
KTH Electrical Engineering

# Coding for the Wiretap Channel

MATTIAS ANDERSSON

Licentiate Thesis in Telecommunications
Stockholm, Sweden 2011

Coding for the Wiretap Channel



TRITA-EE 2011:026
ISSN 1653-5146
ISBN 978-91-7415-927-1

Communication Theory
School of Electrical Engineering
Royal Institute of Technology (KTH)
SE-100 44 Stockholm, Sweden

Tel. +46 8 790 7516, Fax. +46 8 790 7260
http://www.ee.kth.se

# Abstract

We consider code design for Wyner's wiretap channel. Optimal coding schemes for this channel require an overall code that is capacity achieving for the main channel, partitioned into smaller subcodes, all of which are capacity achieving for the wiretapper's channel. To accomplish this we introduce two edge type low density parity check (LDPC) ensembles for the wiretap channel. For the scenario when the main channel is error free and the wiretapper's channel is a binary erasure channel (BEC) we find secrecy capacity achieving code sequences based on standard LDPC code sequences for the BEC. However, this construction does not work when there are also erasures on the main channel. For this case we develop a method based on linear programming to optimize two edge type degree distributions. Using this method we find code ensembles that perform close to the secrecy capacity of the binary erasure wiretap channel (BEC-WT). We generalize a method of Méasson, Montanari, and Urbanke in order to compute the conditional entropy of the message at the wiretapper. This conditional entropy is a measure of how much information is leaked to the wiretapper. We apply this method to relatively simple ensembles and find that they show very good secrecy performance.

Based on the work of Kudekar, Richardson, and Urbanke, which showed that regular spatially coupled codes are capacity achieving for the BEC, we construct a regular two edge type spatially coupled ensemble. We show that this ensemble achieves the whole capacity-equivocation region for the BEC-WT.

We also find a coding scheme using Arıkans polar codes. These codes achieve the whole capacity-equivocation region for any symmetric binary input wiretap channel where the wiretapper's channel is degraded with respect to the main channel.

# Acknowledgments

I want to express my gratitude to my supervisors Prof. Mikael Skoglund and Asst. Prof. Ragnar Thobaben. I am grateful to Mikael for welcoming me to his research group and for introducing me to the world of information theory. Ragnar has always gone out of his way to help me with any aspect of research. Both of their doors have always been open and I thank them for their great patience.

This thesis would not have been written without the help of Dr. Vishwambhar Rathi. Doing research with him has been nothing less than spectacular. He has shared not only parts of his great knowledge about channel coding, but also many laughs with me. Most of all I am grateful to call him my friend.

I have collaborated with Asst. Prof. Jörg Kliewer on many papers in this thesis. I want to thank him for his insightful comments and inspiring ideas.

I have shared an office with Zhongwei Si for most of my time here. She truly makes every day brighter. I am also thankful to all my other friends and colleagues on the fourth floor for interesting discussions on life and research. I am indebted to Ricardo Blasco Serrano, Vish, Zhongwei, Mikael, Dr. Alan Sola, and especially Ragnar for helping me proofread my thesis. I would like to thank Annika Augustsson for handling all administrative matters with ease.

I wish to thank Dr. Michael Lentmaier for taking the time to act as an opponent to this thesis.

I want to express my deepest gratitude to my parents Jan and Agneta, my sisters Emma and Johanna and my brother Frans for their endless love and support. I also want to thank Vincent and Lorna Agnesi for welcoming me into their family.

Last but not least I want to thank Carla Agnesi for all the love, joy and happiness she brings to me from half a world away.

# Contents

# Chapter 1

# Introduction

Wireless communication is ubiquitous in today's society. Indeed, cell phones and Wifi networks are everywhere. Regrettably, wireless transmissions are by their broadcast nature open to eavesdropping. Everyone has the possibility to listen in to the communication between for example a computer and a wireless router. Such connections are usually secured through encryption protocols, relying on pre-shared keys and the computational difficulty of solving certain problems, for example, the prime factorization of large integers, or the calculation of discrete logarithms. This is not entirely satisfactory. Encryption protocols may have undiscovered weaknesses, and, perhaps a smaller concern, the computational hardness of these problems is only conjectured.

An example of the first problem is the Wired Equivalent Privacy (WEP) protocol. It was introduced in 1997 as part of the original IEEE 802.11 protocol but has since then been found wanting [FMS01]. Today there exist readily available tools that can break any WEP key in minutes, and that run on an off-the-shelf personal computer. WEP was declared deprecated in 2004 and has been replaced with newer protocols like WPA and WPA2 that do not share its flaws, but it is still in wide use.

The assumption that prime factorization and calculation of discrete logarithms is hard is not as big a concern as poorly implemented or designed protocols. Today no efficient algorithms for solving these problems on regular computers have been found, and it is widely believed that no such algorithms exist. However, there exist algorithms for both of these problems that run in polynomial time on *quantum computers* [Sho99]. There is a lot of research into quantum computing, and there have been

experimental demonstrations of Shor's algorithm for integer factorization [LBYP07, LWL+07].

In the field of Information Theoretic Security we take a different view of the problem. We assume that the eavesdropper has unlimited computational powers, rendering the approach of public-key cryptography useless. Instead we assume that the legitimate receiver of the message has a physical advantage over the eavesdropper. In the example of a wireless network we will assume that the legitimate receiver has a higher signal to noise ratio than the eavesdropper. One way of assuring this is by assuming that the eavesdropper is situated further from the transmitter than the legitimate receiver, for example that the eavesdropper is outside the building in which the wireless network is located. Based on this physical advantage we then use a randomized coding scheme to transmit information. The legitimate receiver has a better channel than the eavesdropper and is able to determine which information we send. The eavesdropper however is unable to obtain any information at all from her received signals.

## 1.1 Outline and Contributions

This section outlines the thesis and summarizes its contributions.

### Chapter 2

This chapter contains a review of fundamental results in information theory and coding needed for the rest of the thesis. It is divided into three parts. First we give an information theoretic overview of channel coding and in particular Wyner's wiretap channel. We also review previous work on coding for the wiretap channel. The second part is an overview of LDPC codes with a section devoted to spatially coupled LDPC codes. The third part is an introduction to polar codes.

### Chapter 3

In this chapter we introduce a two edge type LDPC ensemble for the wiretap channel. We give a construction that achieves the secrecy capacity when the main channel is noise-free. In the case of a noisy main channel we numerically optimize the ensemble and find codes

that operate close to the secrecy capacity. We also generalize a result from [MMU08] in order to be able to calculate the equivocation at the eavesdropper. Using this result we find relatively simple ensembles that have very good secrecy performance. Finally we introduce a spatially coupled two edge type LDPC ensemble. Based on the result shown in [KRU10], that one edge type spatially coupled LDPC codes are capacity achieving for the BEC we show that our construction achieves the whole capacity-equivocation region for the BEC wiretap channel. This chapter is based on the following published/submitted papers:

[RAT+09] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund. Two edge type LDPC codes for the wiretap channel. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pages 834 –838, 2009

[ART+10a] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund. Equivocation of Eve using two edge type LDPC codes for the erasure wiretap channel. In *Proceedings of Asilomar Conference on Signals, Systems and Computers (to appear)*, Nov. 2010

[RAT+10] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund. Performance Analysis and Design of Two Edge Type LDPC Codes for the BEC Wiretap Channel. *Submitted to IEEE Trans. on Inf. Theory*, Sep. 2010

[RUAS11] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund. Rate-Equivocation Optimal Spatially Coupled LDPC Codes for the BEC Wiretap Channel. *Submitted to Proc. IEEE Int. Sympos. Information Theory (ISIT)*, Jul. 2011

where [RAT+10] is a journal version of [RAT+09] and [ART+10a].

## Chapter 4

In this chapter we construct polar codes for binary input symmetric wiretap channels where the wiretapper's channel is degraded with respect to the main channel. We show that the construction achieves the whole rate-equivocation region. This chapter is based on the following published paper:

[ART+10b] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund. Nested polar codes for wiretap and relay channels. *IEEE Communications Letters*, 14(8):752 –754, Aug. 2010

## Chapter 5

In this chapter we conclude the thesis and point out some directions for possible future work.

## 1.2 Notation and Abbreviations

We will use the following notation and abbreviations throughout the thesis.

| | |
|---|---|
| $X$ | A random variable |
| $x$ | A realization of the random variable $X$ |
| $\mathcal{X}$ | The set (alphabet) which $X$ takes values in |
| $\|\mathcal{X}\|$ | The cardinality of $\mathcal{X}$ |
| $p_X(x)$ | The probability mass/density function of $X$ |
| $p_{Y\|X}(y\|x)$ | The conditional probability mass/density function of $Y$ conditioned on $X$ |
| $\mathbb{E}[X]$ | The expectation of $X$ |
| $H(X)$ | The entropy of $X$ |
| $H(X\|Y)$ | The conditional entropy of $X$ conditioned on $Y$ |
| $I(X;Y)$ | The mutual information between $X$ and $Y$ |
| $I(X;Y\|S)$ | The conditional mutual information between $X$ and $Y$ conditioned on $S$ |
| BEC($\epsilon$) | The binary erasure channel with erasure |

| | |
|---|---|
| | probability $\epsilon$ |
| $\mathrm{BEC}(\epsilon_m, \epsilon_w)$ | A wiretap channel where the main channel is a $\mathrm{BEC}(\epsilon_m)$ and the wiretapper's channel is a $\mathrm{BEC}(\epsilon_w)$ |
| $\log(x)$ | The logarithm to base 2 |
| $h(x)$ | The binary entropy function to base 2 |
| $\mathbb{1}_{\{S\}}$ | The indicator variable which is 1 if $S$ is true and 0 otherwise |
| $\mathrm{coef}\left\{\sum_i F_i D^i, D^j\right\}$ | The coefficient of $D^j$ in $\sum_i F_i D^i$ |
| $x^N$ | A vector with $N$ elements |
| $x_i^j$ | The vector $[x_i \; x_{i+1} \; \ldots \; x_{j-1} \; x_j]$ |
| $x_e^N$ | The vector consisting of the elements in $x^N$ with even indices |
| $x_o^N$ | The vector consisting of the elements in $x^N$ with odd indices |
| LDPC | Low Density Parity Check |
| b.p.c.u. | bits per channel use |

# Chapter 2

# Fundamentals

In this chapter we will review results used in later parts of the thesis. We will begin by a short introduction to channel coding and the classic result by Shannon [Sha48]. We will then give an overview of the wiretap channel as introduced by Wyner in [Wyn75]. We will give an introduction to LDPC codes, spatially coupled LDPC codes, and polar codes, which will be used in later chapter to construct codes for the wiretap channel.

## 2.1 Channel Coding

Channel coding is concerned with the communication problem depicted in Figure 2.1. At the source there is a message that we want to replicate at the destination. To do this we have a channel available. The channel can in general be any medium, for example a telephone line, the air, the Internet or a hard drive. Shannon studied this problem from a mathematical viewpoint in his revolutionary paper [Sha48] and quantified how much information the source can reliably, i.e. with low probability of error, transmit to the destination.


Figure 2.1: A Communication System

We define the channel by the triple $(\mathcal{X}, \mathcal{Y}, P_{Y^N|X^N})$, where $\mathcal{X}$ and $\mathcal{Y}$ are two sets called the *input alphabet* and the *output alphabet* respectively,

and $P_{Y^N|X^N}(y^N|x^N)$ are the channel transition probabilities for different number of channel uses $N$. $P_{Y^N|X^N}(y^N|x^N)$ is the probability of seeing the output $y^N$ at the channel when the input is $x^N$.

Note that in general we let the channel transition probability $P_{Y^N|X^N}$ depend on the block length $N$. If the channel transition probabilities factorize as

$$P_{Y^N|X^N}(y^N|x^N) = \prod_{i=1}^{N} P_{Y|X}(y_i|x_i)$$

we say that the channel is memoryless and write $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$.

An $(M, N)$ code for the channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ consists of a message set

$$\mathcal{M} = \{1, \ldots, M\}$$

of cardinality $M$, an encoder

$$f: \ \mathcal{M} \to \mathcal{X}^N,$$

and a decoder

$$g: \ \mathcal{Y}^N \to \mathcal{M}.$$

The rate $R$ of the code is defined as the logarithm of the number of codewords normalized with the length:

$$R = \frac{\log M}{N}.$$

The average decoding error probability is defined as

$$P_e^N = \frac{1}{M} \sum_{i=1}^{M} \Pr(g(Y^N) \neq i | X^N = f(i)),$$

and it is the probability of the decoder making an error when all of the possible messages in $\mathcal{M}$ are used with equal probability.

We say that a rate $R$ is achievable if there exists a sequence of $(\lceil 2^{NR_N} \rceil, N)$ codes such that for every $\epsilon > 0$

$$\liminf_{N\to\infty} R_N > R - \epsilon,$$
$$\lim_{N\to\infty} P_e^N < \epsilon.$$

We call the supremum of all achievable rates the *capacity* $C$ of the channel

$$C = \sup\{R : \ R \text{ is achievable}\}.$$

Shannon showed that the capacity is equal to the maximum mutual information $I(X;Y)$ between the input and the output of the channel, where the maximization is taken over all possible input distributions $P_X$:

$$C = \max_{P_X} I(X;Y). \tag{2.1}$$

We also define the *symmetric capacity* $I(P_{Y|X})$ of a channel as

$$I(P_{Y|X}) = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} \frac{1}{|\mathcal{X}|} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{\frac{1}{|\mathcal{X}|}\sum_{x'\in\mathcal{X}} p_{Y|X}(y|x')}.$$

This is the maximum achievable rate when all channel inputs $x$ are used with the same probability. If the maximizing distribution $P_X$ in (2.1) is the uniform distribution then the symmetric capacity is equal to the capacity.

One class of channels for which this is the case is the class of symmetric discrete memoryless channels. In order to define a symmetric discrete memoryless channel we note that we can write the transition probabilities of a discrete and memoryless channel in matrix form. Each row $i$ of the matrix correspond to a different input $x_i$ and each column $j$ corresponds to a different output $y_j$. The element in position $(i,j)$ is the channel transition probability $p_{Y|X}(y_j|x_i)$. Based on this matrix we have the following definition:

**Definition 2.1.1** (Symmetric discrete memoryless channel [Gal68])**.** *A discrete and memoryless channel is said to be symmetric if we can partition the set of outputs $y$ so that for each subset the matrix of transition probabilities corresponding to this subset fulfills:*

1. *The rows of the matrix are permutations of each other,*
2. *The columns of the matrix are permutations of each other.*

For an example of a symmetric channel see the following subsection, in which we define the binary erasure channel, a channel model that we will use frequently throughout the rest of the thesis.

### 2.1.1 The Binary Erasure Channel

The Binary Erasure Channel was introduced by Elias [Eli55] as a toy example. The practical interest in it, or rather in its generalization the packet erasure channel, has risen since the introduction of the Internet. The binary erasure channel with erasure probability $\epsilon$, or BEC($\epsilon$), is a memoryless channel with binary input alphabet $\mathcal{X} = \{0, 1\}$, a ternary output alphabet $\mathcal{Y} = \{0, 1, ?\}$ and channel transition probabilities given by:

$$\begin{aligned} P_{Y|X}(0|0) &= 1 - \epsilon \\ P_{Y|X}(1|0) &= 0 \\ P_{Y|X}(?|0) &= \epsilon \\ P_{Y|X}(0|1) &= 0 \\ P_{Y|X}(1|1) &= 1 - \epsilon \\ P_{Y|X}(?|1) &= \epsilon. \end{aligned}$$

In Figure 2.2 we see a representation of the different possible channel transitions and their probabilities. We see that the input is either reconstructed perfectly at the output, with probability $1 - \epsilon$, or erased, with probability $\epsilon$.

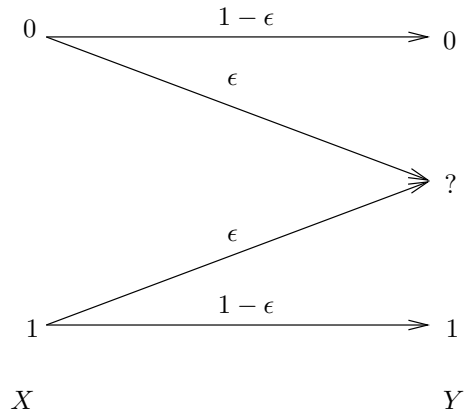


Figure 2.2: Binary Erasure Channel

We can write the channel transition probability matrix as

$$\begin{bmatrix} 1-\epsilon & \epsilon & 0 \\ 0 & \epsilon & 1-\epsilon \end{bmatrix}.$$

Rows one and two correspond to the inputs 0 and 1 respectively, and columns one, two, and three correspond to the outputs 0, ?, and 1 respectively. We now partition the output alphabet into the sets $\{0, 1\}$ and $\{?\}$. This gives us the following two transition probability matrices:

$$\begin{bmatrix} 1-\epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix}, \quad \begin{bmatrix} \epsilon \\ \epsilon \end{bmatrix}.$$

Since for both of these matrices the rows (and the columns) are a permutation of each other the BEC($\epsilon$) is a symmetric channel. Thus the maximizing input distribution is the uniform distribution, and the capacity, as well as the symmetric capacity, is $1 - \epsilon$.

In the next section we give a short information theoretic introduction to the wiretap channel. We also present a code construction method based on linear nested codes which will be used in the main part of the thesis.

## 2.2 The Wiretap Channel

In [Wyn75] Wyner introduced the notion of a wiretap channel which is depicted in Figure 2.3. It is the most basic channel model that takes security into account. A wiretap channel consists of an input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}$, and $\mathcal{Z}$, and a transition probability $P_{YZ|X}(y, z|x)$. We call the marginal channels $P_{Y|X}$ and $P_{Z|X}$ the main channel and the wiretapper's channel respectively.

In a wiretap channel, Alice communicates a message $S$, which is chosen uniformly at random from the message set $\mathcal{S}$, to Bob through the main channel. Alice performs this task by encoding $S$ as a vector $X^N$ of length $N$ and transmitting $X^N$. Bob and Eve receive noisy versions of $X^N$, which we denote by $Y^N$ and $Z^N$, via their respective channels.

The encoding of a message $S$ by Alice should be such that Bob is able to decode $S$ reliably and $Z^N$ provides as little information as possible to Eve about $S$. To measure the amount of information that Eve receives about $S$ we use the following normalized conditional entropy $H(S|Z^N)/N$ which we call the *equivocation rate*.

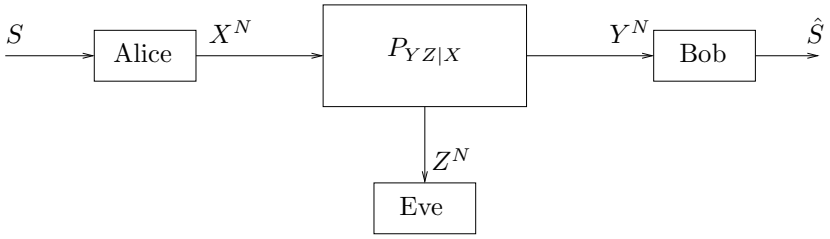


Figure 2.3: Wiretap channel.

A code of rate $R_N$ with block length $N$ for the wiretap channel is given by a message set $\mathcal{S}$ of cardinality $|\mathcal{S}| = 2^{\lceil NR_N \rceil}$, and a collection of disjoint subcodes $\{\mathcal{C}_s \subset \mathcal{X}^N\}_{s\in\mathcal{S}}$. To encode the message $s \in \mathcal{S}$, Alice chooses one of the codewords in $\mathcal{C}_s$ uniformly at random and transmits it. Bob uses a decoder $\phi : \mathcal{Y}^N \to \mathcal{S}$ to determine which message was sent. We assume that all messages are equally likely. Let $P_e^N$ be the average decoding error probability for Bob

$$P_e^N = \Pr(\phi_n(Y^N) \neq S)$$

and let $R_e^N$ be the equivocation rate of Eve

$$R_e^N = \frac{1}{N} H(S|Z^N).$$

The equivocation rate is a measure of how much uncertainty Eve has about the message $S$ after observing $Z^N$. We want $R_e^N$ to be as high as possible, and ideally it should equal the rate $R$. For ease of notation, whenever we say equivocation in the rest of the thesis we will mean the equivocation rate.

A rate-equivocation pair $(R, R_e)$ is said to be achievable if, for every $\epsilon > 0$, there exists a sequence of codes of rate $R_N$ and length $N$, and decoders $\phi_N$ such that the following reliability and secrecy criteria are satisfied:

$$\text{Rate}: \liminf_{N\to\infty} R_N > R - \epsilon, \tag{2.2}$$

$$\text{Reliability}: \lim_{N\to\infty} P_e^N < \epsilon, \tag{2.3}$$

$$\text{Secrecy}: \liminf_{N\to\infty} R_e^N > R_e - \epsilon. \tag{2.4}$$

The capacity-equivocation region is the closure of all achievable pairs $(R, R_e)$.

For a general wiretap channel the capacity-equivocation region is given by the rate-equivocation pairs $(R, R_e)$ satisfying

$$\bigcup_{P_{QU}P_{X|U}P_{YZ|X}} \left\{ \begin{array}{c} (R, R_e): \\ 0 \leq R \leq I(U;Y|Q), \\ 0 \leq R_e \leq R, \\ R_e \leq I(U;Y|Q) - I(U;Z|Q) \end{array} \right\}, \tag{2.5}$$

for some random variables $U, Q$ that form the Markov chain $Q \to U \to X \to (Y, Z)$ [CK78]. $U$ corresponds to the message, and it is split into two parts. One part is $Q$ which can be decoded by Eve, while the other part can be kept secret from her. We also see that the capacity-equivocation region only depends on the marginal transition probabilities $P_{Y|X}$ and $P_{Z|X}$.

The highest $R$, such that the pair $(R, R)$ is achievable, is called the *secrecy capacity*. In this case $R = R_e$, which we call *perfect secrecy*. This is equivalent to $\lim_{N\to\infty} I(S, Z^N)/N = 0$, or $\lim_{N\to\infty} H(S|Z^N)/N = R$, and means that the information leakage to the wire-tapper goes to zero rate-wise. The secrecy capacity for a general wiretap channel is

$$C_S = \max_{P_{UX}} I(U;Y) - I(U;Z),$$

where $U$ satisfies the Markov chain $U \to X \to (Y, Z)$. As expected there is no common part $Q$ that can be decoded by Eve. Note that the secrecy capacity is always non-negative since we can choose $U$ and $X$ to be independent. This will ensure that $I(U;Y) - I(U;Z) = 0$.

One could also consider the case where the mutual information between $S$ and $X^N$ is required to go to zero instead of just the mutual information rate, i.e

$$\lim_{N\to\infty} I(S|Z^N) = 0$$

instead of

$$\lim_{N\to\infty} \frac{I(S|Z^N)}{N} = 0.$$

This constraint is called *strong secrecy*, whereas the constraint given in (2.4) is called *weak secrecy*. Maurer and Wolf showed that the secrecy capacity using the strong notion of secrecy is the same as the weak secrecy

capacity if Alice and Bob are allowed to communicate over a noiseless public channel in addition to the wiretap channel [MW00]. We will only consider the case of weak secrecy in this thesis.

If there exists a channel transition probability $P_{Z|Y'}$ with input alphabet $\mathcal{Y}$ such that

$$P_{Z|X}(z|x) = \sum_{y' \in \mathcal{Y}} P_{Y|X}(y'|x) P_{Z|Y'}(z|y') \quad \forall z, x$$

we say that the wiretapper's channel is *stochastically degraded* with respect to the main channel. If the channel transition probability $P_{ZY|X}$ factorizes as

$$P_{YZ|X}(y, z|x) = P_{Y|X}(y|x) P_{Z|Y}(z|y)$$

we say that the wiretapper's channel is *physically degraded* with respect to the main channel. Since the capacity-equivocation region only depends on the marginal probabilities, the capacity-equivocation region for physically and stochastically degraded wiretap channels is the same and is given by [CK78]:

$$\bigcup_{P_X P_{YZ|X}} \left\{ \begin{array}{c} (R, R_e): \\ 0 \leq R \leq I(X;Y) \\ 0 \leq R_e \leq R \\ R_e \leq I(X;Y) - I(X;Z) \end{array} \right\}. \tag{2.6}$$

In this case the secrecy capacity is

$$C_S = \max_{P_X} I(X;Y) - I(X;Z).$$

The simplified region in (2.6) actually holds for more general channels than degraded channels. Assume that $I(U;Z) \leq I(U;Y)$ for all $U$ such that $U \to X \to (Y, Z)$ is a Markov chain. If this condition holds we say that the channel to Bob is *less noisy* than the channel to Eve. Degradedness is a stronger condition than less noisy. It is straightforward to show that every bound in (2.5) is smaller than the corresponding bound in (2.6) using that $I(U;Z) \leq I(U;Y)$. The less noisy region is also easy to achieve by choosing $U = X$ and $Q = \emptyset$.

In the less noisy case, if the same input distribution $P_X$ maximizes both $I(X;Y)$ and $I(X;Z)$, for example when both $P_{Y|X}$ and $P_{Z|X}$ are symmetric channels, the capacity-equivocation region is given by

$$R_e \leq R \leq C_M, \quad 0 \leq R_e \leq C_M - C_W, \tag{2.7}$$

and the secrecy capacity is

$$C_s = \max(0, C_M - C_W),$$

where $C_M$ and $C_W$ are the capacities of the main and the wiretapper's channels respectively. The rate region described by (2.7) is depicted in Figure 2.4. The line AB corresponds to points with perfect secrecy, and the point C corresponds to using the main channel at full rate.

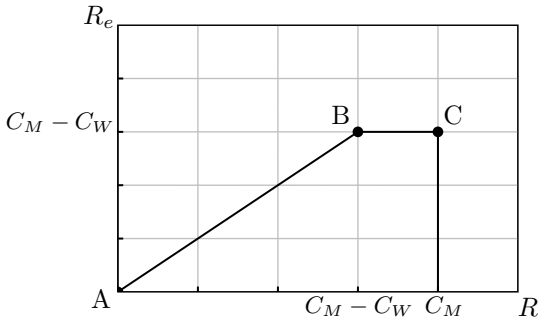


Figure 2.4: Capacity-equivocation region for a degraded symmetric wiretap channel.

When both the main channel and the wiretapper's channel are binary erasure channels we call the resulting wiretap channel the binary erasure wiretap channel, and we denote it by BEC-WT($\epsilon_m, \epsilon_w$). Here $\epsilon_m$ and $\epsilon_w$ are the erasure probabilities of the main channel and the wiretapper's channel respectively. If $\epsilon_w \geq \epsilon_m$, the BEC-WT($\epsilon_m, \epsilon_w$) is a symmetric degraded wiretap channel and its capacity-equivocation region is given by

$$R_e \leq R \leq 1 - \epsilon_m, \quad 0 \leq R_e \leq \epsilon_w - \epsilon_m,$$

and the secrecy capacity is

$$C_s = \epsilon_w - \epsilon_m.$$

A detailed information theoretic overview of general wiretap channels can be found in [LPSS09].

In the next subsection we present a coding strategy based on cosets of linear codes introduced by Wyner.

### 2.2.1 Nested Codes

Wyner and Ozarow used the following coset encoding strategy [Wyn75, OW84] to show that perfect secrecy can be achieved when the main chan-

nel is error free and the input alphabet is binary. Similar nested code structures for other multiterminal setups were considered in [ZSE02]. The secrecy capacity of the wiretap channel considered by Wyner and Ozarow is $1 - C_W$. Let $\mathcal{C}_0$ be the binary linear code of rate $R_0$ defined by the parity check check equation $Hx^N = 0$. The coset $\mathcal{C}_s$ is the set

$$\mathcal{C}_s = \{x^N : Hx^N = s\}.$$

To transmit the binary message $s$, Alice chooses one of the messages in $\mathcal{C}_s$ uniformly at random. Since there are $2^N/2^{NR_0}$ different cosets, the rate of the coding scheme is $1 - R_0$. Bob decodes by multiplying $H$ with $x$. If $\mathcal{C}_0$ comes from a capacity approaching sequence of linear codes both the rate and the equivocation can be made as close to $1 - C_W$ as wanted. To see this we consider the similar code construction method for a noisy main channels using nested codes introduced in [TDC+07]:

**Definition 2.2.1** (Wiretap code $\mathcal{C}_N$ with coset encoding)**.** *Let $H$ be an $N(1 - R^{(1,2)}) \times N$ parity check matrix with full rank, and let $\mathcal{C}^{(1,2)}$ be the code whose parity-check matrix is $H$. Let $H_1$ and $H_2$ be the sub-matrices of $H$ such that*

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix},$$

*where $H_1$ is an $N(1-R^{(1)}) \times N$ matrix and $H_2$ is an $NR \times N$ matrix. We see that $R = R^{(1)} - R^{(1,2)}$. Let $\mathcal{C}^{(1)}$ be the code with parity-check matrix $H_1$. Alice uses the following* coset encoding method *to communicate her message to Bob.*

**Coset Encoding Method:** *Assume that Alice wants to transmit a message whose binary representation is given by an $NR$-bit vector $S$. To do this she transmits $X^N$, which is a randomly chosen member of the coset*

$$\mathcal{C}_S = \left\{ X^N : \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} X^N = \begin{bmatrix} 0 \\ S \end{bmatrix} \right\}.$$

*Bob uses the following* syndrome decoding *to retrieve the message from Alice.*

**Syndrome Decoding:** *After observing $Y^N$, Bob obtains an estimate $\hat{X^N}$ for $X^N$ using the parity check equations $H_1 X^N = 0$. Then he computes an estimate $\hat{S}$ for $S$ as $\hat{S} = H_2 \hat{X^N}$.*

*We call this the wiretap code $\mathcal{C}_N$.*

We see that $\mathcal{C}^{(1)}$ can be partitioned into $2^{NR}$ disjoint subsets given by the cosets of $\mathcal{C}^{(1,2)}$. This is a generalization of Wyner's construction

above. To see this note that in Wyner's construction $\mathcal{C}^{(1,2)}$ is the set of all binary vectors of length $N$, and $\mathcal{C}^{(1)} = \mathcal{C}_0$.

