsilon,D\right) = -R_\varepsilon^{\text{sh}}(D), \tag{IV.23}$$

$$\frac{\partial}{\partial \nu}F(\delta(\nu,d_c),\varepsilon,D)\Big|_{\nu=\frac{1}{2}} = 0, \tag{IV.24}$$

$$\frac{\partial^2}{\partial \nu^2}F(\delta(\nu,d_c),\varepsilon,D)\Big|_{\nu=\frac{1}{2}} = 0. \tag{IV.25}$$

Proof. Note that $\delta\left(\frac{1}{2}\right)=\frac{1}{2}$, $\forall d_c \geq 1$. Thus,

$$G\left(\tau,\lambda,\frac{1}{2},\beta,D\right) = (1-\beta)\log_2\left(f_1\left(\frac{1}{2},\lambda\right)\right) - \frac{\lambda D}{\log 2} + (1-\beta)\left[h\left(\frac{\tau}{1-\beta}\right) - h\left(\frac{D}{1-\beta}\right)\right]. \tag{IV.26}$$

Moreover, $\max_{\tau\in[0,D]}\left\{h\left(\frac{\tau}{1-\beta}\right)-h\left(\frac{D}{1-\beta}\right)\right\}=0$, since $\frac{\tau}{1-\beta}\leq\frac{D}{1-\beta}\leq\frac{1}{2}$ and $h(\cdot)$ is an increasing function on $\left[0,\frac{1}{2}\right]$. Finally, the infimum of the first two terms in (IV.26) is attained at $\lambda=-\log\left(\frac{1-\beta}{D}-1\right)$. Consequently we have

$$F\left(\delta\left(\frac{1}{2},d_c\right),\varepsilon,D\right)=-R_{\varepsilon}^{\text{sh}}(D).$$

Since $\delta(\nu,d_c)$ is twice differentiable in ν, $F(\delta(\nu,d_c),\varepsilon,D)$ is also twice differentiable in ν. Since $\frac{\partial}{\partial\nu}\delta(\nu,d_c)\Big|_{\nu=\frac{1}{2}}=\frac{\partial^2}{\partial\nu^2}\delta(\nu,d_c)\Big|_{\nu=\frac{1}{2}}=0$ if $d_c\geq 4$, then by using chain rule we obtain (IV.24) and (IV.25). □

We are now ready to prove our main result.

Theorem IV. 4.1 - ***Optimality of the Compound Construction*****:**
Consider lossy compression of BES(ε) using the compound construction. Given any distortion D, $D\leq(1-\varepsilon)/2$, $\forall\eta>0$, let R be the desired rate of compression with $R<R_{\varepsilon}^{\text{sh}}(D)+\eta$. Then there exist degrees (d_c,d_v,d_c') (independent of the blocklength) and a compound code $\mathbb{C}(\mathbf{G},\mathbf{H})\in\mathfrak{C}(d_c,d_v,d_c',m,n)$ with rate R which achieves average distortion D.

Proof. To complete the proof of the theorem we need to show that compound codes with a top degree d_c independent of the blocklength are sufficient. We will restrict ourselves to even d_c'. Similar to (IV.21) but for the compound construction case, equation (IV.9) is equivalent to

$$\Delta\geq\max_{\nu\in[0,1]}\{K(\nu,d_c)\},$$

where $\Delta\triangleq R-R_{\varepsilon}^{\text{sh}}(D)$ and $K(\nu,d_c)\triangleq\frac{R}{R_H}B(\nu,d_v,d_c')+F\left(\delta(\nu,d_c),\varepsilon,D\right)$. We now divide the proof into three steps.

1. $\exists\mu_1>0$, independent of d_c, such that $\forall\nu\in[0,\mu_1]$, $K(\nu,d_c)\leq\Delta$.
2. $\exists\mu_2>0$, independent of d_c, such that $\forall\nu\in\left[\frac{1}{2}-\mu_2,\frac{1}{2}\right]$, $K(\nu,d_c)\leq\Delta$.
3. $\exists d_c^*<\infty$ such that $\forall\nu\in[\mu_1,1/2-\mu_2]$, $K(\nu,d_c^*)\leq\Delta$.

