KTH Electrical Engineering

Coding and Transmission Strategies for Secrecy

MATTIAS ANDERSSON

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2014

TRITA-EE 2014:011
ISSN 1653-5146
ISBN 978-91-7595-051-8

KTH School of Electrical Engineering
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i telekommunikation fredagen den 4 april 2014 klockan 14.15 i hörsal F3, Kungliga Tekniska högskolan, Lindstedtsvägen 26, Stockholm.

Tryck: Universitetsservice US AB

Sammanfattning

I den här avhandlingen behandlar vi flera problem relaterade till informationsteoretisk säkerhet. Wiretap-kanalen är den enklaste informationsteoretiska modellen som behandlar säkerhet och i de första kapitlen av avhandlingen designar vi praktiska koder för wiretap-kanalen.

Först designar vi glesa paritetskontrollkoder (LDPC) med två kanttyper för den binära erasure-wiretap-kanalen (BEC-WT). För scenariot där huvudkanalen är felfri och avlyssnarens kanal är en binär erasure-kanal (BEC) konstruerar vi en följd av koder som uppnår säkerhetkapaciteten. Dessa koder är baserade på vanliga LDPC-koder för BEC. Vår konstruktion fungerar dock inte när huvudkanalen inte är felfri. Om så inte är fallet använder vi en metod baserad på linjär programmering för att optimera gradfördelningen hos våra koder, vilket låter oss designa kodensembler som har prestanda nära säkerhetskapaciteten hos BEC-WT. Vi generaliserar sedan en av Méassons, Montanaris och Urbankes metoder för att räkna ut den betingade entropin av meddelandet hos avlyssnaren.

Vi visar sedan att Arıkans polära koder kan användas för att uppnå hela kapacitets-ekvivokationsregionen för en degraderad symmetrisk wiretap-kanal med binärt inalfabet. Vi designar också polära koder för decode-and-forward-protokollet för den fysiskt degraderade reläkanalen och för den bidirektionella broadcastkanalen med gemensamma och konfidentiella meddelanden. Vi visar att koderna uppnår kapaciteten och kapacitets-ekvivokationsregionen för dessa kanalmodeller.

I nästföljande kapitel behandlar vi en gaussisk kanalmodell. Vi visar att Josephs och Barrons glesa regressionskoder (SPARCs) kan användas för att uppnå säkerhetskapaciteten för wiretapkanaler med gaussiskt brus och för decode-and-forward-protokollet för reläkanalen. Vi behandlar också generering av hemliga nycklar från korrelerade gaussiska källor med hjälp av en publik kanal av begränsad kapacitet. Vi visar att SPARC-koder uppnår kapacitetsregionen för detta problem.

I det sista kapitlet behandlar vi generering av hemliga nycklar över fädande kanaler. Vi behandlar först ett scenario med flera antenner och högt signal-till-brusförhållande (SNR) och föreslår ett protokoll baserat på träning och slumpdelning. Vi behandlar sedan ett scenario med en antenn hos varje terminal och lågt SNR, där vi begränsar den ena terminalen till att endast sända pilotsignaler. Vi föreslår ett protokoll baserat på sporadisk träning och opportunistisk sändning med en wiretap-kod och visar att det är optimalt.

Abstract

In this thesis we consider several problems relating to information theoretic security. The wiretap channel is the simplest information theoretic setting which takes security into account, and in the first chapters of the thesis we design some practical coding schemes for this channel model.

First we consider the design of two edge type low density parity check (LDPC) codes for the binary erasure wiretap channel (BEC-WT). 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 BEC-WT. We generalize a method of Méasson, Montanari, and Urbanke in order to compute the conditional entropy of the message at the wiretapper. We apply this method to relatively simple ensembles and find very good secrecy performance.

We then show that Arıkan's polar codes can be used to achieve the whole capacity-equivocation region of for any degraded symmetric binary input wiretap channel. We also design capacity achieving polar codes for the decode-and-forward scheme for the physically degraded relay channel, and for the bidirectional broadcast channel with common and confidential messages.

In the subsequent chapter we consider a Gaussian system model. We show that sparse regression codes (SPARCS) as introduced by Joseph and Barron achieve the secrecy capacity of the additive white Gaussian noise (AWGN) wiretap channel, and can be used to implement the decode-and-forward scheme for the Gaussian relay channel. We also consider secret key agreement using correlated Gaussian random variables and a rate-limited public channel. We show that SPARCs attain the capacity region also for this problem.

Finally we consider secret key agreement over reciprocal fading channels. We first consider a multiple-antenna setup in the high signal-to-noise-ratio (SNR) regime and propose a scheme based on training and randomness sharing. We then consider a single antenna setup in the low SNR regime, where one of the terminals is only allowed to transmit pilot signals. We propose a bursty transmission scheme based on training and opportunistic transmission using a wiretap channel code, and show that this scheme is optimal.

Till Mamma och Pappa.

Acknowledgments

I want to express my deepest gratitude to my supervisors Prof. Mikael Skoglund and Assoc. Prof. Ragnar Thobaben. I am grateful to Mikael for welcoming me to his research group and for introducing me to, and teaching me, information theory. Mikael is the kindest advisor imaginable, and has always let me pursue my own research interests. 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 dearly for their great patience.

I wish to extend my gratitude to Asst. Prof. Ashish Khisti at the University of Toronto for generously allowing me to visit his research group and our many discussions on secret key agreement long after my visit officially ended. I also wish to thank the Ericsson Research Foundation for partially funding my stay in Toronto.

The first part of this thesis could not have been written without the help of Dr. Vishwambhar Rathi. He has shared not only parts of his great knowledge about channel coding, but also many laughs with me, and I am happy to call him my friend.

My discussions and my collaboration with Assoc. Prof. Tobias J. Oechtering have always been very enjoyable, and, perhaps unfortunately for him, due to the location of his office next to mine he has certainly taught me a lot.

I have shared an office with Dr. Zhongwei Si for most of my time here, and my discussions with her always brightened my day. I also especially want to thank Dr. Ricardo Blasco Serrano. We have gotten lost in countless information-theoretic and probabilistic labyrinths together, but hopefully we managed to find our way out in the end. The same also holds for Frédéric Gabry and our game theoretic escapades. Dr. Nicolas Schrammar is probably the one who has most patiently listened to my random ramblings, and for this I am very grateful. I also want to thank Dr. Emil Björnsson, Leefke Grosjean, Dr. Johannes Kron, Dennis Sundman, Dr. Dave Zachariah, and all my other friends and colleagues on the fourth floor for interesting discussions on life and research.

I am indebted to Ricardo, Frédéric, Mikael, Dennis, Carla Agnesi, and especially

Ragnar for their diligent proofreading of my thesis.

I want to thank Annika Augustsson, Irène Kindblom, and Raine Tiivel for handling all administrative matters with ease.

I would like to thank Asst. Prof. Matthieu Bloch from Georgia Institute of Technology for acting as an opponent for this thesis. Thanks are also due to Assoc. Prof. Alexandre Graell i Amat from Chalmers University of Technology, Assoc. Prof. Joakim Jaldén from KTH, and Prof. Thomas Johansson from Lund University for acting on the grading committee.

Outside of the academic world I would like to thank Mattias Blennow, Merle Breyer, James Drake, Kristin Fahlberg, Christina Enblom Falk, Andreas Eriksson, Julien Grosjean, Kerstin Holmström, Klas Ingesson, Magnus Linderoth, Rikard Olofsson, Odd Runevall, Katarina Olsson, Aline Schrammar, Sebastian Sahl, Martin Singh-Blom, Amrita Singh-Blom, Alan Sola, Per Sundelin, Ana Rodriguez, and Andreas Ziethén for the distractions, the food and all the fun.

Words can not express my gratitude to my family. I want to thank my sisters Emma and Johanna and my brother Frans for their endless love and support. I dedicate this thesis to my parents Jan and Agneta. I also want to thank Vincent and Lorna for giving me a home away from home.

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

Mattias Andersson
Stockholm, March 2014

Contents

Chapter 1

Introduction

Secure communication is essential when considering not only communication between people or between a person and an electronic device, but also machine-to-machine communication. The recent large rise in the number of devices communicating wirelessly is not expected to slow at any time in the foreseeable future, and therefore the analysis of cyber physical systems, in which different systems communicate, form networks, and interact with the physical world, is needed in order to make tomorrow's power grids, transportation systems, and manufacturing plants more efficient, safer, and sustainable.

One example of such a system could be a manufacturing plant with many sensors and actuators distributed over a large area, in which wireless communication protocols allow for cheap deployment and easy reconfiguration. On the other hand, wireless communication opens up the possibility for industrial espionage or even sabotage.

There are also many environments where wired communication is not feasible at all. One example is health monitoring via sensors embedded in the patient's body, or even the control of implanted medical devices such as pacemakers. Here there are privacy concerns around the leakage of sensitive medical data, and in the case of sabotage, the consequences could be fatal.

Another example where wireless communication is needed is Automated Highway Systems, in which several vehicles form platoons in order to increase fuel efficiency and reduce congestion. If the communication between trucks and cars traveling at 100 mph is compromised, the outcome could once again be severe.

Secure communication is also important in smart grids, where on one end of the spectrum, unsecure communication could result in a blackout of a large area due to sabotage, and on the other end there are privacy concerns in reporting the detailed electricity usage patterns of a single household.

It is clear from these examples that security has a large part to play in future wireless communication systems. Traditionally, security has been implemented

in higher layers using methods based on secret key or public key cryptography [MVO96]. This solution is not ideal for all applications considered above, as pointed out by Liang, Poor, and Shamai [LPS08]. The methods based on physical layer security which we consider in this thesis can often be implemented with less computational overhead than cryptographic solutions. This is essential, for example, when extending the battery life of remotely situated sensors, or medical devices inside the body. Another advantage of these methods is that they can be better suited for networks without infrastructure, or rapidly changing networks, where the distribution of keys needed for cryptography-based methods could be impractical.

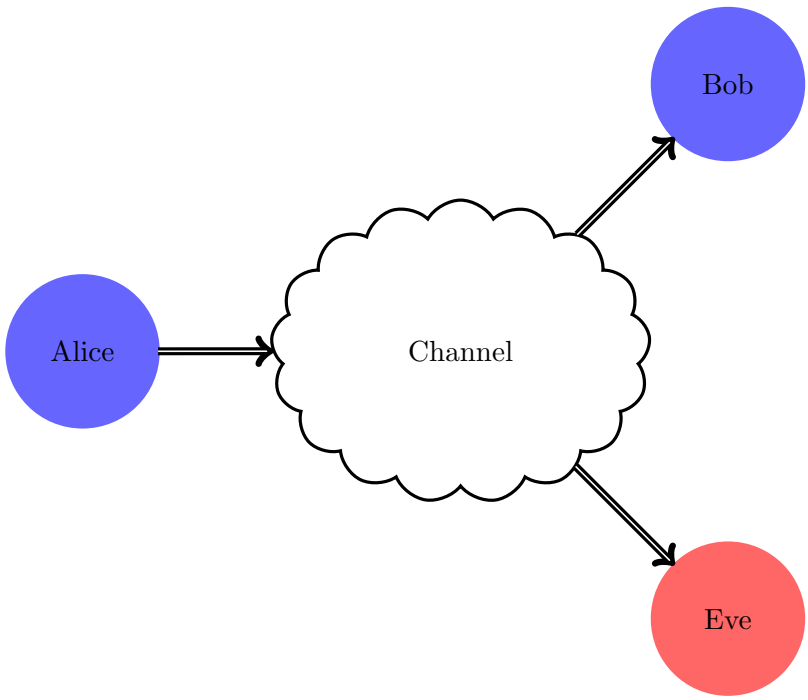


Figure 1.1: A wiretap channel.

We will mostly consider the type of system depicted in Figure 1.1. Here Alice and Bob are two trusted users that want to exchange messages over a network, while keeping their communication secret from an untrusted entity Eve. In public key cryptography, Alice encrypts her message using Bob's public key, which is known to everyone, and transmits the ciphertext over the network. After receiving the ciphertext, Bob then decrypts it using his private key. If Eve somehow gains access to the ciphertext she is unable to decode it since she only has access to Bob's public key. The reason that Eve cannot decode the message without access to the private key is the conjectured difficulty of solving certain computational problems,

and Eve's limited computational powers. The absence of this assumption on Eve's computational ability is one reason that makes physical layer security attractive, in addition to those mentioned above. We instead rely on Eve's physical limitations compared to Bob's. For example, let Alice be a wireless router and Bob a computer situated in the same room, and assume that Eve is located outside the building. In this case the channel from Alice to Eve is noisier than the channel from Alice to Bob, and Wyner showed that this makes it possible to transmit a secret message from Alice to Bob without using any pre-shared keys [Wyn75]. In Chapter 3–5 we design practical coding schemes for similar setups.