Now assume that $\mathcal{C}^{(1)}$ comes from a capacity achieving sequence over the main channel and that $\mathcal{C}^{(1,2)}$ comes from a capacity achieving sequence over the wiretapper's channel[1]. Thangaraj *et al.* [TDC+07] showed that in this case the coset encoding scheme achieves $\lim_{N\to\infty} P_e^N = 0$ and $\lim_{N\to\infty} I(S;Z^N)/N = 0$.

It is easy to see that the error probability over the main channel goes to zero. Since $\mathcal{C}^{(1)}$ is capacity achieving over the main channel Bob can determine which codeword $X^N$ was sent with arbitrarily low probability of error, and then multiply $H_2$ by $X^N$ to obtain $S$.

To bound the mutual information $I(S;Z^N)$, we use the chain rule of mutual information on $I(X^N, S; Z^N)$ in two ways:

$$I(X^N;Z^N) + I(S;Z^N|X^N) = I(S;Z^N) + I(X^N;Z^N|S).$$

Since $S \to X^N \to Z^N$ is a Markov chain, $I(S;Z^N|X^N) = 0$, and we get

$$\begin{aligned} I(S;Z^N) =& I(X^N;Z^N) - I(X^N;Z^N|S) \\ =& I(X^N;Z^N) - H(X^N|S) + H(X^N|Z^N,S) \\ \leq& NC_W - NR^{(1,2)} + H(X^N|Z^N,S), \end{aligned}$$

where we have used that $I(X^N;Z^N) \leq NC_W$ and that $H(X^N|S) = NR^{(1,2)}$ in the last step. Since $\mathcal{C}^{(1,2)}$ is capacity achieving we must have $\lim_{N\to\infty} R^{(1,2)} = C_W$. To bound $H(X^N|Z^N,S)$ we use Fano's inequality:

$$H(X^N|Z^N,S) \leq h(P_e^{N,S}) + P_e^{N,S} N R^{(1,2)},$$

where $P_e^{N,S}$ is the error probability of decoding $X^N$ when knowing $Z^N$ and the coset $S$, and $h(x)$ is the binary entropy function. Since all the cosets $\mathcal{C}_S$ are capacity achieving over the wiretapper's channel we have $\lim_{N\to\infty} P_e^{N,S} = 0$. In total we get

$$\lim_{N\to\infty} \frac{I(S;Z^N)}{N} \leq \lim_{N\to\infty} \left( C_W - R^{(1,2)} + \frac{h(P_e^{N,S})}{N} + P_e^{N,S} R^{(1,2)} \right) = 0.$$

□

In the next subsection we give a short overview of previous work on coding for the wiretap channel.

[1]Since the cosets are just translations of each other, this implies that all cosets $C_s$ are capacity achieving over the wiretapper's channel. Equivalently, conditioned on which coset $S$ a codeword $x^N$ belongs to, the error probability of the wiretapper can be made arbitrarily small.

### 2.2.2 Previous Work

Thangaraj *et al.* [TDC+07] considered nested LDPC codes for the case when the main channel is noiseless, but no explicit construction was given for the case of a noisy main channel. Liu *et al.* also considered noiseless main channels in [LLPS07], with a BEC, BSC, or an AWGN channel to the wiretapper.

In [LPSL08] Liu *et al.* considered nested codes designed for the BEC-WT used over general binary input symmetric channels for transmission at rates below the secrecy capacity. In [CV10] Chen and Vinck showed that nested random linear codes can achieve the secrecy capacity over the binary symmetric wiretap channel and an upper bound on the information leakage was derived.

In [SST+10] Suresh *et al.* suggested a coding scheme for the BEC-WT that guarantees strong secrecy for a noiseless main channel and some range of $\epsilon_w$ using duals of sparse graph codes.

That nested polar codes are capacity achieving for the wiretap channel was shown by several research groups independently. The results by Hof and Shamai [HS10], Mahdavifar and Vardy [MV10], and Koyluoglu and El Gamal [OE10] are closely related to the results we show in Chapter 4.

In the next section we introduce LDPC codes. They are the building blocks for the wiretap codes we consider in Chapter 3.

## 2.3 LDPC Codes

Low Density Parity Check codes, or LDPC codes, were introduced by Gallager in his PhD thesis [Gal63]. Following the success of Turbo codes they were studied in the 1990's in work by MacKay and Neal [MN95], Luby, Mitzenmacher, Shokrollahi, Spielman, and Stemann [LMS+97], Richardson and Urbanke [RSU01], and many others. We will give a short introduction and give the results we need. For a detailed overview see [RU08].

Low density parity check codes are linear codes defined by a parity check matrix. We will consider binary codes, where all operations are carried out in the binary field. Consider the linear code $\mathcal{C}$ defined by the parity check matrix $H$, that is

$$\mathcal{C} = \{x^N : Hx^N = 0\}.$$

To each parity check matrix we associate a bipartite *Tanner graph* in the following way [Tan81]. We refer to the two types of nodes in the

bipartite graph as *variable nodes* and *check nodes* respectively. Each row in $H$ corresponds to a check node, and each column in $H$ corresponds to a variable node. The check node $i$ and the variable node $j$ are connected with an edge if element $(i, j)$ in $H$ is 1. The Tanner graph in Figure 2.5 corresponds to the check matrix

$$H = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

and has the variable node names and check equations written out.

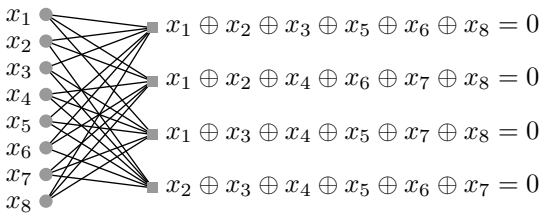


Figure 2.5: Tanner graph of an LDPC code of length $N = 8$.

The following compact notation for the degree sequences of an LDPC code was introduced by Luby *et al.* in [LMSS01a]. Let $\Lambda_{\mathtt{l}}$ be the fraction of variable nodes of degree $\mathtt{l}$ and let $\Gamma_{\mathtt{r}}$ be the fraction of check nodes of degree $\mathtt{r}$ in the Tanner graph, and let $\Lambda(x)$ and $\Gamma(x)$ be the polynomials defined by

$$\Lambda(x) = \sum_{\mathtt{l}=1}^{\mathtt{l}_{\max}} \Lambda_{\mathtt{l}} x^{\mathtt{l}}, \ \Gamma(x) = \sum_{\mathtt{r}=1}^{\mathtt{r}_{\max}} \Gamma_{\mathtt{r}} x^{\mathtt{r}},$$

where $\mathtt{l}_{\max}$ and $\mathtt{r}_{\max}$ are the largest variable node and check node degrees respectively. For the graph in Figure 2.5 we have $\Lambda(x) = x^3$ and $\Gamma(x) = x^6$.

We call $(\Lambda(x), \Gamma(x))$ the degree distribution from the node perspective of the Tanner graph. We also define the degree distribution from the edge perspective. Let $\lambda_{\mathtt{l}}$ be the fraction of edges in the graph connected to a variable node of degree $\mathtt{l}$ and $\rho_{\mathtt{r}}$ be the fraction of edges connected to a

check node of degree $\mathtt{r}$. Define the polynomials

$$\lambda(x) = \sum_{\mathtt{l}=1}^{\mathtt{l}_{\max}} \lambda_{\mathtt{l}} x^{\mathtt{l}-1}, \ \rho(x) = \sum_{\mathtt{r}=1}^{\mathtt{r}_{\max}} \rho_{\mathtt{r}} x^{\mathtt{r}-1}.$$

For the graph in Figure 2.5 we have $\lambda(x) = x^2$ and $\rho(x) = x^5$.

Let $N$ be the number of variable nodes in a Tanner graph, $M$ the number of check nodes, and $E$ the number of edges. We can find the following relations

$$\begin{aligned} E = N\Lambda'(1) &= M\Gamma'(1), \\ \lambda_{\mathtt{l}} = \frac{\mathtt{l}\Lambda_{\mathtt{l}}}{\sum_{\mathtt{k}=1}^{\mathtt{l}_{\max}} \mathtt{k}\Lambda_{\mathtt{k}}}, \ \rho_{\mathtt{r}} &= \frac{\mathtt{r}\Gamma_{\mathtt{r}}}{\sum_{\mathtt{k}=1}^{\mathtt{r}_{\max}} \mathtt{k}\Gamma_{\mathtt{k}}}, \\ \lambda(x) = \frac{\Lambda'(x)}{\Lambda'(1)}, \ \rho(x) &= \frac{\Gamma'(x)}{\Gamma'(1)}, \\ \Lambda_{\mathtt{l}} = \frac{\frac{\lambda_{\mathtt{l}}}{\mathtt{l}}}{\sum_{\mathtt{k}=1}^{\mathtt{l}_{\max}} \frac{\lambda_{\mathtt{k}}}{\mathtt{k}}}, \ \Gamma_{\mathtt{r}} &= \frac{\frac{\rho_{\mathtt{r}}}{\mathtt{r}}}{\sum_{\mathtt{k}=1}^{\mathtt{r}_{\max}} \frac{\rho_{\mathtt{k}}}{\mathtt{k}}}, \end{aligned}$$

where $f'(x)$ denotes the derivative of the function $f(x)$.

If all rows of the parity check matrix $H$ are linearly independent, then the rate of the code defined by $H$ is

$$R_{\text{des}} = 1 - \frac{M}{N} = 1 - \frac{\Lambda'(1)}{\Gamma'(1)} = 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx}.$$

We call this the design rate of the code. Note that when the connections in the Tanner graph are chosen randomly the check equations might not be independent, and the true rate of the code might be larger than the design rate. Both the actual rate and the design rate of the graph in Figure 2.5 are 1/2.

Given a degree distribution $(\Lambda(x), \Gamma(x))$ and a block length $N$ define the standard ensemble of LDPC codes as follows:

**Definition 2.3.1** (LDPC$(N, \Lambda(x), \Gamma(x))$)**.** *The LDPC$(N, \Lambda(x), \Gamma(x))$ ensemble is the collection of all bipartite graphs that have $N\Lambda_{\mathtt{l}}$ variable nodes of degree $\mathtt{l}$ and $N\frac{\Lambda'(1)}{\Gamma'(1)}\Gamma_{\mathtt{r}}$ check nodes of degree $\mathtt{r}$ for all $\mathtt{l}$ and $\mathtt{r}$. We allow multiple edges between two nodes. We impose a probability distribution on the ensemble by fixing one member of it and then permuting the endpoints of all edges on the check node side using a permutation of $E$ objects chosen uniformly at random.*

Note that we allow multiple edges between a variable and check node. To create a parity check matrix from a Tanner graph with multiple edges let the corresponding entry in $H$ be one if the variable and check node are connected with an odd number of edges and zero otherwise.

In the following subsection we describe the belief propagation decoder when the LDPC code is used over a BEC.

### 2.3.1 The Belief Propagation Decoder for the BEC

The belief propagation decoder is a message passing decoder. This means that the nodes in the Tanner graph exchange messages with their neighbors[2]. For general channels these messages are related to the probabilities of the variable nodes being 1 or 0, but for the BEC these messages take a simple form. A node can send the message 0, 1, or ? to its neighbor. We call ? the erasure message.

1. We first look at a message from a variable node to a check node. If a variable node knows its value, either from the channel observation or from incoming messages from other check nodes in previous iterations, it sends that value to the check node, otherwise it sends the erasure message.

2. Now look at a message from a check node to a variable node. If any incoming messages to the check node from other variable nodes are the erasure message, then the check node sends the erasure message. Otherwise it calculates the XOR of all incoming messages from other variable nodes and sends this value as the message.

3. In the final step we update the values of all variable nodes. If an unknown variable node receives an incoming message which is not the erasure message it becomes known.

4. If any unknown variable nodes were recovered in this iteration go to step 1. Otherwise, if all variable nodes are known, return the decoded codeword. Otherwise stop and declare an error.

Luby *et al.* analyzed the BP decoder for the BEC($\epsilon$) using the following density evolution method in [LMS$^+$97] and [LMSS01a]. Consider transmission over the BEC($\epsilon$) using a code from the LDPC($\lambda(x), \rho(x)$) ensemble.

[2]We say that two nodes are neighbors if they are connected by an edge.

Let $x^{(k)}$ be the probability that a variable node sends the erasure message in iteration $k$. Clearly $x^{(1)} = \epsilon$. Similarly let $y^{(k)}$ be the probability that a check node sends the erasure message in iteration $k$. Consider an edge connected to a variable node of degree $\mathtt{l}$. This outgoing message is an erasure if the incoming message from the channel, and all incoming messages on the other edges are erasures. This happens with probability $\epsilon(y^{(k-1)})^{\mathtt{l}-1}$. Averaging over all incoming edges we get

$$x^{(k)} = \sum_{\mathtt{l}} \lambda_{\mathtt{l}} \epsilon (y^{(k-1)})^{\mathtt{l}-1} = \epsilon \lambda(y^{(k-1)}) \tag{2.8}$$

Now consider an edge connected to a check node of degree $\mathtt{r}$. The outgoing message on this edge is an erasure unless all the incoming $\mathtt{r}-1$ messages are not erasures. Thus the probability that this outgoing message is an erasure is $1-(1-x^{(k)})^{\mathtt{r}-1}$. Averaging over all incoming messages we get

$$y^{(k)} = \sum_{\mathtt{r}} \rho_{\mathtt{r}} (1-(1-x^{(k)})^{\mathtt{r}-1}) = 1-\rho(1-x^{(k)}). \tag{2.9}$$

Putting (2.8) and (2.9) together we get

$$x^{(k+1)} = \epsilon \lambda(1-\rho(1-x^{(k)})),$$

which we call the density evolution recursion equation. This equation will correctly predict the erasure probability if the neighborhood of a variable node up to distance $k+1$ is a tree. For any fixed $k$ the probability that this neighborhood is not a tree goes to zero as $N$ goes to infinity.

Successful decoding is equivalent to $x^{(k)} \to 0$. This happens if the function

$$f_\epsilon(x) = \epsilon \lambda(1-\rho(1-x))$$

has no fixed points for $x$ in the range $(0, \epsilon)$.

Let

$$\epsilon^{\mathrm{BP}} = \sup_{\epsilon \in (0,1)} \left\{ f_\epsilon(x) \text{ has no fixed point for } x \in (0,\epsilon) \right\}.$$

If $\epsilon < \epsilon^{\mathrm{BP}}$ then the average error probability when communicating over the BEC($\epsilon$) using a randomly chosen code from LDPC($N, \lambda(x), \Gamma(x)$) and using the belief propagation decoding method goes to zero almost surely as $N \to \infty$. Conversely, if $\epsilon > \epsilon^{BP}$ the average error probability is always

bounded away from zero. $\epsilon^{BP}$ is called the belief propagation threshold for the degree distribution $(\lambda, \rho)$.

In the following subsection we describe a method to calculate the conditional entropy $H(X^N|Y^N)$ introduced by Méasson, Montanari and Urbanke in [MMU08].

### 2.3.2 MAP Decoding

In [MMU08], Méasson, Montanari and Urbanke considered the conditional entropy $H(X^N|Y^N)$ of the transmitted codeword $X^N$ conditioned on the received sequence $Y^N$ when using LDPC codes over the BEC. They found a criterion on the degree distribution $(\lambda(x), \rho(x))$ and the erasure probability $\epsilon$, that when satisfied allows the calculation of $\lim_{N\to\infty} H(X^N|Y^N)/N$.

Consider transmission over the BEC using an LDPC code. The *Peeling decoder* introduced by Luby *et al.* in [LMS+97] is an iterative message passing decoder equivalent to belief propagation. The peeling decoder removes edges and nodes from the graph as the variables get recovered. When no more recovery is possible it returns the resulting graph. We call this the residual graph $G_{\text{res}}$ and an empty residual graph corresponds to successful decoding. We now describe the decoding algorithm.

At each check node we introduce a book-keeping bit. The value of this bit is the sum of all known neighbouring nodes.

1. Initialize all variable nodes to the received value and calculate the book-keeping bit at each check node.

2. For each variable node $v$ in $G$. If $v$ is known, update the book-keeping bits of all connected check nodes. Then remove $v$ and all its edges from $G$. Otherwise do nothing.

3. For each check node $c$ in $G$. If $c$ has degree one, declare its neighboring variable node known and give it the value of the book-keeping bit. Then remove $c$ and its edge from $G$. Otherwise do nothing.

4. If no changes were made to the graph in the last iteration return $G$, otherwise go to 2.

In Figure 2.6 we show the peeling decoder applied to the code defined by the Tanner graph in Figure 2.5. The sent codeword is 11101101 and the received word is 1??01?01. In the initialization step it removes all known variable nodes and their edges from the graph. In the first iteration the

decoder manages to recover $x_3$ since the third check node has degree 1, but then it gets stuck since all remaining check nodes have degree at least 2. The resulting residual graph is the one on the right in Figure 2.6.

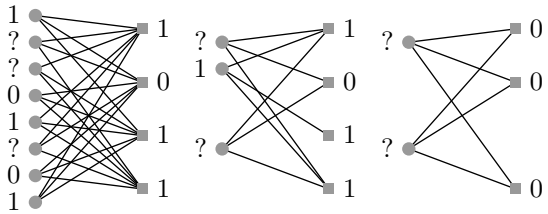


Figure 2.6: Peeling decoder

Now consider the ensemble of residual graphs defined as follows. Choose a graph at random from the ensemble LDPC$(N, \Lambda(x), \Gamma(x))$, transmit a codeword over the BEC$(\epsilon)$, and decode it using the peeling decoder. Call the resulting residual graph $G$ and its degree distribution from the node perspective $(\Omega, \Phi)$. It was shown in [LMSS01b] that conditioned on the degree distribution $(\Omega, \Phi)$ all residual graphs $G$ are equally likely. It was shown in [MMU08] that the residual degree distribution $(\Omega, \Phi)$ is concentrated around its expected value. This expected value converges to $(\Lambda_\epsilon(z), \Gamma_\epsilon(z))$ as $N$ goes to infinity, where

$$\Lambda_\epsilon(z) = \epsilon\Lambda(zy),$$
$$\Gamma_\epsilon(z) = \Gamma(1 - x + zx) - \Gamma(1 - x) - zx\Gamma'(1 - x),$$

where $x$ is the fixed point of the density evolution equation $x_k = \epsilon\lambda(1 - \rho(1 - x_{k-1}))$ when initialized with $x_0 = \epsilon$, and $y = \rho(1 - x)$. Here the degree distributions $(\Lambda_\epsilon, \Gamma_\epsilon)$ and $(\Omega, \Phi)$ are normalized with respect to the number of variable nodes $N$ in the original graph.

Now consider the residual graph. The number of different assignments of ones and zeros to the variable nodes that satisfy all the check equations is equal to the number of codewords of the original code that are consistent with the received sequence $Y^N$. This means that $H(X^N|Y^N)/N$ is equal to the rate of the residual graph. Lemma 7 from [MMU08] gives a condition on the degree distribution $(\Lambda, \Gamma)$ that when satisfied guarantees that the rate of a randomly chosen code from the ensemble LDPC$(N, \Lambda, \Gamma)$ is close to its design rate:

**Lemma 2.3.2** (Lemma 7 from [MMU08])**.** *Let $\mathcal{C}$ be a code chosen uniformly at random from the ensemble LDPC$(N, \Lambda, \Gamma)$ and let $r_\mathcal{C}$ be its rate.*

*Let $r = 1 - \Lambda'(1)/\Gamma'(1)$ be the design rate of the ensemble. Consider the function $\Psi_{\Lambda,\Gamma}(u)$*

$$\begin{aligned}\Psi_{\Lambda,\Gamma}(u) = &- \Lambda'(1)\log\left(\frac{1+uv}{1+v}\right) + \sum_{\mathtt{l}}\log\left(\frac{1+u^{\mathtt{l}}}{2}\right) \\ &+ \frac{\Lambda'(1)}{\Gamma'(1)}\sum_{\mathtt{r}}\log\left[1+\left(\frac{1-v}{1+v}\right)^{\mathtt{r}}\right],\end{aligned} \tag{2.10}$$

*where*

$$v = \left(\sum_{\mathtt{l}}\frac{\lambda_{\mathtt{l}}}{1+u^{\mathtt{l}}}\right)^{-1}\left(\sum_{\mathtt{l}}\frac{\lambda_{\mathtt{l}}u^{\mathtt{l}-1}}{1+u^{\mathtt{l}}}\right). \tag{2.11}$$

*Assume that $\Psi_{\Lambda,\Gamma}(u)$ takes on its global maximum in the range $u \in [0,\infty)$ at $u = 1$. Then there exists $B > 0$ such that, for any $\xi > 0$, and $N > N_0(\xi,\Lambda,\Gamma)$*

$$\Pr|r_G - r| > \xi \le e^{-BN\xi}.$$

*Moreover, there exists $C > 0$ such that, for $N > N_0(\xi,\Lambda,\Gamma)$*

$$\mathbb{E}[|r_G - r|] \le C\frac{\log N}{N}.$$

*Proof.* The lemma is proved using the following idea. The expected number of codewords where $e$[3] edges are connected to a variable node assigned a one is given by

$$\mathbb{E}[N_W(e)] = \frac{\operatorname{coef}\left\{\prod_{\mathtt{l}}(1+u^{\mathtt{l}})^{N\Lambda_{\mathtt{l}}}\prod_{\mathtt{r}}q_{\mathtt{r}}(v)^{M\Gamma_r}, u^e, v^e\right\}}{\binom{N\Lambda'(1)}{e}}, \tag{2.12}$$

where $\operatorname{coef}\left\{\sum_j D_j v^j, v^k\right\}$ is the coefficient of $v^k$ in the polynomial $\sum_j D_j v^j$ and $q_{\mathtt{r}}(v) = ((1+v)^{\mathtt{r}} + (1-v)^{\mathtt{r}})/2$. To see this, note that

$$\operatorname{coef}\left\{\prod_{\mathtt{l}}(1+u^{\mathtt{l}})^{N\Lambda_{\mathtt{l}}}, u^e\right\}$$

[3]Here $e$ is a variable and not the constant $e$.

is equal to the number of ways of assigning ones and zeros to the variable nodes so that $e$ edges are connected to a variable node assigned a one. Also

$$\operatorname{coef}\left\{\prod_{\mathtt{r}} q_{\mathtt{r}}(v)^{M\Gamma_r}, v^e\right\}$$

is equal to the number of ways of assigning $e$ ones to the sockets on the check node side so that each check node has an even number of incoming ones. The number of ways of connecting the sockets together is given by $e!(N\Lambda'(1)-e)!$. Thus the total number of codewords involving $e$ edges in the ensemble is given by

$$\operatorname{coef}\left\{\prod_{\mathtt{l}}(1+u^{\mathtt{l}})^{N\Lambda_{\mathtt{l}}}\prod_{\mathtt{r}} q_{\mathtt{r}}(v)^{M\Gamma_r}, u^e, v^e\right\} e!(N\Lambda'(1)-e)!.$$

Dividing by the number of graphs in the ensemble $(N\Lambda'(1))!$ yields (2.12).

Since the expected rate

$$\mathbb{E}[r_G] = \mathbb{E}\left[\frac{1}{N}\log\sum_e N_W(e)\right]$$

is hard to calculate we instead calculate

$$\frac{1}{N}\log\left(\mathbb{E}\left[\sum_e N_W(e)\right]\right)$$

which by Jensen's inequality is an upper bound on the expected rate. If $\lim_{N\to\infty}\frac{1}{N}\log\left(\mathbb{E}\left[\sum_e N_W(e)\right]\right) = r_{\mathrm{des}}$ the rate of a code will be close to the design rate.

Since the number of possible different values of $e$ only grows linearly with $N$ we get

$$\lim_{N\to\infty}\frac{1}{N}\log\left(\mathbb{E}\left[\sum_e N_W(e)\right]\right) = \sup_{e\in[0,1]}\lim_{N\to\infty}\frac{1}{N}\log\left(\mathbb{E}\left[N_W(eN\Lambda'(1))\right]\right)$$

From the Hayman approximations

$$\operatorname{coef}\left\{F(D)^N, D^k\right\} \le \inf_{x>0} F(x)^N/x^k,$$

and

$$\lim_{N\to\infty} \frac{1}{N} \log\left[\binom{\alpha N}{e\alpha N}\right] = \alpha h(e)$$

in [RU08, Appendix D] we get

$$\lim_{N\to\infty} \frac{1}{N} \log\left(\mathbb{E}\left[N_W(eN\Lambda'(1))\right]\right) = \inf_{u,v>0} \phi(e,u,v)$$

where

$$\phi(e,u,v) = \sum_{\mathtt{l}} \Lambda_{\mathtt{l}} \log(1+u^{\mathtt{l}}) - \Lambda'(1) e \log(u) + \\ + \frac{\Lambda'(1)}{\Gamma'(1)} \sum_{\mathtt{r}} \Gamma_{\mathtt{r}} \log(q_{\mathtt{r}}(v)) - \Lambda'(1) e \log(v) - \Lambda'(1) h(e).$$

We now bound the exponent $\sup_{e\in[0,1]} \inf_{u,v} \phi(e,u,v)$ from above as follows. The exponent is given by a stationary point of $\phi(e,u,v)$. Taking the derivative of $\phi$ with respect to $e$ and equating it to zero gives

$$e = \frac{uv}{1+uv}.$$

Inserting this value for $e$ into $\phi$ and taking the derivative with respect to $u$ gives the expression (2.11) for $v$. If we subtract the design rate $r_{\text{des}}$ from the resulting expression we get $\Psi_{\Lambda,\Gamma}(u)$, which is an upper bound on

$$\lim_{N\to\infty} \frac{1}{N} \log(\mathbb{E}[N]) - r_{\text{des}}.$$

If $\sup_{u>0} \Psi_{\Lambda,\Gamma}(u) = 0$, then the expected value of the rate is equal to the design rate and we can use Markov's inequality to get the bounds in the lemma. □

We now use the above lemma to check that the residual graph has rate equal to its design rate. If this is the case we can calculate the conditional entropy as the design rate of this ensemble, making sure to normalize its rate to the original block length $N$. This what is done in [MMU08, Theorem 10]:

**Theorem 2.3.3** (Theorem 10 from [MMU08])**.** *Let $\mathcal{C}$ be a code picked uniformly at random from the ensemble LDPC($N,\Lambda,\Gamma$) and let $H_{\mathcal{C}}(X|Y)$*

*be the conditional entropy of the transmitted message when the code is used for communicating over BEC($\epsilon$). Let $(\Lambda_\epsilon, \Gamma_\epsilon)$ be the typical degree distribution of the residual graph and let $\Psi_{\Lambda_\epsilon,\Gamma_\epsilon}(u)$ be as defined in Lemma 2.3.2. Assume that $\Psi_{\Lambda_\epsilon,\Gamma_\epsilon}(u)$ achieves its global maximum for $u \in [0,\infty)$ at $u = 1$, that $\Psi''_{\Lambda_\epsilon,\Gamma_\epsilon}(1) < 0$, and that $\epsilon$ is nonexceptional. Then*

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}[H_{\mathcal{C}}(X|Y)] = \Lambda'(1)x(1-y) - \frac{\Lambda'(1)}{\Gamma'(1)}(1-\Gamma(1-x)) + \epsilon\Gamma(y)$$

*where $x \in [0,1]$ is the largest solution of $x = \epsilon\lambda(1-\rho(1-x))$ and $y = 1-\rho(1-x)$.*

As noted before, Theorem 2.3.3 can be used to calculate the MAP decoding threshold of an ensemble. We call this the MMU method in acknowledgement of the authors of [MMU08], and we will use it in a generalized form in Chapter 3 to calculate the equivocation rate of Eve when using two edge type LDPC codes over the BEC-WT($\epsilon_m, \epsilon_w$). The MMU method was extended to non-binary LDPC codes for transmission over the BEC in [Rat08, RA10].