Proof of 1): By the last property of $B(\nu)$ in Lemma IV.4.1, $\exists\mu_1>0$ such that $B(\nu)\leq 0$ for all $\nu\in[0,\mu_1]$. Since $F(\delta(\nu,d_c),\varepsilon,D)\leq 0$ for all ν we have $K(\nu,d_c)\leq 0<\Delta$. Note that μ_1 is independent of the LDGM part.

Proof of 2): We will use Taylor expansion of $K(\nu,d_c)$ around $\nu=\frac{1}{2}$ up to second order. Note that $K(\nu,d_c)\leq K(\nu,4)$, $\forall d_c\geq 4$ since $\delta(\nu,d_c)$ is increasing in d_c and $F(\gamma,\varepsilon,D)$ is decreasing in γ. Thus it suffices to show that $K(\nu,4)\leq\Delta$, $\forall\nu\in\left[\frac{1}{2}-\mu_2,\frac{1}{2}\right]$. Using Lemma IV.4.1 and IV.4.2 we can calculate the derivative of $K(\nu,d_c)$ with respect to ν.

$$K\left(\frac{1}{2},d_c\right)=R-R_{\varepsilon}^{\text{sh}}(D)=\Delta,$$

$$\frac{\partial}{\partial\nu}K(\nu,d_c)\Big|_{\nu=\frac{1}{2}}=0,$$

$$\frac{\partial^2}{\partial\nu^2}K(\nu,d_c)\Big|_{\nu=\frac{1}{2}}=\frac{R}{R_H}B''\left(\frac{1}{2}\right)<0.$$

By continuity of second derivative of $K(\nu,d_c)$, for some $\mu_2>0$ we have for any $\nu\in\left[\frac{1}{2}-\mu_2,\frac{1}{2}\right]$ the second derivative is negative. For any $\nu\in\left[\frac{1}{2}-\mu_2,\frac{1}{2}\right]$, there exists $\tilde{\nu}\in\left[\nu,\frac{1}{2}\right]$ such that

$$\begin{aligned} K(\nu,4) &= \Delta+\frac{1}{2}\left(\tilde{\nu}-\frac{1}{2}\right)^2\frac{\partial^2}{\partial\nu^2}K(\nu,d_c)\bigg|_{\nu=\frac{1}{2}}, \\ &\leq \Delta. \end{aligned}$$

Thus we have

$$\forall\nu\in\left[\frac{1}{2}-\mu_2,\ \frac{1}{2}\right],\ K(\nu,d_c)\leq K(\nu,4)\leq\Delta.$$

Proof of 3): Using Lemma IV.4.1 there exists a function $\sigma(\mu_2)$ such that

$$B(\nu)\leq R_H\left[1-\sigma(\mu_2)\right],\ \text{for all } \nu\leq\frac{1}{2}-\mu_2. \tag{IV.27}$$

Since $F(\gamma,\varepsilon,D)$ is continue in γ and $\lim_{d_c\to\infty}\delta(\mu_1,d_c)=\frac{1}{2}$ we have

$$\lim_{d_c\to\infty}F(\delta(\mu_1,d_c),\varepsilon,D)=-R_\varepsilon^{\text{sh}}(D).$$

As $F(\gamma,\varepsilon,D)$ is a decreasing function in γ, for any $\mu_3>0$, $\exists d_c^*<\infty$ such that

$$F(\delta(\mu_1,d_c^*),\varepsilon,D)\leq -R_\varepsilon^{\text{sh}}(D)+\mu_3. \tag{IV.28}$$

Combining the results of both equations (IV.27) and (IV.28) we have

$$\begin{aligned} K(\nu,d_c^*) &= \frac{R}{R_H}B(\nu)+F(\delta(\nu,d_c^*),\varepsilon,D), \\ &\leq R\left[1-\sigma(\mu_2)\right]-R_\varepsilon^{\text{sh}}(D)+\mu_3, \\ &= \Delta+(\mu_3-R\sigma(\mu_2)). \end{aligned}$$