Key-based cryptography is still possible without assuming that Eve has bounded computational powers. Shannon studied this problem [Sha49] and found that in order to guarantee secrecy in this case the key needs to act as a One Time Pad. This means that the key needs to be the same size as the message, and key reuse weakens the secrecy considerably. Due to the large size of the key needed this is not easy to realize in practice because of the difficulty in distributing large keys, especially in the type of rapidly changing ad-hoc networks we envision.

A related problem we consider is one in which the wireless channel connecting Alice, Bob, and Eve changes in a random manner. In this case we can use the random state of the channel itself to generate a secret key K at both Alice and Bob, without needing to agree on it beforehand. This key can then be used as a One Time Pad to communicate secretly in the manner mentioned above. This is a problem which has been studied extensively, but, surprisingly, relatively little is known about the fundamental limits on which key sizes can be achieved, and which schemes are optimal. In Chapter 6 we study this problem.

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, and the secret-key agreement problem. We also briefly introduce the relay channel and the bidirectional broadcast channel. The second part is an overview of LDPC codes, polar codes, and sparse regression codes, which are practical coding schemes that we will use to construct optimal coding schemes for these problems. Finally we give an overview of previous work on practical coding schemes for secrecy. Parts of this chapter also appeared in the author's licentiate thesis [And11].

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. This chapter also appeared in the author's licentiate thesis [And11] and is based on the following published papers:

[RAT+09] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund. Two edge type LDPC codes for the wiretap channel. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, pages 834 –838, 2009, © 2009 IEEE.

[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 *Proc. Asilomar Conf. Signals, Systems, and Computers*, November 2010, © 2010 IEEE.

[RAT+13] 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. *IEEE Transactions on Information Theory*, 59(2):1048–1064, February 2013, © 2013 IEEE.

Here [RAT+13] is an extended journal version of [RAT+09] and [ART+10a].

Chapter 4

In this chapter we construct polar codes for degraded wiretap channels, the physically degraded relay channel, and the bidirectional broadcast channel with common and confidential messages. We show that these constructions achieve the fundamental limits of these channel models. This chapter is based on the following published papers:

[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, August 2010, © 2010 IEEE.

[AWOS12] M. Andersson, R. Wyrembelski, T. J. Oechtering, and M. Skoglund. Polar codes for bidirectional broadcast channels with common and confidential messages. In *Proc. Int. Symp. on Wireless Communication Systems (ISWCS)*, pages 1014 –1018, August 2012, © 2012 IEEE.

[ASOS13] M. Andersson, R. Schaefer, T. J. Oechtering, and M. Skoglund. Polar coding for bidirectional broadcast channels with common and confidential messages. *IEEE Journal on Selected Areas in Communications*, 31(9):1901–1908, September 2013, © 2013 IEEE.

Here [ASOS13] is an extended journal version of [AWOS12]. Parts of this chapter also appeared in the author's licentiate thesis [And11], and some results from [ART+10b] were also included in [BSTA+12].

Chapter 5

In this chapter we construct sparse regression codes for the secret key agreement problem with degraded correlated Gaussian sources, the Gaussian wiretap channel, and the physically degraded Gaussian relay channel with orthogonal receivers. We show that these codes achieve the whole capacity region of the studied problems. The material in this chapter has not yet been submitted for publication.

Chapter 6

In this chapter we consider secure key agreement over a reciprocal non-coherent fading channel. First we consider a scenario where the terminals have multiple antennas. We propose a scheme based on training and randomness sharing, and characterize its achievable secure degrees of freedom in the high SNR regime. In the second part we consider a single antenna scenario in the low SNR regime. We constrain one of the terminals to only transmit pilot symbols, and find the secret key capacity and the secrecy capacity. In particular, we show that both the secret key capacity and the secrecy capacity scales as the channel capacity without an eavesdropper. We also note that in both the high SNR and the low SNR schemes studied no knowledge about Eve's channel is needed. This chapter is based on the following published papers:

[AKS12] M. Andersson, A. Khisti, and M. Skoglund. Secret-key agreement over a non-coherent block-fading MIMO wiretap channel. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 153 –157, September 2012, © 2012 IEEE.

[AKS13] M. Andersson, A. Khisti, and M. Skoglund. Secure key agreement over reciprocal fading channels in the low SNR regime. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 674–678, June 2013, © 2013 IEEE.

Chapter 7

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

1.2 Contributions outside the Thesis

In addition to the material covered in this thesis, the author has also contributed to the following works.

[OAS09] T. J. Oechtering, M. Andersson, and M. Skoglund. Arimoto-Blahut algorithm for the bidirectional broadcast channel with side information. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 394–398, October 2009

[SAS11] N. Schrammar, M. Andersson, and M. Skoglund. Approximate capacity of the general Gaussian parallel relay network. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 89–93, July 2011

[RUAS11] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund. Rate-equivocation optimal spatially coupled LDPC codes for the BEC wiretap channel. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 2393–2397, July 2011

[SATS11] Z. Si, M. Andersson, R. Thobaben, and M. Skoglund. Rate-compatible LDPC convolutional codes for capacity-approaching hybrid ARQ. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 513–517, October 2011

[AZWS11] M. Andersson, A. Zaidi, N. Wernersson, and M. Skoglund. Nonlinear distributed sensing for closed-loop control over gaussian channels. In *Communication Technologies Workshop (Swe-CTW), 2011 IEEE Swedish*, pages 19–23, October 2011

[BSTA+12] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund. Polar codes for cooperative relaying. *IEEE Transactions on Communications*, 60(11):3263 –3273, November 2012

1.3 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
$\lvert\mathcal{X}\rvert$	The cardinality of $\mathcal{X}$
$p_X(x)$	The probability mass function of X
$p_{Y\lvert X}(y\lvert x)$	The conditional probability mass function of X

	conditioned on Y
$f_X(x)$	The probability density function of X
$f_{X\|Y}(x\|y)$	The conditional probability density function of X conditioned on Y
$\mathbb{E}[X]$	The expectation of X
$H(X)$	The entropy of X
$H(X\|Y)$	The conditional entropy of X conditioned on Y
$h(X)$	The differential entropy of X
$h(X\|Y)$	The conditional differential 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
$X \to Y \to Z$	(X, Y, Z) form a Markov chain in this order
BEC(ϵ)	The binary erasure channel with erasure probability ϵ
BEC-WT(ϵ_m, ϵ_w)	A wiretap channel where the main channel is a BEC(ϵ_m) and the wiretapper's channel is a BEC(ϵ_w)
$\log(x)$	The logarithm to base 2
$\ln(x)$	The natural logarithm
$h_2(x)$	The binary entropy function to base 2
$\mathbb{1}_{\{S\}}$	The indicator variable which is 1 if S is true and 0 otherwise
$\operatorname{coef}\left\{\sum_i F_i D^i, D^j\right\}$	The coefficient of D^j in $\sum_i F_i D^i$
x^N	A vector with N elements
x_i^j	The vector $[x_i \ x_{i+1} \ \dots \ x_{j-1} \ x_j]$
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
b.p.c.u.	bits per channel use
LDPC code	Low Density Parity Check code
R-S code	Reed-Solomon code
SPARC	Sparse Regression Code
s.d.o.f.	secure degrees of freedom

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], and the related problem of secret key agreement studied by Maurer [Mau93], and by Ahlswede and Csiszár [AC93]. We then briefly discuss the relay channel introduced by Cover and El-Gamal [CG79] and the bidirectional broadcast channel first studied by Larsson, Johansson, and Sunell [LJS05]. We then give an introduction to Gallager's LDPC codes [Gal63], Arıkan's polar codes [Arı09], and sparse regression codes as introduced by Joseph and Barron [JB12], which will be used in later chapters to construct practical codes for the channel models mentioned above.

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 finite 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 $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$.

A $(2^{NR}, N)$ code of rate R for the channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ consists of a message set

$$\mathcal{M} = \{1, \ldots, \lceil 2^{NR} \rceil\}$$

of cardinality $\lceil 2^{NR} \rceil$, an encoder

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

and a decoder

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

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 $(2^{NR_N}, N)$ 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 (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.

$\Diamond$

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.

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 ϵ, or BEC(ϵ), 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:

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

$$P_{Y|X}(?|1) = \epsilon.$$

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 ϵ.

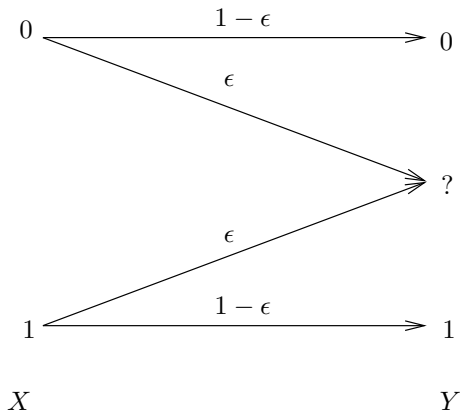


Figure 2.2: Binary erasure channel.

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(ϵ) is a symmetric channel. Thus the maximizing input distribution is the uniform distribution, and the capacity, as well as the symmetric capacity, is found to be $1 - \epsilon$.

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.

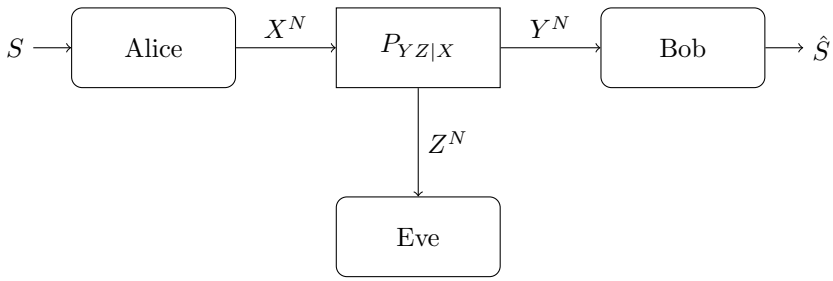


Figure 2.3: Wiretap channel.

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.

We define an $(2^{nR_N}, N)$ code for the wiretap channel by

- a message set $\mathcal{S} = \{1, \ldots, \lceil 2^{nR_N} \rceil\}$,
- a (randomized) encoding function at Alice $f_N : \mathcal{S} \to \mathcal{X}^N$,
- a decoding function at Bob $g_N : \mathcal{Y}^N \to \mathcal{S}$.

The structure of the codebook is as follows. The codebook $\mathcal{C}$ is made up of disjoint subcodes $\mathcal{C}_S$, each labelled by one of the possible messages. To encode the message $S \in \mathcal{S}$, Alice chooses one of the codewords in $\mathcal{C}_S$ uniformly at random and transmits it. We assume that all messages are equally likely. Let P_e^N be the average decoding error probability for Bob

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

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 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), and was found by Csiszár and Körner:

Theorem 2.2 (Corollary 2 from [CK78])**.** *The capacity-equivocation region of the wiretap channel is the set of rate-equivocation pairs $(R, R_e) \in \mathbb{R}_+^2$ that satisfy*

$$R_e \leq R, \tag{2.5}$$

$$R_e \leq I(V;Y|U) - I(V;Z|U), \tag{2.6}$$

$$R \leq I(V;Y), \tag{2.7}$$

for random variables $U \to V \to X \to (Y, Z)$. The cardinalities of the ranges of U and V can be bounded by

$$|\mathcal{U}| \leq |\mathcal{X}| + 3, \qquad |\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3.$$

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 $\limsup_{N\to\infty} I(S;Z^N)/N = 0$, or $\liminf_{N\to\infty} H(S|Z^N)/N = R$, and means that the information leakage to the wiretapper goes to zero rate-wise. From Theorem 2.2 we get

Corollary 2.3. *The secrecy capacity for a general wiretap channel is*

$$C_S = \max_{P_{VX}} \left[I(V;Y) - I(V;Z) \right],$$

where V satisfies the Markov chain $V \to X \to (Y, Z)$. □

Note that the secrecy capacity is always non-negative since we can choose V and X to be independent which will ensure that $I(V;Y) - I(V;Z) = 0$.

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 $P_{YZ|X}$ factorizes as

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

or equivalently the Markov chain $X \to Y \to Z$ holds, we say that the wiretapper's channel is *physically degraded* with respect to the main channel. It is easy to show that the capacity-equivocation region only depends on the marginal probabilities, which means that the capacity-equivocation region for physically and stochastically degraded wiretap channels is the same. We have:

Corollary 2.4 (Theorem 3 from [CK78])**.** *The capacity-equivocation region of the degraded wiretap channel is the set of rate-equivocation pairs* $(R, R_e) \in \mathbb{R}_+^2$ *that satisfy*

$$\begin{aligned} R_e &\leq R, \\ R_e &\leq I(X;Y) - I(X;Z), \\ R &\leq I(X;Y), \end{aligned}$$

for some input probability distribution P_X*. In particular, the secrecy capacity is given by*

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

□

In the degraded 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

$$R_e \leq R \leq C_M, \quad 0 \leq R_e \leq C_M - C_W, \tag{2.8}$$

and the secrecy capacity is

$$C_s = \left[C_M - C_W \right]^+ = \max(0, C_M - C_W),$$

where C_M and C_W are the capacities of the main and the wiretapper's channels respectively. The rate region described by (2.8) 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.