### 2.3.3 Spatially Coupled Codes

Convolutional LDPC codes were introduced by Felström and Zigangirov and were shown to have excellent thresholds [FZ99]. There has been a significant amount of work done on convolutional-like LDPC ensembles [EZ99, LTZ99, TSS$^+$04, SLCZ04, LSZC05, LFZC09, LSCZ10], and see in particular the literature review in [KRU10]. The explanation for the excellent performance of convolutional-like or "spatially coupled" codes over the BEC was given by Kudekar, Richardson, and Urbanke in [KRU10]. (In the following, we also use the term spatially coupled codes when we refer to convolutional like codes.) More precisely, it was shown in [KRU10] that the phenomenon of spatial coupling has the effect of converting the MAP threshold of the underlying ensemble into the BP threshold for the BEC and regular LDPC codes. This phenomenon has been observed to hold in general over Binary Memoryless Symmetric (BMS) channels, see [KMRU10, LMFC10].

Thus, when point-to-point transmission is considered over BMS channels, regular convolutional-like LDPC ensembles are conjectured to be *universally* capacity achieving. This is because the MAP threshold of regular LDPC ensembles converges to the Shannon threshold for BMS

channels as their left and right degrees are increased by keeping the rate fixed. To date there is only empirical evidence for this conjecture.

In [KRU10] two ensembles of spatially coupled codes are defined. The $(\mathtt{l}, \mathtt{r}, L)$ ensemble, which is similar to the ensemble in [LFZC09], and the $(\mathtt{l}, \mathtt{r}, L, w)$ ensemble, which shows worse performance empirically, but is easier to analyze. We will introduce the parameters $\mathtt{l}, \mathtt{r}, L$, and $w$ in the following.

In order to introduce the $(\mathtt{l}, \mathtt{r}, L)$ ensemble we first look at a coupled ensemble of protograph codes. Protograph codes were introduced by Thorpe, Andrews and Dolinar in [TAD04] as a way of designing structured LDPC codes. Consider the $(3, 6)$ protograph in Figure 2.7. Copy this graph $M$ times, so that there are $M$ variable nodes at the top, $M$ check nodes, and $M$ variable nodes at the bottom. There are six edge bundles going between the check nodes and the variable nodes. To construct an ensemble of protograph codes permute the edges within each edge bundle choosing a permutation uniformly at random. In Figure 2.8 we show this procedure for the (3,6) protograph with $M = 5$.

Figure 2.7: Protograph of a (3,6) code.

To get a spatially coupled graph start with $2L + 1$ copies of the protograph next to each other at positions numbered from $-L$ to $L$. Then switch the connections so that each variable node has one connection going to a check node at the position on the left, one connection going to a check node at its own position, and one edge going to a check node at the position to its right. Introduce extra check nodes at the boundary to make each variable node have the same degree. Such a protograph is shown in Figure 2.9. Now copy this spatially coupled protograph $M$ times and connect the edges using a permutation as described above. To generalize this protograph based ensemble, which needs $\mathtt{r}$ to be a multiple of $\mathtt{l}$, Kudekar, Richardson, and Urbanke introduced the $(\mathtt{l}, \mathtt{r}, L)$ ensemble which is defined for odd $\mathtt{l}$ and general $\mathtt{r}$.

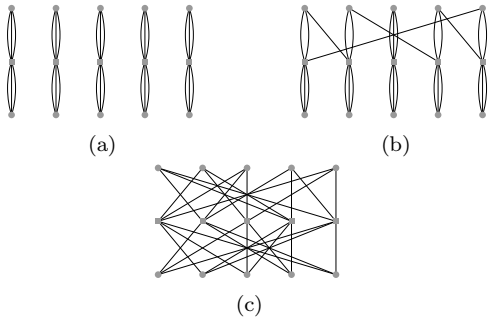


Figure 2.8: (a) 5 copies of a (3,6) protograph. (b) One edge bundle permuted. (c) All edge bundles permuted.

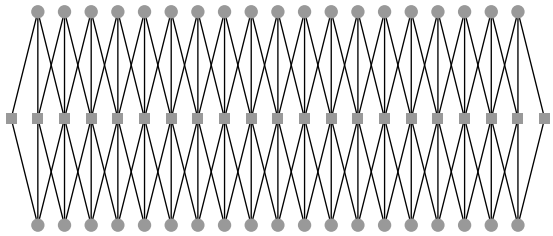

Figure 2.9: Spatially coupled (3,6) protograph with $L = 9$.

**Definition 2.3.4** (The $(\mathtt{l}, \mathtt{r}, L)$ ensemble)**.** *Place $M$ variable nodes at each position in the interval $[-L, L]$. Let $\hat{\mathtt{l}} = (\mathtt{l}-1)/2$ and place $M\mathtt{l}/\mathtt{r}$ check nodes at each position in the interval $[-L-\hat{\mathtt{l}}, L+\hat{\mathtt{l}}]$. Connect one edge from each variable at position $i$ to a check node at positions $[i-\hat{\mathtt{l}}, i+\hat{\mathtt{l}}]$. At the boundary there are fewer incoming connections to each check node, so decrease the degree of the check nodes at the boundary linearly according to their position. Impose a probability distribution on the codes in the ensemble by choosing a random permutation of the incoming edges at each check node position.*

The above ensemble is difficult to analyze, so Kudekar, Richardson and Urbanke introduced the $(\mathtt{l}, \mathtt{r}, L, w)$ ensemble. Before giving this definition, we define $\mathcal{T}(\mathtt{l})$ to be the set of $w$-tuples of non-negative integers

which sum to $\mathtt{l}$. More precisely,

$$\mathcal{T}(\mathtt{l}) = \{(t_0, \cdots, t_{w-1}) : \sum_{j=0}^{w-1} t_j = \mathtt{l}\}.$$

**Definition 2.3.5** ($\{\mathtt{l}, \mathtt{r}, L, w\}$ Spatially Coupled LDPC Ensemble)**.** *As above there are $M$ variable nodes at each of the positions $[-L, L]$. Now place $M\mathtt{l}/\mathtt{r}$ check nodes at each of the positions $[L, L+w-1]$. Not all of these check nodes will be connected to variable nodes. Now connect each variable node at position $i$ to check nodes at position $[i, i+w-1]$ in the following way.*

*For each variable node choose a* constellation *$c = (c_1, \ldots, c_{\mathtt{l}})$ with $c_j \in [0, w-1]$ uniformly at random. If a variable node at position $i$ has constellation $c$ then its $k$th edge is connected to a check node at position $i + c_k$. We denote the set of all constellations by $\mathfrak{C}$. Let $\tau(c)$ be the $w$-tuple which counts the occurrence of $0, 1, \cdots, w-1$ in $c$. Clearly $\tau(c) \in \mathcal{T}(\mathtt{l})$. We impose a uniform distribution over all constellations in $\mathfrak{C}$. This imposes the following distribution over $t \in \mathcal{T}(\mathtt{l})$*

$$p(t) = \frac{|\{c \in \mathfrak{C} : \tau(c) = t\}|}{w^{\mathtt{l}}}.$$

*Now we pick $M$ so that $Mp^{(1)}(t_1)$ is a natural number for all $t \in \mathcal{T}(\mathtt{l})$. For each position $i$ pick $Mp^{(1)}(t)$ variable nodes. For each of these variable nodes we use a random permutation over $\mathtt{l}$ letters to map $t$ to a constellation $c$. We then assign the edges of the variable nodes according to the constellation $c$.*

*Finally, at each check node position connect the incoming $M\mathtt{l}$ edges to the $M\mathtt{l}/\mathtt{r}$ check node edges using a permutation chosen uniformly at random.*

In [KRU10] the following was shown:

**Theorem 2.3.6** (Part of [KRU10] Theorem 12)**.** *Consider transmission over the BEC($\epsilon$) using random elements from the ensemble $(\mathtt{l}, \mathtt{r}, L, w)$. Let $\epsilon^{\mathrm{BP}}(\mathtt{l}, \mathtt{r}, L, w)$ and $\epsilon^{\mathrm{MAP}}(\mathtt{l}, \mathtt{r}, L, w)$ be the BP and MAP thresholds and let $R(\mathtt{l}, \mathtt{r}, L, w)$ be the design rate of this ensemble. Then*

$$\begin{aligned}
\lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} R(\mathtt{l}, \mathtt{r}, L, w) &= 1 - \frac{\mathtt{l}}{\mathtt{r}}, \\
\lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} \epsilon^{\mathrm{BP}}(\mathtt{l}, \mathtt{r}, L, w) &= \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} \epsilon^{\mathrm{MAP}}(\mathtt{l}, \mathtt{r}, L, w) \\
&= \epsilon^{\mathrm{MAP}}(\mathtt{l}, \mathtt{r}),
\end{aligned}$$

*where* $\epsilon^{\mathrm{MAP}}(\mathtt{l},\mathtt{r})$ *is the MAP threshold of the* $(\mathtt{l},\mathtt{r})$ *regular LDPC ensemble.*

Note that, since the MAP threshold of the $(\mathtt{l},\mathtt{r})$ regular LDPC ensemble approaches $\mathtt{l}/\mathtt{r}$ as $\mathtt{l}$ and $\mathtt{r}$ increase while keeping the ratio $\mathtt{l}/\mathtt{r}$ fixed [KRU10, Lemma 8], this means that the $(\mathtt{l},\mathtt{r},L,w)$ ensemble achieves capacity on the BEC.

## 2.4 Polar Codes

Polar codes were introduced by Arıkan and were shown to be capacity achieving for a large class of channels [Arı09]. Let $W$ be a binary input channel with discrete output alphabet $\mathcal{Y}$. Denote the channel transition probability of $W$ by $W(y|x)$. Let $I(W)$ denote the symmetric capacity

$$I(W) = \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} \frac{1}{2} W(y|x) \log \frac{2W(y|x)}{W(y|0)+W(y|1)},$$

and recall that $I(W)$ is the capacity of $W$ when the input distribution is constrained to be uniform. If $W$ is a symmetric channel, then $I(W)$ equals the Shannon capacity of $W$.

Polar codes rely on a phenomenon called *channel polarization*, which is achieved in a two-step process called *channel combining* and *channel splitting*. Channel combining takes $N$ copies of the channel $W$ and creates a vector channel $W_N(y^N|u^N)$ in a recursive manner. The vector channel $W_N$ is then split into $N$ binary input channels $W_N^{(i)}$. The channels $W_N^{(i)}$ are polarized in the sense that their symmetric capacities are either close to 0 or 1, and the idea behind polar codes is to send information only over the channels with $I(W)$ close to 1. We now describe the channel combining and channel splitting steps in detail.

Channel combining is a recursive transformation that takes two copies of a vector channel $W_{N/2}(y_1^{N/2}|u_1^{N/2})$ and creates a new vector channel $W_N(y_1^N|u_1^N)$ according to

$$W_N(y_1^N|u_1^N) = W_{N/2}(y_1^{N/2}|u_{1,o}^N \oplus u_{1,e}^N) W_{N/2}(y_{N/2+1}^N|u_{1,e}^N), \tag{2.13}$$

where $u_{1,o}^N = (u_1, u_3, \ldots, u_{N-1})$ and $u_{1,e}^N = (u_2, u_4, \ldots, u_N)$.

For the first two steps $N = 2$ and 4, (2.13) becomes

$$W_2(y_1, y_2|u_1, u_2) = W(y_1|u_1 \oplus u_2) W(y_2|u_2)$$

and

$$W_4(y_1^4|u_1^4) = W_2(y_1, y_2|u_1 \oplus u_2, u_3 \oplus u_4)W_2(y_3, y_4|u_2, u_4)$$

respectively, as illustrated in Figures 2.10 and 2.11.

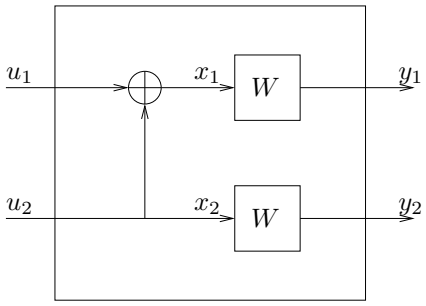


Figure 2.10: The channel $W_2$ constructed from two copies of $W$.

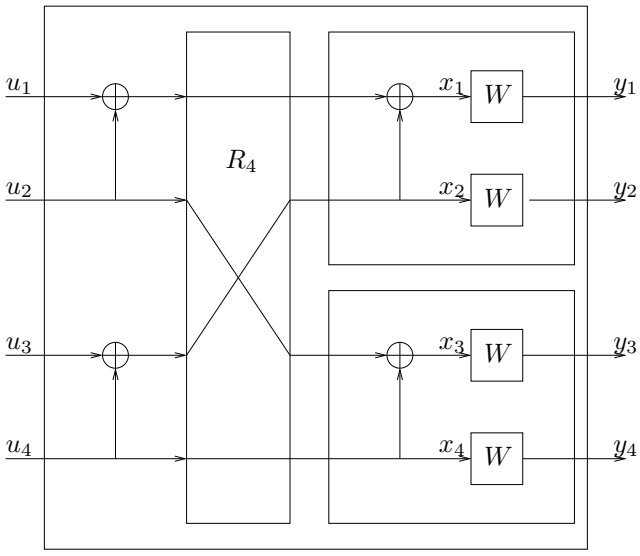


Figure 2.11: The channel $W_4$ constructed from two copies of $W_2$.

Note that the inputs $(x_1, \ldots, x_N)$ to the individual copies of the chan-

nel $W$ can be written as $u_1^N G_N$ where

$$G_N = B_N F^{\otimes n}. \tag{2.14}$$

Here $B_N$ is a bit-reversal permutation matrix where the output is generated from the input by writing the indices of the bits $u_i$ in bit format and reversing the indices. For example

$$B_8 : \ (u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8) \mapsto (u_1, u_5, u_3, u_7, u_2, u_5, u_4, u_8)$$

since in bit format

$$(u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8) = \\ (u_{000}, u_{001}, u_{010}, u_{011}, u_{100}, u_{101}, u_{110}, u_{111}),$$

and

$$(u_1, u_5, u_3, u_7, u_2, u_5, u_4, u_8) = \\ (u_{000}, u_{100}, u_{010}, u_{110}, u_{001}, u_{101}, u_{011}, u_{111}).$$

The matrix $F^{\otimes n}$ is the $n$th Kronecker power of the matrix

$$F = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

This means that in general we have $W_N(y_1^N|u_1^N) = W^N(y_1^N|u_1^N G_N)$, where $W^N(y_1^N|x_1^N) = \prod_{i=1}^N W(y_i|x_i)$.

Channel splitting is done by converting the combined vector channel $W_N(y_1^N|u_1^N)$ into $N$ binary input channels $W_N^{(i)}(y_1^N, u_1^{i-1}|u_i)$.

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) = \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-i}} W_N(y_1^N|u_1^N). \tag{2.15}$$

Note that $W_N^{(i)}$ has $y_1^N$ as well as the previous inputs $u_1^{i-1}$ as output. The successive cancellation decoder proposed by Arıkan gets around this problem by decoding $W_N^{(i)}$ before $W_N^{(j)}$ if $i < j$, and thus obtaining an estimate $\hat{u}_i$ of $u_i$. If these estimates are correct we will have all outputs of $W_N^{(j)}$ available before decoding.

Arıkan showed that the channels $\{W_N^{(i)}\}$ polarize as $N$ goes to infinity, that is for any $\delta \in (0, 1)$, the fraction of indices $i$ for which $I(W_N^{(i)}) \in$

$(1-\delta, 1]$ goes to $I(W)$ and the fraction for which $I(W_N^{(i)}) \in [0, \delta)$ goes to $1 - I(W)$.

The idea behind polar coding is to send information only over the good channels, while keeping the input to the bad channels fixed. Let $\mathcal{A}$ be a subset of $\{1, \ldots, N\}$ and let $u_\mathcal{A}$ be a binary vector of length $|\mathcal{A}|$. We call $\mathcal{A}$ and $\mathcal{A}^\mathcal{C}$ the information set and the frozen set respectively. Similarly we call $u_\mathcal{A}$ and $u_{\mathcal{A}^c}$ the information bits and the frozen bits. We now define the polar code $\mathcal{P}(N, \mathcal{A}, u_{\mathcal{A}^c})$ as follows:

**Definition 2.4.1** (The polar code $\mathcal{P}(N, \mathcal{A}, u_{\mathcal{A}^c})$)**.** *Let $G$ be the matrix $G_N$ as defined in (2.14) and let $G_\mathcal{A}$ be the submatrix composed of the columns of $G$ whose indices belong to the index set $\mathcal{A}$. The polar code $\mathcal{P}(N, \mathcal{A}, u_{\mathcal{A}^c})$ is the set of codewords $x^N$ of the form*

$$x^N = u_\mathcal{A} G_\mathcal{A} \oplus u_{\mathcal{A}^c} G_{\mathcal{A}^c}.$$

We see that the polar code fixes the input to the channels $W_n^{(i)}$ where $i$ is in the frozen set, and sends information over the channels where $i \in \mathcal{A}$. The rate of the polar code is equal to

$$R = \frac{|\mathcal{A}|}{N}.$$

The decoder that Arıkan proposed uses the following successive cancellation decoding rule

$$\hat{u}_i = \begin{cases} u_i & i \in \mathcal{A}^\mathcal{C} \\ 0 & \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | u_i = 0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | u_i = 1)} \geq 1 \text{ and } i \in \mathcal{A} \\ 1 & \text{otherwise} \end{cases}$$

to decode the transmitted bits. The decoder decodes the bits in increasing order and thus has the estimates $\hat{u}_i^{i-1}$ available when decoding $u_i$.

The average error probability $P_e^N$ of the successive cancellation decoder, averaged over all possible frozen sets, can be bounded from above in the following way

$$\begin{aligned} P_e^N &\leq \sum_{i \in \mathcal{A}} \Pr(\hat{u}_i \neq u_i) \\ &= \sum_{i \in \mathcal{A}} \sum_{y_1^N, u_1^{i-1}} p_{u_i} W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) \mathbb{1}_{\left\{ \frac{W_N^{(i)}(y_1^N, u_1^{i-1} | u_i \oplus 1)}{W_N^{(i)}(y_1^N, u_1^{i-1} | u_i)} \geq 1 \right\}} \end{aligned}$$

$$\leq \sum_{i \in \mathcal{A}} \sum_{y_1^N, u_1^{i-1}} p_{u_i} W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) \sqrt{\frac{W_N^{(i)}(y_1^N, u_1^{i-1}|u_i \oplus 1)}{W_N^{(i)}(y_1^N, u_1^{i-1}|u_i)}}$$
$$= \sum_{i \in \mathcal{A}} Z_N^{(i)}. \tag{2.16}$$

Here $Z_N^{(i)}$ is the Bhattacharyya parameter of the channel $W_N^{(i)}$, defined as

$$Z_N^{(i)} = \sum_{y_1^N} \sum_{u_1^{i-1}} \sqrt{W_N^{(i)}(y_1^N, u_1^{i-1}|0) W_N^{(i)}(y_1^N, u_1^{i-1}|1)}.$$

In [AT09] Arıkan and Telatar showed the following result on the rate of the polarization process:

**Theorem 2.4.2** (Rate of Polarization [AT09])**.** *For any* $0 < \beta < 1/2$

$$\lim_{n \to \infty} \frac{1}{N} |\{i : Z_N^{(i)} < 2^{-N^\beta}\}| = I(W).$$

This result shows us how to choose the frozen set when using the successive cancellation decoder.

**Theorem 2.4.3** ([Arı09], [AT09])**.** *Let* $W$ *be a discrete memoryless channel with binary input, and let* $R < I(W)$*. For any* $0 < \beta < 1/2$ *there exists a sequence of polar codes of block lengths* $N = 2^n$*, with rates* $R_N$ *such that*

$$\lim_{n \to \infty} R_N > R$$

*and there exists an* $n_0$ *such that the error probability under successive cancellation decoding satisfies*

$$P_e^N < 2^{-N^\beta} \ \forall n > n_0.$$

*Proof.* Let $\beta < \beta' < 1/2$ and choose the the non-frozen set $\mathcal{A}_N$ as

$$\mathcal{A}_N = \{i : \ Z_N^{(i)} < 2^{-N^{\beta'}}\}.$$

Then due to Theorem 2.4.2

$$\lim_{n \to \infty} R_N = I(W) > R.$$

For large enough $N$ we have

$$N2^{-N^{\beta'}} < 2^{-N^{\beta}},$$

which together with (2.16) implies that there exists an $n_0$ such that

$$P_e^N \leq \sum_{i \in \mathcal{A}_\mathcal{N}} Z_N^{(i)} < N2^{-N^{\beta'}} < 2^{-N^{\beta}}$$

provided that $n > n_0$. Finally since this is the error probability averaged over all frozen sets there must exist a frozen set with error probability at most $N2^{-N^{\beta}}$. □

If the channel $W$ is symmetric, then the symmetric capacity $I(W)$ is equal to the capacity $C$, and further, the error probability does not depend on the values of the frozen bits $u_{\mathcal{A}^c}$ [Arı09].

# Chapter 3

# LDPC Codes for the Wiretap Channel

In this chapter we consider LDPC codes for the BEC-WT channel. We propose a code construction method using two edge type LDPC codes based on the coset encoding scheme. Using a standard LDPC ensemble with a given threshold over the BEC, we give a construction for a two edge type LDPC ensemble with the same threshold. Thus if the standard LDPC ensemble is capacity achieving over the wiretapper's channel, our construction guarantees perfect secrecy.

However, our construction cannot guarantee reliability over the main channel if $\epsilon_m > 0$ and the given standard LDPC ensemble has degree two variable nodes. This is because our approach gives rise to degree one variable nodes in the code used over the main channel. This results in zero threshold over the main channel. In order to circumvent this problem, we numerically optimize the degree distribution of the two edge type LDPC ensemble. We find that the resulting codes approach the rate-equivocation region of the wiretap channel. For example, for the BEC-WT$(0.5, 0.6)$ we find ensembles that achieve the points $(R, R_e) = (0.0999064, 0.0989137)$ and $(R, R_e) = (0.498836, 0.0989137)$ which are very close to the best achievable points $B = (0.1, 0.1)$ and $C = (0.5, 0.1)$ as depicted in Figure 3.1.

Note that reliability, which corresponds to the probability of decoding error for the intended receiver, can be easily measured using density evolution recursion. However secrecy, which is given by the equivocation

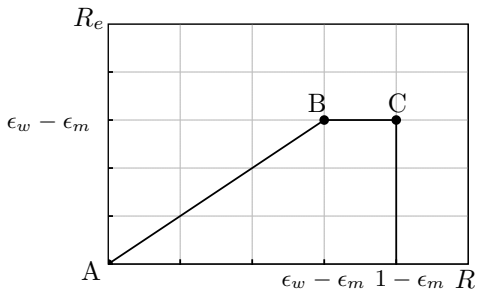


Figure 3.1: Capacity-equivocation region for the BEC-WT($\epsilon_m, \epsilon_w$).

of the message conditioned on the wiretapper's observation, can not be easily calculated. By generalizing the MMU method from [MMU08] to two edge type LDPC ensembles, we show how the equivocation for the wiretapper can be computed. We find that relatively simple constructions give very good secrecy performance and are close to the secrecy capacity.

We also introduce a spatially coupled two edge type LDPC ensemble. By showing that the density evolution recursion for the two edge type ensemble is the same as for the regular spatially coupled ensemble of [KRU10] we show that the spatially coupled two edge type LDPC ensemble achieves the whole capacity-equivocation region for the BEC. Since spatially coupled ensembles are conjectured to be capacity achieving not only for the BEC but also for general binary input channels we conjecture that our construction is optimal for general binary input degraded wiretap channels.

The chapter is organized in the following way. In Section 3.1, we define two edge type LDPC ensembles and give the density evolution recursion for them. Section 3.2 contains the code design and optimization for the BEC wiretap channel BEC-WT($\epsilon_m, \epsilon_w$). The MMU method and its extension to compute the equivocation of Eve for two edge type LDPC codes is given in Section 3.3. In Section 3.4 we present various examples to elucidate the computation of equivocation and show that our optimized degree distributions also approach the information theoretic equivocation limit. In Section 3.5 we introduce the spatially coupled ensemble and show that it achieves the whole capacity-equivocation region.

## 3.1 Two Edge Type LDPC Ensembles

We will use the coset encoding scheme introduced in Section 2.2.1. A natural candidate for coset encoding is a two edge type LDPC code since a two edge type parity check matrix $H$ has the form

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}. \tag{3.1}$$

The two types of edges are the edges connected to check nodes in $H_1$ and those connected to check nodes in $H_2$. An example of a two edge type LDPC code is shown in Figure 3.2.

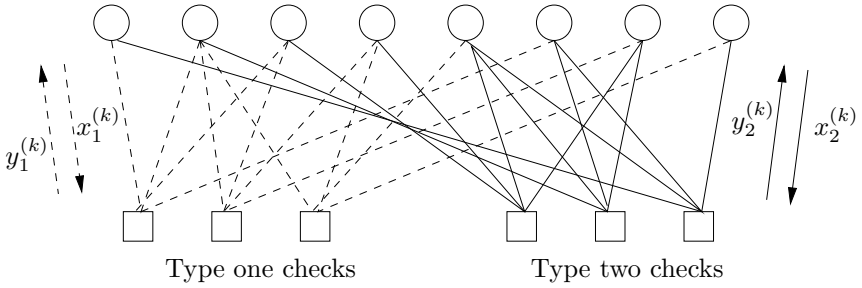


Figure 3.2: Two edge type LDPC code.

We now define the degree distribution of a two edge type LDPC ensemble. Let $\lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2}$ denote the fraction of type $j$ ($j = 1$ or 2) edges connected to variable nodes with $\mathtt{l}_1$ outgoing type one edges and $\mathtt{l}_2$ outgoing type two edges. The fraction $\lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2}$ is calculated with respect to the total number of type $j$ edges. Let $\Lambda_{\mathtt{l}_1\mathtt{l}_2}$ be the fraction of variable nodes with $\mathtt{l}_1$ outgoing edges of type one and $\mathtt{l}_2$ outgoing edges of type two. This gives the following relationships between $\Lambda, \lambda^{(1)}$, and $\lambda^{(2)}$:

$$\lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2} = \frac{\mathtt{l}_1\Lambda_{\mathtt{l}_1\mathtt{l}_2}}{\sum_{\mathtt{k}_1,\mathtt{k}_2} \mathtt{k}_1\Lambda_{\mathtt{k}_1\mathtt{k}_2}}, \tag{3.2}$$

$$\lambda^{(2)}_{\mathtt{l}_1\mathtt{l}_2} = \frac{\mathtt{l}_2\Lambda_{\mathtt{l}_1\mathtt{l}_2}}{\sum_{\mathtt{k}_1,\mathtt{k}_2} \mathtt{k}_2\Lambda_{\mathtt{k}_1\mathtt{k}_2}}, \tag{3.3}$$

$$\Lambda_{\mathtt{l}_1\mathtt{l}_2} = \frac{\frac{\lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_1}}{\sum_{\mathtt{k}_1,\mathtt{k}_2} \frac{\lambda^{(1)}_{\mathtt{k}_1\mathtt{k}_2}}{\mathtt{k}_1}} = \frac{\frac{\lambda^{(2)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_2}}{\sum_{\mathtt{k}_1,\mathtt{k}_2} \frac{\lambda^{(2)}_{\mathtt{k}_1\mathtt{k}_2}}{\mathtt{k}_2}}. \tag{3.4}$$

Similarly, let $\rho^{(j)}$ and $\Gamma^{(j)}$ denote the degree distribution of type $j$ edges on the check node side from the edge and node perspective respectively. Note that only one type of edges is connected to a particular check node. An equivalent definition of the degree distribution is given by the following polynomials:

$$\begin{aligned}
\Lambda(x,y) &= \sum_{\mathtt{l}_1,\mathtt{l}_2} \Lambda_{\mathtt{l}_1\mathtt{l}_2} x^{\mathtt{l}_1} y^{\mathtt{l}_2}, \\
\lambda^{(1)}(x,y) &= \sum_{\mathtt{l}_1,\mathtt{l}_2} \lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2} x^{\mathtt{l}_1-1} y^{\mathtt{l}_2}, \\
\lambda^{(2)}(x,y) &= \sum_{\mathtt{l}_1,\mathtt{l}_2} \lambda^{(2)}_{\mathtt{l}_1\mathtt{l}_2} x^{\mathtt{l}_1} y^{\mathtt{l}_2-1}, \\
\Gamma^{(j)}(x) &= \sum_{\mathtt{r}} \Gamma^{(j)}_{\mathtt{r}} x^{\mathtt{r}}, \quad j = 1,2, \\
\rho^{(j)}(x) &= \sum_{\mathtt{r}} \rho^{(j)}_{\mathtt{r}} x^{\mathtt{r}-1}, \quad j = 1,2.
\end{aligned}$$

Like the standard LDPC ensemble of Definition 2.3.1, the two edge type LDPC ensemble with block length $N$ and degree distribution $\left\{\lambda^{(1)}, \lambda^{(2)}, \rho^{(1)}, \rho^{(2)}\right\}$ ($\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ from the node perspective) is the collection of all bipartite graphs satisfying the degree distribution constraints, where we allow multiple edges between two nodes. We will call a two edge type LDPC ensemble for which $\Lambda(x,y) = x^{\mathtt{l}_1} y^{\mathtt{l}_2}$, *left regular*, and denote it by $\{\mathtt{l}_1, \mathtt{l}_2, \Gamma^{(1)}, \Gamma^{(2)}\}$.