We complete the proof by choosing any μ_3 less than $R\sigma(\mu_2)$. □

IV.5 Conclusion

In this chapter, we derived upper bounds on the rate-distortion performance of the CRP LDGM ensemble and the compound LDGM-LDPC ensemble for the BES case. Furthermore, contrary to CRP LDGM codes, we showed that the compound construction can achieve the Shannon rate-distortion function for lossy compression of the BES with degrees remaining independent of the blocklength.

These compound LDGM-LDPC codes have a natural nested structure, that is a compound code can be described as the union of compound subcodes, which make them well suited for channel and source coding with side information [31]. Indeed, compound codes were proved to be optimal for the Wyner-Ziv and Gelfand-Pinsker problems [17, 18]. In addition, it is is fairly easy to show that compound codes are also optimal for the one helper problem [32], and lossy compression with side information available at both the encoder and decoder. An interesting question would be to know if they are likewise optimal for the problem of channel and source coding with two-sided information [33], which can be viewed as a generalization of channel and source coding with side information.

V

Conclusions

V.1 Summary

In this thesis, we focused on lossy compression of a binary erasure source, which is a discrete memoryless source with ternary output alphabet, and can be viewed as a generalization of a binary symmetric source. The lossy source coding itself was based on low-density generator matrix (LDGM) codes, and two random ensembles based on them. The first one, called the check regular Poisson LDGM (CRP LDGM) ensemble, is generated by considering LDGM codes constructed in a specific random manner. The second one, generated by compounding codes from this latter ensemble and the standard regular low-density parity-check (LDPC) ensemble, is called the compound LDGM-LDPC ensemble.

As our main contributions, we bounded the rate-distortion performance of the aforementioned sparse-graph codes for lossy source coding of BES. More specifically we derived two lower bounds on the rate-distortion performance of LDGM codes for lossy compression of BES using two different methods. The fist one used a counting argument, whereas the second one was based on a probabilistic argument. We then showed that both methods lead to the same bound. It is noteworthy that these lower bounds are valid for any LDGM codes of a given rate and generator node degree distribution, and any encoding function.

Last but not least, we derived upper bounds on the rate-distortion performance of codes from the CRP LDGM ensemble and from the compound LDGM-LDPC ensemble for lossy compression of a BES. Contrary to the lower bounds part, these upper bounds were obtained assuming optimal encoding. This is because a low complexity iterative decoding algorithm for the compound construction is still unknown. Finally, we showed that the compound LDGM-LDPC ensemble is optimal for lossy compression of a BES, i.e., given a point on the Shannon rate-distortion curve, there exists a code from this ensemble which can achieve it asymptotically as the blocklength is increasing. In addition to this optimality, we proved for the BES case that it was sufficient to consider compound codes whose degrees remain independent of the blocklength. This is in contrast with LDPC codes, which are known to be optimal for source coding of BES if their degrees are increasing at least logarithmically with the blocklength [8].

V.2 Future Research

In the following we list a few topics which might be of interest for future research.

Improving the Lower Bound on the Rate-Distortion Performance of LDGM Codes
As shown by Figure III.2, LDGM codes with finite degrees are suboptimal. Since the lower bound we derived in Theorem III.2.1 is close to the Shannon rate-distortion curve for high rates, there might exist a tighter bound which moves away from the Shannon's limit. By looking back at Chapter III, the biggest approximation seems to be (III.11), where we lower bounded the distortion by D if the corresponding source sequence was not in $\mathscr{C}_b^n(D)$ and by 0 otherwise. If a better lower bound on the distortion was available, we could obtain a tighter bound.

Although probably more difficult, an other way of improving this lower bound would be by proving the conjecture stated in (III.29). This result would be of huge importance since it would allow to fundamentally understand the link between the test-channel model and the rate-distortion problem, as well as providing a lower bound in the continuity of Gallager's work.