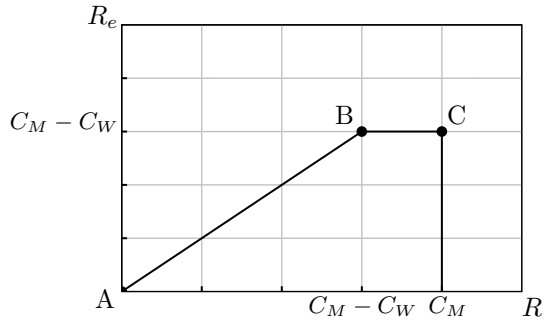


Figure 2.4: Capacity-equivocation region for a degraded symmetric wiretap channel.

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(ϵ_m, ϵ_w). Here ϵ_m and ϵ_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] and [BB11].

Weak versus Strong Secrecy

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.

$$\limsup_{N\to\infty} I(S; Z^N) = 0$$

instead of

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

This constraint is called *strong secrecy*, whereas the constraint given in (2.4) is called *weak secrecy*. Csiszár showed that the secrecy capacity for discrete memoryless channels under the strong and the weak secrecy criterion is the same [Csi96], a result which was recently extended by Bloch and Lanemann to a more general class

of channels [BL13] using the concept of channel resolvability introduced by Han and Verdú in [HV93]. We will mostly consider the case of weak secrecy in the rest of the thesis.

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 channel 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 2.5 (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:

$$H(X^N | Z^N S) \leq h_2(P_e^{N,S}) + P_e^{N,S} N R^{(1,2)},$$

where $P_e^{N,S}$ is the error probability of decoding X^N when knowing Z^N and the coset S, and $h_2(x)$ 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

$$\lim_{N\to\infty} \frac{I(S; Z^N)}{N} \leq \lim_{N\to\infty} \left(C_W - R^{(1,2)} + \frac{h_2(P_e^{N,S})}{N} + P_e^{N,S} R^{(1,2)} \right) = 0.$$

■

[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.

2.3 Secret Key Agreement

Secret key agreement is a related problem to secret message transmission over the wiretap channel. The goal of secret key agreement is for Alice and Bob to agree on a key K, which is to be kept secret from Eve. In Chapter 5 we construct sparse regression codes for secret key agreement, and in Chapter 6 we consider secret key agreement over non-coherent fading channels. We will consider the source model and the channel model for secret key agreement as introduced by Ahlswede and Csiszár [AC93].

2.3.1 Source Model

The setup in Figure 2.5 is the source model for secret key agreement. Alice, Bob and Eve observe $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, and $Z \in \mathcal{Z}$ respectively, where (X, Y, Z) is a discrete memoryless source distributed according to P_{XYZ}. Alice and Bob are allowed to exchange messages over a public channel, the output of which is also observed by Eve. We assume that Alice and Bob will use the public channel for q rounds, and without loss of generality we assume that Alice uses the channel in odd rounds, Bob uses the channel in even rounds, and that q is even. A q-round key agreement scheme of length N is then given by

- a finite message set for the public channel $\mathcal{P}$ and a finite key set $\mathcal{K}$,
- $q/2$ encoding functions at Alice

$$f_i : \mathcal{X}^N \times \mathcal{P}^{(i-1)/2} \to \mathcal{P} \text{ for odd } i,$$

- $q/2$ encoding functions at Bob

$$g_i : \mathcal{Y}^N \times \mathcal{P}^{i/2} \to \mathcal{P} \text{ for even } i,$$

- A key generating function at Alice

$$k_A : \mathcal{X}^N \times \mathcal{P}^{q/2} \to \mathcal{K},$$

- A key generating function at Bob

$$k_B : \mathcal{Y}^N \times \mathcal{P}^{q/2} \to \mathcal{K}.$$

Let P_i denote the message transmitted over the public channel in round i, and let K_A and K_B denote the keys generated at Alice and Bob after q rounds respectively. We say that a key rate R is achievable if $\forall \epsilon > 0$, there exists a sequence of key agreement schemes that satisfies

$$\limsup_{N \to \infty} \Pr(K_A \neq K_B) < \epsilon, \tag{2.9}$$

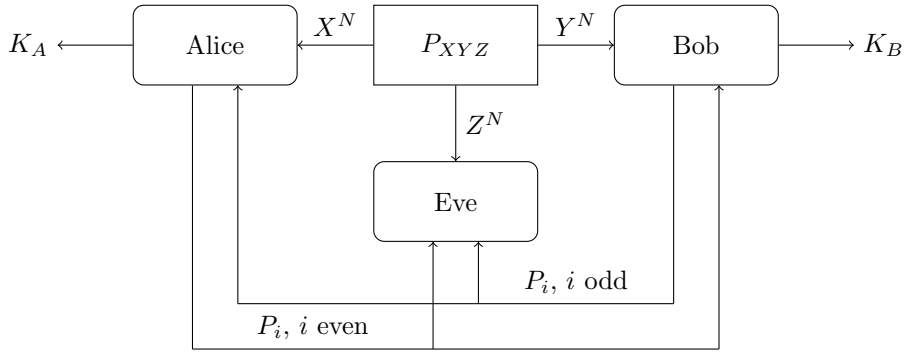


Figure 2.5: Source model for secret key agreement.

$$\liminf_{N\to\infty} \frac{1}{N} H(K_A) > R - \epsilon, \tag{2.10}$$

$$\liminf_{N\to\infty} \max\left(\frac{1}{N} I(K_A; Z^N P^q), \frac{1}{N} I(K_B; Z^N P^q)\right) < \epsilon. \tag{2.11}$$

As before we call the supremum of all achievable secret key rates the secret key capacity C_K, and note that the secret key capacity is not known in general. The following upper bound was found by Maurer [Mau93] and Ahlswede and Csiszár [AC93]:

$$C_K \leq \min\left[I(X;Y), I(X;Y|Z)\right],$$

together with a lower bound

$$C_K \geq \max\left[I(X;Y) - I(X;Z), I(X;Y) - I(Y;Z)\right].$$

These bounds are not tight in general, but they match if (X, Y, Z) form a Markov chain in any order.

2.3.2 Channel Model

The other similar setup we consider is the channel model for secret key agreement, see Figure 2.6. In this setup, instead of a source generating (X, Y, Z) we let Alice, Bob, and Eve be connected by a memoryless broadcast channel $P_{YZ|X}$, and let Alice control the input X to the channel. We also allow Alice and Bob access to two independent sources of randomness M_A and M_B. In this case a q-round secret key agreement scheme of length N consists of

- a finite message set $\mathcal{P}$ and a finite key set $\mathcal{K}$ as before.
- $Nq/2$ encoding functions for the public channel at Alice

$$f_{i,j} : \mathcal{M}_A \times \mathcal{P}^{q(i-1)/2+(j-1)/2} \to \mathcal{P},$$

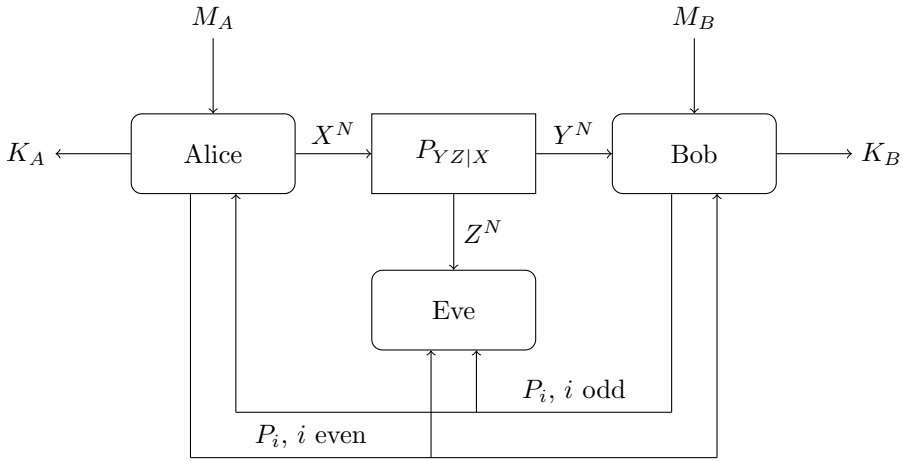


Figure 2.6: Channel model for secret key agreement.

- $Nq/2$ encoding functions for the public channel at Bob

$$g_{i,j} : \mathcal{M}_B \times \mathcal{P}^{q(i-1)/2+j/2} \to \mathcal{P},$$

- N encoding functions for the broadcast channel at Alice

$$h_i : \mathcal{M}_A \times \mathcal{P}^{q(i-1)/2} \to \mathcal{X},$$

- A key generating function at Alice

$$k_A : \mathcal{M}_A \times \mathcal{P}^{qN/2} \to \mathcal{K},$$

- A key generating function at Bob

$$k_B : \mathcal{M}_B \times \mathcal{Y}^N \times \mathcal{P}^{qN/2} \to \mathcal{K}.$$

Alice's input to the broadcast channel at time i is a function of M_A and the communication P^{i-1} over the public channel up to that point. After the ith use of the broadcast channel Alice generates a public message $P_{i,1} = f_{i,1}(M_A, P^{i-1})$, Bob then generates a public message $P_{i,2} = g_{i,2}(M_B, P^{i-1}, P_{i,1})$. This message exchange takes place over q rounds, after which Alice generates a new input $X_{i+1} = h_{i+1}(M_A, P^i)$. After N uses of the public channel and a final exchange of public messages Alice and Bob generate their respective keys K_A and K_B using their key generating functions. As in the source model, we say that a key rate R is achievable if $\forall \epsilon > 0$ there exists a sequence of key agreement schemes that satisfies (2.9), (2.10), but (2.11) is replaced with

$$\liminf_{N\to\infty} \max\left(\frac{1}{N} I(K_A; Z^N P^N), \frac{1}{N} I(K_B; Z^N P^N)\right) > R - \epsilon. \tag{2.12}$$

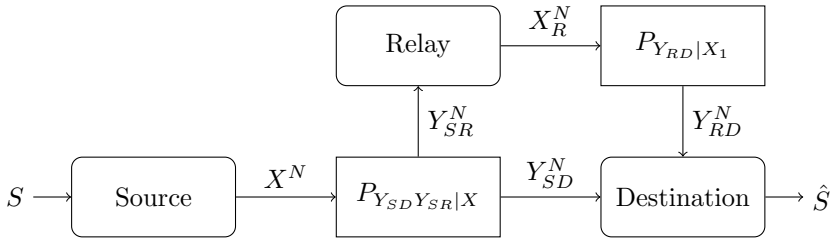


Figure 2.7: Relay channel with orthogonal receivers.

As for the source model, Ahlswede and Csiszár [AC93] found upper and lower bounds on the secret key capacity C_K:

$$C_K \leq \max_{P_X} \min \left[I(X;Y), I(X;Y|Z) \right],$$

and a lower bound was also found

$$C_K \geq \max \left[\max_{P_X} \left(I(X;Y) - I(X;Z) \right), \max_{P_X} \left(I(X;Y) - I(Y;Z) \right) \right].$$

These are not tight in general, but if (X, Y, Z) form a Markov chain in any order they match.

2.4 Multiuser Channels with a Relay

The same nested coding schemes used to achieve secrecy over the wiretap channel can also be used for other multiuser channels. Here we present two such channels that make use of a relay to facilitate communication between two users, the relay channel introduced by Cover and El-Gamal [CG79], and the bidirectional broadcast channel introduced by Larsson, Johansson, and Sunell [LJS05]. We will construct polar codes for these two channels in Chapter 4, and sparse regression codes for the relay channel in Chapter 5.

2.4.1 The Relay Channel

The relay channel consists of three nodes, a sender, a relay, and a destination. The sender wishes to convey a message to the destination with the aid of the relay. We consider the discrete memoryless relay channel with orthogonal receivers, which consists of finite input sets $\mathcal{X}$ and $\mathcal{X}_R$ at the source and the relay respectively, two channel transition probabilities $P_{Y_{SD}Y_{SR}|X}$ and $P_{Y_{RD}|X_R}$, and three finite output sets $\mathcal{Y}_{SR}$, $\mathcal{Y}_{SD}$, and $\mathcal{Y}_{RD}$, corresponding to the received signal at the relay, the received signal at the destination from the source, and the received signal at the destination from the relay respectively.