Consider the two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$. If we consider the ensemble of the subgraph induced by one particular type of edges it is easy to see that the resulting ensemble is the standard LDPC ensemble and we can easily calculate its degree distribution. Let $\{\Lambda^{(j)}, \Gamma^{(j)}\}$ be the degree distribution of the ensemble induced by type $j$ edges, $j = 1, 2$. Then $\Lambda^{(j)}$, for $j = 1, 2$, is given by

$$\Lambda^{(1)}_{\mathtt{l}_1} = \sum_{\mathtt{l}_2} \Lambda_{\mathtt{l}_1\mathtt{l}_2}, \quad \Lambda^{(2)}_{\mathtt{l}_2} = \sum_{\mathtt{l}_1} \Lambda_{\mathtt{l}_1\mathtt{l}_2}. \tag{3.5}$$

We now derive the density evolution equations for two edge type LDPC ensembles, assuming that transmission takes place over the BEC($\epsilon$). Let $x_j^{(k)}$ denote the probability that a message from a variable node to a check node on an edge of type $j$ in iteration $k$ is erased. Clearly,

$$x_j^{(1)} = \epsilon, \quad j = 1, 2. \tag{3.6}$$

In the same way, let $y_j^{(k)}$ be the probability that a message from a check node to a variable node on an edge of type $j$ in iteration $k$ is erased. This probability is

$$y_j^{(k)} = 1 - \rho^{(j)}(1 - x_j^{(k)}), \quad j = 1, 2. \tag{3.7}$$

Using this we can write down the following recursions for $x_j^{(k)}$:

$$x_1^{(k+1)} = \epsilon\lambda^{(1)}(y_1^{(k)}, y_2^{(k)}), \tag{3.8}$$
$$x_2^{(k+1)} = \epsilon\lambda^{(2)}(y_1^{(k)}, y_2^{(k)}). \tag{3.9}$$

In the next section, we show how the degree distribution of a two edge type LDPC ensemble can be chosen such that it has the same density evolution recursion as that of a given standard LDPC ensemble. We also numerically optimize the degree distributions of two edge type LDPC ensembles and show that we can approach points on the boundary of the capacity-equivocation region.

## 3.2 Optimization

As the density evolution recursion for two edge type LDPC ensembles is two dimensional, it is difficult to analyze. Thus we look for degree distributions which reduce the two dimensional recursion to a single dimension. This will enable us to use the density evolution recursion for standard LDPC ensembles over the BEC, which has been very well studied. In the following theorem, we accomplish this task.

**Theorem 3.2.1.** *Let $(\lambda, \rho)$ be a standard LDPC degree distribution with design rate $R$ and threshold $\epsilon^\star$ over the BEC. Then the following assignment,*

$$\rho^{(1)}(x) = \rho^{(2)}(x) = \rho(x), \tag{3.10}$$
$$\lambda_{\mathtt{l},\mathtt{l}}^{(1)} = \lambda_{\mathtt{l},\mathtt{l}}^{(2)} = \lambda_{2\mathtt{l}}, \tag{3.11}$$
$$\lambda_{\mathtt{l},\mathtt{l}+1}^{(1)} = \lambda_{\mathtt{l}+1,\mathtt{l}}^{(2)} = \frac{\mathtt{l}}{2\mathtt{l}+1}\lambda_{2\mathtt{l}+1}, \tag{3.12}$$
$$\lambda_{\mathtt{l}+1,\mathtt{l}}^{(1)} = \lambda_{\mathtt{l},\mathtt{l}+1}^{(2)} = \frac{\mathtt{l}+1}{2\mathtt{l}+1}\lambda_{2\mathtt{l}+1}, \tag{3.13}$$
$$\lambda_{\mathtt{l}_1,\mathtt{l}_2}^{(1)} = \lambda_{\mathtt{l}_1,\mathtt{l}_2}^{(2)} = 0, \quad |\mathtt{l}_1 - \mathtt{l}_2| > 1, \tag{3.14}$$

*ensures that the two edge type LDPC ensemble* $\left\{\lambda^{(1)}, \lambda^{(2)}, \rho^{(1)}, \rho^{(2)}\right\}$ *also has design rate* $R$ *and threshold* $\epsilon^\star$.

*Proof.* Assume that we choose $\lambda^{(1)}, \lambda^{(2)}, \rho^{(1)}$, and $\rho^{(2)}$ such that (3.10) and the following relation

$$\lambda^{(1)}(x,x) = \lambda^{(2)}(x,x) = \lambda(x). \tag{3.15}$$

is satisfied. Note that since

$$\begin{aligned}\lambda^{(j)}(x,x) &= \sum_{\mathtt{l}_1,\mathtt{l}_2} \lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2} x^{\mathtt{l}_1+\mathtt{l}_2-1} \\ &= \sum_{\mathtt{k}} \left( \sum_{\mathtt{l}_1+\mathtt{l}_2=\mathtt{k}} \lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2} \right) x^{\mathtt{k}-1},\end{aligned}$$

(3.15) implies

$$\sum_{\mathtt{l}_1+\mathtt{l}_2=\mathtt{k}} \lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2} = \sum_{\mathtt{l}_1+\mathtt{l}_2=\mathtt{k}} \lambda^{(2)}_{\mathtt{l}_1\mathtt{l}_2} \quad \forall \mathtt{k}.$$

From the density evolution recursion for two edge type LDPC ensembles given in (3.6)-(3.9), we see that (3.10) ensures that $y_1^{(k)} = y_2^{(k)}$ whenever $x_1^{(k)} = x_2^{(k)}$, and (3.15) ensures that $x_1^{(k+1)} = x_2^{(k+1)}$ whenever $y_1^{(k)} = y_2^{(k)}$. Since $x_j^{(1)} = \epsilon$, we see by induction that $x_1^{(k)} = x_2^{(k)}$ and $y_1^{(k)} = y_2^{(k)}$ for $k \geq 1$. Thus we can reduce the two dimensional density evolution recursion to the one dimensional density evolution recursion for the standard LDPC ensemble

$$x^{(k+1)} = \epsilon\lambda(1 - \rho(1 - x^{(k)})), \tag{3.16}$$

where $\lambda(x) = \sum_{\mathtt{l}} \lambda_{\mathtt{l}} x^{\mathtt{l}-1}$,

$$\lambda_{\mathtt{l}} = \sum_{\mathtt{l}_1+\mathtt{l}_2=\mathtt{l}} \lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2}, \tag{3.17}$$

and we have dropped the subscript of $x^{(k)}$ as $x_1^{(k)} = x_2^{(k)}$. Note that by (3.11)-(3.14)

$$\frac{\lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_1} = \frac{\lambda^{(2)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_2} \quad \forall \mathtt{l}_1, \mathtt{l}_2. \tag{3.18}$$

This ensures that (3.4) is fulfilled.

We now show that (3.11)-(3.14) guarantee that $\lambda^{(1)}(x,x) = \lambda^{(2)}(x,x) = \lambda(x)$. Then the two dimensional density evolution recursion becomes the one dimensional recursion in (3.16) and the two edge type ensemble will have the same threshold as the standard LDPC ensemble. We have

$$\begin{aligned}
\lambda^{(1)}(x,x) &= \sum_{\mathtt{l}_1,\mathtt{l}_2} \lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2} x^{\mathtt{l}_1+\mathtt{l}_2-1} \\
&\overset{(a)}{=} \sum_{\mathtt{l}} \left( \lambda^{(1)}_{\mathtt{l},\mathtt{l}+1} x^{2\mathtt{l}} + \lambda^{(1)}_{\mathtt{l},\mathtt{l}} x^{2\mathtt{l}-1} + \lambda^{(1)}_{\mathtt{l}+1,\mathtt{l}} x^{2\mathtt{l}} \right) \\
&\overset{(b)}{=} \sum_{\mathtt{l}} \left( \frac{\mathtt{l}}{2\mathtt{l}+1} \lambda_{2\mathtt{l}+1} x^{2\mathtt{l}} + \lambda_{2\mathtt{l}} x^{2\mathtt{l}-1} \right) \\
&\quad + \sum_{\mathtt{l}} \frac{\mathtt{l}+1}{2\mathtt{l}+1} \lambda_{2\mathtt{l}+1} x^{2\mathtt{l}} \\
&= \sum_{\mathtt{l}} \left( \lambda_{2\mathtt{l}+1} x^{2\mathtt{l}} + \lambda_{2\mathtt{l}} x^{2\mathtt{l}-1} \right) \\
&= \lambda(x),
\end{aligned}$$

where (a) is due to (3.14) and (b) is due to (3.11)–(3.13). The proof for $\lambda^{(2)}(x,x)$ is done in the same way.

We now show that the design rate of the resulting two edge type LDPC ensemble is the same as the design rate of the given standard LDPC ensemble. The design rate of the two edge type ensemble is

$$R_{\text{des}} = 1 - (M_1 + M_2)/N$$

where $M_j$ is the number of parity checks of type $j$ and $N$ is the number of variable nodes. If we let $d_{\text{avg}}$ denote the average check node degree (this is the same for both types because of (3.10)) and count the number of type $j$ edges in two different ways, we get

$$N \sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_j \Lambda_{\mathtt{l}_1\mathtt{l}_2} = M_j d_{\text{avg}}, \quad j = 1, 2,$$

or

$$\begin{aligned}
\frac{M_j}{N} &= \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_j \Lambda_{\mathtt{l}_1\mathtt{l}_2}}{d_{\text{avg}}}, \\
&\overset{(a)}{=} \frac{1}{d_{\text{avg}}} \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_j \frac{\lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_j}}{\sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_j}},
\end{aligned}$$

$$\stackrel{\text{(b)}}{=} \frac{1}{d_{\text{avg}}} \frac{1}{\sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_j}},$$

where (a) is due to (3.4) and (b) follows since the $\lambda^{(j)}_{\mathtt{l}_1\mathtt{l}_2}$ sum to 1. The design rate then becomes

$$\begin{aligned}
R_{\text{des}} &= 1 - (M_1 + M_2)/N, \\
&= 1 - \frac{1}{d_{\text{avg}}} \left( \frac{1}{\sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_1}} + \frac{1}{\sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\lambda^{(2)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_2}} \right) \\
&\stackrel{\text{(a)}}{=} 1 - \frac{2}{d_{\text{avg}}} \left( \frac{1}{\sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2}}{\mathtt{l}_1}} \right) \\
&\stackrel{\text{(b)}}{=} 1 - \frac{2}{d_{\text{avg}}} \left( \frac{1}{\sum_{\mathtt{l}} \left( \frac{\lambda_{2\mathtt{l}+1}}{2\mathtt{l}+1} + \frac{\lambda_{2\mathtt{l}}}{\mathtt{l}} + \frac{\lambda_{2\mathtt{l}+1}}{2\mathtt{l}+1} \right)} \right) \\
&= 1 - \frac{1}{d_{\text{avg}}} \frac{1}{\sum_{\mathtt{l}} \left( \frac{\lambda_{2\mathtt{l}+1}}{2\mathtt{l}+1} + \frac{\lambda_{2\mathtt{l}}}{2\mathtt{l}} \right)} \\
&= 1 - \frac{1}{d_{\text{avg}}} \frac{1}{\sum_{\mathtt{l}} \frac{\lambda_{\mathtt{l}}}{\mathtt{l}}},
\end{aligned}$$

where (a) is due to (3.18) and (b) follows using (3.11) - (3.14). Since this expression is the same as the design rate of the standard LDPC ensemble $(\lambda, \rho)$, we have shown that the two edge type LDPC ensemble has design rate $R$. This completes the proof of the theorem. □

To compute the threshold achievable on the main channel, we need to compute the threshold of the ensemble of parity-check matrices $H_1$ induced by type one edges. The ensemble of matrices $H_1$ is a standard LDPC ensemble, and its degree distribution can be easily calculated from the degree distribution of the two edge type ensemble. Hence we can easily compute its threshold.

Since all capacity approaching sequences of degree distributions have some degree two variable nodes, because of (3.11) we see that our construction will have some degree one variable nodes in the matrix $H_1$. This means that the threshold over the main channel will be zero. To get around this problem we use linear programming methods to find good

degree distributions for two edge type LDPC ensembles based on their two dimensional density evolution recursion.

First we optimize the degree distribution of $H_1$ for the main channel using the methods described in [RU08] and obtain a good ensemble $(\Lambda^{(1)}, \Gamma^{(1)})$.

For a given two edge type ensemble we can find the corresponding one edge type ensemble for $H_1$ by summing over the second index, since the fraction of variable nodes with $\mathtt{l}_1$ outgoing type one edges is given by $\sum_{\mathtt{l}_2} \Lambda_{\mathtt{l}_1\mathtt{l}_2}$. To fix the degree distribution of $H_1$ we then impose the constraint

$$\sum_{\mathtt{l}_2} \Lambda_{\mathtt{l}_1\mathtt{l}_2} = \Lambda_{\mathtt{l}_1}^{(1)} \text{ for all } \mathtt{l}_1.$$

For successful decoding we further impose the two constraints $x_1^{(k+1)} \leq x_1^{(k)}$ and $x_2^{(k+1)} \leq x_2^{(k)}$ which can be written as

$$\begin{aligned} x_1 &\geq \epsilon \lambda^{(1)}(y_1, y_2) \\ &= \epsilon \sum_{\mathtt{l}_1,\mathtt{l}_2} \lambda_{\mathtt{l}_1\mathtt{l}_2}^{(1)} y_1^{\mathtt{l}_1-1} y_2^{\mathtt{l}_2} \\ &= \epsilon \sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\mathtt{l}_1 \Lambda_{\mathtt{l}_1,\mathtt{l}_2}}{\sum_{\mathtt{k}_1,\mathtt{k}_2} \mathtt{k}_1 \Lambda_{\mathtt{k}_1,\mathtt{k}_2}} y_1^{\mathtt{l}_1-1} y_2^{\mathtt{l}_2}, \end{aligned}$$

where we have used (3.2) in the last step, and $y_1, y_2$ are given by

$$y_j = 1 - \rho^{(j)}(1 - x_j), \ j = 1, 2.$$

This simplifies to the linear constraint

$$0 \leq \sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 (x_1 - \epsilon y_1^{\mathtt{l}_1-1} y_2^{\mathtt{l}_2}) \Lambda_{\mathtt{l}_1\mathtt{l}_2}.$$

The corresponding constraint for $x_2$ is

$$0 \leq \sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 (x_2 - \epsilon y_1^{\mathtt{l}_1} y_2^{\mathtt{l}_2-1}) \Lambda_{\mathtt{l}_1\mathtt{l}_2}.$$

The design rate can be written as

$$R_{\text{des}} = 1 - \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 \Lambda_{\mathtt{l}_1\mathtt{l}_2}}{\sum_{\mathtt{l}_1} \mathtt{l}_1 \Gamma_{\mathtt{l}_1}^{(1)}} - \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 \Lambda_{\mathtt{l}_1\mathtt{l}_2}}{\sum_{\mathtt{l}_2} \mathtt{l}_2 \Gamma_{\mathtt{l}_2}^{(2)}},$$

where the term $\frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 \Lambda_{\mathtt{l}_1\mathtt{l}_2}}{\sum_{\mathtt{l}_1} \mathtt{l}_1 \Gamma^{(1)}_{\mathtt{l}_1}}$ is a constant because of the fixed degree distribution of $H_1$. If $\Gamma^{(2)}$ is fixed, we see that maximizing the design rate is the same as minimizing $\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 \Lambda_{\mathtt{l}_1\mathtt{l}_2}$. Thus we end up with the following linear program:

$$\text{minimize} \sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 \Lambda_{\mathtt{l}_1\mathtt{l}_2}$$

subject to

$$\sum_{\mathtt{l}_2} \Lambda_{\mathtt{l}_1\mathtt{l}_2} = \Lambda^{(1)}_{\mathtt{l}_1},\ \mathtt{l}_1 = 2, \ldots, \mathtt{l}_{1,\max}$$

$$\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 (x_1 - \epsilon y_1^{\mathtt{l}_1 - 1} y_2^{\mathtt{l}_2}) \Lambda_{\mathtt{l}_1\mathtt{l}_2} \geq 0,\ \forall\ x_1, y_1, y_2 \in [0,1] \tag{3.19}$$

$$\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 (x_2 - \epsilon y_1^{\mathtt{l}_1} y_2^{\mathtt{l}_2 - 1}) \Lambda_{\mathtt{l}_1\mathtt{l}_2} \geq 0,\ \forall\ x_2, y_1, y_2 \in [0,1] \tag{3.20}$$

where $\mathtt{l}_{1,\max}$ is the largest degree in $\Lambda^{(1)}(x)$. Since (3.19) and (3.20) represent infinitely many constraints we replace them with

$$\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 (x_1(k) - \epsilon y_1(k)^{\mathtt{l}_1 - 1} y_2(k)^{\mathtt{l}_2}) \Lambda_{\mathtt{l}_1\mathtt{l}_2} \geq 0,\ k = 1, \ldots, K$$

$$\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 (x_2(k) - \epsilon y_1(k)^{\mathtt{l}_1} y_2(k)^{\mathtt{l}_2 - 1}) \Lambda_{\mathtt{l}_1\mathtt{l}_2} \geq 0,\ k = 1, \ldots, K,$$

in order to have a finite number of constraints. The points $\{x_1(k), x_2(k)\}_{k=1}^K$ are chosen by generating a distribution $\Lambda$ and then running the density evolution recursion

$$\begin{aligned} x_1^{(1)} &= x_2^{(1)} = \epsilon \\ x_1^{(k+1)} &= \epsilon \lambda^{(1)}(y_1^{(k)}, y_2^{(k)}) \\ x_2^{(k+1)} &= \epsilon \lambda^{(2)}(y_1^{(k)}, y_2^{(k)}) \end{aligned}$$

$K$ times. The program is then solved repeatedly, each time updating $\{x_1(k), x_2(k)\}_{k=1}^K$. This process is repeated several times for different check node degree distributions $\Gamma^{(2)}$ until there is negligible improvement in rate.

We now present some optimized degree distributions obtained by this method. We use the following degree distribution

*Standard LDPC Degree Distribution* 1.

$$\begin{aligned}\Lambda^{(1)}(x) &= 0.5572098x^2 + 0.1651436x^3 + 0.07567923x^4 \\ &\quad + 0.0571348x^5 + .043603x^7 + 0.02679802x^8 \\ &\quad + 0.013885518x^{13} + 0.0294308x^{14} + 0.02225301x^{31} \\ &\quad + 0.00886105x^{100}, \\ \Gamma^{(1)}(x) &= 0.25x^9 + 0.75x^{10}\end{aligned}$$

as the ensemble $(\Lambda^{(1)}, \Gamma^{(1)})$ for the main channel. It has rate 0.498826 bits per channel use (b.p.c.u.), threshold 0.5, and multiplicative gap to capacity $(1 - \epsilon - R_{\text{des}})/(1 - \epsilon) = 0.00232857$. We use it to obtain two optimized degree distributions, one for $\epsilon_w = 0.6$, and one for $\epsilon_w = 0.75$.

The degree distribution for the ensemble optimized for the BEC-WT$(0.5, 0.6)$ is given by

*Two Edge Type Degree Distribution* 1.

$$\begin{aligned}\Lambda(x,y) &= 0.463846x^2 + 0.0814943x^2y + 0.0118691x^2y^2 \\ &\quad + 0.14239x^3 + 0.0201658x^3y + 0.00258812x^3y^2 \\ &\quad + 0.0292241x^4 + 0.0464551x^4y + 0.0564162x^5 \\ &\quad + 0.000718585x^5y + 0.0436039x^7y \\ &\quad + 0.0258926x^8y + 0.000905503x^8y^2 \\ &\quad + 0.00631474x^{13}y^2 + 0.00757076x^{13}y^5 \\ &\quad + 0.011051x^{14}y + 0.0173718x^{14}y^2 \\ &\quad + 0.00100807x^{14}y^5 + 0.00240762x^{31} \\ &\quad + 0.0012626x^{31}y^4 + 0.0185828x^{31}y^5 \\ &\quad + 0.000326117x^{100}y^4 + 0.00383319x^{100}y^{17} \\ &\quad + 0.00470174x^{100}y^{18}, \\ \Gamma^{(1)}(x) &= 0.25x^9 + 0.75x^{10}, \\ \Gamma^{(2)}(x) &= x^6.\end{aligned}$$

This ensemble has design rate 0.39893 b.p.c.u., threshold 0.6, and the multiplicative gap to capacity is 0.00267632. The rate $R$ from Alice to Bob is 0.099906 b.p.c.u. and $R_e$, the equivocation of Eve, is 0.0989137 b.p.c.u. Thus there is a small information leakage of 0.0009923 b.p.c.u.

However both $R$ and $R_e$ are very close to the secrecy capacity $C_S = 0.1$ b.p.c.u.

The degree distribution for the ensemble optimized for the BEC-WT$(0.5, 0.75)$ is given by

*Two Edge Type Degree Distribution* 2.

$$\begin{aligned}
\Lambda(x,y) = {} & 0.367823x^2 + 0.166244x^2y + 0.0231428x^2y^2 \\
& + 0.125727x^3 + 0.0394166x^3y + 0.00286773x^4 \\
& + 0.0728115x^4y + 0.0571348x^5y \\
& + 0.0300989x^7y^2 + 0.013505x^7y^3 \\
& + 0.0196622x^8y^3 + 0.00713582x^8y^4 \\
& + 0.000565918x^{13}y^2 + 0.0133196x^{13}y^5 \\
& + 0.0149732x^{14}y^2 + 0.0132215x^{14}y^5 \\
& + 0.0012361x^{14}y^6 + 0.00490831x^{31}y^8 \\
& + 0.0173447x^{31}y^9 + 0.00130606x^{100}y^{17} \\
& + 0.00498932x^{100}y^{30} + 0.00256567x^{100}y^{31}, \\
\Gamma^{(1)}(x) = {} & 0.25x^9 + 0.75x^{10}, \\
\Gamma^{(2)}(x) = {} & 0.25x^4 + 0.75x^5.
\end{aligned}$$

This ensemble has design rate 0.248705 b.p.c.u. and threshold 0.75. The multiplicative gap to capacity is 0.00518359. The rate $R$ from Alice to Bob is 0.250131 b.p.c.u. and $R_e$, the equivocation of Eve, is 0.248837 b.p.c.u. Note that the secrecy capacity $C_s$ for this channel is 0.25 b.p.c.u. Thus the obtained point is slightly to the right of and below point B in Figure 3.1.

As mentioned earlier, computing the equivocation of Eve is not as straightforward as computing the reliability on the main channel. In the next section we show how to compute the equivocation of Eve by generalizing the methods from [MMU08] to two edge type LDPC codes.

## 3.3 Analysis of Equivocation

In order to compute the average equivocation of Eve over the erasure pattern and ensemble of codes, we generalize the MMU method of [MMU08] to two edge type LDPC codes. In [MMU08], the equivocation of standard LDPC ensembles for point-to-point communication over BEC$(\epsilon)$ was

computed. More precisely, let $\tilde{X}^N$ be a randomly chosen codeword of a randomly chosen code $\mathcal{C}$ from the standard LDPC ensemble. Let $\tilde{X}^N$ be transmitted over BEC($\epsilon$) and let $\tilde{Z}^N$ be the channel output. Then the MMU method computes

$$\lim_{N\to\infty} \frac{\mathbb{E}\left(H_{\mathcal{C}}(\tilde{X}^N|\tilde{Z}^N)\right)}{N},$$

where $H_{\mathcal{C}}(\tilde{X}^N|\tilde{Z}^N)$ is the conditional entropy of the transmitted codeword given the channel observation for the code $\mathcal{C}$, and we do the averaging over the ensemble. Note that we need not average over the codewords as the analysis can be carried out under the assumption that the all-zero codeword is transmitted [RU08, Chap. 3]. The MMU method is described below.

1. Consider decoding using the peeling decoder. The peeling decoder gets stuck in the largest stopping set contained in the set of erased variable nodes. The subgraph induced by this stopping set is again a code whose codewords are compatible with the erasure set. We call this subgraph the *residual graph*. Thus the peeling decoder associates to every graph and erasure set a residual graph. If the erasure probability is above the BP threshold, then almost surely the residual graph has a degree distribution close to the *average residual degree distribution* [LMSS01a]. The average residual degree distribution can be computed by the asymptotic analysis of the peeling decoder.

2. Conditioned on the residual degree distribution, the induced probability distribution is uniform over all the graphs with the given degree distribution. This implies that almost surely a residual graph is an element of the standard LDPC ensemble with degree distribution equal to the average residual degree distribution.

3. One can easily compute the design rate of the average residual degree distribution. However, the design rate is only a lower bound on the rate. A criterion was derived in [MMU08], which, when satisfied, guarantees that the actual rate is equal to the design rate. If the actual rate is equal to the design rate, then the equivocation is given by the design rate of the standard LDPC ensemble with degree distribution equal to the average residual degree distribution.

In order to compute the equivocation of Eve $H(S|Z^N)$, using the chain rule we write $H(S, X^N|Z^N)$ in two different ways and obtain

$$H(X^N|Z^N) + H(S|X^N, Z^N) = H(S|Z^N) + H(X^N|S, Z^N). \quad (3.21)$$

By noting that $H(S|X^N, Z^N) = 0$ and substituting it in (3.21), we obtain

$$\frac{H(S|Z^N)}{N} = \frac{H(X^N|Z^N)}{N} - \frac{H(X^N|S, Z^N)}{N}. \quad (3.22)$$

In the following two subsections we show how the normalized average of $H(X^N|Z^N)$ and $H(X^N|S, Z^N)$ can be computed. The next subsection deals with $H(X^N|Z^N)$.

### 3.3.1 Computing the Normalized $H(X^N|Z^N)$

In the following lemma we show that the average of $\lim_{N\to\infty} H(X^N|Z^N)/N$ can be computed by the MMU method.

**Lemma 3.3.1.** *Consider transmission over the BEC-WT($\epsilon_m, \epsilon_w$) using the syndrome encoding method with a two edge type LDPC code* $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$, *where the dimensions of* $H$, $H_1$, *and* $H_2$ *are* $N(1 - R^{(1,2)}) \times N$, $N(1 - R^{(1)}) \times N$, *and* $NR \times N$ *respectively. Let* $S$ *be a randomly chosen message from Alice for Bob and let* $X^N$ *be the transmitted vector which is a randomly chosen solution of* $HX^N = \begin{bmatrix} 0 \\ S \end{bmatrix}$. *Let* $Z^N$ *be the channel observation of Eve. Consider a point-to-point communication set-up over the BEC($\epsilon_w$) using a standard LDPC code* $H_1$. *Let* $\hat{X}^N$ *be a randomly chosen transmitted codeword, i.e.,* $\hat{X}^N$ *is a randomly chosen solution of* $H_1\hat{X}^N = 0$. *Further let* $\hat{Z}^N$ *be the channel output. Then*

$$H\left(X^N|Z^N\right) = H\left(\hat{X}^N|\hat{Z}^N\right).$$

*Proof.* We prove the lemma by showing that $(X^N, Z^N)$ and $(\hat{X}^N, \hat{Z}^N)$ have the same joint distribution. Clearly, $\Pr(Z^N = z^N|X^N = x^N) = \Pr(\hat{Z}^N = z^N|\hat{X}^N = x^N)$ as transmission takes place over the BEC($\epsilon_w$)

in both cases. Now

$$\begin{aligned}\Pr(X^N = x^N) &= \sum_s \Pr\left(X^N = x^N, S = s\right), \\ &\overset{(a)}{=} \frac{1}{2^{NR}} \sum_s \Pr\left(X^N = x^N | S = s\right), \\ &\overset{(b)}{=} \frac{1}{2^{NR}} \sum_s \frac{1}{2^{NR^{(1,2)}}} \mathbb{1}_{\{H_1 x^N = 0\}} \mathbb{1}_{\{H_2 x^N = s\}}, \\ &\overset{(c)}{=} \frac{\mathbb{1}_{\{H_1 x^N = 0\}}}{2^{NR^{(1)}}}, \end{aligned} \tag{3.23}$$

where $\mathbb{1}_{\{S\}}$ is the indicator function for the statement $S$, (a) follows from the uniform *a priori* distribution on $S$, (b) follows since conditioned on $s$ there are $2^{NR^{(1,2)}}$ equally likely solutions to $Hx^N = [0\ s]^T$, and (c) follows because for a fixed $x^N$,

$$\sum_s \mathbb{1}_{\{H_2 x^N = s\}} = 1.$$

Now the *a priori* distribution of $\hat{X}^N$ is also the RHS of (3.23). This is because $\hat{X}^N$ is a randomly chosen solution of $H_1 \hat{X}^N = 0$. This proves the lemma. □

From Lemma 3.3.1, we see that when we consider transmission over the BEC-WT($\epsilon_m, \epsilon_w$) using the two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$, we can compute the average of $\lim_{N\to\infty} H(X^N|Z^N)/N$ by applying the MMU method to the standard LDPC ensemble $\{\Lambda^{(1)}, \Gamma^{(1)}\}$ for transmission over the BEC($\epsilon_w$). We formally state this in the following theorem.