Iterative Encoding and Decoding Algorithm for Compound Codes
The current available analysis on the compound construction is made assuming optimal encoding and decoding due to the lack of efficient iterative algorithms for doing such tasks. Since compound LDGM-LDPC codes have a bounded graphical complexity, an important step toward the use of these codes in practical situations, would be to develop a variation of the message passing algorithm adapted to decode them [18].

Spatially Coupled LDGM and Compound Codes
In [34], the performance of spatially coupled LDPC codes was studied for the binary erasure channel (BEC) case. In particular, the authors showed that the belief propagation (BP) threshold for these spatially coupled codes converge toward the maximum a posteriori (MAP) threshold of single LDPC code, leading to a much faster saturation of Shannon's capacity. An interesting future research direction would be to study in a first step the impact of spatial coupling of LDGM codes on lossy compression for a variety of sources. Moreover, assuming the existence of a low complexity iterative decoding algorithm for the compound construction, a more long term project would be to analyze the properties of spatial coupling of compound codes for both source and channel coding.

A
The Royal Institute of Technology (KTH)

The aim of this chapter is to present the university and the laboratory which hosted me during this gap year. This chapter is organized as follows. In Section A.1 we provide a general presentation to KTH, then in Section A.2 we will present in more details the School of Electrical Engineering and finally we will present in Section A.3 the laboratory of Communication Theory in which I did my internship.

A.1 Presentation of KTH

The materials for this section mainly comes from [35].

KTH which stands for "Kungliga Tekniska Högskolan" in Swedish, is the country's largest university for technical and engineering education at university level. It was founded in 1827 in Stockholm, and mainly focused on applied technology as a way to merge academic and industrial perspectives.

Nowadays the university has 4 different campuses with around 12,000 full-year undergraduate students, more than 1,400 active postgraduate students and 2,800 full time employees. KTH is organized into 11 eleven schools each of them consisting of a certain number of departments:

- Architecture and Built Environment
- Biotechnology
- KTH Business Liaison
- Chemical Science and Engineering
- Computer Science and Communication
- **Electrical Engineering**
- Information and Communication Technology
- Industrial Engineering and Management
- Engineering Sciences
- Technology and Health
- Scientific Information and Learning

Although these figures are quite impressive, especially in comparison with our French universities, it does not reflect the reality of my daily work which is constituted by the Communication Theory group which belongs to the School of Electrical Engineering.

A.2 THE SCHOOL OF ELECTRICAL ENGINEERING

The figures and facts of this section mainly come from the 2008 annual report concerning the School of Electrical Engineering [36].

The School of Electrical Engineering is one of the eleven KTH schools and is organized in twelve laboratories:

- Automatic Control
- Communication Networks
- **Communication Theory**
- Electric Power Systems
- Electrical Machines and Power Electronics
- Electromagnetic Engineering
- Fusion Plasma Physics
- Industrial Information and Control Systems
- Microsystem Technology
- Signal Processing
- Sound and Image Processing
- Space and Plasma Physics

A.3 THE COMMUNICATION THEORY LABORATORY

The Communication Theory Laboratory is one of the twelve departments in the KTH School of Electrical Engineering and was founded in 2003. It is formed by two professors, one of them being my supervisor, around 20 PhD students and 5 postdocs [37]. In addition to its research activities, the group gives also several courses at PhD level on various subjects close to digital communications. To satisfy my personal curiosity, I attended two courses *Coding for Wireless Communications* and *Multiuser Information Theory*, as well as the different weekly internal seminars.

I am working in an open space among other PhD and master thesis students. Each week, I have usually several informal meetings with my co-supervisor Dr. Vishwambhar Rathi. These meetings are mainly made so that I can report my advances or problems, as well as going deep into various technical details to ensure that we have both the same understanding of the problem considered. Finally, we report our main advances to my supervisor Lars K. Rasmussen, who gives us some feedback and points out possible flaws.