We define an $(2^{nR}, N)$ code for the relay channel by

- a message set $\mathcal{M} = \{1, \ldots, \lceil 2^{nR} \rceil\}$,
- an encoding function at the source $f : \mathcal{M} \to \mathcal{X}^N$,
- a set of encoding functions at the relay $f_{R,i} : \mathcal{Y}_{SR}^{i-1} \to \mathcal{X}_R$,
- a decoding function at the destination $g : \mathcal{Y}_{SD}^N \times \mathcal{Y}_{RD}^N \to \mathcal{M}$.

Assuming that the message S is transmitted, the inputs to the channels at time i are given by

$$X_i = f(S)_i \tag{2.13}$$

$$X_{R,i} = f_{R,i}(Y_{SR}^{i-1}). \tag{2.14}$$

At time N the destination produces an estimate $\hat{S} = g(Y_{SD}^N, Y_{RD}^N)$, and we denote the error probability by $P_e^N = \Pr(S \neq \hat{S})$, where we assume that S is uniformly distributed. We say that a rate R is achievable if $\forall \epsilon > 0$ there exists a sequence of codes $(2^{NR_N}, N)$ such that

$$\lim_{N\to\infty} P_e^N < \epsilon \tag{2.15}$$

$$\liminf_{N\to\infty} R_N > R - \epsilon. \tag{2.16}$$

The capacity C is the supremum of all achievable rates.

In general the capacity is not known for the relay channel. We will consider the special case of a physically degraded relay channel, where the channel transition probability factors as $P_{Y_{SR}Y_{SD}|X} = P_{Y_{SR}|X} P_{Y_{SD}|Y_{SR}}$. In this case the *Decode-and-Forward* scheme is optimal. In this scheme the relay decodes the message M, and transmits extra information over the relay-to-destination channel which helps the destination decode the message. The capacity is given by

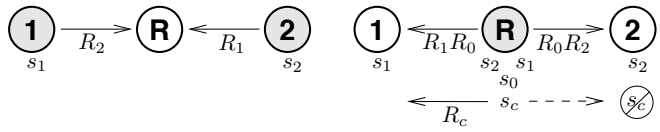
Theorem 2.6 (Theorem 1 from [CG79])**.** *The capacity of the physically degraded relay channel is*

$$C = \max_{P_X P_{X_R}} \min \left\{ I(X; Y_{SD}) + I(X_R; Y_{RD}), I(X; Y_{SD}, Y_{SR}) \right\}. \tag{2.17}$$

If the marginal channels $P_{Y_{SR}|X}$, $P_{Y_{SD}|X}$, and $P_{Y_{RD}|X_R}$ are symmetric, this simplifies to

$$C = \min \left\{ C_{SD} + C_{RD}, C_{SR} \right\}, \tag{2.18}$$

where C_{SD}, C_{SR}, and C_{RD} are the capacities of the source-to-destination, source-to-relay, and relay-to-destination channels respectively.



(a) MAC phase (b) BBC phase

Figure 2.8: Physical layer service integration in bidirectional relay networks. In the initial MAC phase, nodes 1 and 2 transmit their messages m_1 and m_2 with rates R_2 and R_1 to the relay node. Then, in the BBC phase, the relay forwards the messages m_1 and m_2 and adds a common message m_0 with rate R_0 to the communication and further a confidential message m_c for node 1 with rate R_c which should be kept secret from node 2. (© 2013 IEEE. Reused with permission.)

2.4.2 Bidirectional Broadcast Channel

The bidirectional broadcast channel consists of three nodes; two users and a relay. We assume that the two users wish to communicate with one another using the relay, and that there is no direct channel between the two users. The communication takes place over two phases, the multiple access (MAC) phase, and the bidirectional broadcast phase (BBC). In the MAC phase the two users communicate their messages to the relay, and in the BBC phase the relay transmits the two messages to the users simultaneously. This phase is different from the normal broadcast channel since the two users know the messages they transmitted in the first phase. Perhaps surprisingly, this allows the relay to transmit to the two users at the full capacity of their marginal channels [OSBB08, KMT08, KS07].

Here we consider the second phase with two additional messages from the relay, one common message intended for both users, and one confidential message intended for user 1 which should be kept secret from user 2.

The BBC is given by a finite input alphabet $\mathcal{X}$, two finite output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a channel transition probability $P_{Y_1Y_2|X}$.

A $(2^{NR_c}, 2^{NR_0}, 2^{NR_1}, 2^{NR_2}, N)$ code for the BBC with common and confidential messages is given by

- four message sets

$$\begin{aligned}\mathcal{M}_C &= \{1, \ldots, \lceil 2^{NR_c} \rceil\},\\ \mathcal{M}_0 &= \{1, \ldots, \lceil 2^{NR_0} \rceil\},\\ \mathcal{M}_1 &= \{1, \ldots, \lceil 2^{NR_2} \rceil\},\\ \mathcal{M}_2 &= \{1, \ldots, \lceil 2^{NR_1} \rceil\},\end{aligned}$$

  for the confidential, common, and individual messages respectively.

- an encoding function $f : \mathcal{M}_C \times \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^N$,

- two decoding functions

$$g_1 : \mathcal{M}_1 \times \mathcal{Y}_1^N \to \mathcal{M}_c \times \mathcal{M}_0 \times \mathcal{M}_2 \tag{2.19}$$

$$g_2 : \mathcal{M}_2 \times \mathcal{Y}_2^N \to \mathcal{M}_0 \times \mathcal{M}_1. \tag{2.20}$$

We say that a rate-equivocation tuple $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ is achievable if $\forall \epsilon > 0$ there exists a sequence of $(2^{NR_{cN}}, 2^{NR_{0N}}, 2^{NR_{1N}}, 2^{NR_{2N}}, N)$ codes such that the error probability

$$P_e^N = \Pr((g_1(S_1, Y_1^N), g_2(S_2, Y_2^N)) \neq (S_C, S_0, S_2, S_0, S_1)),$$

and the equivocation rate

$$\frac{H(S_c|Y_2^N S_2)}{N}$$

satisfy

$$\limsup_{N\to\infty} P_e^N < \epsilon \tag{2.21}$$

$$\limsup_{N\to\infty} \frac{H(S_c|Y_2^N S_2)}{N} > R_e - \epsilon. \tag{2.22}$$

We call the closure of the set of achievable rate-equivocation tuples the capacity-equivocation region, and it was found by Wyrembelski and Boche in [WB11].

Theorem 2.7 (Theorem 1 from [WB11])**.** *The capacity-equivocation region of the BBC with common and confidential messages is the set of rate-equivocation tuples $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ that satisfy*

$$\begin{aligned} R_e &\leq R_c \\ R_e &\leq I(V;Y_1|U) - I(V;Y_2|U) \\ R_c + R_0 + R_k &\leq I(V;Y_1|U) + I(U;Y_k), \quad k = 1,2 \\ R_0 + R_k &\leq I(U;Y_k), \quad k = 1,2 \end{aligned}$$

for random variables $U \to V \to X \to (Y_1, Y_2)$. The cardinalities of the ranges of U and V can be bounded by

$$|\mathcal{U}| \leq |\mathcal{X}| + 3, \qquad |\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3.$$

2.5 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]. In Chapter 3 we construct codes for the BEC-WT using LDPC codes.

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 2.9 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 2.9: 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 $\mathtt{l}$, 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 2.9 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 2.9 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 2.9 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 2.8 (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. $\Diamond$

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.

2.5.1 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(ϵ) using the following density evolution method in [LMS+97] and [LMSS01a]. Consider transmission over the BEC(ϵ) using a code from the LDPC($\lambda(x), \rho(x)$) ensemble.

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

$$x^{(k)} = \sum_{\mathtt{l}} \lambda_{\mathtt{l}} \epsilon (y^{(k-1)})^{\mathtt{l}-1} = \epsilon \lambda(y^{(k-1)}) \tag{2.23}$$

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

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

$$y^{(k)} = \sum_{\mathtt{r}} \rho_{\mathtt{r}}(1-(1-x^{(k)})^{\mathtt{r}-1}) = 1-\rho(1-x^{(k)}). \tag{2.24}$$

Putting (2.23) and (2.24) 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

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

If $\epsilon < \epsilon^{\mathrm{BP}}$ then the average error probability when communicating over the BEC(ϵ) 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. ϵ^{BP} is called the belief propagation threshold for the degree distribution (λ, ρ).

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].

2.5.2 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 ϵ, 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_{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 2.10 we show the peeling decoder applied to the code defined by the Tanner graph in Figure 2.9. 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 2.10.

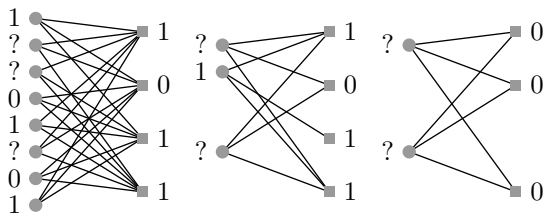


Figure 2.10: Peeling decoder.

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(ϵ), and decode it using the peeling decoder. Call the resulting residual graph G and its degree distribution from the node perspective (Ω, Φ). It was shown in [LMSS01b] that conditioned on the degree distribution (Ω, Φ) all residual graphs G are equally likely. It was shown in [MMU08] that the residual degree distribution (Ω, Φ) 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 (Ω, Φ) 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 (Λ, Γ) that when satisfied guarantees that the rate of a randomly chosen code from the ensemble LDPC(N, Λ, Γ) is close to its design rate:

Lemma 2.9 (Lemma 7 from [MMU08])**.** *Let $\mathcal{C}$ be a code chosen uniformly at random from the ensemble LDPC(N, Λ, Γ) 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 $\Psi_{\Lambda,\Gamma}(u)$*

$$\begin{aligned}\Psi_{\Lambda,\Gamma}(u) = &- \Lambda'(1)\log\left(\frac{1+uv}{1+v}\right) + \sum_l \log\left(\frac{1+u^l}{2}\right) \\ &+ \frac{\Lambda'(1)}{\Gamma'(1)}\sum_r \log\left[1 + \left(\frac{1-v}{1+v}\right)^r\right],\end{aligned} \tag{2.25}$$

where

$$v = \left(\sum_l \frac{\lambda_l}{1+u^l}\right)^{-1}\left(\sum_l \frac{\lambda_l u^{l-1}}{1+u^l}\right). \tag{2.26}$$

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 $N > N_0(\xi, \Lambda, \Gamma)$

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

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

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

□

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

$$\mathbb{E}[N_W(e)] = \frac{\operatorname{coef}\left\{\prod_l (1+u^l)^{N\Lambda_l}\prod_r q_r(v)^{M\Gamma_r}, u^e, v^e\right\}}{\binom{N\Lambda'(1)}{e}}, \tag{2.27}$$

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

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

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

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 (2.27).

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

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

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 ϕ with respect to e and equating it to zero gives

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

Inserting this value for e into ϕ and taking the derivative with respect to u gives the expression (2.26) for v. If we subtract the design rate r_{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 is what is done in [MMU08, Theorem 10]:

Theorem 2.10 (Theorem 10 from [MMU08])**.** *Let $\mathcal{C}$ be a code picked uniformly at random from the ensemble LDPC(N,Λ,Γ) and let $H_{\mathcal{C}}(X|Y)$ be the conditional entropy of the transmitted message when the code is used for communicating over BEC(ϵ). 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 2.9. 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 ϵ 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 2.10 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(ϵ_m, ϵ_w). The MMU method was extended to non-binary LDPC codes for transmission over the BEC in [Rat08, RA11].

2.6 Polar Codes

Polar codes were introduced by Arıkan and were shown to be capacity achieving for a large class of channels [Arı09]. In Chapter 4 we construct polar coding schemes for the wiretap channel, the relay channel and the bidirectional broadcast channel with common and confidential messages. 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

$$W_N(y_1^N|u_1^N) = W_{N/2}(y_1^{N/2}|u_{1,o}^N \oplus u_{1,e}^N) W_{N/2}(y_{N/2+1}^N|u_{1,e}^N), \tag{2.28}$$

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, (2.28) 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 2.11 and 2.12.

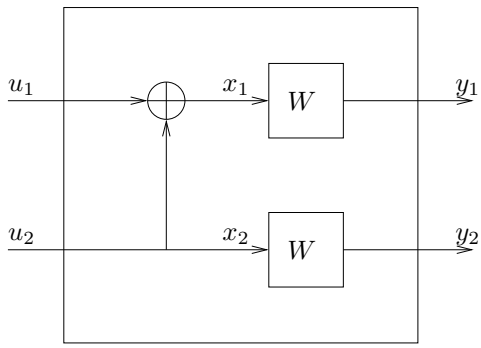


Figure 2.11: The channel W_2 constructed from two copies of W.