**Theorem 3.3.2.** *Consider transmission over the BEC-WT($\epsilon_m, \epsilon_w$) using a randomly chosen code $\mathcal{C}$ from the two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ and the coset encoding method. Let $X^N$ be the transmitted word and $Z^N$ be the wiretapper's observation.*

*Consider a point-to-point communication setup for transmission over BEC($\epsilon_w$) using the standard LDPC ensemble $\{\Lambda^{(1)}, \Gamma^{(1)}\}$. Let $\{\Omega, \Phi\}$ (from the node perspective) be the average residual degree distribution of the residual ensemble given by the peeling decoder and let $R^r_{\text{des}}$ be the design rate of the average residual ensemble $\{\Omega, \Phi\}$. If almost every element of the average residual ensemble $\{\Omega, \Phi\}$ has its rate equal to the*

*design rate* $R^r_{\text{des}}$*, then*

$$\lim_{N\to\infty} \frac{\mathbb{E}\left(H_{\mathcal{C}}(X^N|Z^N)\right)}{N} = \epsilon_w \Lambda^{(1)} \left(1-\rho^{(1)}(1-x)\right) R^r_{\text{des}},$$

*where* $x$ *is the fixed point of the density evolution recursion for* $\{\Lambda^{(1)}, \Gamma^{(1)}\}$ *initialized with erasure probability* $\epsilon_w$*, and* $\rho^{(1)}$ *is the check node degree distribution of* $H_1$ *from the edge perspective.*

*Proof.* Note that the condition that almost every element of the average residual ensemble $\{\Omega, \Phi\}$ has its rate equal to the design rate can be verified by using Lemma 2.3.2.

The proof is a straightforward consequence of Lemma 3.3.1 and Theorem 2.3.3. The factor $\epsilon_w \Lambda^{(1)} \left(1-\rho^{(1)}(1-x)\right)$, which is the ratio of the block length of the average residual ensemble $\{\Omega, \Phi\}$ to the initial ensemble $\{\Lambda^{(1)}, \Gamma^{(1)}\}$, takes care of the fact that we are normalizing $H_{\mathcal{C}}(X^N|Z^N)$ by the block-length of the initial ensemble $\{\Lambda^{(1)}, \Gamma^{(1)}\}$. □

In the following subsection we generalize the MMU method to two edge type LDPC ensembles in order to compute $H(X^N|S, Z^N)$.

### 3.3.2 Computing the Normalized $H(X^N|S, Z^N)$ by Generalizing the MMU method to Two Edge Type LDPC Ensembles

Similarly to Lemma 3.3.1, in the following lemma we show that computing $H(X^N|S, Z^N)$ for the BEC-WT$(\epsilon_m, \epsilon_w)$ using the coset encoding method and two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ is equivalent to computing the equivocation of the same ensemble for point-to-point communication over the BEC$(\epsilon_w)$.

**Lemma 3.3.3.** *Consider transmission over BEC-WT*$(\epsilon_m, \epsilon_w)$ *using the syndrome encoding method with a two edge type LDPC code* $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$*, where the dimensions of* $H$*,* $H_1$*, and* $H_2$ *are* $N(1-R^{(1,2)}) \times N$*,* $N(1-R^{(1)}) \times N$*, and* $NR \times N$ *respectively. Let* $S$ *be a randomly chosen message from Alice for Bob and let* $X^N$ *be the transmitted vector which is a randomly chosen solution of* $HX^N = \begin{bmatrix} 0 \\ S \end{bmatrix}$*. Let* $Z^N$ *be the channel observation of Eve.*

*Consider a point-to-point communication set-up for transmission over the BEC($\epsilon_w$) using the two edge type LDPC code* $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$. *Let* $\hat{X}^N$ *be the transmitted codeword which is a randomly chosen solution of* $H\hat{X}^N = 0$ *and let* $\hat{Z}^N$ *be the channel output. Then*

$$H(X^N|S, Z^N) \overset{\text{(a)}}{=} H(X^N|S = 0, Z^N) \overset{\text{(b)}}{=} H(\hat{X}^N|\hat{Z}^N).$$

*Proof.* Equality (b) is obvious. To prove equality (a), note that for a solution $x^N$ of $Hx^N = \begin{bmatrix} 0 \\ s \end{bmatrix}$ we can write $x^N = x'^N \oplus x_s^N$, where $Hx'^N = 0$ and $Hx_s^N = \begin{bmatrix} 0 \\ s \end{bmatrix}$. Let $z^N$ be a specific received vector and let $z'^N$ be the vector that has the same erased positions as $z^N$ and is equal to the corresponding position in $x'^N$ in the unerased positions. The proof is completed by noting that

$$\Pr(X^N = x^N, Z^N = z^N|S = s) = \Pr(X^N = x'^N, Z^N = z'^N|S = 0).$$

□

Thus from Lemma 3.3.3 we see that $H(X^N|S, Z^N)$ can be computed by generalizing the MMU method to two edge type LDPC ensembles. The proof of Step 1 and 2 of the MMU method for two edge type LDPC ensemble is the same as for the standard LDPC ensemble. We state it in the following two lemmas.

**Lemma 3.3.4.** *Consider transmission over the BEC($\epsilon_w$) using the two edge type LDPC ensemble* $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ *and decoded via the peeling decoder. Let* $G$ *be a random residual graph. Conditioned on the event that* $G$ *has degree distribution* $\{\Omega, \Phi^{(1)}, \Phi^{(2)}\}$*, it is equally likely to be any element of the two edge type ensemble* $\{\Omega, \Phi^{(1)}, \Phi^{(2)}\}$.

*Proof.* The proof is the same as for standard LDPC codes [LMSS01b]. □

**Lemma 3.3.5.** *Consider transmission over the BEC($\epsilon_w$) using the two edge type LDPC ensemble* $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ *which is decoded using the peeling decoder. Let* $\{\Omega, \Phi^{(1)}, \Phi^{(2)}\}$ *be the average residual degree distribution. Let* $\{\Omega_G, \Phi_G^{(1)}, \Phi_G^{(2)}\}$ *be the residual degree distribution of a random residual graph* $G$*. Then, for any* $\delta > 0$

$$\lim_{N\to\infty} \Pr\left\{ d\left( \left(\Omega, \Phi^{(1)}, \Phi^{(2)}\right), \left(\Omega_G, \Phi_G^{(1)}, \Phi_G^{(2)}\right) \right) \geq \delta \right\} = 0.$$

*The distance $d(\cdot,\cdot)$ is the $L_1$ distance*

$$d\left(\left(\Omega,\Phi^{(1)},\Phi^{(2)}\right),\left(\tilde{\Omega},\tilde{\Phi}^{(1)},\tilde{\Phi}^{(2)}\right)\right) = \sum_{\mathtt{l}_1\mathtt{l}_2}|\Omega_{\mathtt{l}_1\mathtt{l}_2}-\tilde{\Omega}_{\mathtt{l}_1\mathtt{l}_2}|+\sum_{\mathtt{r}_1}|\Phi^{(1)}_{\mathtt{r}_1}-\tilde{\Phi}^{(1)}_{\mathtt{r}_1}|+\sum_{\mathtt{r}_2}|\Phi^{(2)}_{\mathtt{r}_2}-\tilde{\Phi}^{(2)}_{\mathtt{r}_2}|.$$

*Proof.* The proof is the same as that for standard LDPC ensembles [LMSS98, LMSS01b], [RU08, Theorem 3.106]. □

In the following lemma we compute the average residual degree distribution of the two edge type LDPC ensemble.

**Lemma 3.3.6.** *Consider transmission over BEC($\epsilon_w$) using the two type LDPC ensemble $\{\Lambda,\Gamma^{(1)},\Gamma^{(2)}\}$ which is decoded by the peeling decoder. Let $(x_1,x_2)$ be the fixed points of (3.8) and (3.9) when initialized with channel erasure probability $\epsilon_w$. Let $y_j = 1-\rho^{(j)}(1-x_j)$, $j=1,2$, where $\rho^{(j)}$ is the degree distribution of check nodes of type $j$ from edge perspective. Then the average residual degree distribution $\{\Omega,\Phi^{(1)},\Phi^{(2)}\}$ is given by*

$$\begin{aligned}\Omega(z_1,z_2) &= \epsilon\Lambda(z_1y_1,z_2y_2),\\ \Phi^{(j)}(z) &= \Gamma^{(j)}(1-x_j+x_jz)-x_jz\Gamma'^{(j)}(1-x_j)\\ &\quad -\Gamma^{(j)}(1-x_j),\quad j=1,2,\end{aligned}$$

*where $\Gamma'^{(j)}(x)$ is the derivative of $\Gamma^{(j)}(x)$. Note that the degree distributions are normalized with respect to the number of variable (check) nodes in the original graph.*

*Proof.* The proof follows by the analysis of the peeling decoder for general multi-edge type LDPC ensembles in [HW10]. However, as we are interested in only two edge type LDPC ensembles, the proof also follows from the analysis for the standard LDPC case [LMSS01b]. □

Lemma 3.3.4, 3.3.5, and 3.3.6 generalize Step 1 and 2 of the MMU method for two edge type LDPC ensembles. The key technical task in extending Step 3 to two edge type LDPC ensemble is to derive a criterion, which when satisfied, guarantees that almost every code in the residual ensemble has its rate equal to the design rate. The rate is equal to the normalized logarithm of the total number of codewords. However, as the average of the logarithm of the total number of codewords is hard

to compute, we compute the normalized logarithm of the average of the total number of codewords. By Jensen's inequality this is an upper bound on the average rate. If this upper bound is equal to the design rate, then by the same arguments as in Lemma 2.3.2 we can show that almost every code in the ensemble has its rate equal to the design rate.

In the following lemma we derive the average of the total number of codewords of a two edge type LDPC ensemble.

**Lemma 3.3.7.** *Let $N_W$ be the total number of codewords of a randomly chosen code from the two edge type LDPC ensemble $(\Lambda, \Gamma^{(1)}, \Gamma^{(2)})$. Then the average of $N_W$ over the ensemble is given by*

$$\mathbb{E}(N_W) = \sum_{E_1=0,E_2=0}^{N\Lambda'_1(1,1),N\Lambda'_2(1,1)} \operatorname{coef}\left\{\prod_{\mathtt{l}_1,\mathtt{l}_2} (1+u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2})^{N\Lambda_{\mathtt{l}_1,\mathtt{l}_2}}, u_1^{E_1}u_2^{E_2}\right\} \times \frac{\operatorname{coef}\left\{\prod_{\mathtt{r}_1,\mathtt{r}_2} q_{\mathtt{r}_1}(v_1)^{\frac{N\Lambda'_1(1,1)}{\Gamma'^{(1)}(1)}\Gamma^{(1)}_{\mathtt{r}_1}} q_{\mathtt{r}_2}(v_2)^{\frac{N\Lambda'_2(1,1)}{\Gamma'^{(2)}(1)}\Gamma^{(2)}_{\mathtt{r}_2}}, v_1^{E_1}v_2^{E_2}\right\}}{\binom{N\Lambda'_1(1,1)}{E_1}\binom{N\Lambda'_2(1,1)}{E_2}},$$

*where $\Lambda'_j(1,1) = \sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_j \Lambda_{\mathtt{l}_1,\mathtt{l}_2}$, $\Gamma'^{(j)}(1) = \sum_{\mathtt{r}_j} \mathtt{r}_j \Gamma^{(j)}_{\mathtt{r}_j}$, $j \in \{1,2\}$. The polynomial $q_{\mathtt{r}}(v)$ is defined as*

$$q_{\mathtt{r}}(v) = \frac{(1+v)^{\mathtt{r}} + (1-v)^{\mathtt{r}}}{2}.$$

*Proof.* Let $\mathcal{W}(E_1, E_2)$ be the set of assignments of ones and zeros to the variable nodes which result in $E_1$ (resp. $E_2$) type one (resp. type two) edges connected to variable nodes assigned value one. Denote the cardinality of $\mathcal{W}(E_1, E_2)$ by $|\mathcal{W}(E_1, E_2)|$. For an assignment $w$, let $\mathbb{1}_w$ be a random indicator variable which evaluates to one if $w$ is a codeword of a randomly chosen code and zero otherwise. Let $N_W(E_1, E_2)$ be the number of codewords belonging to the set $\mathcal{W}(E_1, E_2)$. Then we have the following relationships

$$N_W(E_1, E_2) = \sum_{w \in \mathcal{W}(E_1,E_2)} \mathbb{1}_w,$$

$$N_W = \sum_{E_1=0,E_2=0}^{N\Lambda'_1(1,1),N\Lambda'_2(1,1)} N_W(E_1, E_2).$$

By linearity of expectation we obtain

$$\mathbb{E}(N_W(E_1, E_2)) = \sum_{w \in \mathcal{W}(E_1, E_2)} \mathbb{E}(\mathbb{1}_w),$$

and

$$\mathbb{E}(N_W) = \sum_{E_1=0, E_2=0}^{N\Lambda_1'(1,1), N\Lambda_2'(1,1)} \mathbb{E}(N_W(E_1, E_2)). \tag{3.24}$$

From the symmetry of code generation, we observe that $\mathbb{E}(\mathbb{1}_w)$, for $w \in \mathcal{W}(E_1, E_2)$, is independent of $w$. Thus we can fix $w$ to any one element of $\mathcal{W}(E_1, E_2)$ and obtain

$$\mathbb{E}(N_W(E_1, E_2)) = |\mathcal{W}(E_1, E_2)| \Pr(w \text{ is a codeword}). \tag{3.25}$$

Note that $|\mathcal{W}(E_1, E_2)|$ is given by

$$|\mathcal{W}(E_1, E_2)| = \text{coef}\left\{ \prod_{\mathtt{l}_1, \mathtt{l}_2} (1 + u_1^{\mathtt{l}_1} u_2^{\mathtt{l}_2})^{N\Lambda_{\mathtt{l}_1, \mathtt{l}_2}}, u_1^{E_1} u_2^{E_2} \right\}. \tag{3.26}$$

We now evaluate the probability that an assignment $w$, $w \in \mathcal{W}(E_1, E_2)$, is a codeword, which is given by

$$\Pr(w \text{ is a codeword}) = \frac{\text{Total number of graphs for which } w \text{ is a codeword}}{\text{Total number of graphs}}. \tag{3.27}$$

Similar to the arguments for the standard LDPC ensemble in the proof of Lemma 2.3.2, the total number of graphs for which $w$ is a codeword is given by

$$E_1! E_2! (N\Lambda_1'(1,1) - E_1)! (N\Lambda_2'(1,1) - E_2)! \\ \text{coef}\left\{ \prod_{\mathtt{r}_1, \mathtt{r}_2} q_{\mathtt{r}}(v_1)^{\frac{N\Lambda_1'(1,1)}{\Gamma'^{(1)}(1)} \Gamma^{(1)}_{\mathtt{r}_1}} q_{\mathtt{r}}(v_2)^{\frac{N\Lambda_2'(1,1)}{\Gamma'^{(2)}(1)} \Gamma^{(2)}_{\mathtt{r}_2}}, v_1^{E_1} v_2^{E_2} \right\}. \tag{3.28}$$

By noting that the total number of graphs is equal to $(N\Lambda_1'(1,1))!(N\Lambda_2'(1,1))!$, and combining (3.24)-(3.28), we obtain the expression for the average of the total number of codewords.

□

*Remark:* Note that related problems of computing the weight distribution of two edge type and more generally multi-edge type LDPC ensembles have been addressed in [IKS+05, KAD+09].

**Lemma 3.3.8.** *Let $\mathcal{E}(N)$ be the set of $(e_1, e_2)$ such that*

$$\operatorname{coef}\left\{\prod_{\mathtt{l}_1,\mathtt{l}_2}(1+u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2})^{N\Lambda_{\mathtt{l}_1,\mathtt{l}_2}}, u_1^{e_1 N\Lambda_1'(1,1)}u_2^{e_2 N\Lambda_2'(1,1)}\right\} \neq 0. \tag{3.29}$$

*Then $\lim_{N\to\infty}\mathcal{E}(N)$ is the set of $(e_1, e_2)$ such that*

$$(e_1, e_2) = \left(\frac{\sum_{\mathtt{l}_1,\mathtt{l}_2}\mathtt{l}_1\Lambda_{\mathtt{l}_1,\mathtt{l}_2}\sigma(\mathtt{l}_1,\mathtt{l}_2)}{\Lambda_1'(1,1)}, \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2}\mathtt{l}_2\Lambda_{\mathtt{l}_1,\mathtt{l}_2}\sigma(\mathtt{l}_1,\mathtt{l}_2)}{\Lambda_2'(1,1)}\right),$$

*where $0 \le \sigma(\mathtt{l}_1,\mathtt{l}_2) \le 1$. We call this set $\mathcal{E}$.*

*$\mathcal{E}$ can also be represented as the subset of the unit square enclosed between two piecewise linear curves. Order the pairs $(\mathtt{l}_1,\mathtt{l}_2)$ for which $\Lambda_{\mathtt{l}_1,\mathtt{l}_2} > 0$ in decreasing order of $\mathtt{l}_1/\mathtt{l}_2$ and assume that there are $D$ distinct such values. Let*

$$\sigma_d(\mathtt{l}_1,\mathtt{l}_2) = \begin{cases} 1 & \text{if } \mathtt{l}_1/\mathtt{l}_2 \text{ takes the } d\text{th largest possible value,} \\ 0 & \text{otherwise,} \end{cases}$$

*and let*

$$p_d = \left(\frac{\sum_{\mathtt{l}_1,\mathtt{l}_2}\mathtt{l}_1\Lambda_{\mathtt{l}_1,\mathtt{l}_2}\sigma_d(\mathtt{l}_1,\mathtt{l}_2)}{\Lambda_1'(1,1)}, \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2}\mathtt{l}_2\Lambda_{\mathtt{l}_1,\mathtt{l}_2}\sigma_d(\mathtt{l}_1,\mathtt{l}_2)}{\Lambda_2'(1,1)}\right).$$

*Then $\mathcal{E}$ is the set above the piecewise linear curve connecting the points $\{(0,0), p_1, p_1+p_2, \ldots, (1,1)\}$ and below the piecewise linear curve connecting the points $\{(0,0), p_D, p_D+p_{D-1}, \ldots, (1,1)\}$, where addition of points $p_1+p_2$ is the point obtained by component wise addition of $p_1$ and $p_2$.*

*Proof.* The proof is given in Appendix 3.A. □

Before stating our next result we need the following definition. For a two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ with design rate $R_{\text{des}}$ we

define the function $\theta(e_1, e_2)$ for $(e_1, e_2) \in \mathcal{E}$ as

$$\begin{aligned}\theta(e_1,e_2) = &\sum_{\mathtt{l}_1,\mathtt{l}_2} \Lambda_{\mathtt{l}_1,\mathtt{l}_2} \log(1+u_1^{\mathtt{l}_2}u_2^{\mathtt{l}_2}) - \Lambda_1'(1,1)e_1\log u_1\\ &- \Lambda_2'(1,1)e_2\log u_2 + \frac{\Lambda_1'(1,1)}{\Gamma'^{(1)}(1)}\sum_{\mathtt{r}_1}\Gamma^{(1)}_{\mathtt{r}_1}\log q_{\mathtt{r}_1}(v_1)\\ &- \Lambda_1'(1,1)e_1\log v_1 + \frac{\Lambda_2'(1,1)}{\Gamma'^{(2)}(1)}\sum_{\mathtt{r}_2}\Gamma^{(2)}_{\mathtt{r}_2}\log q_{\mathtt{r}_2}(v_2)\\ &- \Lambda_2'(1,1)e_2\log v_2 - \Lambda_1'(1,1)h(e_1) - \Lambda_2'(1,1)h(e_2)\\ &- R_{\mathrm{des}}, \end{aligned} \tag{3.30}$$

where $u_1, u_2, v_1$, and $v_2$ are positive solutions to the following equations

$$\frac{v_1}{\Gamma^{(1)'}(1)}\sum_{\mathtt{r}_1}\mathtt{r}_1\Gamma^{(1)}_{\mathtt{r}_1}\frac{(1+v_1)^{\mathtt{r}_1-1}-(1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1}+(1-v_1)^{\mathtt{r}_1}} = e_1, \tag{3.31}$$

$$\frac{v_2}{\Gamma^{(2)'}(1)}\sum_{\mathtt{r}_2}\mathtt{r}_2\Gamma^{(2)}_{\mathtt{r}_2}\frac{(1+v_2)^{\mathtt{r}_2-1}-(1-v_2)^{\mathtt{r}_2-1}}{(1+v_2)^{\mathtt{r}_2}+(1-v_2)^{\mathtt{r}_2}} = e_2, \tag{3.32}$$

$$\frac{1}{\Lambda_1'(1,1)}\sum_{\mathtt{l}_1,\mathtt{l}_2}\Lambda_{\mathtt{l}_1,\mathtt{l}_2}\mathtt{l}_1\frac{u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2}}{1+u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2}} = e_1, \tag{3.33}$$

$$\frac{1}{\Lambda_2'(1,1)}\sum_{\mathtt{l}_1,\mathtt{l}_2}\Lambda_{\mathtt{l}_1,\mathtt{l}_2}\mathtt{l}_2\frac{u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2}}{1+u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2}} = e_2. \tag{3.34}$$

In the following theorem, we present a criterion for two edge type LDPC ensembles, which, when satisfied, guarantees that the actual rate is equal to the design rate.

**Theorem 3.3.9.** *Consider the two edge type LDPC ensemble* $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ *with design rate* $R_{\mathrm{des}}$. *Let* $N_W$ *be the total number of codewords of a randomly chosen code* $\mathcal{C}$ *from this ensemble and let* $R_{\mathcal{C}}$ *be the actual rate of the code* $\mathcal{C}$. *Then*

$$\lim_{N\to\infty}\frac{\log(\mathbb{E}[N_W])}{N} = \sup_{(e_1,e_2)\in\mathcal{E}}\theta(e_1,e_2) + R_{\mathrm{des}},$$

*where the set* $\mathcal{E}$ *is defined in Lemma 3.3.8 and* $\theta(e_1, e_2)$ *is defined in (3.30). Also, if* $\sup_{(e_1,e_2)\in\mathcal{E}}\theta(e_1,e_2) = 0$, *i.e.,* $\theta(1/2,1/2) \geq \theta(e_1,e_2), \forall (e_1,e_2)\in\mathcal{E}$, *then for any* $\delta > 0$

$$\lim_{N\to\infty}\Pr\left(R_{\mathcal{C}} \geq R_{\mathrm{des}} + \delta\right) = 0.$$

*Proof.* By (3.24), we have

$$\lim_{N\to\infty} \frac{\log(\mathbb{E}[N_W])}{N} = \sup_{(e_1,e_2)\in\mathcal{E}} \lim_{N\to\infty} \frac{\log(\mathbb{E}[N(e_1 N\Lambda_1'(1,1), e_2 N\Lambda_2'(1,1))])}{N}.$$

Using Stirling's approximation for the binomial coefficients and [BM04, Theorem 2] for the coefficient growths in Lemma 3.3.7 we know that

$$\lim_{N\to\infty} \frac{\log(\mathbb{E}[N(e_1 N\Lambda_1'(1,1), e_2 N\Lambda_2'(1,1))])}{N} = \sup_{(e_1,e_2)\in\mathcal{E}} \inf_{u_1,u_2,v_1,v_2>0} \psi(e_1,e_2,u_1,u_2,v_1,v_2)$$

where $\psi(e_1,e_2,u_1,u_2,v_1,v_2)$ is given by

$$\begin{aligned}
&\sum_{\mathtt{l}_1,\mathtt{l}_2} \Lambda_{\mathtt{l}_1,\mathtt{l}_2} \log(1+u_1^{\mathtt{l}_2}u_2^{\mathtt{l}_2}) - \Lambda_1'(1,1)e_1\log u_1 \\
&-\Lambda_2'(1,1)e_2\log u_2 + \frac{\Lambda_1'(1,1)}{\Gamma'^{(1)}(1)} \sum_{\mathtt{r}_1} \Gamma^{(1)}_{\mathtt{r}_1} \log q_{\mathtt{r}_1}(v_1) \\
&-\Lambda_1'(1,1)e_1\log v_1 + \frac{\Lambda_2'(1,1)}{\Gamma'^{(2)}(1)} \sum_{\mathtt{r}_2} \Gamma^{(2)}_{\mathtt{r}_2} \log q_{\mathtt{r}_2}(v_2) \\
&-\Lambda_2'(1,1)e_2\log v_2 - \Lambda_1'(1,1)h(e_1) - \Lambda_2'(1,1)h(e_2).
\end{aligned}$$

Further, the infimum of $\psi$ with respect to $u_1, u_2, v_1$, and $v_2$ can be found by solving the following saddle point equations

$$\frac{\partial\psi}{\partial u_1} = \frac{\partial\psi}{\partial u_2} = \frac{\partial\psi}{\partial v_1} = \frac{\partial\psi}{\partial v_2} = 0,$$

which are equivalent to (3.31) - (3.34). The second claim of the theorem follows from Lemma 2.3.2. □

Note that in general for a two edge type LDPC ensemble, in order to check if the actual rate is equal to the design rate, we need to compute the maximum of a two variable function over the set $\mathcal{E}$. However, the set $\mathcal{E}$ is just a line for left regular two edge type LDPC ensembles. Thus we deal with the case of left regular LDPC ensembles in the following lemma.

**Lemma 3.3.10.** *Consider the left regular two edge type LDPC ensemble* $\{\mathtt{l}_1, \mathtt{l}_2, \Gamma^{(1)}, \Gamma^{(2)}\}$ *with design rate* $R_{\text{des}}$*. Let* $N$ *be the total number of codewords of a randomly chosen code* $\mathcal{C}$ *from this ensemble and* $R_{\mathcal{C}}$ *be its actual rate. Then*

$$\lim_{N\to\infty} \frac{\log(\mathbb{E}[N_W])}{N} = \sup_{e\in(0,1)} \theta(e) + R_{\text{des}}.$$

*If* $\sup_{e\in(0,1)} \theta(e) = 0$ *i.e.* $\theta(1/2) \geq \theta(e), \forall e \in (0,1)$*, then for any* $\delta > 0$

$$\lim_{N\to\infty} \Pr\left(R_{\mathcal{C}} > R_{\text{des}} + \delta\right) = 0$$

*The function* $\theta(e)$ *is defined as*

$$\begin{aligned}\theta(e) = (1-\mathtt{l}_1-\mathtt{l}_2)h(e) &+ \frac{\mathtt{l}_1}{\Gamma^{(1)'}(1)} \sum_{\mathtt{r}} \Gamma^{(1)}_{\mathtt{r}} \log q_{\mathtt{r}}(v_1) \\ &+ \frac{\mathtt{l}_2}{\Gamma^{(2)'}(1)} \sum_{\mathtt{r}} \Gamma^{(2)}_{\mathtt{r}} \log q_{\mathtt{r}}(v_2) - e\mathtt{l}_1 \log v_1 - e\mathtt{l}_2 \log v_2 - R_{\text{des}},\end{aligned}$$

*where* $v_1$ *(resp.* $v_2$*) is the unique positive solution of (3.31) (resp. (3.32)) with* $e_1$ *(resp.* $e_2$*) substituted by* $e$ *on the RHS.*

*Proof.* Most of the arguments in this lemma are the same as those of Theorem 3.3.9, so we will omit them. First note that the cardinality of the set $\mathcal{W}(E_1, E_2)$, as defined in Lemma 3.3.7, is given by

$$\begin{aligned}|\mathcal{W}(E_1,E_2)| &= \operatorname{coef}\left\{(1+u_1^{\mathtt{l}_1}u_2^{\mathtt{l}_2})^N, u_1^{E_1}u_2^{E_2}\right\} \\ &= \begin{cases} 0 & \frac{E_2}{\mathtt{l}_2} \neq \frac{E_1}{\mathtt{l}_1}, \\ \binom{N}{E_1/\mathtt{l}_1} & \text{otherwise.}\end{cases}\end{aligned}$$

Let $e = E_1/(N\mathtt{l}_1) = E_2/(N\mathtt{l}_2)$. By Stirling's approximation and the saddle point approximation for the coefficient terms [RU08, pp. 517], we obtain

$$\begin{aligned}\lim_{N\to\infty} \frac{\log(\mathbb{E}[N_W])}{N} &= \lim_{N\to\infty} \sup_{e\in(0,1)} \frac{\log(\mathbb{E}[N(eN\mathtt{l}_1, eN\mathtt{l}_2)])}{N} \\ &= \sup_{e\in(0,1)} \inf_{y_1,y_2>0} \Big\{(1-\mathtt{l}_1-\mathtt{l}_2)h(e) \\ &\quad + \frac{\mathtt{l}_1}{\Gamma^{(1)'}(1)} \sum_{\mathtt{r}_1} \Gamma^{(1)}_{\mathtt{r}_1} \log q_{\mathtt{r}_1}(v_1) - e\mathtt{l}_1 \log v_1\end{aligned}$$

$$+ \frac{\mathtt{l}_1}{\Gamma^{(2)\prime}(1)} \sum_{\mathtt{r}_2} \Gamma^{(2)}_{\mathtt{r}_2} \log q_{\mathtt{r}_2}(v_2) - e\mathtt{l}_2 \log v_2 \Bigg\}$$
$$= \sup_{e\in(0,1)} \inf_{y_1,y_2>0} \psi(e, v_1, v_2)$$

The saddle point equations are obtained by taking the partial derivatives of $\psi$ with respect to $v_j, j \in \{1, 2\}$ and setting them equal to 0. These equations are the same as (3.31) (resp. (3.32)) with $e_1$ (resp. $e_2$) substituted by $e$ on the RHS.