References

[1] H. Nyquist, “Certain factors affecting telegraph speed,” *Bell System Technical Journal*, vol. 3, pp. 324–352, Apr. 1924.

[2] R. V. L. Hartley, “Transmission of information,” *Bell System Technical Journal*, vol. 7, pp. 535–563, Jul. 1928.

[3] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July and October 1948.

[4] C. Berrou, A. Glavieux, and P. Thitmajshima, “Near shannon limit error correcting coding and decoding: Turbo codes,” in *IEEE International Conference on Communications*, vol. 2, May 1993, pp. 1063–1070.

[5] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices,” *IEEE Transactions on Information Theory*, vol. 45, no. 2, Mar. 1999.

[6] R. G. Gallager, “Low-density parity-check codes,” Ph.D. dissertation, M.I.T., 1963.

[7] R. M. Tanner, “A recursive approach to low complexity codes,” *IEEE Transactions on Information Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.

[8] 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.

[9] S. Cilberti, M. Mezard, and R. Zecchina, “Lossy data compression with random gates,” *Physics Review Letters*, vol. 95, 2005.

[10] M. J. Wainwright and E. Maneva, “Lossy source encoding via message-passing and decimation over generalized codewords of LDGM codes,” in *Proc. of the IEEE International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1493–1497.

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

[12] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of LDGM codes," in *Proc. of the Allerton Conference on Communication, Control, and Computing*, Sep. 2007.

[13] E. Arikan, "Polar codes: A deterministic construction of capacity achieving codes," *IEEE Transactions on Information Theory*, vol. 50, no. 6, Jul. 2009.

[14] S. Korada and R. Urbanke, "Optimality of polar codes for lossy compression," submitted to IEEE Transactions on Information Theory.

[15] E. Martinian and M. J. Wainwright, "Analysis of LDGM and compound codes for lossy compression and binning," in *Proc. of the IEEE Information Theory Workshop*, San Diego, CA, USA, Feb. 2006.

[16] ——, "Low-density codes achieve the rate-distortion bound," in *Proc. of the Data Compression Conference*, Snowbird, UT, Mar. 2006.

[17] ——, "Low-density construction can achieve the Wyner-Ziv and Gelfand-Pinsker bounds," in *Proc. of the IEEE International Symposium on Information Theory*, Jul. 2006, pp. 484–488.

[18] M. J. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1061–1079, Mar. 2009.

[19] A. Dimakis, M. J. Wainwright, and K. Ramchandran, "Lower bounds on the rate-distortion function of LDGM codes," in *Proc. of the IEEE Information Theory Workshop*, 2007.

[20] S. Kudekar and R. Urbanke, "Lower bounds on the rate-distortion function of individual LDGM codes," in *5th International Symposium on Turbo Codes and Related Topics*, Lausanne, Switzerland, 2008.

[21] G. Demay, V. Rathi, and L. K. Rasmussen, "Rate distortion bounds for binary erasure source using sparse graph codes," in *Proc. of the Data Compression Conference*, Snowbird, UT, Mar. 2010.

[22] ——, "Optimality of LDGM-LDPC compound codes for lossy compression of binary erasure source," in *accepted to the International Symposium on Information Theory and its Applications*, Taichung, Taiwan, Oct. 2010.

[23] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2000.

[24] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[26] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[27] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs and belief propagation," in *Proc. of the International Symposium on Information Theory*, 1998, p. 117.

[28] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE National Convention Record*, pp. 142–163, 1959.

[29] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge University Press, 2009.

[30] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Transactions on Information Theory*, vol. 50, no. 6, Jun. 2004.

[31] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[32] A. D. Wyner, "A theorem on the entropy of certain binary sequences and applications: Part II," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 772–777, Nov. 1973.

[33] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1629–1638, Jun. 2002.

[34] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional ldpc ensembles perform so well over the bec," in *Proceedings of the International Symposium on Information Theory*, 2010, pp. 684–688.

[35] "About KTH," http://www.kth.se/om?l=en_UK.