Note that the inputs $(x_1, \ldots, x_N)$ to the individual copies of the channel W can be written as $u_1^N G_N$ where

$$G_N = B_N F^{\otimes n}. \tag{2.29}$$

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

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

and

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

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

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

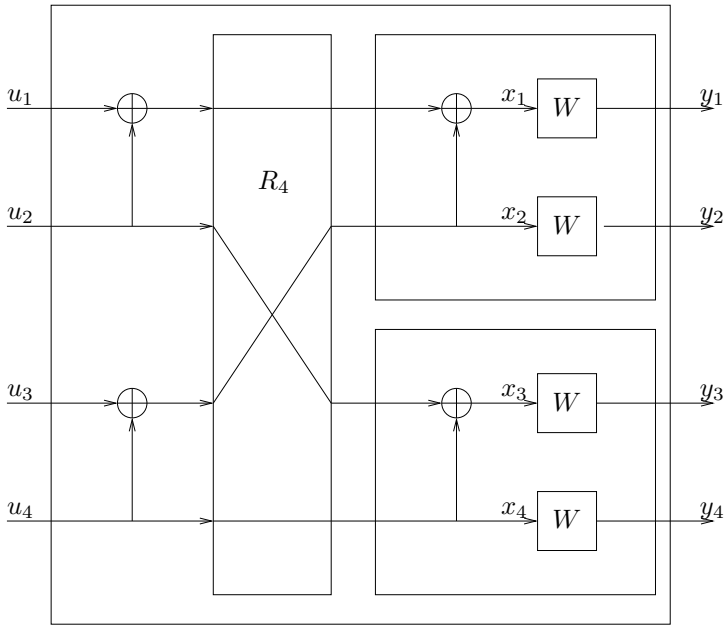


Figure 2.12: The channel W_4 constructed from two copies of W_2.

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 $W_N^{(i)}(y_1^N, u_1^{i-1}|u_i)$.

$$W_N^{(i)}(y_1^N, u_1^{i-1}|u_i) = \sum_{u_{i+1}^N \in \mathcal{X}^{N-i}} \frac{1}{2^{N-i}} W_N(y_1^N|u_1^N). \tag{2.30}$$

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 2.11 (The polar code $\mathcal{P}(N, \mathcal{A}, u_{\mathcal{A}^c})$)**.** Let G be the matrix G_N as defined in (2.29) 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

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

$\lozenge$

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

$$\hat{u}_i = \begin{cases} u_i & i \in \mathcal{A}^{\mathcal{C}} \\ 0 & \frac{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | u_i = 0)}{W_N^{(i)}(y_1^N, \hat{u}_1^{i-1} | u_i = 1)} \geq 1 \text{ and } i \in \mathcal{A} \\ 1 & \text{otherwise} \end{cases} \tag{2.31}$$

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

$$\begin{aligned} P_e^N &\leq \sum_{i \in \mathcal{A}} \Pr(\hat{u}_i \neq u_i) \\ &= \sum_{i \in \mathcal{A}} \sum_{y_1^N, u_1^{i-1}} p_{u_i} W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) \mathbb{1}_{\left\{ \frac{W_N^{(i)}(y_1^N, u_1^{i-1} | u_i \oplus 1)}{W_N^{(i)}(y_1^N, u_1^{i-1} | u_i)} \geq 1 \right\}} \\ &\leq \sum_{i \in \mathcal{A}} \sum_{y_1^N, u_1^{i-1}} p_{u_i} W_N^{(i)}(y_1^N, u_1^{i-1} | u_i) \sqrt{\frac{W_N^{(i)}(y_1^N, u_1^{i-1} | u_i \oplus 1)}{W_N^{(i)}(y_1^N, u_1^{i-1} | u_i)}} \\ &= \sum_{i \in \mathcal{A}} Z_N^{(i)}. \end{aligned} \tag{2.32}$$

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

$$\lim_{n\to\infty} \frac{1}{N} |\{i : Z_N^{(i)} < 2^{-N^\beta}\}| = I(W). \tag{2.33}$$

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

Theorem 2.13 ([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 2.12

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

For large enough N we have

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

which together with (2.32) implies that there exists an n_0 such that

$$P_e^N \leq \sum_{i\in\mathcal{A}_\mathcal{N}} Z_N^{(i)} < N2^{-N^{\beta'}} < 2^{-N^\beta} \tag{2.34}$$

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].

2.7 Sparse Regression Codes

Sparse Regression Codes (SPARCs) are non-linear codes introduced by Joseph and Barron in [JB12]. They were shown to achieve capacity of the AWGN channel when decoded using minimum distance decoding. Varying the code construction slightly, they were also shown to be capacity achieving using a less computationally demanding decoding algorithm in [JB14]. For lossy source coding, Venkataramanan, Joseph, and Tatikonda showed in [VJT12] that SPARCs with optimal encoding attain the rate-distortion function $R_{RD}(D)$ with optimal error exponent for D below a certain threshold. As when used for channel coding, a slight variation of the construction allows for computationally efficient encoding [VST13]. Venkataraman and Tatikonda constructed sparse regression codes for several multi-terminal problems in [VT12]. We will use nested SPARCs in Chapter 5 to construct codes for the Gaussian wiretap channel, the decode-and-forward scheme for Gaussian relay channels, and for secret key agreement from Gaussian sources.

We define an ensemble of sparse regression codes (SPARCs) in the following way. Let A be an $N \times ML$ *design matrix*, and divide A into L sections containing M columns each. We assign a probability distribution to A by generating each entry independently from a $\mathcal{N}(0,1)$ distribution.

Each codeword of the code is given by choosing one column X_l from each section of the design matrix, multiplying them by a fixed weight $c_l > 0$ and adding them together:

$$X = \sum_{l=1}^{L} c_l X_l,$$

or equivalently, $X = A\beta$, where β is a vector of length ML which has exactly one nonzero element in each section of M elements. The nonzero elements of β are given by the weights $\{c_l\}$. We will choose all weights to be equal to P/L, to satisfy the power constraint of the code. Let the set of all such vectors β be denoted by $\mathcal{B}$.

Note that by choosing $L = 1$ this gives the usual ensemble of random codes used in the proof of Shannon's coding theorem for the Gaussian channel. We will instead choose the number of columns M in each section as $M = L^b$ for some $b > 1$, in order to hopefully develop computationally feasible encoders and decoders. Let R denote the rate of the code in nats. The number of codewords in the code is given by M^L, which implies that

$$e^{NR} = M^L, \tag{2.35}$$

or

$$NR = bL \ln L. \tag{2.36}$$

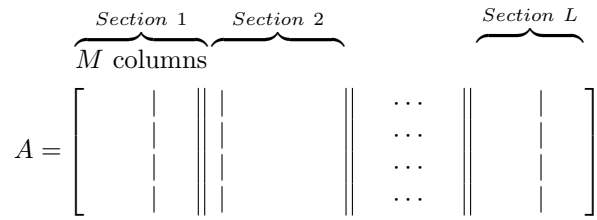


Figure 2.13: The design matrix of a sparse regression code.

Channel Coding using SPARCs

The minimum distance decoder is given by

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} |Y - A\beta|^2. \tag{2.37}$$

Let $v^\star \approx 15.8$ be the solution to the equation $(1+v^\star)\ln(1+v^\star) = 3v^\star$, and let

$$b_0(v) = \begin{cases} \frac{4v(1+v)\ln(1+v)}{((1+v)\ln(1+v)-v)^2} & \text{if } v < v^\star, \\ \frac{(1+v)\ln(1+v)}{(1+v)\ln(1+v)-2v} & \text{if } v \geq v^\star. \end{cases} \tag{2.38}$$

Joseph and Barron showed that if $R < C = \frac{1}{2}\ln(1+v)$, and if $b > b_0(v)$, then the probability of $\hat{\beta}$ differing from β in more than a small fraction α_0 of the sections decays exponentially in the block length N. In order to make the total error probability $P_e^N(\mathcal{C}_N)$ small, they suggested concatenating a SPARC with an outer Reed-Solomon code (R-S code) [RS60] of high rate. In particular they choose a sequence of SPARCs with rates $R = C - \delta_N$, with $\delta_N \leq \frac{1}{\ln N}$, and a sequence of outer codes of rate $1 - 2\delta_N$. We have the following result:

Theorem 2.14 (Proposition 2 from [JB12]). *Let $R = C - \delta_N = \frac{1}{2}\ln(1+v) - \delta_N$, and $b > b_0(v)$. Then there exists a sequence of sparse regression codes $\mathcal{C}_N(R, b)$, and R-S codes of rate $1 - 2\delta_N$, such that the block error probability $P_e^N(\mathcal{C}_N)$ of the concatenated code satisfies*

$$P_e^N(\mathcal{C}_N) \leq e^{-Nc(C-R)^2} \tag{2.39}$$

for some constant $c > 0$.

Lossy Source Coding using SPARCs

Consider the following lossy source coding problem studied by Shannon [Sha48]. The encoder tries to compress a Gaussian source S^N with power σ^2 to a quantized codeword $\hat{S}^N$ coming from a codebook $\mathcal{C}_N$ of rate R. We define an error as the event that the distortion between S^N and $\hat{S}^N$ exceeds the maximum allowable distortion D, where the distortion is given by the normalized distance between S^N and $\hat{S}^N$ squared:

$$d(S^N, \hat{S}^N) = \frac{1}{N}\sum_{i=1}^{N}(S_i - \hat{S}_i)^2. \tag{2.40}$$

Let $P_e^N(\mathcal{C}_N, D) = \Pr\left(d(S^n, \hat{S}^N) > D\right)$ denote the error probability of the code $\mathcal{C}_N$ at distortion-level D. Note that the encoder that maps the sequence S^N to the closest codeword in $\mathcal{C}_N$ minimizes $P_e^N(\mathcal{C}_N, D)$ for a given codebook $\mathcal{C}_N$. The minimum rate R such that $P_e^N(\mathcal{C}_N, D)$ can be made arbitrarily small is given by the rate-distortion function [CT91]

$$R_{RD}(D) = \frac{1}{2}\ln\frac{\sigma^2}{D}. \tag{2.41}$$

Venkataramanan, Joseph, and Tatikonda showed that if the distortion D is small enough, SPARCs can achieve the rate-distortion bound $R_{RD}(D)$ with optimal error exponent $\frac{1}{N}\ln P_e^N(\mathcal{C}_N, D)$. The error exponent is a measure of how fast P_e^N decays with the block length N, and if it is positive P_e^N decays at least exponentially fast in N. They showed the following:

Theorem 2.15 (Theorem 1 from [VJT12])**.** *Fix a target distortion D that satisfies*

$$D < \sigma^2/x^\star,$$

where $x^\star \approx 4.913$ is the solution to $\frac{1}{2}\ln x = 1 - \frac{1}{x}$, and fix a rate $R > R_{RD}(D)$. If $b > \frac{3.5R}{R-(1-D/\rho^2)}$, where ρ^2 satisfies $R = \frac{1}{2}\ln\frac{\rho^2}{D}$, then there exists a sequence of sparse regression codes $\{\mathcal{C}_N(R, b)\}$ which satisfies

$$-\limsup_{N\to\infty}\frac{1}{N}\ln P_e^N(\mathcal{C}_N, D) = \frac{1}{2}\left(\frac{\rho^2}{\sigma^2} - 1 - \ln\frac{\rho^2}{\sigma^2}\right). \tag{2.42}$$

2.8 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 ϵ_w using duals of sparse graph codes. In [STBM11] Subraminian *et al.* constructed large girth LDPC codes for a BEC-WT with a noiseless main channel that achieve strong secrecy, albeit at a lower rate than the secrecy capacity. Rathi *et al.* constructed spatially coupled LDPC codes that achieve weak secrecy for the general BEC-WT in [RUAS11].

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 [KEG12] are closely related to the results we show in Chapter 4. Recently this scheme was extended to provide strong secrecy by Sasoglu and Vardy [SV13]. A secret key agreement scheme providing strong secrecy based on polar codes was suggested by Chou, Bloch, and Abbe in [CBA13].

Lattice codes have been proposed for the Gaussian wiretap channel and for secret key agreement using Gaussian sources, and can generally be used to achieve strong secrecy. See for example the works by Ling *et al.* [LLBS12] and by Ling, Luzzi, and Bloch [LLB13].

Bellare, Tessaro, and Vardy investigated a stronger version of security called *semantic security* in [BTV12], and designed a scheme based on seeded extractors.

Secret key agreement over fading channels has been extensively studied [WTS07, DSC09, CDS10, SP08, YMR+10, WBS09, LLP12, LLD12, PCB13]. In [WBS09] the secret key capacity for the coherent fast fading MIMO wiretap channel was found. The non-coherent fast fading case was studied in [ARKA11]. The related problem of secret message transmission over fading channels was studied in e.g. [GLEG08, LPS08, LYT07]. Secret-key agreement for non-coherent block-fading SISO channels was considered in [LLP12], where a two-phase scheme with training and secret message transmission was proposed. This approach is extended in Chapter 6 to MIMO block fading channels using a two-phase scheme involving channel training and randomness sharing. Furthermore, in [Khi12] it was shown that if imperfect reciprocity is assumed between the forward and reverse channel gains, then the two-phase scheme consisting of channel training and randomness sharing is optimal in the high SNR regime for SISO channels. In Chapter 6 we complement [Khi12] by studying the capacity scaling behaviour in the low SNR regime.