□

*Remark:* Note that as in [MMU08], we can change the order of inf and sup. Taking the derivatives after changing the order gives a function which is an upper bound on $\theta(e)$. The advantage of this upper bound is that it can be computed without solving any saddle point equations. However, as opposed to the standard LDPC ensembles, for two edge type LDPC ensembles this upper bound is not tight and does not provide a meaningful criterion to check if the rate is equal to the design rate.

The following two lemmas show that in the case of a left regular ensemble where $\Gamma^{(1)}$ and $\Gamma^{(2)}$ both have only either odd or even degrees, the function $\theta(e)$ attains its maximum inside the interval $[0, 1/2]$.

**Lemma 3.3.11.** *Consider the left regular two edge type LDPC ensemble* $\{\mathtt{l}_1, \mathtt{l}_2, \Gamma^{(1)}, \Gamma^{(2)}\}$. *Let* $\theta(e)$ *be the function as defined in Lemma 3.3.10. If both* $\Gamma^{(1)}$ *and* $\Gamma^{(2)}$ *are such that both the type of check nodes only have odd degrees, then for* $e > 1/2$

$$\theta(e) < \theta(1/2).$$

*Proof.* The proof is given in Appendix 3.B. □

**Lemma 3.3.12.** *Consider the left regular two edge type LDPC ensemble* $\{\mathtt{l}_1, \mathtt{l}_2, \Gamma^{(1)}, \Gamma^{(2)}\}$. *Let* $\theta(e)$ *be the function as defined in Lemma 3.3.10. If both* $\Gamma^{(1)}$ *and* $\Gamma^{(2)}$ *are such that both the type of check nodes only have even degrees, then for* $e \in (0, 1/2)$

$$\theta(e) = \theta(1 - e).$$

*Proof.* The proof is given in Appendix 3.C. □

In the following theorem we state how we can compute the conditional entropy $H(X^N|S, Z^N)$ appearing in (3.22).

**Theorem 3.3.13.** *Consider transmission over the BEC-WT($\epsilon_m, \epsilon_w$) using a random code $\mathcal{C}$ from the two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ and the coset encoding method. Let $S$ be the message from Alice for Bob, $X^N$ be the transmitted word, and $Z^N$ be the wiretapper's observation.*

*Also consider a point-to-point communication setup for transmission over the BEC($\epsilon_w$) using the two edge type LDPC ensemble $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ Assume that the erasure probability $\epsilon_w$ is above the BP threshold of the ensemble. Let $\{\Omega, \Phi^{(1)}, \Phi^{(2)}\}$ be the residual ensemble resulted from the peeling decoder. Let $R^r_{\text{des}}$ be the design rate of the residual ensemble $\{\Omega, \Phi^{(1)}, \Phi^{(2)}\}$. If $\{\Omega, \Phi^{(1)}, \Phi^{(2)}\}$ satisfies the condition of Theorem 3.3.9, i.e. if the design rate of the residual ensemble is equal to the rate then*

$$\lim_{N\to\infty} \frac{\mathbb{E}(H_{\mathcal{C}}(X^N|S, Z^N))}{N} = \epsilon_w \Lambda(y_1, y_2) R^r_{\text{des}}, \tag{3.35}$$

*where $y_1$, and $y_2$ are the fixed points of the density evolution equations (3.8) and (3.9) obtained when initializing them with $x_1^{(1)} = x_2^{(2)} = \epsilon_w$.*

*Proof.* From Lemma 3.3.3, we know that the conditional entropy in the point-to-point set-up is identical to $H(X^N|S, Z^N)$. The conditional entropy in the point-to-point case is equal to the RHS of (3.35). This follows from the same arguments as in [MMU08, Theorem 10]. The quantity $\epsilon_w \Lambda(y_1, y_2)$ on the RHS of (3.35) is the ratio of the number of variable nodes in the residual ensemble to that in the initial ensemble. □

This gives us the following method to calculate the equivocation of Eve when using two edge type LDPC ensembles for the BEC-WT($\epsilon_m, \epsilon_w$) based on the coset encoding method.

1. If the threshold of the two edge type LDPC ensemble is lower than $\epsilon_w$, calculate the residual degree distribution for the two edge type LDPC ensemble for transmission over the BEC($\epsilon_w$). Check that the rate of this residual ensemble is equal to the design rate using Theorem 3.3.9. Calculate $H(X^N|S, Z^N)$ using Theorem 3.3.13. If the threshold is higher than $\epsilon_w$, $H(X^N|S, Z^N)$ is trivially zero.

2. If the threshold of the standard LDPC ensemble induced by type one edges is higher than $\epsilon_w$, calculate the residual degree distribution of this ensemble for transmission over the BEC($\epsilon_w$). Check

that its rate is equal to the design rate using Lemma 2.3.2. Calculate $H(X^N|Z^N)$ using Theorem 3.3.2. If the threshold is higher than $\epsilon_w$, $H(X^N|Z^N)$ is trivially zero.

3. Finally calculate $H(S|Z^N)$ using
$$H(S|Z^N) = H(X^N|Z^N) - H(X^N|S, Z^N).$$

In the following section we demonstrate this procedure by computing the equivocation of Eve for various two edge type LDPC ensembles.

## 3.4 Examples

*Example* 1. Consider using the ensemble defined by
*Standard LDPC Degree Distribution* 1.

$$\begin{aligned}\Lambda^{(1)}(x) = {} & 0.5572098x^2 + 0.1651436x^3 + 0.07567923x^4 \\ & + 0.0571348x^5 + .043603x^7 + 0.02679802x^8 \\ & + 0.013885518x^{13} + 0.0294308x^{14} + 0.02225301x^{31} \\ & + 0.00886105x^{100}, \\ \Gamma^{(1)}(x) = {} & 0.25x^9 + 0.75x^{10}\end{aligned}$$

from Section 3.2 for transmission over the BEC-WT$(0.5, 0.6)$ at rate $R = 0.498836$ b.p.c.u. (the full rate of the ensemble), without using the coset encoding scheme. Here every possible message $s$ corresponds to a single codeword $x^N$, and encoding and decoding is done as with a standard LDPC code. Since the threshold is 0.5, Bob can decode with error probability approaching zero. The equivocation of Eve is given by $H(S|Z^N) = H(X^N|Z^N)$ which can be calculated using the MMU method. In Figure 3.3 we plot the function $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ defined in Lemma 2.3.2 corresponding to the standard LDPC ensemble $\{\Omega^{(1)}, \Phi^{(1)}\}$, which is the average residual degree distribution of the ensemble induced by type one edges for transmission over BEC$(\epsilon_w)$.

From Lemma 2.3.2, if the maximum of $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ over the unit interval occurs at $u = 1$, which holds in this case, the design rate of the residual graph is equal to the actual rate. Thus we can calculate the average equivocation $\lim_{N\to\infty} H(X^N|Z^N)/N = 0.0989137$ b.p.c.u. Using this ensemble we can achieve the point $(R, R_e) = (0.498836, 0.0989137)$ in the rate-equivocation region which is very close to the point C $= (0.5, 0.1)$ in Figure 3.1.

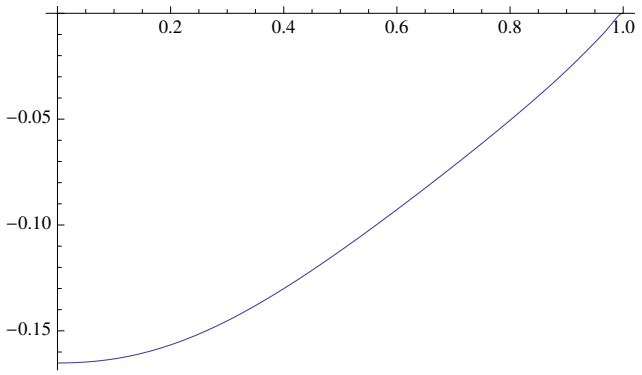


Figure 3.3: $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for Example 1 and 2.

*Example* 2. Now consider the two edge type ensemble defined by *Two Edge Type Degree Distribution* 1.

$$\begin{aligned}
\Lambda(x,y) = {} & 0.463846x^2 + 0.0814943x^2y + 0.0118691x^2y^2 \\
& + 0.14239x^3 + 0.0201658x^3y + 0.00258812x^3y^2 \\
& + 0.0292241x^4 + 0.0464551x^4y + 0.0564162x^5 \\
& + 0.000718585x^5y + 0.0436039x^7y \\
& + 0.0258926x^8y + 0.000905503x^8y^2 \\
& + 0.00631474x^{13}y^2 + 0.00757076x^{13}y^5 \\
& + 0.011051x^{14}y + 0.0173718x^{14}y^2 \\
& + 0.00100807x^{14}y^5 + 0.00240762x^{31} \\
& + 0.0012626x^{31}y^4 + 0.0185828x^{31}y^5 \\
& + 0.000326117x^{100}y^4 + 0.00383319x^{100}y^{17} \\
& + 0.00470174x^{100}y^{18}, \\
\Gamma^{(1)}(x) = {} & 0.25x^9 + 0.75x^{10}, \\
\Gamma^{(2)}(x) = {} & x^6,
\end{aligned}$$

from Section 3.2, for transmission over the BEC-WT$(0.5, 0.6)$ using the coset encoding scheme. Again Bob can decode since the threshold of the ensemble induced by type one edges is 0.5. Since the threshold of the two edge type ensemble is 0.6, we get $H(X^N|S, Z^N) = 0$, and

$H(S|Z^N) = H(X^N|Z^N)$. The degree distribution of type one edges is the same as the degree distribution in Example 1, so we again get $\lim_{N\to\infty} \mathbb{E}(H(X^N|Z^N))/N = 0.0989137$ b.p.c.u. Using this scheme we achieve the point $(R, R_e) = (0.0999064, 0.0989137)$ in the rate-equivocation region which is very close to point B $= (0.1, 0.1)$ in Figure 3.1.

*Example* 3. Consider transmission over the BEC-WT$(0.429, 0.75)$ using the coset encoding scheme and the regular two edge type ensemble defined by

*Two Edge Type Degree Distribution* 3.

$$\Lambda(x, y) = x^3 y^3$$
$$\Gamma^{(1)}(x) = x^6$$
$$\Gamma^{(2)}(x) = x^{12}.$$

The design rate of this ensemble is 0.25 b.p.c.u. and the threshold is 0.469746. The threshold for the ensemble induced by type one edges is 0.4294, so it can be used for reliable communication if $\epsilon_m < 0.4294$.

To calculate the equivocation of Eve, we first calculate $H(X^N|Z^N)/N$ by the MMU method. We calculate the average residual degree distribution $\{\Omega^{(1)}, \Phi^{(1)}\}$ of the ensemble induced by type one edges for erasure probability $\epsilon_w$ and plot $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ in Figure 3.4. As in Examples 1 and 2, we see that it takes its maximum at $u = 1$. Thus, by Lemma 2.3.2, we obtain that the conditional entropy is equal to the design rate of the residual ensemble, that is, $\lim_{N\to\infty} \mathbb{E}(H(X^N|Z^N))/N = 0.250124$ b.p.c.u.

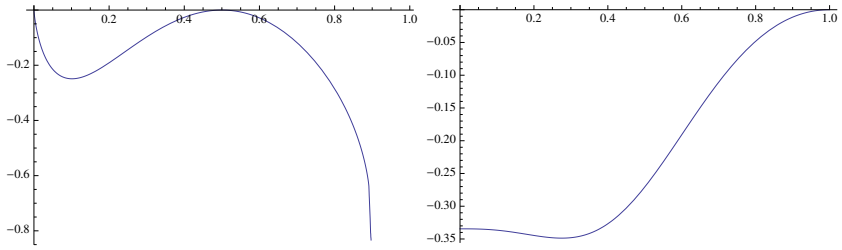


Figure 3.4: $\theta(e)$ and $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for Example 3.

We now calculate the residual degree distribution $(\Omega, \Phi^{(1)}, \Phi^{(2)})$ of the two edge type ensemble corresponding to erasure probability $\epsilon_w$ and plot

the function $\theta(e)$ defined in Lemma 3.3.10. If $\theta(e)$ is less than or equal to zero for $e \in [0,1]$, then the rate of the residual ensemble is equal to the design rate by Lemma 3.3.10. Then we can calculate $H(X^N|S,Z^N)$ using Lemma 3.3.13. In Figure 3.4 we see that $\sup_{e\in[0,1]} \theta(e) = 0$, and we get $\lim_{N\to\infty} \mathbb{E}(H(X^N|S,Z^N))/N = 0.000124297$ b.p.c.u.

Finally, using (3.22) we get $\lim_{N\to\infty} \mathbb{E}(H(S|Z^N))/N = 0.24999998$ b.p.c.u. We thus achieve the point $(R, R_e) = (0.25, 0.24999998)$ in the rate-equivocation region. We see that we are very close to perfect secrecy. The reason that we are so far away from the secrecy capacity $C_s = 0.321$ is that the $(3,6)$ ensemble for the main channel is far from being capacity achieving.

*Example* 4. Consider the two edge type ensemble

*Two Edge Type Degree Distribution* 4.

$$\begin{aligned}
\Lambda(x,y) &= 0.5572098x^2y^3 + 0.1651436x^3y^3 + 0.07567923x^4y^3 \\
&\quad + 0.0571348x^5y^3 + \ .043603x^7y^3 + 0.02679802x^8y^3 \\
&\quad + 0.013885518x^{13}y^3 + 0.0294308x^{14}y^3 \\
&\quad + 0.02225301x^{31}y^3 + 0.00886105x^{100}y^3, \\
\Gamma^{(1)}(x) &= 0.25x^9 + 0.75x^{10}, \\
\Gamma^{(2)}(x) &= x^{12}
\end{aligned}$$

where the graph induced by type one edges has the same degree distribution as Standard LDPC Degree Distribution 1 and the graph induced by type two edges is $(3,12)$ regular. The rate of the overall ensemble is 0.248836 b.p.c.u. and the rate from Alice to Bob is $R = 0.25$ b.p.c.u. Consider transmission over the BEC-WT$(0.5, 0.751164)$.

In Figure 3.5, we plot $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for the residual ensemble $\{\Omega^{(1)},\Phi^{(1)}\}$ induced by type one edges for transmission over BEC$(\epsilon_w)$. Since the maximum of $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ over the unit interval occurs at $u = 1$, we obtain by Lemma 2.3.2 that the rate is equal to the design rate for this residual ensemble. In Figure 3.5 we plot $\theta(e_1,e_2)$ for the residual ensemble $(\Omega,\Phi^{(1)},\Phi^{(2)})$ of the two edge type LDPC ensemble for transmission over BEC$(\epsilon_w)$. Since the maximum of $\theta(e_1,e_2)$ over the set $\mathcal{E}$ is zero, we obtain by Theorem 3.3.9 that the rate is equal to the design rate for this residual two edge type ensemble. In this case we can calculate the equivocation of Eve and find it to be 0.24999999 b.p.c.u., which is very close to the rate. Thus this ensemble achieves the point

$(R, R_e) = (0.25, 0.24999999)$ in the capacity-equivocation region in Figure 3.1. Note that the secrecy capacity is 0.251164 b.p.c.u.

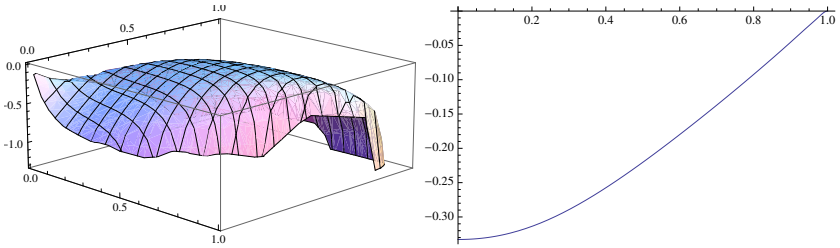


Figure 3.5: $\theta(e_1, e_2)$ and $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for Example 4.

These examples demonstrate that there exist simple ensembles with very good secrecy performance.

In the following section we consider spatially coupled two edge type LDPC codes. They are the extension of the spatially coupled LDPC codes of [KRU10] to two edge type codes. We show that regular two edge type spatially coupled codes are optimal for the BEC-WT$(\epsilon_m, \epsilon_w)$.

## 3.5 Spatially Coupled Codes

In this section we describe spatially coupled LDPC codes for the wiretap channel. They are the two edge type equivalent to the spatially coupled codes of Section 2.3.3. As for the standard LDPC codes of the previous sections we will use Wyner's coset encoding method described in Definition 2.2.1.

In the previous sections we considered irregular two edge type ensembles, but for our purposes here it is sufficient to focus on regular two edge type LDPC ensembles.

**Definition 3.5.1** ($\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2\}$ Two Edge Type LDPC Ensemble)**.** *A $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2\}$ two edge type LDPC ensemble of block length $N$ contains all the bipartite graphs (allowing multiple edges between a variable node and a check node) where all the $N$ variable nodes are connected to $\mathtt{l}_i$ check nodes of type $i$ and all the type $i$ check nodes have degree $\mathtt{r}_i$, $i \in \{1, 2\}$.*

A protograph of a regular two edge type LDPC code with $\mathtt{l}_1 = \mathtt{l}_2 = 3$ and $\mathtt{r}_1 = \mathtt{r}_2 = 6$ is shown in Figure 3.6.

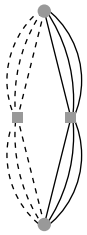

Figure 3.6: A protograph of a two edge type LDPC ensemble with $\mathtt{l}_1 = \mathtt{l}_2 = 3$ and $\mathtt{r}_1 = \mathtt{r}_2 = 6$.

Based on the definition of an $\{\mathtt{l}, \mathtt{r}, L, w\}$ ensemble from [KRU10], we define the regular spatially coupled two edge type LDPC ensemble. Before giving this definition, we define $\mathcal{T}(\mathtt{l})$ to be the set of $w$-tuples of non-negative integers which sum to $\mathtt{l}$. More precisely,

$$\mathcal{T}(\mathtt{l}) = \{(t_0, \cdots, t_{w-1}) : \sum_{j=0}^{w-1} t_j = \mathtt{l}\}.$$

*Remark:* Note that the $w$-tuple $(t_0, \cdots, t_{w-1})$ is called a *type* in [KRU10]. We avoid this terminology as we refer to different edges in the two edge type LDPC ensemble by their type.

**Definition 3.5.2** ($\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2, L, w\}$ Spatially Coupled Two Edge Type LDPC Ensemble)**.** *Assume that there are $M$ variable nodes each at the positions $[-L, L]$, $L \in \mathbb{N}$. The block length of a code in the ensemble is $N = M(2L+1)$. Every variable node has degree $\mathtt{l}_1$ with respect to type 1 edges and $\mathtt{l}_2$ with respect to type 2 edges. At each position $[-L, L+w-1]$ there are $\frac{\mathtt{l}_1}{\mathtt{r}_1}M$ type 1 check nodes and $\frac{\mathtt{l}_2}{\mathtt{r}_2}M$ type 2 check nodes. All type 1 check nodes have degree $\mathtt{r}_1$ and all type 2 check nodes have degree $\mathtt{r}_2$.*

*Assume that for each variable node we order its edges in an arbitrary but fixed order. A type $j$ constellation $c$ is an $\mathtt{l}_j$-tuple, $c = (c_1, \cdots, c_{\mathtt{l}_j})$ with elements in $\{0, 1, \cdots, w-1\}^{\mathtt{l}_j}$. Its operational significance is that if a variable node at position $i$ has type $j$ constellation $c$ then its $k$-th edge of type $j$ is connected to a check node at position $i + c_k$, $j \in \{1, 2\}$. We denote the set of all type $j$ constellations by $\mathfrak{C}_j$. Let $\tau(c)$ be the $w$-tuple which counts the occurrence of $0, 1, \cdots, w-1$ in $c$. Clearly, if $c$ is a type $j$ constellation then $\tau(c) \in \mathcal{T}(\mathtt{l}_j)$. We impose a uniform distribution over both types of constellations. This imposes the following distribution over*

$t \in \mathcal{T}(\mathtt{l}_j)$

$$p^{(j)}(t) = \frac{|\{c \in \mathfrak{C}_j : \tau(c) = t\}|}{w^{\mathtt{l}_j}}, \quad j \in \{1, 2\}.$$

*Now we pick $M$ so that $Mp^{(1)}(t_1)p^{(2)}(t_2)$ is a natural number for $\forall t_1 \in \mathcal{T}(\mathtt{l}_1), \forall t_2 \in \mathcal{T}(\mathtt{l}_2)$. For each position $i$ pick $Mp^{(1)}(t_1)p^{(2)}(t_2)$ variable nodes. For each of these variable nodes we use a random permutation over $\mathtt{l}_j$ letters to map $t_j$ to a type $j$ constellation $c$. We then assign the type $j$ edges of the variable nodes according to the constellation $c$. We do this for both type 1 and 2 edges. Ignoring boundary effects, for each check position $i$, the number of type $j$ edges that come from variables at position $i-k$, $k \in \{0, \cdots, w-1\}$, is $M\frac{\mathtt{l}_j}{w}$. This implies that exactly a fraction $\frac{1}{w}$ of the total number $M\mathtt{l}_j$ of type $j$ sockets at position $i$. At the check nodes, we distribute these edges by randomly choosing a permutation over $M\mathtt{l}_j$ letters, to the $M\frac{\mathtt{l}_j}{\mathtt{r}_j}$ check nodes of type $j$, $j \in \{1, 2\}$.*

*Remark:* Each of the $\mathtt{l}_1$ (resp. $\mathtt{l}_2$) type 1 (resp. 2) connections of a variable node at position $i$ is uniformly and independently chosen from the range $[i, \ldots, i + w - 1]$, where $w$ is a "smoothing" parameter. Similarly, as was remarked in [KRU10], for each check node each edge is roughly independently chosen to be connected to one of its nearest $w$ "left" neighbors. More precisely, the corresponding probability deviates at most by a term of order $1/M$ from the uniform distribution.

To summarize, a $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2, L, w\}$ spatially coupled two edge type LDPC ensemble is obtained by replacing the standard regular LDPC ensemble in the $(\mathtt{l}, \mathtt{r}, L, w)$ ensemble (defined in [KRU10]) by a $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2\}$ two edge type LDPC ensemble. The spatial coupling is done such that only edges of the same type are coupled together. An

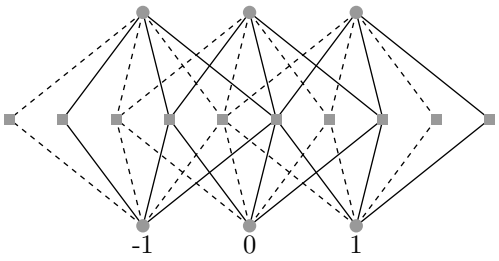


Figure 3.7: A coupled chain of protographs of a two edge type LDPC code with $L = 1$ for $\mathtt{l}_1 = \mathtt{l}_2 = 3$ and $\mathtt{r}_1 = \mathtt{r}_2 = 6$.

example of a protograph of a two edge type LDPC code is shown in Figure 3.6 and its spatially coupled version is shown in Figure 3.7. Solid lines correspond to type one edges and dashed lines to type two edges.

In the next lemma we show that if the degrees of the two types of check nodes are the same, i.e. if $\mathtt{r}_1 = \mathtt{r}_2 = \mathtt{r}$, then the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ spatially coupled two edge type LDPC ensemble has the same asymptotic performance as that of the spatially coupled ensemble $(\mathtt{l}_1 + \mathtt{l}_2, \mathtt{r}, L, w)$.

**Lemma 3.5.3.** *The $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ spatially coupled two edge type LDPC ensemble has the same BP threshold as the spatially coupled ensemble $(\mathtt{l}_1 + \mathtt{l}_2, \mathtt{r}, L, w)$.*

*Proof.* Let $x_i^{(k,j)}$ be the average erasure probability which is emitted by a variable node at position $i$ in the $k$th iteration along an edge of type $j$, $j \in \{1, 2\}$. For $i \notin [-L, L]$, we set $x_i^{(k,j)} = 0$. For $i \in [-L, L]$, $j \in \{1, 2\}$, and $k = 1$, we set $x_i^{(k,j)} = \epsilon$.

As in [KRU10], the density evolution recursion for the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ two edge type spatially coupled LDPC ensemble is given by

$$x_i^{(k,1)} = \epsilon \left(1 - \frac{1}{w} \sum_{p=0}^{w-1} \left(1 - \frac{1}{w} \sum_{m=0}^{w-1} x_{i+p-m}^{(k-1,1)}\right)^{\mathtt{r}-1}\right)^{\mathtt{l}_1 - 1} \left(1 - \frac{1}{w} \sum_{p=0}^{w-1} \left(1 - \frac{1}{w} \sum_{m=0}^{w-1} x_{i+p-m}^{(k-1,2)}\right)^{\mathtt{r}-1}\right)^{\mathtt{l}_2}, \tag{3.36}$$

$$x_i^{(k,2)} = \epsilon \left(1 - \frac{1}{w} \sum_{p=0}^{w-1} \left(1 - \frac{1}{w} \sum_{m=0}^{w-1} x_{i+p-m}^{(k-1,1)}\right)^{\mathtt{r}-1}\right)^{\mathtt{l}_1} \left(1 - \frac{1}{w} \sum_{p=0}^{w-1} \left(1 - \frac{1}{w} \sum_{m=0}^{w-1} x_{i+p-m}^{(k-1,2)}\right)^{\mathtt{r}-1}\right)^{\mathtt{l}_2 - 1}. \tag{3.37}$$

Here $x_i^{(k,1)} = x_i^{(k,2)}$ if $x_i^{(k-1,1)} = x_i^{(k-1,2)}$. Indeed, for $k = 1$ and $i \in [-L, L]$ , $x_i^{(1,1)} = x_i^{(1,2)} = \epsilon$ and for $i \notin [-L, L]$, $x_i^{(1,1)} = x_i^{(1,2)} = 0$. Thus, by induction on the number of iterations $k$, $x_i^{(k,1)} = x_i^{(k,2)}$. Hence we drop the superscript corresponding to the type of edge and write the

density evolution recursion as

$$x_i^{(k)} = \epsilon \left( 1 - \frac{1}{w} \sum_{p=0}^{w-1} \left( 1 - \frac{1}{w} \sum_{m=0}^{w-1} x_{i+p-m}^{(k-1)} \right)^{\mathtt{r}-1} \right)^{\mathtt{l}_1+\mathtt{l}_2-1}. \tag{3.38}$$

This recursion is same as that of $\{\mathtt{l}_1 + \mathtt{l}_2, \mathtt{r}, L, w\}$ spatially coupled ensemble given in [KRU10]. This proves the lemma. □

Before proving the main result, we show that regular two edge type LDPC ensembles $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}\}$ have the same growth rate of the average stopping set distribution as that of the standard regular $\{\mathtt{l}_1 + \mathtt{l}_2, \mathtt{r}\}$ LDPC ensemble.