[36] "Annual report 2008 - school of electrical engineering," http://www.kth.se/ees/omskolan/annualreport?l=en_uk, 2008.

[37] "Communication theory laboratory," http://www.kth.se/ees/omskolan/organisation/avdelningar/commth?l=en_UK.

GLOSSARY

achievable rate-distortion pair (X, 17)
Consider a source S, a source alphabet $\mathscr{S}$, a reproduction alphabet $\hat{\mathscr{S}}$, and a distortion measure d defined over $\mathscr{S} \times \hat{\mathscr{S}}$. A rate-distortion pair (R,D) is said to be achievable if for any $\varepsilon > 0$ there exists for a sufficiently large blocklength n_ε, a $(n_\varepsilon, M_\varepsilon)$ rate-distortion code defined over $\hat{\mathscr{S}}$ such that

$$\frac{1}{n}\log_{|\hat{\mathscr{S}}|} M_\varepsilon \leq R + \varepsilon,$$
$$\mathbb{E}\left[d\left(S^n, g\left(f\left(S^n\right)\right)\right)\right] \leq D + \varepsilon.$$

block code (11)
A *code* $\mathscr{C}$ of length n and cardinality M over a finite field $\mathbb{F}$ is a collection of M elements from $\mathbb{F}^n$, i.e.,

$$\mathscr{C}\left(n,M\right) = \left\{c_1^n, \cdots, c_M^n\right\}, c_i^n \in \mathbb{F}^n, 1 \leq i \leq M.$$

The elements of a code $\mathscr{C}\left(n,M\right)$ are called *codewords* and the parameter n is called the *blocklength*.

capacity (11)
Consider two discrete random variables X, Y defined over the finite sets $\mathscr{X}$, and $\mathscr{Y}$ with probability mass function p_X, and p_Y respectively. Let $\mathscr{Q}$ be the set of probability mass functions p_X defined on $\mathscr{X}$. Consider a discrete memoryless channel with input random variable X, and output random variable Y. Then the *capacity* C for this channel is defined by

$$C = \max_{p_X \in \mathscr{Q}} I(X;Y).$$

compound LDGM-LDPC code (VIII, 1, 37, 39, 40, 46, 49, 52)
Let $\mathbb{L}\left(\mathbf{G}\right)$, $\mathbb{M}\left(\mathbf{H}\right)$ be a LDGM code of rate R_G and a LDPC code of rate R_H respectively, where $\mathbf{G} \in \{0,1\}^{m \times n}$ and $\mathbf{H} \in \{0,1\}^{k \times m}$. Then, $\mathbb{M}\left(\mathbf{H}\right)\mathbf{G}$ defines a new linear block code with blocklength n and rate $R = R_G R_H$, called a compound LDGM-LDPC code and denoted by $\mathbb{C}\left(\mathbf{G},\mathbf{H}\right)$,

$$\mathbb{C}\left(\mathbf{G},\mathbf{H}\right) = \left\{c^n \in \{0,1\}^n : \exists w^m \in \{0,1\}^m \text{ s.t. } c^n = w^m\mathbf{G} \text{ and } \mathbf{H}(w^m)^T = \mathbf{0}\right\}.$$

compound LDGM-LDPC ensemble (1, 37, 39, 40, 42, 45, 49, 51)

The compound LDGM-LDPC ensemble denoted by $\mathfrak{C}(d_c, d_v, d'_c, m, n)$, is the set of compound LDGM-LDPC codes $\mathbb{C}(\mathbf{G}, \mathbf{H})$, where $\mathbf{G}$ and $\mathbf{H}$ are both chosen uniformly at random in the CRP LDGM ensemble $\mathfrak{L}_P(d_c, m, n)$ and in the standard (m, d_v, d'_c)-regular LDPC ensemble respectively.

conditional entropy (10)