Chapter 3

Two Edge Type LDPC Codes

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.

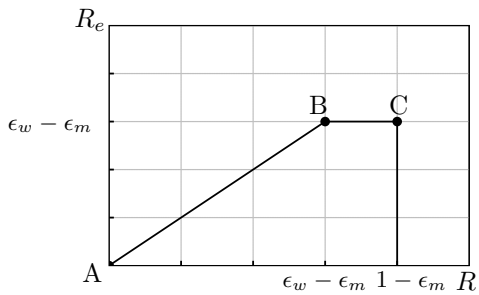


Figure 3.1: Capacity-equivocation region for the BEC-WT(ϵ_m, ϵ_w). (© 2013 IEEE. Reused with permission.)

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 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.

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(ϵ_m, ϵ_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.

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.

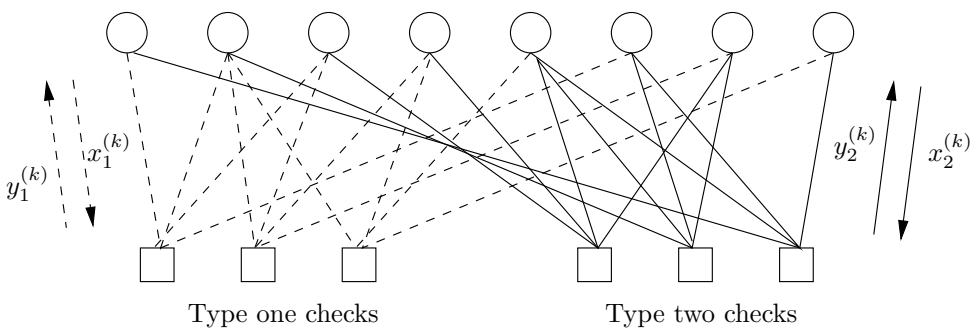


Figure 3.2: Two edge type LDPC code. (© 2013 IEEE. Reused with permission.)

We now define the degree distribution of a two edge type LDPC ensemble. Let $\lambda_{1_1 1_2}^{(j)}$ 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 2.8, the two edge type LDPC ensemble with block length N and degree distribution $\{\lambda^{(1)}, \lambda^{(2)}, \rho^{(1)}, \rho^{(2)}\}$ ($\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ 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(ϵ). 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 3.1. *Let (λ, ρ) be a standard LDPC degree distribution with design rate R and threshold $\epsilon^\star$ over the BEC. Then the following assignment,*

$$\rho^{(1)}(x) = \rho^{(2)}(x) = \rho(x), \tag{3.10}$$

$$\lambda_{l,l}^{(1)} = \lambda_{l,l}^{(2)} = \lambda_{2l}, \tag{3.11}$$

$$\lambda_{l,l+1}^{(1)} = \lambda_{l+1,l}^{(2)} = \frac{l}{2l+1}\lambda_{2l+1}, \tag{3.12}$$

$$\lambda_{l+1,l}^{(1)} = \lambda_{l,l+1}^{(2)} = \frac{l+1}{2l+1}\lambda_{2l+1}, \tag{3.13}$$

$$\lambda_{l_1,l_2}^{(1)} = \lambda_{l_1,l_2}^{(2)} = 0, \quad |l_1 - l_2| > 1, \tag{3.14}$$

ensures that the two edge type LDPC ensemble $\{\lambda^{(1)}, \lambda^{(2)}, \rho^{(1)}, \rho^{(2)}\}$ *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

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

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_{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

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

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

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

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

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 (λ, ρ), 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

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

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

$$x_1 \geq \epsilon \lambda^{(1)}(y_1, y_2)$$

$$= \epsilon \sum_{\mathtt{l}_1,\mathtt{l}_2} \lambda^{(1)}_{\mathtt{l}_1\mathtt{l}_2} y_1^{\mathtt{l}_1-1} y_2^{\mathtt{l}_2}$$
$$= \epsilon \sum_{\mathtt{l}_1,\mathtt{l}_2} \frac{\mathtt{l}_1 \Lambda_{\mathtt{l}_1,\mathtt{l}_2}}{\sum_{\mathtt{k}_1,\mathtt{k}_2} \mathtt{k}_1 \Lambda_{\mathtt{k}_1,\mathtt{k}_2}} y_1^{\mathtt{l}_1-1} y_2^{\mathtt{l}_2},$$

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

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

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

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

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

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 Λ 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.

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

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.

$$\begin{aligned} \Lambda(x,y) = {} & 0.463846x^2 + 0.0814943x^2y + 0.0118691x^2y^2 \\ & + 0.14239x^3 + 0.0201658x^3y + 0.00258812x^3y^2 \\ & + 0.0292241x^4 + 0.0464551x^4y + 0.0564162x^5 \\ & + 0.000718585x^5y + 0.0436039x^7y \\ & + 0.0258926x^8y + 0.000905503x^8y^2 \\ & + 0.00631474x^{13}y^2 + 0.00757076x^{13}y^5 \\ & + 0.011051x^{14}y + 0.0173718x^{14}y^2 \\ & + 0.00100807x^{14}y^5 + 0.00240762x^{31} \end{aligned}$$

$$
\begin{aligned}
&+ 0.0012626x^{31}y^4 + 0.0185828x^{31}y^5 \\
&+ 0.000326117x^{100}y^4 + 0.00383319x^{100}y^{17} \\
&+ 0.00470174x^{100}y^{18}, \\
\Gamma^{(1)}(x) &= 0.25x^9 + 0.75x^{10}, \\
\Gamma^{(2)}(x) &= x^6.
\end{aligned}
$$

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(ϵ) 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(ϵ) 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 $H(X^N S|Z^N)$ in two different ways and obtain

$$H(X^N|Z^N) + H(S|X^N Z^N) = H(S|Z^N) + H(X^N|Z^N S). \tag{3.21}$$

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

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

In the following two subsections we show how the normalized average of $H(X^N|Z^N)$ and $H(X^N|Z^N S)$ 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 3.2. *Consider transmission over the BEC-WT(ϵ_m, ϵ_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(ϵ_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(ϵ_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 3.2, we see that when we consider transmission over the BEC-WT(ϵ_m, ϵ_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(ϵ_w). We formally state this in the following theorem.

Theorem 3.3. *Consider transmission over the BEC-WT(ϵ_m, ϵ_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(ϵ_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_{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_{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 ϵ_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 2.9.

The proof is a straightforward consequence of Lemma 3.2 and Theorem 2.10. 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 $H(X^N|Z^N S)$.

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

Similarly to Lemma 3.2, in the following lemma we show that computing $H(X^N|S,Z^N)$ for the BEC-WT(ϵ_m,ϵ_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(ϵ_w).

Lemma 3.4. *Consider transmission over BEC-WT(ϵ_m,ϵ_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(ϵ_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

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

□

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 3.4 we see that $H(X^N|Z^NS)$ 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 3.5. *Consider transmission over the BEC(ϵ_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 3.6. *Consider transmission over the BEC(ϵ_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

$$d\left(\left(\Omega, \Phi^{(1)}, \Phi^{(2)}\right), \left(\tilde{\Omega}, \tilde{\Phi}^{(1)}, \tilde{\Phi}^{(2)}\right)\right) = \sum_{l_1 l_2} |\Omega_{l_1 l_2} - \tilde{\Omega}_{l_1 l_2}| + \sum_{r_1} |\Phi_{r_1}^{(1)} - \tilde{\Phi}_{r_1}^{(1)}| + \sum_{r_2} |\Phi_{r_2}^{(2)} - \tilde{\Phi}_{r_2}^{(2)}|.$$

□

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 3.7. *Consider transmission over BEC(ϵ_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 ϵ_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*

$$\begin{aligned} \Omega(z_1, z_2) &= \epsilon\Lambda(z_1 y_1, z_2 y_2), \\ \Phi^{(j)}(z) &= \Gamma^{(j)}(1 - x_j + x_j z) - x_j z \Gamma'^{(j)}(1 - x_j) \\ &\quad - \Gamma^{(j)}(1 - x_j), \quad j = 1, 2, \end{aligned}$$

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 3.5, 3.6, and 3.7 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 2.9 we can show that almost every code in the ensemble has its rate equal to the design rate.

Recall that $\text{coef}\left\{\sum_j D_j v^j, v^k\right\}$ is the coefficient of v^k in the polynomial $\sum_j D_j v^j$. In the following lemma we derive the average of the total number of codewords of a two edge type LDPC ensemble.

Lemma 3.8. *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*

$$\mathbb{E}(N_W) = \sum_{E_1=0, E_2=0}^{N\Lambda_1'(1,1), N\Lambda_2'(1,1)} \text{coef}\left\{\prod_{\imath_1,\imath_2}(1+u_1^{\imath_1}u_2^{\imath_2})^{N\Lambda_{\imath_1,\imath_2}}, u_1^{E_1}u_2^{E_2}\right\} \times \frac{\text{coef}\left\{\prod_{r_1,r_2} q_{r_1}(v_1)^{\frac{N\Lambda_1'(1,1)}{\Gamma'^{(1)}(1)}\Gamma_{r_1}^{(1)}} q_{r_2}(v_2)^{\frac{N\Lambda_2'(1,1)}{\Gamma'^{(2)}(1)}\Gamma_{r_2}^{(2)}}, v_1^{E_1}v_2^{E_2}\right\}}{\binom{N\Lambda_1'(1,1)}{E_1}\binom{N\Lambda_2'(1,1)}{E_2}},$$

where $\Lambda_j'(1,1) = \sum_{\imath_1,\imath_2} \imath_j \Lambda_{\imath_1,\imath_2}$, $\Gamma'^{(j)}(1) = \sum_{r_j} r_j \Gamma_{r_j}^{(j)}$, $j \in \{1,2\}$. The polynomial $q_r(v)$ is defined as

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

□

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

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

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

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

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

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

Similar to the arguments for the standard LDPC ensemble in the proof of Lemma 2.9, the total number of graphs for which w is a codeword is given by

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

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 3.9. 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 3.10. *Let $\mathcal{E}(N)$ be the set of (e_1, e_2) such that*

$$\operatorname{coef}\left\{\prod_{l_1,l_2}(1+u_1^{l_1}u_2^{l_2})^{N\Lambda_{l_1,l_2}}, u_1^{e_1 N\Lambda_1'(1,1)}u_2^{e_2 N\Lambda_2'(1,1)}\right\} \neq 0. \tag{3.29}$$

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

$$(e_1,e_2)=\left(\frac{\sum_{l_1,l_2} l_1\Lambda_{l_1,l_2}\sigma(l_1,l_2)}{\Lambda_1'(1,1)},\frac{\sum_{l_1,l_2} l_2\Lambda_{l_1,l_2}\sigma(l_1,l_2)}{\Lambda_2'(1,1)}\right),$$

where $0\le\sigma(l_1,l_2)\le 1$. 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 (l_1, l_2) for which $\Lambda_{l_1,l_2}>0$ in decreasing order of l_1/l_2 and assume that there are D distinct such values. Let

$$\sigma_d(l_1,l_2)=\begin{cases}1 & \text{if } l_1/l_2 \text{ takes the } d\text{th largest possible value,}\\ 0 & \text{otherwise,}\end{cases}$$

and let

$$p_d=\left(\frac{\sum_{l_1,l_2} l_1\Lambda_{l_1,l_2}\sigma_d(l_1,l_2)}{\Lambda_1'(1,1)},\frac{\sum_{l_1,l_2} l_2\Lambda_{l_1,l_2}\sigma_d(l_1,l_2)}{\Lambda_2'(1,1)}\right).$$

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

$$\begin{aligned}\theta(e_1,e_2)=&\sum_{l_1,l_2}\Lambda_{l_1,l_2}\log(1+u_1^{l_2}u_2^{l_2})-\Lambda_1'(1,1)e_1\log u_1\\&-\Lambda_2'(1,1)e_2\log u_2+\frac{\Lambda_1'(1,1)}{\Gamma'^{(1)}(1)}\sum_{r_1}\Gamma_{r_1}^{(1)}\log q_{r_1}(v_1)\\&-\Lambda_1'(1,1)e_1\log v_1+\frac{\Lambda_2'(1,1)}{\Gamma'^{(2)}(1)}\sum_{r_2}\Gamma_{r_2}^{(2)}\log q_{r_2}(v_2)\\&-\Lambda_2'(1,1)e_2\log v_2-\Lambda_1'(1,1)h(e_1)-\Lambda_2'(1,1)h(e_2)\\&-R_{\text{des}},\end{aligned} \tag{3.30}$$

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 3.11. *Consider the two edge type LDPC ensemble* $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ *with design rate* R_{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*

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

where the set $\mathcal{E}$ *is defined in Lemma 3.10 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* $\delta > 0$