**Lemma 3.5.4.** *Consider the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}\}$ regular two edge type LDPC ensemble with block length $N$, $\mathtt{l}_1 \geq 3$, and positive design rate. Let $N_{SS}(N, \omega N)$ be the stopping set distribution of a randomly chosen code from this ensemble and let $\mathbb{E}(N_{SS}(N, \omega N))$ be its average. Then the growth rate of $\mathbb{E}(N_{SS}(N, \omega N))$ is the same as that of the standard regular $\{\mathtt{l}_1 + \mathtt{l}_2, \mathtt{r}\}$ ensemble. In particular, the minimum stopping set weight of the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}\}$ regular two edge type LDPC ensemble grows linearly in $N$.*

*Proof.* A stopping set is a subset of the variable nodes such that all check nodes connected to this subset is connected to it by at least two edges. We first show that

$$\mathbb{E}(N_{SS}(N, \omega N)) = \binom{N}{N\omega} \frac{\operatorname{coef}\left\{ p^{(\mathtt{r})}(x)^{\frac{\mathtt{l}_1 N}{\mathtt{r}}}, x^{\omega \mathtt{l}_1 N} \right\} \operatorname{coef}\left\{ p^{(\mathtt{r})}(x)^{\frac{\mathtt{l}_2 N}{\mathtt{r}}}, x^{\omega \mathtt{l}_2 N} \right\}}{\binom{\mathtt{l}_1 N}{\omega \mathtt{l}_1 N} \binom{\mathtt{l}_2 N}{\omega \mathtt{l}_2 N}}, \tag{3.39}$$

where $p^{(\mathtt{r})}(x) = (1+x)^{\mathtt{r}} - \mathtt{r}x$. The proof of this is similar to the proof of Lemma 3.3.7. Let $v^N$ be a binary vector of weight $\omega N$. From the symmetry of code generation we see that the probability that a vector $v^N$ is a stopping set of a randomly chosen code is independent of $v^N$. Thus we can fix $v^N$ to a specific vector and get

$$\mathbb{E}(N_{SS}(N, \omega N)) = \binom{N}{\omega N} \times \Pr(v^N \text{ is a stopping set}), \tag{3.40}$$

where $\binom{N}{\omega N}$ is the number of vectors of weight $\omega N$. The probability that $v^N$ is a stopping set is

$$\Pr\left(v^N \text{ is a stopping set}\right) = \frac{\text{Total number of graphs for which } v^N \text{ is a stopping set}}{\text{Total number of graphs}}. \quad (3.41)$$

There are $\mathtt{l}_j\omega N$ outgoing edges of type $j$ from $v^N$, and the number of ways of connecting them to the $\frac{dl_j}{\mathtt{r}}N$ type $j$ check nodes, so that no check node is connected exactly once is given by

$$\operatorname{coef}\left\{\left(\sum_{k=0,2,3,\ldots}^{\mathtt{r}}\binom{r}{k}x^k\right)^{\frac{\mathtt{l}_j N}{\mathtt{r}}}, x^{\omega \mathtt{l}_j N}\right\} = \operatorname{coef}\left\{p^{(\mathtt{r})}(x)^{\frac{\mathtt{l}_j N}{\mathtt{r}}}, x^{\omega \mathtt{l}_j N}\right\}. \quad (3.42)$$

The number of ways of permuting the $\mathtt{l}_j\omega N$ type $j$ edges connected to $v^N$, and the $\mathtt{l}_j N - \mathtt{l}_j\omega N$ type $j$ edges not connected to $v^N$ is given by $(\mathtt{l}_j\omega N)!(\mathtt{l}_j N - \mathtt{l}_j\omega N)!$ and the total number of graphs is given by $(\mathtt{l}_1 N)!(\mathtt{l}_2 N)!$. Combining these results we get

$$\Pr\left(v^N \text{ is a stopping set}\right) = \frac{\operatorname{coef}\left\{p^{(\mathtt{r})}(x)^{\frac{\mathtt{l}_1 N}{\mathtt{r}}}, x^{\omega \mathtt{l}_1 N}\right\}\operatorname{coef}\left\{p^{(\mathtt{r})}(x)^{\frac{\mathtt{l}_2 N}{\mathtt{r}}}, x^{\omega \mathtt{l}_2 N}\right\}}{\binom{\mathtt{l}_1 N}{\omega \mathtt{l}_1 N}\binom{\mathtt{l}_2 N}{\omega \mathtt{l}_2 N}}, \quad (3.43)$$

which gives us (3.39).

Using Stirling's approximation for the binomial terms and the Hayman expansion for the coefficient term, see [RU08, Appendix D], we obtain

$$\begin{aligned}\lim_{N\to\infty}\frac{\log\left(E(N_{SS}(N,\omega N))\right)}{N} &= (1-\mathtt{l}_1-\mathtt{l}_2)h(\omega)\\ &\quad + \frac{\mathtt{l}_1}{\mathtt{r}}\log\left(p^{(\mathtt{r})}(t)\right) - \omega\mathtt{l}_1\log(t)\\ &\quad + \frac{\mathtt{l}_2}{\mathtt{r}}\log\left(p^{(\mathtt{r})}(t)\right) - \omega\mathtt{l}_2\log(t), \quad (3.44)\end{aligned}$$

where $t$ is a positive solution of

$$x\frac{(1+x)^{\mathtt{r}-1}-1}{(1+x)^{\mathtt{r}}-\mathtt{r}x} = \omega. \quad (3.45)$$

From (3.44), we see that the growth rate is the same as that of the average stopping set distribution of the standard $\{\mathtt{l}_1 + \mathtt{l}_2, \mathtt{r}\}$ regular LDPC ensemble [OVZ05, Theorem 2]. Now, the linearity of minimum stopping set distance immediately follows from [OVZ05, Cor. 7]. □

*Remark:* We could have come to this conclusion by specializing the general result contained in [KAD+09, Theorem 5]. But for the convenience of the reader, and since the above proof is so short, we decided to include a complete proof.

Lemma 3.5.4 and [KRU10, Lemma 1] imply that $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ spatially coupled two edge type LDPC ensembles with variable node degree at least three have a linear minimum stopping set distance. This gives us the following lemma on the block error probability of the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ ensemble under iterative decoding.

**Lemma 3.5.5.** *Consider transmission over the BEC($\epsilon$) using the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ spatially coupled two edge type LDPC ensemble with BP threshold $\epsilon^{\mathrm{BP}}$ and block length $N$. Let $\mathtt{l}_1 + \mathtt{l}_2 \geq 3$. Assume that $\epsilon < \epsilon^{\mathrm{BP}}$. Denote by $P_e^N$ the block error probability under iterative decoding. Then*

$$\lim_{N\to\infty} N P_e^N = 0.$$

*Proof.* In fact, a much stronger result is true – the block error probability converges to 0 exponentially fast. But for our purpose we only need that it converges to zero faster than linearly.

To see why this is correct, fix $\epsilon < \epsilon^{\mathrm{BP}}$. Then, for any $\delta > 0$, there exists a $k$ so that after $k$ iterations of DE, the bit error probability is below $\delta/3$. Further, for $N = N(k)$, sufficiently large, the expected behavior over all instances of the code and the channel deviates from the density evolution predictions by at most $\delta/3$. Finally, by standard concentration results (see [RU08, Theorem 3.30]) it follows that the probability that a particular instance deviates more than $\delta/3$ from its average decays exponentially fast in the block length.

We summarize, with a probability which converges exponentially fast (in the block length) to 1, an individual instance will have reached a bit error probability of at most $\delta$ after a fixed number of iterations.

If $\delta$ is chosen sufficiently small, in particular smaller than the relative minimum stopping set weight, then we know that the decoder can correct the remaining erasures with probability 1. □

In the following lemma we calculate the design rate of the spatially coupled two edge type ensemble.

**Lemma 3.5.6** (Design Rate)**.** *The design rate of the spatially coupled two edge type ensemble* $(\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2, L, w\})$ *with* $w \leq 2L$ *is given by*

$$R(\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2, L, w) = \tag{3.46}$$

$$\left(1 - \frac{\mathtt{l}_1}{\mathtt{r}_1} - \frac{\mathtt{l}_2}{\mathtt{r}_2}\right) - \left(\frac{\mathtt{l}_1}{\mathtt{r}_1} + \frac{\mathtt{l}_2}{\mathtt{r}_2}\right) \frac{w + 1 - 2\sum_{i=0}^{w} \left(\frac{i}{w}\right)^r}{2L+1}. \tag{3.47}$$

*The design rate of the coset encoding scheme for the wiretap channel is given by*

$$R_{\text{des}} = \frac{\mathtt{l}_2}{\mathtt{r}_2} - \frac{\mathtt{l}_2}{\mathtt{r}_2} \frac{w + 1 - 2\sum_{i=0}^{w} \left(\frac{i}{w}\right)^r}{2L+1}. \tag{3.48}$$

*Proof.* Let $C_1(C_2)$ be the number of type one (two) check nodes connected to variable nodes and let $V$ be the number of variable nodes. Then $R(\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2, L, w) = 1 - C_1/V - C_2/V$ and $R_{\text{des}} = C_2/V$. The calculations then follow from the proof of [KRU10, Lemma 3]. □

The number of possible messages $s$ of the coset encoding scheme is given by the number of cosets of $\mathcal{C}_N^{(1,2)}$ in $\mathcal{C}_N^{(1)}$. For a standard LDPC ensemble the design rate is a lower bound on the rates of the codes in the ensemble. This is not true for the coset encoding scheme for the wiretap channel. For example, suppose that the rate of $\mathcal{C}_N^{(1)}$ equals the design rate, but the rate of $\mathcal{C}_N^{(1,2)}$ is higher than its design rate. Then there will be fewer cosets than the maximum possible value. This corresponds to the equation

$$\begin{bmatrix} H_1 \\ H_2 \end{bmatrix} X^N = [0 \cdots 0 \; S]^T$$

not having solutions for some $S$.

Now, we are ready to state one of our main theorems. It shows that, by spatial coupling of two edge type LDPC codes, we can achieve perfect secrecy (the branch AB in Figure 3.1), and in particular the secrecy capacity (the point B in Figure 3.1) of the binary erasure wiretap channel.

**Theorem 3.5.7.** *Consider transmission over the BEC-WT*$(\epsilon_m, \epsilon_w)$ *using the spatially coupled regular* $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ *two edge type LDPC ensemble. Assume that the desired rate of information transmission from Alice to Bob is* $R$, $R \leq C_m - C_w$. *Let* $\mathtt{l}_1 = \lceil (1 - C_w - R)\mathtt{r} \rceil$ *and* $\mathtt{l}_2 = \lceil (1 - C_w)\mathtt{r} \rceil - \lceil (1 - C_w - R)\mathtt{r} \rceil$. *Let* $R_e^N$ *be the average (over the channel and ensemble) equivocation achieved for the wiretapper. Then,*

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} \mathbb{E}\left(P_e^N(\mathcal{C}_N)\right) = 0,$$

$$\lim_{\mathtt{r}\to\infty}\lim_{w\to\infty}\lim_{L\to\infty}\lim_{M\to\infty} R_e^N = R.$$

*Let $R(\mathcal{C}_N)$ be the rate from Alice to Bob of a randomly chosen code in the ensemble. Then*

$$\lim_{\mathtt{r}\to\infty}\lim_{w\to\infty}\lim_{L\to\infty}\lim_{M\to\infty} \Pr(R(\mathcal{C}_N) < R) = 0.$$

*Proof.* We first show that the rate from Alice to Bob is $R$ almost surely. Let $\mathcal{C}_N^{(1,2)}$ be a two edge type spatially coupled code, and let $\mathcal{C}_N^{(1)}$ be the code induced by its type 1 edges only. Then

$$R(\mathcal{C}_N) = R(\mathcal{C}_N^{(1)}) - R(\mathcal{C}_N^{(1,2)}). \tag{3.49}$$

Since both the two edge type spatially coupled ensemble and the ensemble induced by its type 1 edges are capacity achieving we must have

$$\lim_{\mathtt{r}\to\infty}\lim_{w\to\infty}\lim_{L\to\infty}\lim_{M\to\infty} \Pr(R(\mathcal{C}_N^{(1)}) > C_w + R) = 0, \tag{3.50}$$

$$\lim_{\mathtt{r}\to\infty}\lim_{w\to\infty}\lim_{L\to\infty}\lim_{M\to\infty} \Pr(R(\mathcal{C}_N^{(1,2)}) > C_w) = 0. \tag{3.51}$$

This implies that

$$\lim_{\mathtt{r}\to\infty}\lim_{w\to\infty}\lim_{L\to\infty}\lim_{M\to\infty} \Pr(R(\mathcal{C}_N) < R) = 0. \tag{3.52}$$

The reliability part easily follows from the capacity achieving property of the spatially coupled ensemble. This is because the rate of the ensemble corresponding to type 1 edges approaches $C_w + R$. As this ensemble is capacity achieving, its threshold is $1 - C_w - R$. As $R < C_m - C_w$, we see that the threshold is greater than $\epsilon_m$. This proves reliability.

To bound the equivocation of Eve, we expand the mutual information $I(X^N, S; Z^N)$ in two different ways using the chain rule

$$I(X^N, S; Z^N) = I(X^N; Z^N) + I(S; Z^N \mid X^N) \tag{3.53}$$

$$= I(S; Z^N) + I(X^N; Z^N \mid S). \tag{3.54}$$

As $S \to X^N \to Z^N$ is a Markov chain, $I\left(S; Z^N \mid X^N\right) = 0$. Using that

$I(S;Z^N) = H(S) - H(S \mid Z^N)$, we obtain,

$$\begin{aligned}\frac{1}{N}H(S \mid Z^N) &= \frac{1}{N}\big(H(S) + I(X^N; Z^N \mid S) - I(X^N; Z^N)\big) \\ &= \frac{1}{N}\big(H(S) + H(X^N \mid S) - H(X^N \mid Z^N, S)\big) \\ &\quad - \frac{I(X^N; Z^N)}{N} \\ &\geq \frac{1}{N}\big(H(X^N) - H(X^N \mid Z^N, S)\big) - C_w, \end{aligned} \tag{3.55}$$

where we have used that $H(S) + H(X^N \mid S) = H(S, X^N) = H(X^N)$ and that $I(X^N; Z^N)/N \leq C_w$.

Since the ensemble induced by type 1 edges is capacity achieving its rate must equal its design rate asymptotically, so

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{N\to\infty} H(X^N)/N = R + C_w. \tag{3.56}$$

Denote the block error probability of decoding $X^N$ from $Z^N$ and $S$ by $P_e^{N,S}$. From Fano's inequality we obtain,

$$\frac{H(X^N \mid S, Z^N)}{N} \leq \frac{h(P_e^{N,S})}{N} + P_e^{N,S}(1 - \epsilon_w). \tag{3.57}$$

Note that, as the two edge type spatially coupled construction is capacity achieving over the wiretapper's channel,

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} P_e^{N,S} = 0.$$

We now obtain the desired bound on the equivocation by substituting (3.57) and (3.56) in (3.55), and taking the limit $\mathtt{r}, w, L, M \to \infty$. □

Note that in the previous theorem our requirement was to have perfect secrecy. Hence we constructed a spatially coupled two edge type code which was capacity achieving over the wiretapper's channel. In the next theorem we prove that using spatially coupled two edge LDPC codes, it is possible to achieve an information rate equal to $C_m$, the capacity of the main channel, and equivocation equal to $C_m - \epsilon_w$.

**Theorem 3.5.8.** *Consider transmission over the BEC-WT($\epsilon_m, \epsilon_w$) using the spatially coupled regular $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ two edge type LDPC ensemble. Assume that the desired rate of information transmission from*

*Alice to Bob is $R$, $R > C_m - C_w$ and $R \leq C_m$. Let $\mathtt{l}_1 = \lceil (1 - C_m)\mathtt{r} \rceil$ and $\mathtt{l}_2 = \lceil R\mathtt{r} \rceil$. Let $R_e^N$ be the average (over the channel and ensemble) equivocation achieved for the wiretapper. Then,*

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} \mathbb{E}\left(P_e^N\right) = 0,$$

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} R_e^N = C_m - C_w.$$

*Let $R(\mathcal{C}_N)$ be the rate from Alice to Bob of a randomly chosen code in the ensemble. Then*

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} \Pr(R(\mathcal{C}_N) < R) = 0.$$

*Proof.* The proof that the rate is $R$ asymptotically is the same as in the proof of Theorem 3.5.7.

The reliability part easily follows from the capacity achieving property of the spatially coupled ensemble corresponding to type 1 edges. This is because the rate of the ensemble corresponding to type 1 edges approaches $C_m$. As this ensemble is capacity achieving, its threshold is $\epsilon_m$. This proves reliability.

The proof for equivocation is very similar to that of Theorem 3.5.7. From (3.55), we know

$$\frac{1}{N} H(S \mid Z^N) \geq \frac{1}{N} \left( H(X^N) - H(X^N \mid Z^N, S) \right) - C_w. \tag{3.58}$$

Since the code induced by type 1 edges is capacity achieving we have

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{N\to\infty} H(X^N)/N = C_m. \tag{3.59}$$

Note that as the two edge type code has rate $C_m - R$ and is capacity achieving, its threshold for the BEC is $1 - C_m + R$. As $R > C_m - C_w$, the threshold is higher than $\epsilon_w$. As in Theorem 3.5.7, given $S$ the error probability of decoding $X^N$ from $Z^N$, denoted by, $P_e^{N,S}$ goes to zero. Thus (3.57) holds and we obtain

$$\lim_{\mathtt{r}\to\infty} \lim_{w\to\infty} \lim_{L\to\infty} \lim_{M\to\infty} \frac{H(X^N \mid S, Z^N)}{N} = 0. \tag{3.60}$$

We obtain the desired bound on the equivocation by substituting (3.59) and (3.60) in (3.58), and taking the limit $\mathtt{r}, w, L, M \to \infty$. □

In the next subsection we show some simulation results measuring the equivocation of Eve for some different ensembles.

### 3.5.1 Simulation Results

We have showed that the ensemble $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}, \mathtt{r}, L, w\}$ achieves the secrecy capacity of the BEC-W$(\epsilon_m, \epsilon_w)$. In this section we show some simulation results for the $\{\mathtt{l}_1, \mathtt{l}_2, \mathtt{r}_1, \mathtt{r}_2, L\}$ ensemble over the BEC-WT$(0.5, 0.75)$. This ensemble is the two edge type equivalent of the $\{\mathtt{l}, \mathtt{r}, L\}$ ensemble in [KRU10]. In Table 3.1 we show the design rate and the equivocation for the $\{3, 3, 6, 12, L\}$ ensemble for different values of $L$ and $M = 1000$. We see that as $L$ increases the equivocation approaches the rate, and the rate approaches the secrecy capacity $C_S = 0.25$.

| $L$ | 20 | 30 | 40 | 50 | 60 | 70 |
|---|---|---|---|---|---|---|
| $R$ | 0.2622 | 0.2582 | 0.2562 | 0.255 | 0.2541 | 0.2535 |
| $R_e$ | 0.2276 | 0.235 | 0.2387 | 0.241 | 0.2425 | 0.2436 |

Table 3.1: Rate and equivocation for the $\{3, 3, 6, 12, L\}$ ensemble with $M = 1000$.

## Appendix 3.A Proof of Lemma 3.3.8

The terms in the expansion of $\prod_{\mathtt{l}_1,\mathtt{l}_2}(1 + u_1^{\mathtt{l}_1} u_2^{\mathtt{l}_2})^{N\Lambda_{\mathtt{l}_1,\mathtt{l}_2}}$ have the form

$$u_1^{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 k(\mathtt{l}_1,\mathtt{l}_2)\Lambda_{\mathtt{l}_1,\mathtt{l}_2}} u_2^{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 k(\mathtt{l}_1,\mathtt{l}_2)\Lambda_{\mathtt{l}_1,\mathtt{l}_2}},$$

where $0 \leq k(\mathtt{l}_1, \mathtt{l}_2) \leq N$. If the coefficient of $u_1^{e_1 N\Lambda_1'(1,1)} u_2^{e_2 N\Lambda_2'(1,1)}$ is non-zero, there exist $\{k(\mathtt{l}_1, \mathtt{l}_2)\}_{\mathtt{l}_1,\mathtt{l}_2}$ such that

$$\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 k(\mathtt{l}_1, \mathtt{l}_2)\Lambda_{\mathtt{l}_1,\mathtt{l}_2} = e_1 N\Lambda_1'(1, 1)$$

and

$$\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 k(\mathtt{l}_1, \mathtt{l}_2)\Lambda_{\mathtt{l}_1,\mathtt{l}_2} = e_2 N\Lambda_2'(1, 1)$$

which is the same as

$$(e_1, e_2) = \left( \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_1 \Lambda_{\mathtt{l}_1,\mathtt{l}_2}\sigma(\mathtt{l}_1, \mathtt{l}_2)}{\Lambda_1'(1, 1)}, \frac{\sum_{\mathtt{l}_1,\mathtt{l}_2} \mathtt{l}_2 \Lambda_{\mathtt{l}_1,\mathtt{l}_2}\sigma(\mathtt{l}_1, \mathtt{l}_2)}{\Lambda_2'(1, 1)} \right),$$

where $0 \leq \sigma(\mathtt{l}_1, \mathtt{l}_2) = k(\mathtt{l}_1, \mathtt{l}_2)/N \leq 1$. When $N$ grows this is the same as (3.29).

Now we show that $\mathcal{E}$ is the set between the two piecewise linear curves described in the statement of this lemma. We show this by varying the $\sigma(\mathtt{l}_1, \mathtt{l}_2)$ between 0 and 1 while trying to make the ratio $e_1/e_2$ as large as possible. Start by letting $\sigma(\mathtt{l}_1, \mathtt{l}_2) = 0$ if $\mathtt{l}_1/\mathtt{l}_2$ is not maximal, and letting $\sigma(\mathtt{l}_1, \mathtt{l}_2)$ increase to 1 if $\mathtt{l}_1/\mathtt{l}_2$ is maximal. This traces out the line between $(0, 0)$ and $p_1$, and clearly we can not have $(e_1, e_2)$ below this line for $(e_1, e_2) \in \mathcal{E}$. Then increase $\sigma(\mathtt{l}_1, \mathtt{l}_2)$ for $\mathtt{l}_1, \mathtt{l}_2$ such that $\mathtt{l}_1/\mathtt{l}_2$ takes the second largest value. This traces out the line between $p_1$ and $p_1 + p_2$ and again it is clear that we can not have $(e_1, e_2)$ below this line for $(e_1, e_2) \in \mathcal{E}$. We continue like this until we have $\sigma(\mathtt{l}_1, \mathtt{l}_2) = 1$ for all $\mathtt{l}_1, \mathtt{l}_2$, which corresponds to the point $(1, 1)$. The upper curve is obtained by reversing the order and starting with the line between $(0, 0)$ and $p_D$. □

# Appendix 3.B Proof of Lemma 3.3.11

Take the derivative of $\theta(e)$ with respect to $e$ to get

$$\begin{aligned}\frac{d\theta}{de} =& (1 - \mathtt{l}_1 - \mathtt{l}_2) \log\left(\frac{1-e}{e}\right) - \mathtt{l}_1 \log v_1 - \mathtt{l}_2 \log v_2 \\ =& \log\left(\frac{1-e}{e}\right) - \mathtt{l}_1 \log\left(\frac{(1-e)v_1}{e}\right) \\ & - \mathtt{l}_2 \log\left(\frac{(1-e)v_2}{e}\right).\end{aligned}$$

We can now write

$$\begin{aligned}\frac{1-e}{e} &= \frac{1 - \frac{v_1}{\Gamma^{(1)'}(1)} \sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} \frac{(1+v_1)^{\mathtt{r}_1-1} - (1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1} + (1-v_1)^{\mathtt{r}_1}}}{\frac{v_1}{\Gamma^{(1)'}(1)} \sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} \frac{(1+v_1)^{\mathtt{r}_1-1} - (1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1} + (1-v_1)^{\mathtt{r}_1}}} \\ &= \frac{\sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} \left(1 - v_1 \frac{(1+v_1)^{\mathtt{r}_1-1} - (1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1} + (1-v_1)^{\mathtt{r}_1}}\right)}{\sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} v_1 \frac{(1+v_1)^{\mathtt{r}_1-1} - (1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1} + (1-v_1)^{\mathtt{r}_1}}} \\ &= \frac{\sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} \frac{(1+v_1)^{\mathtt{r}_1-1} + (1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1} + (1-v_1)^{\mathtt{r}_1}}}{\sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} v_1 \frac{(1+v_1)^{\mathtt{r}_1-1} - (1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1} + (1-v_1)^{\mathtt{r}_1}}}\end{aligned}$$

or

$$\frac{(1-e)v_1}{e} = \frac{\sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} \frac{(1+v_1)^{\mathtt{r}_1-1}+(1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1}+(1-v_1)^{\mathtt{r}_1}}}{\sum_{\mathtt{r}_1} \mathtt{r}_1 \Gamma^{(1)}_{\mathtt{r}_1} \frac{(1+v_1)^{\mathtt{r}_1-1}-(1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1}+(1-v_1)^{\mathtt{r}_1}}}. \tag{3.61}$$

We obtain a similar expression for $(1-e)v_2/e$. Note that $v_j(e)$ are increasing functions of $e$ and $v_j(1/2) = 1$. Thus for $e > 1/2$, $v_j > 1$ which together with (3.61) implies $\frac{(1-e)v_j}{e} > 1$ when all $\mathtt{r}$ are odd. This in turn implies that $\frac{d\theta}{de} < 0$ for $e > 1/2$.

□

## Appendix 3.C Proof of Lemma 3.3.12

First we show that $v(1-e) = 1/v(e)$ if there are only even check degrees. Let $v_j(e) = v$ and $1/v = \tilde{v}$. Then

$$\begin{aligned} e &= \frac{1/\tilde{v}}{\Gamma^{(j)'}(1)} \sum_{\mathtt{r}} \mathtt{r}\Gamma^{(j)}_{\mathtt{r}} \frac{(1+1/\tilde{v})^{\mathtt{r}-1} - (1-1/\tilde{v})^{\mathtt{r}-1}}{(1+1/\tilde{v})^{\mathtt{r}} + (1-1/\tilde{v})^{\mathtt{r}}} \\ &= \frac{1}{\Gamma^{(j)'}(1)} \sum_{\mathtt{r}} \mathtt{r}\Gamma^{(j)}_{\mathtt{r}} \frac{(1+\tilde{v})^{\mathtt{r}-1} + (1-\tilde{v})^{\mathtt{r}-1}}{(1+\tilde{v})^{\mathtt{r}} + (1-\tilde{v})^{\mathtt{r}}} \end{aligned}$$

and

$$\begin{aligned} 1-e &= 1 - \frac{v}{\Gamma^{(j)'}(1)} \sum_{\mathtt{r}} \mathtt{r}\Gamma^{(j)}_{\mathtt{r}} \frac{(1+v))^{\mathtt{r}-1} - (1-v)^{\mathtt{r}-1}}{(1+v)^{\mathtt{r}} + (1-v)^{\mathtt{r}}} \\ &= \frac{1}{\Gamma^{(j)'}(1)} \sum_{\mathtt{r}} \Gamma^{(j)}_{\mathtt{r}} \left(1 - v\frac{(1+v)^{\mathtt{r}-1} - (1-v)^{\mathtt{r}-1}}{(1+v)^{\mathtt{r}} + (1-v)^{\mathtt{r}}}\right) \\ &= \frac{1}{\Gamma^{(j)'}(1)} \sum_{\mathtt{r}} \mathtt{r}\Gamma^{(j)}_{\mathtt{r}} \frac{(1+v)^{\mathtt{r}-1} + (1-v)^{\mathtt{r}-1}}{(1+v)^{\mathtt{r}} + (1-v)^{\mathtt{r}}} \end{aligned}$$

These two equations imply that $v(1-e) = 1/v(e)$. Now note that

$$q_{\mathtt{r}}(1/v) = \frac{q_{\mathtt{r}}(v)}{v^{\mathtt{r}}}$$

for $\mathtt{r}$ even, so

$$\begin{aligned}\theta(1-e) &= (1-\mathtt{l}_1-\mathtt{l}_2)h(1-e) + \frac{\mathtt{l}_1}{\Gamma^{(1)\prime}(1)}\sum_{\mathtt{r}}\Gamma^{(1)}_{\mathtt{r}}\log q_{\mathtt{r}}(v_1)\\ &\quad - \mathtt{l}_1\log v_1 + \frac{\mathtt{l}_2}{\Gamma^{(2)\prime}(1)}\sum_{\mathtt{r}}\Gamma^{(2)}_{\mathtt{r}}\log q_{\mathtt{r}}(v_2) - \mathtt{l}_2\log v_2\\ &\quad - (1-e)\mathtt{l}_1\log(1/v_1) - (1-e)\mathtt{l}_2\log(1/v_2) - R_{\mathrm{des}}\\ &= \theta(e).\end{aligned}$$

□

# Chapter 4

# Polar Codes

In this chapter we discuss the application of polar codes to the wiretap channel. Based on a construction of nested polar codes by Korada [Kor09] we construct polar codes that achieve the whole capacity-equivocation region for binary input symmetric wiretap channels.

## 4.1 Nested Polar Codes

For polar codes we will define the nested structure in terms of the frozen set instead of as the solution to a certain parity check equation as we did for LDPC codes. These definitions are equivalent, but the characterization based on the frozen sets makes it particularly easy to prove the results we want.