Consider two discrete random variables (X, Y) defined over the finite set $\mathscr{X} \times \mathscr{Y}$ with joint probability mass function p_{XY}. The *conditional entropy* of the random variable Y given the random variable X is denoted by $H(Y \mid X)$ and is defined as

$$H(Y \mid X) \triangleq \sum_{(x,y) \in \mathscr{X} \times \mathscr{Y}} p_{XY}(x, y) \log_2 \frac{1}{p_{Y|X}(y \mid x)}.$$

CRP LDGM ensemble (VIII, 37, 38, 40, 42, 44, 46, 49, 51)

The CRP LDGM ensemble denoted by $\mathfrak{L}_P(d_c, m, n)$ contains every LDGM code $\mathbb{L}(\mathbf{G})$, where the generator matrix $\mathbf{G}$ lies in $\{0, 1\}^{m \times n}$ and is generated by the following procedure. Each check node is connected to d_c information bits chosen uniformly at random and with replacement.

discrete memoryless channel (9)

A *discrete channel* is a stochastic process characterized by a finite input set $\mathscr{X}$, a finite output set $\mathscr{Y}$ and a transition matrix $W : \mathscr{X} \to \mathscr{Y}$, where $W(y \mid x)$ denotes the probability of observing y, given that x was the channel input, $\forall (x, y) \in \mathscr{X} \times \mathscr{Y}$. The channel is said to be *memoryless* if the probability distribution of the output depends only on the input at that time, and is conditionally independent of past inputs and outputs.

discrete memoryless source (8, 17)

A *discrete source* is a sequence of random variables $\{S_i\}_{i=1}^{\infty}$ taking values in a finite set $\mathscr{S}$, called the *source alphabet*. If the S_i's are i.i.d., we speak of a *discrete memoryless source.*

distortion measure (VII, 16, 17, *see* single-letter distortion measure & multi-letter distortion measure)

dual code (12)

Consider a linear block code $\mathscr{C}(\mathbf{G})$, where $\mathbf{G} \in \mathbb{F}^{k \times n}$. The *dual code* $\mathscr{C}^{\perp}$ associated to $\mathscr{C}(\mathbf{G})$ is

$$\mathscr{C}^{\perp} = \left\{ v^n \in \mathbb{F}^n : \ c^n (v^n)^T = 0, \ \forall c^n \in \mathscr{C}(\mathbf{G}) \right\} = \left\{ v^n \in \mathbb{F}^n : \ \mathbf{G}(v^n)^T = 0 \right\}.$$

entropy (10, 11, 15, 33)

Consider a discrete random variable X defined over a finite set $\mathscr{X}$ with probability mass function p_X. The *entropy* of X is denoted by $H(X)$ and is defined as

$$H(X) \triangleq \sum_{x \in \mathscr{X}} p_X(x) \log_2 \frac{1}{p_X(x)}.$$

generator node degree distribution (IX, 1, 3, 15, 21, 23, 26, 30, 31, 34, 51)

For a given Tanner graph of a LDGM code, let L_i be the proportion of generator nodes having degree i. Then, the generator node degree distribution $L(x)$ is

$$L(x) \triangleq \sum_i L_i x^i.$$

joint entropy (10)

Consider two discrete random variables (X,Y) defined over the finite set $\mathscr{X} \times \mathscr{Y}$ with joint probability mass function p_{XY}. The *joint entropy* of the random variables X and Y is denoted by $H(X,Y)$ and is defined as

$$H(X,Y) \triangleq \sum_{(x,y)\in\mathscr{X}\times\mathscr{Y}} p_{XY}(x,y) \log_2 \frac{1}{p_{XY}(x,y)}.$$

LDGM code (VII, VIII, 14, 38, 39, 52)

A LDGM code is a linear block code defined by a *sparse* generator matrix. More precisely, consider a *sparse* generator matrix $\mathbf{G}$, where $\mathbf{G} \in \mathbb{F}^{k\times n}$. Then, a LDGM code denoted by $\mathbb{L}(\mathbf{G})$, is a linear block code of rate $R_G \leq \frac{k}{n}$ defined by $\mathbf{G}$, i.e.,

$$\mathbb{L}(\mathbf{G}) = \left\{ c^n \in \mathbb{F}^n : \exists w^k \in \mathbb{F}^k \text{ s.t. } c^n = w^k \mathbf{G} \right\}.$$