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

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 3.8 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

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

Further, the infimum of ψ 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 2.9. ■

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 3.12. *Consider the left regular two edge type LDPC ensemble* $\{\mathtt{l}_1, \mathtt{l}_2, \Gamma^{(1)}, \Gamma^{(2)}\}$ *with design rate* R_{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*

$$\theta(e) = (1 - \mathtt{l}_1 - \mathtt{l}_2) h(e) + \frac{\mathtt{l}_1}{\Gamma^{(1)'}(1)} \sum_r \Gamma_r^{(1)} \log q_r(v_1)$$
$$+ \frac{\mathtt{l}_2}{\Gamma^{(2)'}(1)} \sum_r \Gamma_r^{(2)} \log q_r(v_2) - e\,\mathtt{l}_1 \log v_1 - e\,\mathtt{l}_2 \log v_2 - R_{\text{des}},$$

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 3.11, so we will omit them. First note that the cardinality of the set $\mathcal{W}(E_1, E_2)$, as defined in Lemma 3.8, 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

$$\begin{aligned}\lim_{N\to\infty} \frac{\log(\mathbb{E}[N_W])}{N} &= \lim_{N\to\infty} \sup_{e\in(0,1)} \frac{\log(\mathbb{E}[N(eN\mathtt{l}_1, eN\mathtt{l}_2)])}{N} \\ &= \sup_{e\in(0,1)} \inf_{y_1,y_2>0} \Bigg\{(1-\mathtt{l}_1-\mathtt{l}_2)h(e) \\ &\quad + \frac{\mathtt{l}_1}{\Gamma^{(1)'}(1)} \sum_{\mathtt{r}_1} \Gamma^{(1)}_{\mathtt{r}_1} \log q_{\mathtt{r}_1}(v_1) - e\mathtt{l}_1 \log v_1 \\ &\quad + \frac{\mathtt{l}_1}{\Gamma^{(2)'}(1)} \sum_{\mathtt{r}_2} \Gamma^{(2)}_{\mathtt{r}_2} \log q_{\mathtt{r}_2}(v_2) - e\mathtt{l}_2 \log v_2 \Bigg\} \\ &= \sup_{e\in(0,1)} \inf_{y_1,y_2>0} \psi(e, v_1, v_2)\end{aligned}$$

The saddle point equations are obtained by taking the partial derivatives of ψ 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 3.13. *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 3.12. 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 3.14. *Consider the left regular two edge type LDPC ensemble* $\{l_1, l_2, \Gamma^{(1)}, \Gamma^{(2)}\}$. *Let* $\theta(e)$ *be the function as defined in Lemma 3.12. 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 $H(X^N|Z^N S)$ appearing in (3.22).

Theorem 3.15. *Consider transmission over the BEC-WT*(ϵ_m, ϵ_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(ϵ_w) *using the two edge type LDPC ensemble* $\{\Lambda, \Gamma^{(1)}, \Gamma^{(2)}\}$ *Assume that the erasure probability* ϵ_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_{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 3.11, i.e. if the design rate of the residual ensemble is equal to the rate then*

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

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 3.4, we know that the conditional entropy in the point-to-point set-up is identical to $H(X^N|Z^N S)$. 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(ϵ_m, ϵ_w) based on the coset encoding method.

1. If the threshold of the two edge type LDPC ensemble is lower than ϵ_w, calculate the residual degree distribution for the two edge type LDPC ensemble for transmission over the BEC(ϵ_w). Check that the rate of this residual ensemble is equal to the design rate using Theorem 3.11. Calculate $H(X^N|Z^N S)$ using Theorem 3.15. If the threshold is higher than ϵ_w, $H(X^N|Z^N S)$ is trivially zero.

2. If the threshold of the standard LDPC ensemble induced by type one edges is higher than ϵ_w, calculate the residual degree distribution of this ensemble for transmission over the BEC(ϵ_w). Check that its rate is equal to the design rate using Lemma 2.9. Calculate $H(X^N|Z^N)$ using Theorem 3.3. If the threshold is higher than ϵ_w, $H(X^N|Z^N)$ is trivially zero.

3. Finally calculate $H(S|Z^N)$ using

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

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 3.16. 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 2.9 corresponding to the standard LDPC ensemble $\{\Omega^{(1)}, \Phi^{(1)}\}$, which is the average

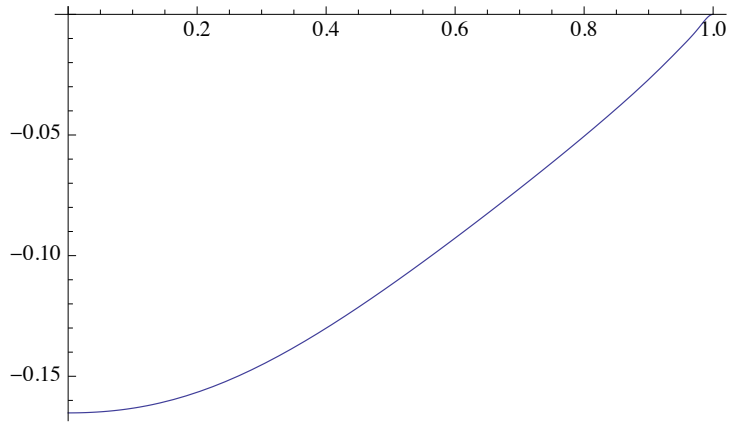


Figure 3.3: $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for Example 3.16 and 3.17. (© 2013 IEEE. Reused with permission.)

residual degree distribution of the ensemble induced by type one edges for transmission over BEC(ϵ_w).

From Lemma 2.9, 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. ◊

Example 3.17. Now consider the two edge type ensemble defined by

Two Edge Type Degree Distribution 1.

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

$$\Gamma^{(1)}(x) = 0.25x^9 + 0.75x^{10},$$
$$\Gamma^{(2)}(x) = x^6,$$

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 $H(X^N|Z^N S) = 0$, 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 3.18. 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 ϵ_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 2.9, 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.

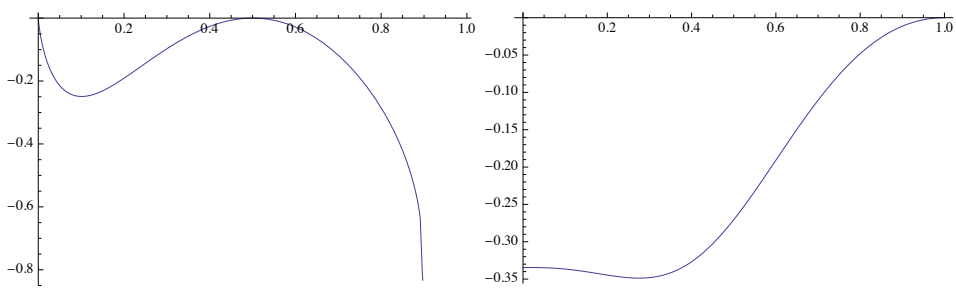


Figure 3.4: $\theta(e)$ and $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for Example 3.18. (© 2013 IEEE. Reused with permission.)

We now calculate the residual degree distribution $(\Omega, \Phi^{(1)}, \Phi^{(2)})$ of the two edge type ensemble corresponding to erasure probability ϵ_w and plot the function $\theta(e)$ defined in Lemma 3.12. 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 3.12. Then we can calculate $H(X^N|Z^N S)$ using Lemma 3.15. In Figure 3.4 we see that $\sup_{e\in[0,1]} \theta(e) = 0$, and we get $\lim_{N\to\infty} \mathbb{E}(H(X^N|Z^N S))/N = 0.000124297$ 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.

$\Diamond$

Example 3.19. 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(ϵ_w). Since the maximum of $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ over the unit interval occurs at $u = 1$, we obtain by Lemma 2.9 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(ϵ_w). Since the maximum of $\theta(e_1, e_2)$ over the set $\mathcal{E}$ is zero, we obtain by Theorem 3.11 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.

$\Diamond$

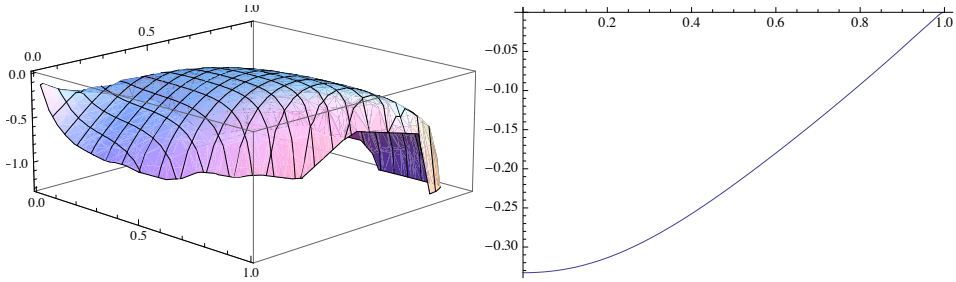


Figure 3.5: $\theta(e_1, e_2)$ and $\Psi_{\{\Omega^{(1)},\Phi^{(1)}\}}(u)$ for Example 4. (© 2013 IEEE. Reused with permission.)

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

3.A Proof of Lemma 3.10

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 $0 \le k(\mathtt{l}_1,\mathtt{l}_2) \le N$. 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 $0 \le \sigma(\mathtt{l}_1,\mathtt{l}_2) = k(\mathtt{l}_1,\mathtt{l}_2)/N \le 1$. 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.

■

3.B Proof of Lemma 3.13

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

$$\frac{(1-e)v_1}{e} = \frac{\sum_{\mathtt{r}_1}\mathtt{r}_1\Gamma^{(1)}_{\mathtt{r}_1}\frac{(1+v_1)^{\mathtt{r}_1-1}+(1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1}+(1-v_1)^{\mathtt{r}_1}}}{\sum_{\mathtt{r}_1}\mathtt{r}_1\Gamma^{(1)}_{\mathtt{r}_1}\frac{(1+v_1)^{\mathtt{r}_1-1}-(1-v_1)^{\mathtt{r}_1-1}}{(1+v_1)^{\mathtt{r}_1}+(1-v_1)^{\mathtt{r}_1}}}. \tag{3.36}$$

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 (3.36) 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 $e > 1/2$.

■

3.C Proof of Lemma 3.14

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

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

■

Chapter 4

Polar Codes

In this chapter we discuss the application of polar codes to the wiretap channel, the decode-and-forward scheme for degraded relay channels and the bidirectional broadcast channel with confidential messages. Based on a construction of nested polar codes by Korada [Kor09] we construct polar codes that achieve the capacity regions for these 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 $\mathcal{P}(N, \mathcal{B}, [0, u_{\mathcal{A}^c}])$. By $[0, u_{\mathcal{A}^c}]$ we mean a binary vector whose elements are zero for the indices i in $\mathcal{A} \setminus \mathcal{B}$, and otherwise they equal the corresponding elements in $u_{\mathcal{A}^c}$. The indices in $\mathcal{A}^{\mathcal{C}}$ are frozen for both codes, but the indices in $\mathcal{B}^{\mathcal{C}}$ are frozen only for $\mathcal{P}(N, \mathcal{B}, [0, u_{\mathcal{A}^c}])$. See Figure 4.1 to see a pictorial representation of the frozen sets. Similarly to Definition 2.11 we now define the nested polar code as follows:

Definition 4.1 (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 (2.29) 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 $|\mathcal{A}|/N$.

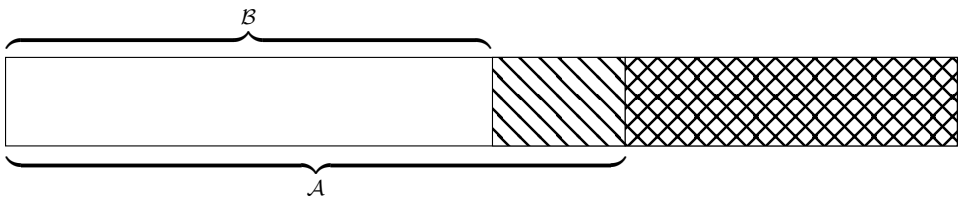


Figure 4.1: A nested polar code. The rectangle corresponds to the whole index set $\{1, \ldots, N\}$. The two frozen sets are $\mathcal{A}^{\mathcal{C}}$ and $\mathcal{B}^{\mathcal{C}}$, and $\mathcal{A}^{\mathcal{C}} \subset \mathcal{B}^{\mathcal{C}}$. (© 2010 IEEE. Reused with permission.)

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 (2.30) by $W_N^{(i)}$ and $\tilde{W}_N^{(i)}$ respectively, and their Bhattacharyya parameters by $Z_N^{(i)}$ and $\tilde{Z}_N^{(i)}$ respectively. The following Lemma from [Kor09] allows us to construct nested polar codes for degraded channels where the overall code is capacity-achieving for W, while the subcodes are capacity achieving for $\tilde{W}$:

Lemma 4.2 (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 4.2 to show that nested polar codes achieve the whole capacity-equivocation region for the degraded wiretap channel.