We will consider binary polar codes of block length $N = 2^n$. Let $\mathcal{A}$ and $\mathcal{B}$ be two index sets such that

$$\mathcal{B} \subset \mathcal{A} \subset \{1, \dots, N\}. \tag{4.1}$$

As for nested parity check codes the the nested structure of polar codes comes from the cosets of a smaller subcode. Consider the polar codes $\mathcal{P}(N, \mathcal{A}, u_{\mathcal{A}^c})$ and $\mathcal{P}(N, \mathcal{B}, [0, u_{\mathcal{A}^c}])$. Here $[0, u_{\mathcal{A}^c}]$ is a binary vector whose elements are zero for the indices $i$ in $\mathcal{A} \backslash \mathcal{B}$, and otherwise they equal the corresponding elements in $u_{\mathcal{A}^c}$. $\mathcal{A}^{\mathcal{C}}$ is a frozen set for both codes, but $\mathcal{B}^{\mathcal{C}}$ is frozen only for $\mathcal{P}(N, \mathcal{B}, [0, u_{\mathcal{A}^c}])$. Similarly to Definition 2.4.1 we now define the nested polar code as follows:

**Definition 4.1.1** (The nested polar code $\mathcal{P}(N, \mathcal{A}, \mathcal{B}, u_{\mathcal{A}^c})$)**.** *Let $G$ be the matrix $G_N$ as defined in (2.14) and let $G_{\mathcal{I}}$ be the submatrix composed of the columns of $G$ whose indices belong to an index set $\mathcal{I}$. The nested polar code $\mathcal{P}(N, \mathcal{A}, \mathcal{B}, u_{\mathcal{A}^c})$ is the set of codewords $x^N$ of the form*

$$x^N = u_{\mathcal{B}} G_{\mathcal{B}} \oplus u_{\mathcal{A}\setminus\mathcal{B}} G_{\mathcal{A}\setminus\mathcal{B}} \oplus u_{\mathcal{A}^c} G_{\mathcal{A}^c}. \tag{4.2}$$

*The vector $u_{\mathcal{A}\setminus\mathcal{B}}$ determines which coset of $\mathcal{P}(N, \mathcal{B}, [0, u_{\mathcal{A}^c}])$ the codeword belongs to.*

The rates of the subcodes $\mathcal{P}(N, \mathcal{B}, [u_{\mathcal{A}\setminus\mathcal{B}}, u_{\mathcal{A}^c}])$ all equal $|\mathcal{B}|/N$, and the rate of the overall code equals $|\mathcal{A}|/N$.

See Figure 4.1 to see a pictorial representation of the frozen sets.

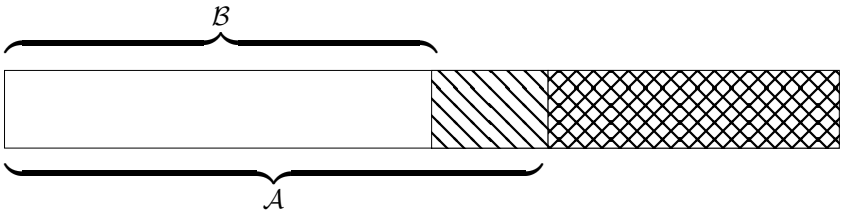


Figure 4.1: A nested polar code. The rectangle corresponds to the whole index set $\{1, \ldots, N\}$. The two frozen sets are $\mathcal{A}^{\mathcal{C}}$ and $\mathcal{B}^{\mathcal{C}}$, and $\mathcal{A}^{\mathcal{C}} \subset \mathcal{B}^{\mathcal{C}}$.

Let $W$ and $\tilde{W}$ be two symmetric binary input memoryless channels and let $\tilde{W}$ be stochastically degraded with respect to $W$. Denote the polarized channels as defined in (2.15) by $W_N^{(i)}$ (resp. $\tilde{W}_N^{(i)}$), and their Bhattacharyya parameters by $Z_N^{(i)}$ (resp. $\tilde{Z}_N^{(i)}$). We will use the following Lemma which is Lemma 4.7 from [Kor09]:

**Lemma 4.1.2** (Lemma 4.7 from [Kor09])**.** *If $\tilde{W}$ is degraded with respect to $W$, then $\tilde{W}_N^{(i)}$ is degraded with respect to $W_N^{(i)}$, and $\tilde{Z}_N^{(i)} \geq Z_N^{(i)}$.*

In the following section we use Lemma 4.1.2 to show that nested polar codes achieve the whole capacity-equivocation region for the degraded wiretap channel.

## 4.2 Nested Polar Wiretap Codes

We consider a wiretap channel where the input alphabet $\mathcal{X}$ is binary, and Bob's and Eve's output alphabets $\mathcal{Y}$ and $\mathcal{Z}$ are discrete. We assume

that the main channel (denoted by $W(y|x)$) and the wiretapper's channel (denoted by $\tilde{W}(z|x)$) are symmetric. We also assume that $\tilde{W}$ is stochastically degraded with respect to $W$, that is, there exists a probability distribution $W'(z|y)$ such that $\tilde{W}(z|x) = \sum_{y\in\mathcal{Y}} W'(z|y)W(y|x)$ for every $z$. Since $W$ and $\tilde{W}$ are symmetric, $C_M = I(W)$ and $C_W = I(\tilde{W})$. For this setup the capacity-equivocation region is given by

$$R_e \le R \le C_M, \quad 0 \le R_e \le C_M - C_W. \tag{4.3}$$

In Theorem 4.2.1 we give a nested polar coding scheme for the wiretap channel that achieves the whole capacity-equivocation region.

**Theorem 4.2.1.** *Let $(R, R_e)$ satisfy (4.3). For every $\epsilon > 0$ and every $0 < \beta < 1/2$ there exists a wiretap polar code of length $N = 2^n$ and rate $R_N$ that satisfies*

$$R_N > R - \epsilon, \tag{4.4}$$

$$P_e^N < 2^{-N^\beta}, \tag{4.5}$$

$$R_e^N > R_e - \epsilon, \tag{4.6}$$

*provided $n$ is large enough.*

*Proof.* Fix $\beta < \beta' < 1/2$. Let

$$\mathcal{A}_N = \{i : Z_N^{(i)} < 2^{-N^{\beta'}}\}$$

and choose the subset $\mathcal{B}_N$ as follows. Order the indices in $\mathcal{A}_N$ by increasing $\tilde{Z}_N^{(i)}$ and choose the $N(C_M - R)$ smallest ones. Since $\lim_{n\to\infty} |\mathcal{A}_N|/N = C_M \ge C_M - R$ a subset of this size exists provided that $n$ is large enough.

Now consider the nested polar code $\mathcal{P}(N, \mathcal{A}_N, \mathcal{B}_N, u_{\mathcal{A}^c})$. Since $W$ and $\tilde{W}$ are symmetric channels the performance of the successive cancellation decoder does not depend on the choice of the frozen bits $u_{\mathcal{A}^c}$. We will therefore set $u_{\mathcal{A}^c} = 0$.

As for the wiretap codes based on LDPC codes we let each coset correspond to a different message. To send the message $s_N$, Alice generates the codeword

$$X^N = T_N G_{\mathcal{B}_N} \oplus s_N G_{\mathcal{A}_N \setminus \mathcal{B}_N}, \tag{4.7}$$

where $T_N$ is a binary vector of length $|\mathcal{B}_N|$ chosen uniformly at random. There are $2^{|\mathcal{A}_N \setminus \mathcal{B}_N|}$ different cosets, so the rate of the coding scheme is

$$R_N = \frac{|\mathcal{A}_N| - |\mathcal{B}_N|}{N} = \frac{|\mathcal{A}_N|}{N} - C_M + R$$

Due to Theorem 2.4.2 we have $\lim_{n\to\infty} |\mathcal{A}_N|/N = C_M$ which implies

$$\lim_{n\to\infty} R_N = R.$$

This proves (4.4).

Since the codewords of the nested code are the same as the ones for the polar code $\mathcal{P}(N, \mathcal{A}_N, 0)$ we can bound $P_e^N$ from above by the corresponding error probability for $\mathcal{P}(N, \mathcal{A}_N, 0)$. Since this error probability is smaller than $2^{-N^\beta}$ provided that $n$ is large enough we get (4.5).

To show (4.6) we look at the equivocation for Eve. We first look at the case where $R \geq C_M - C_W$. We expand $I(X^N, S_N; Z^N)$ in two different ways and obtain

$$\begin{aligned} I(X^N, S_N; Z^N) &= I(X^N; Z^N) + I(S_N; Z^N | X^N) \\ &= I(S_N; Z^N) + I(X^N; Z^N | S_N). \end{aligned} \tag{4.8}$$

Note that $I(S_N; Z^N | X^N) = 0$ as $S_N \to X^N \to Z^N$ is a Markov chain. By (4.8) and noting that $I(S_N; Z^N) = H(S_N) - H(S_N|Z^N)$, we write the equivocation rate $H(S_N|Z^N)/N$ as

$$\begin{aligned} H(S_N|Z^N)/N =& \frac{H(S_N) + I(X^N; Z^N|S_N) - I(X^N; Z^N)}{N} \\ =& \frac{H(S_N)}{N} + \frac{H(X^N|S_N)}{N} \\ & - \frac{H(X^N|Z^N, S_N)}{N} - \frac{I(X^N; Z^N)}{N} \\ \geq& \frac{|\mathcal{A}_N|}{N} - C_W - \frac{H(X^N|Z^N, S_N)}{N}, \end{aligned}$$

where we have used that $H(S_N) + H(X^N|S_N) = H(X^N, S_N) = H(X^N) = |\mathcal{A}_N|$ and that $I(X^N; Z^N)/N \leq C_W$.

We now look at $H(X^N|Z^N, S_N)$. For a fixed $S_N = s_N$ we see that $X^N \in \mathcal{P}(N, \mathcal{B}, [s_N, 0])$. Let $P_e^{N,s_N}$ be the error probability of decoding this code using an SC decoder. By Lemma 4.1.2, the set $\tilde{\mathcal{A}}_N = \{i : \tilde{Z}_N^{(i)} < 2^{-N^{\beta'}}\}$ is a subset of $\mathcal{A}_N$. Also, $\lim_{n\to\infty} \frac{1}{N}|\tilde{\mathcal{A}}_N| = C_W$, so if

$|\mathcal{B}_N| \leq NC_W$ we have $\mathcal{B}_N \subset \tilde{\mathcal{A}}_N$ for large $n$, by the definition of $\mathcal{B}_N$. Since $|\mathcal{B}_N| = N(C_M - R) \leq NC_W$, we have $\tilde{Z}_N^{(i)} < 2^{-N^{\beta'}}$ $\forall i \in \mathcal{B}_N$ for large enough $n$. This implies that

$$P_e^{N,s_N} \leq \sum_{i \in \mathcal{B}_N} \tilde{Z}_N^{(i)} \leq 2^{-N^{\beta}},$$

provided $n$ is large enough. We use Fano's inequality to show that $H(X^N|Z^N, S_N) \to 0$ as $n \to \infty$. We get

$$\lim_{n\to\infty} H(X^N|Z^N, S_N) \leq \lim_{n\to\infty} \left[h(P_e^{N,s_N}) + P_e^{N,s_N}|\mathcal{B}_N|\right] = 0,$$

since $P_e^{N,s_N}|\mathcal{B}_N| = N2^{-N^{\beta}}|\mathcal{B}_N|/N \to 0$ as $n \to \infty$. Thus we have shown that

$$\frac{H(S_N|Z^N)}{N} \geq C_M - C_W - \epsilon \geq R_e - \epsilon$$

for $n$ large enough.

We now consider the case when $R < C_M - C_W$. The only difference from the analysis above is the term $H(X^N|Z^N, S_N)$. Since $|\mathcal{B}_N| = N(C_M - R) > NC_W$, the code defined by (4.7) is not decodable. Instead, let $\mathcal{B}_{1N} = \{i : \tilde{Z}_N^{(i)} < 2^{-N^{\beta'}}\}$, $\mathcal{B}_{2N} = \mathcal{B}_N \setminus \mathcal{B}_{1N}$, and rewrite (4.7) as

$$X^N = T_{1N}G_{\mathcal{B}_{1N}} \oplus T_{2N}G_{\mathcal{B}_{2N}} \oplus S_N G_{\mathcal{A}_N \setminus \mathcal{B}_N}.$$

Note that, since $\lim_{n\to\infty} |\mathcal{B}_{1N}|/N = C_W$, this code is decodable using a successive cancellation decoder given $T_{2N}$. If $T_{2N}$ is unknown we can try all possible combinations and come up with $2^{|\mathcal{B}_{2N}|}$ equally likely solutions (all solutions are equally likely since $T_N$ is chosen uniformly at random). Thus $H(X^N|Z^N, S_N)$ should tend to $H(T_{2N})$. We make this argument precise by bounding $H(X^N|Z^N, S_N)$ as follows:

$$\begin{aligned} H(X^N|Z^N, S_N) &= H(X^N, T_{2N}|Z^N, S_N) \\ &= H(T_{2N}|Z^N, S_N) + H(X^N|Z^N, S_N, T_{2N}) \\ &\leq H(T_{2N}) + H(X^N|Z^N, S_N, T_{2N}) \end{aligned}$$

where in the last step we have used the fact that conditioning reduces entropy. We can show that the second term goes to zero using Fano's inequality as above. Since $\lim_{n\to\infty} \frac{H(T_{2N})}{N} = \lim_{n\to\infty} \frac{|\mathcal{B}_{2N}|}{N} = C_M - R - C_W$, we get $H(S_N|Z^N)/N \geq R - \epsilon$ for $n$ large enough. □

## 4.3 Simulation Results

We show simulation results comparing Eve's equivocation for nested polar wiretap codes and two edge type LDPC codes over a wiretap channel where both the main channel and the wiretapper's channel are binary erasure channels with erasure probabilities $e_m$ and $e_w$ respectively. The LDPC codes are optimized using the methods in Section 3.2 and for the LDPC codes the curve shows the ensemble average. The equivocation of Eve is calculated using an extension of a result in [OW84][1]:

**Lemma 4.3.1.** *Let $H_1$ be a parity check matrix for the overall code ($\mathcal{P}(N, \mathcal{A}_N)$ in the polar case) and let $H$ be a parity check matrix for the subcode ($\mathcal{P}(N, \mathcal{B}_N)$) in a nested coding scheme for the binary erasure channel. Then the equivocation at Eve is rank$(H_\mathcal{E})$ − rank$(H_{1,\mathcal{E}})$, where $H_\mathcal{E}$ is the matrix formed from the columns of $H$ corresponding to erased codeword positions.*

*Proof.* The equivocation at Eve can be written as

$$H(S_N|Z^N) = H(X^N|Z^N) - H(X^N|S_N, Z^N).$$

For a specific received $z$ we have $H_{1,\mathcal{E}} x_\mathcal{E}^T + H_{1,\mathcal{E}^c} x_{\mathcal{E}^c}^T = 0$, where $x_\mathcal{E}^T$ is unknown. The above equation has $2^{N-\text{rank}(H_{1,\mathcal{E}})}$ solutions, all of which are equally likely since the original codewords $X^N$ are equally likely. In the same way $H(X^N|S_N, Z^N) = N - \text{rank}(H_\mathcal{E})$. This implies $H(S_N|Z^N) = \text{rank}(H_\mathcal{E}) - \text{rank}(H_{1,\mathcal{E}})$. □

Figure 4.2 shows the equivocation rate at Eve and also the upper bound for $R_e$ as a function of $e_w$ for fixed $R = 0.25$ and $e_m = 0.25$. It is interesting to note that even with a block length of only 1024 bits the curves are close to the upper bound.

[1] Note that the polar codes $\mathcal{P}(N, \mathcal{A}_N)$ and $\mathcal{P}(N, \mathcal{B}_N)$ are linear codes and we therefore can calculate the corresponding parity check matrices.

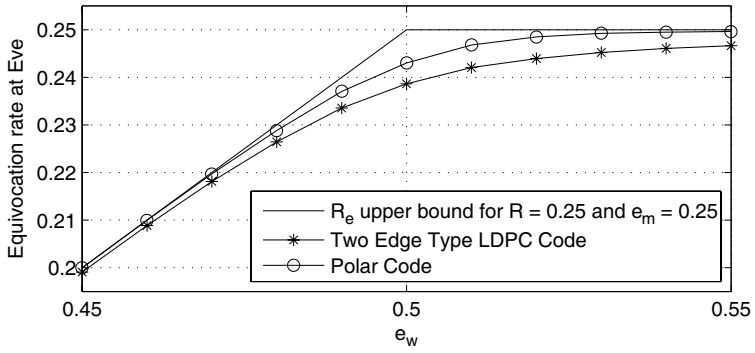


Figure 4.2: Equivocation rate versus $e_w$. Codes designed for $R = 0.25$, $e_m = 0.25$, $e_w = 0.5$, and block length $N = 1024$.

# Chapter 5

# Conclusions

In this thesis we have investigated code design for the wiretap channel.

We have introduced two edge type LDPC ensembles for the wiretap channel. For the scenario when the main channel is error free and the wiretapper's channel is a binary erasure channel (BEC) we find secrecy capacity achieving code sequences based on standard LDPC code sequences for the BEC. Our construction does not work when there are also erasures on the main channel. For this case we have developed a method based on linear programming to optimize two edge type degree distributions. Using this method we have found code ensembles that perform close to the secrecy capacity of the binary erasure wiretap channel (BEC-WT). We have generalized a method of Méasson, Montanari, and Urbanke [MMU08] in order to compute the conditional entropy $\lim_{N\to\infty} H(S|Z^N)/N$. We have applied this method to relatively simple code degree distributions and have found that these degree distributions, which are simpler than the ones found using our numerical method, show very good secrecy performance.

Based on the work of Kudekar, Richardson, and Urbanke [KRU10], which showed that regular spatially coupled codes are capacity achieving for the BEC, we have constructed a regular two edge type spatially coupled ensemble. We have shown that this ensemble achieves the whole capacity-equivocation region for the BEC-WT. Based on the empirical evidence that regular spatially coupled codes perform close to the capacity over a wide range of channels we conjecture that our construction works well over the corresponding wiretap channels, provided that the wiretapper's channel is degraded with respect to the main channel.

Based on Arıkan's [Arı09] polar codes and a lemma by Korada [Kor09] we have constructed nested polar codes that achieve the whole capacity equivocation region for any symmetric binary input wiretap channel where the wiretapper's channel is degraded with respect to the main channel.

In the next section we give some directions for possible future work.

## 5.1 Future Work

Based on the results and methods in the thesis we present some ideas that might be worthy of further study.

### LDPC Codes

From Section 3.4 we see that ensembles where the degree distribution of type two edges is regular show surprisingly good secrecy performance. Therefore it would be interesting to consider such two edge ensembles where the degree distribution for type one edges comes from a capacity achieving sequence for the BEC. Such an ensemble would achieve capacity on the main channel and it might be possible to show, using weight distribution arguments, that it is also optimal from the secrecy perspective.

Another approach could be to generalize our results to other channels than the BEC. Our numerical optimization methods from Section 3.2 should readily generalize to general binary memoryless channels such as binary symmetric channels or binary input additive white Gaussian noise channels.

### Polar Codes

The same nested structure used for wiretap codes is also optimal for the Decode-and-Forward strategy for the relay channel [ART$^+$10b]. It would be interesting to generalize this result to more general relay networks, such as the ones considered in [SK10], with or without security constraints.

# Bibliography


[Arı09] E. Arıkan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Trans. on Inf. Theory*, 55(7):3051 –3073, Jul. 2009.

[ART+10a] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund. Equivocation of Eve using two edge type LDPC codes for the erasure wiretap channel. In *Proceedings of Asilomar Conference on Signals, Systems and Computers (to appear)*, Nov. 2010.

[ART+10b] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund. Nested polar codes for wiretap and relay channels. *IEEE Communications Letters*, 14(8):752 –754, Aug. 2010.

[AT09] E. Arıkan and E. Telatar. On the rate of channel polarization. In *Proc. IEEE Int. Sympos. Information Theory (ISIT)*, pages 1493 –1495, Jul. 2009.

[BM04] D. Burshtein and G. Miller. Asymptotic enumeration methods for analyzing LDPC codes. *IEEE Trans. on Inf. Theory*, 50(6):1115 – 1131, Jun. 2004.

[CK78] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. on Inf. Theory*, 24(3):339 – 348, May 1978.

[CV10] Y. Chen and A. J. H. Vinck. On the binary symmetric wiretap channel. In *Proceedings of International Zurich Seminar on Communications*, pages 17–20, Mar. 2010.

[Eli55] P. Elias. Coding for Two Noisy Channels. In *Information Theory, The 3rd London Symposium*, pages 61–76. Butterworth's Scientific Publications, Sep. 1955.

[EZ99] K. Engdahl and K. S. Zigangirov. On the theory of low-density convolutional codes I. *Problemy Peredachi Informatsii*, 35:12–27, 1999.

[FMS01] S. R. Fluhrer, I. Mantin, and A. Shamir. Weaknesses in the key scheduling algorithm of RC4. In *Selected Areas in Cryptography*, pages 1–24, 2001.

[FZ99] A. J. Felstrom and K. S. Zigangirov. Time-varying periodic convolutional codes with low-density parity-check matrix. *IEEE Trans. on Inf. Theory*, 45(6):2181 –2191, Sep. 1999.

[Gal63] R. G. Gallager. *Low-Density Parity-Check Codes*. PhD thesis, MIT, 1963.

[Gal68] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.

[HS10] E. Hof and S. Shamai. Secrecy-Achieving Polar-Coding for Binary-Input Memoryless Symmetric Wire-Tap Channels. *ArXiv e-prints*, May 2010.

[HW10] R. Hinton and S. Wilson. Analysis of peeling decoder for MET ensembles. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 1 –5, Jan. 2010.

[IKS+05] R. Ikegaya, K. Kasai, Y. Shimoyama, T. Shibuya, and K. Sakaniwa. Weight and stopping set distributions of two-edge type LDPC code ensembles. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, E88-A(10):2745–2761, 2005.

[KAD+09] K. Kasai, T. Awano, D. Declercq, C. Poulliat, and K. Sakaniwa. Weight distributions of multi-edge type LDPC codes. In *Proc. IEEE Int. Sympos. Information Theory (ISIT)*, pages 60–64, 2009.

[KMRU10] S. Kudekar, C. Measson, T. Richardson, and R. Urbanke. Threshold Saturation on BMS Channels via Spatial Coupling. *ArXiv e-prints*, April 2010.

[Kor09] S. B. Korada. *Polar codes for channel and source coding*. PhD thesis, EPFL, 2009.

[KRU10] S. Kudekar, T. Richardson, and R. Urbanke. Threshold Saturation via Spatial Coupling: Why Convolutional LDPC Ensembles Perform so well over the BEC. *ArXiv e-prints*, Jan. 2010.

[LBYP07] C.-Y. Lu, D. E. Browne, T. Yang, and J.-W. Pan. Demonstration of a compiled version of Shor's quantum factoring algorithm using photonic qubits. *Phys. Rev. Lett.*, 99(25):250504, Dec. 2007.

[LFZC09] M. Lentmaier, G. P. Fettweis, K. S. Zigangirov, and D. J. Costello, Jr. Approaching capacity with asymptotically regular LDPC codes. In *Information Theory and Applications Workshop, 2009*, pages 173 –177, 8-13 2009.

[LLPS07] R. Liu, Y. Liang, H. Poor, and P. Spasojević. Secure nested codes for type II wiretap channels. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 337–342, Sep. 2007.

[LMFC10] M. Lentmaier, D. G. M. Mitchel, G. Fettweis, and D. J. Costello, Jr. Asymptotically good LDPC convolutional codes with AWGN channel thresholds close to the Shannon limit. In *6th Intern. Symp. on Turbo Codes and iterative inform. Processing*, 2010.

[LMS+97] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. Spielman, and V. Stemann. Practical loss-resilient codes. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, STOC '97, pages 150–159. ACM, 1997.

[LMSS98] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman. Analysis of low density codes and improved designs using irregular graphs. In *STOC '98: Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 249–258. ACM, 1998.

[LMSS01a] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman. Efficient erasure correcting codes. *IEEE Trans. on Inf. Theory*, 47(2):569–584, Feb. 2001.

[LMSS01b] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman. Improved low-density parity-check codes using irregular graphs. *IEEE Trans. on Inf. Theory*, 47(2):585 –598, Feb. 2001.

[LPSL08] R. Liu, H. V. Poor, P. Spasojevic, and Y. Liang. Nested codes for secure transmission. In *Proc. Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pages 1–5, Sep. 2008.

[LPSS09] Y. Liang, H. V. Poor, and S. Shamai (Shitz). Information theoretic security. *Foundations and Trends in Communications and Information Theory*, 5(4-5):355–580, 2009.

[LSCZ10] M. Lentmaier, A. Sridharan, D. Costello, and K. Zigangirov. Iterative decoding threshold analysis for LDPC convolutional codes. *IEEE Trans. on Inf. Theory*, 56(10):5274 –5289, Oct. 2010.

[LSZC05] M. Lentmaier, A. Sridharan, K. Zigangirov, and D. Costello. Terminated LDPC convolutional codes with thresholds close to capacity. In *Proc. IEEE Int. Sympos. Information Theory (ISIT)*, pages 1372 –1376, Sep. 2005.

[LTZ99] M. Lentmaier, D. V. Truhachev, and K. S. Zigangirov. On the theory of low-density convolutional codes II. *Problemy Peredachi Informatsii*, 35:12–27, 1999.

[LWL+07] B. P. Lanyon, T. J. Weinhold, N. K. Langford, M. Barbieri, D. F. V. James, A. Gilchrist, and A. G. White. Experimental demonstration of a compiled version of Shor's algorithm with quantum entanglement. *Phys. Rev. Lett.*, 99(25):250505, Dec. 2007.

[MMU08] C. Méasson, A. Montanari, and R. Urbanke. Maxwell Construction: The Hidden Bridge Between Iterativew and Maximum a Posteriori Decoding. *IEEE Trans. on Inf. Theory*, 54(12):5277–5307, 2008.

[MN95] D. J. MacKay and R. M. Neal. Good codes based on very sparse matrices. In *Cryptography and Coding. 5th IMA Conference, number 1025 in Lecture Notes in Computer Science*, pages 100–111. Springer, 1995.

[MV10] H. Mahdavifar and A. Vardy. Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes. In *Proc. IEEE Int. Sympos. Information Theory (ISIT)*, Jun. 2010.

[MW00] U. Maurer and S. Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. *Lecture Notes in Computer Science*, 1807:351+, 2000.

[OE10] O. O. Koyluoglu and H. El Gamal. Polar Coding for Secure Transmission and Key Agreement. *ArXiv e-prints*, Mar. 2010.

[OVZ05] A. Orlitsky, K. Viswanathan, and J. Zhang. Stopping set distribution of LDPC code ensembles. *IEEE Trans. on Inf. Theory*, 51:929–953, 2005.

[OW84] L. H. Ozarow and A. D. Wyner. Wire-tap channel II. *AT&T Bell Laboratories Technical Journal*, 63(10):2135–2157, 1984.

[RA10] V. Rathi and I. Andriyanova. Some Results on MAP Decoding of Non-Binary LDPC Codes over the BEC. *Accepted to IEEE Trans. on Inf. Theory*, 2010.

[Rat08] V. Rathi. *Non-binary LDPC codes and EXIT like functions.* PhD thesis, Swiss Federal Institute of Technology (EPFL), Lausanne, 2008.

[RAT+09] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund. Two edge type LDPC codes for the wiretap channel. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pages 834 –838, 2009.

[RAT+10] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund. Performance Analysis and Design of Two Edge Type LDPC Codes for the BEC Wiretap Channel. *Submitted to IEEE Trans. on Inf. Theory*, Sep. 2010.

[RSU01] T. Richardson, A. Shokrollahi, and R. Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Trans. on Inf. Theory*, 47(2):619–637, Feb. 2001.

[RU08] T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

[RUAS11] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund. Rate-Equivocation Optimal Spatially Coupled LDPC Codes for the BEC Wiretap Channel. *Submitted to Proc. IEEE Int. Sympos. Information Theory (ISIT)*, Jul. 2011.

[Sha48] C. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:623–656, 1948.

[Sho99] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2):303–332, 1999.

[SK10] A. Salimi and J. Kliewer. On secure communication over a class of degraded relay networks. In *Proceedings of Asilomar Conference on Signals, Systems and Computers (to appear)*, Nov. 2010.

[SLCZ04] A. Sridharan, M. Lentmaier, D. J. Costello, Jr., and K. S. Zigangirov. Convergence analysis of a class of LDPC convolutional codes for the erasure channel. In *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Oct. 2004.

[SST+10] A. T. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S. W. McLaughlin. Strong secrecy for erasure wiretap channels. In *Proceedings of Information Theory Workshop Dublin*, Aug. 2010.

[TAD04] J. Thorpe, K. Andrews, and S. Dolinar. Methodologies for designing LDPC codes using protographs and circulants. In *Proc. IEEE Int. Sympos. Information Theory (ISIT)*, page 238, Jun.,Jul. 2004.

[Tan81] R. Tanner. A recursive approach to low complexity codes. *IEEE Trans. on Inf. Theory*, 27(5):533 – 547, Sep. 1981.

[TDC+07] A. Thangaraj, S. Dihidar, A. Calderbank, S. McLaughlin, and J.-M. Merolla. Applications of LDPC codes to the wire-tap channel. *IEEE Trans. on Inf. Theory*, 53(8):2933–2945, Aug. 2007.

[TSS+04] R. M. Tanner, D. Sridhara, A. Sridharan, T. E. Fuja, and D. J. Costello, Jr. LDPC block and convolutional codes based on circulant matrices. *IEEE Trans. on Inf. Theory*, 50:2966–2984, 2004.

[Wyn75] A. D. Wyner. The wire-tap channel. *Bell Systems Technical Journal*, 54(8):1355–1387, Oct. 1975.

[ZSE02] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. on Inf. Theory*, 48(6):1250–1276, Jun. 2002.