LDPC code (VII, 13, 39, 45–47, 52)

A LDPC code is a linear block code defined by a *sparse* parity-check matrix. More precisely, consider a *sparse* parity-check matrix $\mathbf{H}$, where $\mathbf{H} \in \mathbb{F}^{(n-k)\times n}$. Then, a LDPC code denoted by $\mathbb{M}(\mathbf{H})$, is a linear block code of rate $R_H \geq \frac{k}{n}$ induced by $\mathbf{H}$, i.e.,

$$\mathbb{M}(\mathbf{H}) = \left\{ c^n \in \mathbb{F}^n : \mathbf{H}(c^n)^T = 0 \right\}.$$

linear block code (VII, IX, X, 12–14, 37, 39, 41, 42)

A *block code* $\mathscr{C}(n,M)$ over the finite field $\mathbb{F}$ is said to be *linear* if the codewords of the code span a *linear subspace* of $\mathbb{F}^n$.

multi-letter distortion measure (IX, 16)

A multi-letter distortion measure between a source sequence $s^n \in \mathscr{S}^n$ and a reconstructed sequence $\hat{s}^n \in \hat{\mathscr{S}}^n$, induced by a single-letter distortion measure d is

$$d^{(n)}(s^n, \hat{s}^n) = \frac{1}{n}\sum_{i=1}^{n} d(s_i, \hat{s}_i).$$

mutual information (11)

Consider two discrete random variables (X,Y) defined over the finite set $\mathscr{X} \times \mathscr{Y}$ with joint probability mass function p_{XY}, and marginals p_X, p_Y. The *mutual information* between the random variables X and Y is denoted by $I(X;Y)$ and is defined as

$$I(X;Y) \triangleq \sum_{(x,y)\in\mathscr{X}\times\mathscr{Y}} p_{XY}(x,y) \log_2 \frac{p_{XY}(x,y)}{p_X(x)\, p_Y(y)}.$$

rate (3, 12, 21)

The *rate* R of a block code $\mathscr{C}\,(n,M)$ defined over $\mathbb{F}$ is

$$R = \frac{1}{n}\log_{|\mathbb{F}|} M.$$

rate-distortion code (VII, X, 17)

A (n,M) rate-distortion code is constituted by an encoding function f, and a decoding function g, where

$$\begin{aligned} f:&\ \mathscr{S}^n \rightarrow \{1,2,\cdots,M\}, \\ g:&\ \{1,2,\cdots,M\} \rightarrow \hat{\mathscr{S}}^n. \end{aligned}$$

rate-distortion function (X, 17, 18)

The rate-distortion function, denoted by $R(D)$, is the minimum of all rates R for a given distortion D, such that the rate-distortion pair (R,D) is achievable.

rate-distortion region (X, 17)

For a given source S, the rate-distortion region is the closure of the set of all achievable rate-distortion pairs (R,D).

regular LDGM code (14, 23, 31)

A (n,p,q)-regular LDGM code is a linear block code of length n characterized by a generator matrix $\mathbf{G}$, $\mathbf{G} \in \mathbb{F}^{k\times n}$, where $\mathbf{G}$ has exactly p ones per column and q ones per row.

regular LDPC code (13, 14)

A (n,p,q)-regular LDPC code is a linear block code of length n characterized by a parity-check matrix **H**, $\mathbf{H} \in \mathbb{F}^{(n-k)\times n}$, where $\mathbf{H}$ has exactly p ones per column and q ones per row.

single-letter distortion measure (IX, X, 16)

A single-letter distortion measure d is a mapping between the source alphabet $\mathscr{S}$, the reconstruction alphabet $\hat{\mathscr{S}}$, and the set of non negative real numbers $\mathbb{R}_+$

$$d:\ \mathscr{S} \times \hat{\mathscr{S}} \rightarrow \mathbb{R}_+.$$