4.2 Polar Codes for the Wiretap Channel

We consider a wiretap channel where Alice's 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

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

In Theorem 4.3 we give a nested polar coding scheme for the wiretap channel that achieves the whole capacity-equivocation region.

Theorem 4.3. *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, and an $n_0 \in \mathbb{N}$ that satisfy*

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

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

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

provided that $n > n_0$. The encoders and decoders can be implemented with complexity $O(N \log N)$.

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 $\lim_{n\to\infty} |\mathcal{A}_N|/N = C_M \geq C_M - R$ 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 2.12 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^β} 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

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

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 4.2, 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

$$\lim_{n\to\infty} H(X^N|Z^N S_N) \leq \lim_{n\to\infty} \max_{s_N} \left[h_2(P_e^{N,s_N}) + P_e^{N,s_N}|\mathcal{B}_N| \right] = 0,$$

since $P_e^{N,s_N}|\mathcal{B}_N| = N2^{-N^\beta}|\mathcal{B}_N|/N \leq N2^{-N^\beta}C_W$ $\forall s_N$.

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 $|\mathcal{B}_N| = N(C_M - R) > NC_W$, Eve cannot decode the code defined by (4.7) with vanishing error probability. 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. Finally, the complexity of the encoder and the decoder is the same as for the point-to-point channel. ■

4.2.1 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 4.4. *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

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

[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.

For a specific received z^N 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 $H(X^N|Z^N S_N) = N - \text{rank}(H_{\mathcal{E}})$. This implies that $H(S_N|Z^N) = \text{rank}(H_{\mathcal{E}}) - \text{rank}(H_{1,\mathcal{E}})$. ■

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.

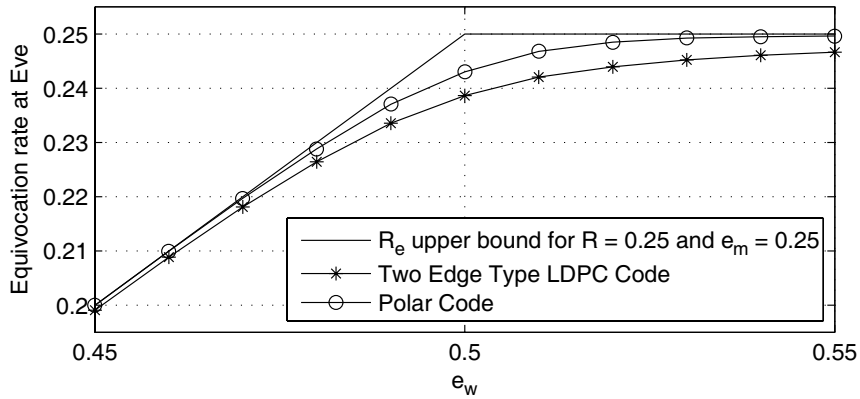


Figure 4.2: Equivocation rate versus e_w. Codes designed for $R = 0.25$, $e_m = 0.25$, $e_w = 0.5$, and block length $N = 1024$.(© 2010 IEEE. Reused with permission.)

4.3 Polar Codes for the physically degraded Relay Channel with orthogonal receivers

Consider the physically degraded relay channel with binary input alphabets $\mathcal{X}$ and $\mathcal{X}_1$ as defined in 2.4.1. We assume that the source to relay (SR), source to destination (SD), and relay to destination (RD) channels are symmetrical. In this case the capacity is given by

$$C = \min\{C_{SD} + C_{RD}, C_{SR}\},$$

and can be achieved using nested polar codes using a block coding scheme. Our result is the following:

Theorem 4.5. *Let $R < C$. For all $\epsilon > 0$ there exists a nested polar code of rate R and length $(B+1)N = (B+1)2^n$, and two integers B_0 and n_0 such that the error probability at the destination is smaller than ϵ provided that $B > B_0$ and $n > n_0$.*

Proof. We use a block coding scheme and transmit B codewords of length N in $B+1$ blocks. Let W and $\tilde{W}$ denote the SR and SD channels respectively. Let $Z_N^{(i)}$ and $\tilde{Z}_N^{(i)}$ be the Bhattacharyya parameters of the corresponding polarized channels.

First assume that $C_{SR} \leq C_{SD} + C_{RD}$. Let $0 < \beta < 1/2$, $\mathcal{A}_N = \{i : Z_N^{(i)} < 2^{-N^\beta}\}$, and let $\mathcal{B}_N = \{i : \tilde{Z}_N^{(i)} < 2^{-N^\beta}\}$. By Lemma 4.2, $\mathcal{B}_N \subset \mathcal{A}_N$. The source will transmit in each block using the nested polar code $P(N, \mathcal{A}_N, \mathcal{B}_N)$. After receiving the whole codeword the relay decodes the bits in $\mathcal{A}_N$. The probability that the relay makes an error when decoding can be made smaller than $\epsilon/(3B)$ by choosing n large enough. The relay then reencodes the bits in $\mathcal{A}_N \setminus \mathcal{B}_N$ and transmits them using a polar code of rate $(|\mathcal{A}_N| - |\mathcal{B}_N|)/N$ in the next block. In general, in block k the source transmits the k^{th} codeword while the relay transmits the bits in $\mathcal{A}_N \setminus \mathcal{B}_N$ from the $(k-1)^{\text{th}}$ block. The destination first decodes the bits in $\mathcal{A}_N \setminus \mathcal{B}_N$ using the transmission from the relay. This can be done with error probability smaller than $\epsilon/(3B)$ provided n is large enough since the rate of the relay to destination code tends to $C_{SR} - C_{SD} \leq C_{RD}$ as n grows. Finally the destination decodes the source transmission from the $(k-1)^{\text{th}}$ block. It uses the bits from the relay transmission in block k to determine which coset of $P(N, \mathcal{B}_N)$ the codeword lies in. If n is large enough, the rate of $P(N, \mathcal{B}_N)$ is smaller than C_{SD} so the destination can decode with block error probability smaller than $\epsilon/(3B)$. By the union bound the overall error probability over all B blocks is then smaller than ϵ. The rate of the scheme is $B|\mathcal{A}_N|/N(B+1)$ which can be made arbitrarily close to C_{SR} provided B and n are large enough since $\liminf_{n\to\infty} |\mathcal{A}_N|/N = C_{SR}$.

Now assume that $C_{SR} > C_{SD} + C_{RD}$. Let $\mathcal{B}_N = \{i : \tilde{Z}_N^{(i)} < 2^{-N^\beta}\}$ and let $\mathcal{A}_N$ be a subset of $\{i : Z_N^{(i)} < 2^{-N^\beta}\}$ of size $N(C_{SD} + C_{RD})$ containing $\mathcal{B}_N$. Such a subset exists provided n is large enough since $C_{SR} > C_{SD} + C_{RD}$. The analysis of the block error probability is the same as in the first case, and the rate of the coding scheme is $B|\mathcal{A}_N|/N(B+1)$ which approaches $C_{SD} + C_{RD}$ when n and B are large. ■

4.4 Polar Codes for the Bidirectional Broadcast Channel

We consider polar codes for the Bidirectional Broadcast Channel introduced in Section 2.4.2. Recall the capacity-equivocation region given by Theorem 2.7

Theorem 2.7. *The capacity-equivocation region of the BBC with common and confidential messages is the set of rate-equivocation tuples* $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ *that satisfy*

$$\begin{aligned} R_e &\leq R_c \\ R_e &\leq I(V;Y_1|U) - I(V;Y_2|U) \\ R_c + R_0 + R_k &\leq I(V;Y_1|U) + I(U;Y_k), \quad k = 1,2 \\ R_0 + R_k &\leq I(U;Y_k), \quad k = 1,2 \end{aligned}$$

for random variables $U \to V \to X \to (Y_1, Y_2)$*. The cardinalities of the ranges of* U *and* V *can be bounded by*

$$|\mathcal{U}| \leq |\mathcal{X}| + 3, \qquad |\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3.$$

For the following analysis of polar codes we need the case where the marginal channels are degraded, i.e., $X \to Y_1 \to Y_2$.

Corollary 4.6. *The capacity-equivocation region of the degraded BBC with common and confidential messages is the set of rate tuples* $(R_c, R_e, R_0, R_1, R_2) \in \mathbb{R}_+^5$ *that satisfy*

$$\begin{aligned} R_e &\leq R_c \\ R_e &\leq I(X;Y_1|U) - I(X;Y_2|U) \\ R_c + R_0 + R_k &\leq I(X;Y_1|U) + I(U;Y_k), \quad k = 1,2 \\ R_0 + R_k &\leq I(U;Y_k), \quad k = 1,2 \end{aligned}$$

for random variables $U \to X \to Y_1 \to Y_2$*. The cardinality of the range of* U *can be bounded by*

$$|\mathcal{U}| \leq |\mathcal{X}|.$$

□

Proof. The achievability follows immediately from the non-degraded case in Theorem 2.7, cf. also [WB11]. We prove the converse and the bound on the cardinality of $\mathcal{U}$ in the appendix. ■

By considering the case of perfect secrecy, i.e. $R_e = R_c$, we obtain the secrecy capacity region.

Corollary 4.7. *The secrecy capacity region of the degraded BBC with common and confidential messages is the set of rate tuples* $(R_c, R_0, R_1, R_2) \in \mathbb{R}_+^4$ *that satisfy*

$$R_c \leq I(X;Y_1|U) - I(X;Y_2|U)$$

$$R_0 + R_k \leq I(U; Y_k), \quad k = 1, 2$$

for random variables $U \to X \to Y_1 \to Y_2$*. The cardinality of the range of* U *can be bounded by*

$$|\mathcal{U}| \leq |\mathcal{X}|.$$

□

Remark 4.8. The improved bound on the cardinality of $\mathcal{U}$ is particularly helpful when designing coding schemes. In the following subsections we will see that it allows us to consider binary input coding schemes when designing codes for a binary input channel, where a looser bound might have required non-binary schemes. ◊

Remark 4.9. Note that by letting $R_e = 0$ in Corollary 4.11 we drop the secrecy constraint on the message s_c. In this case the BBC with common and confidential messages specializes to the broadcast channel with partial receiver side information and degraded message sets considered in [KS07]. Thus the BBC with common and confidential messages is a generalization of the broadcast channel with partial receiver side information and degraded message sets, and any scheme that is capacity achieving for the first is also capacity achieving for the latter. ◊

In the next subsections we design polar coding schemes for the BBC, and then for the BBC with common and confidential messages.

4.4.1 Polar Codes for the BBC

First consider a binary input BBC W with marginal channels W_1 and W_2 with no common and confidential messages. The capacity region is given by

$$R_1 \leq C_1, \tag{4.9}$$
$$R_2 \leq C_2, \tag{4.10}$$

where C_1 and C_2 are the capacities of W_1 and W_2 respectively.

In the following theorem we present a polar coding scheme for this channel. Note how the values of the frozen bits for the two users correspond to the side information available.

Theorem 4.10. *Let* W *be a BBC with binary input alphabet and symmetric marginal channels* W_1 *and* W_2*. For every* $\epsilon > 0$ *and every* $0 < \beta < 1/2$*, there exists a polar coding scheme of length* $N = 2^n$ *and an* $n_0 \in \mathbb{N}$ *that satisfy*

$$R_1 > C_1 - \epsilon, \tag{4.11}$$
$$R_2 > C_2 - \epsilon, \tag{4.12}$$
$$P_e^N < 2^{-N^\beta} \tag{4.13}$$

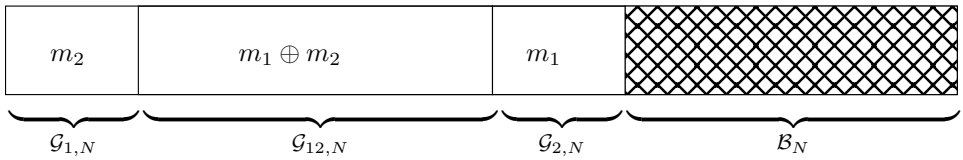


Figure 4.3: Frozen sets and encoding for the BBC. A Part of m_1 (m_2) is transmitted over $\mathcal{G}_{1,N}$ ($\mathcal{G}_2, N$), and the remaining part of m_1 and m_2 are transmitted as $m_1 \oplus m_2$ over $\mathcal{G}_{12,N}$. (© 2013 IEEE. Reused with permission.)

if $n > n_0$. The encoders and decoders can be implemented with complexity $O(N \log N)$.

Proof. Fix $0 < \beta < 1/2$. Let $W_{k,N}^{(i)}$ and $Z_{k,N}^{(i)}$ for $k = 1, 2$ denote the polarized marginal channels and their Bhattacharyya parameters. Now define the following sets:

$$\mathcal{G}_{1,N} = \{i : Z_{1,N}^{(i)} < 2^{-N^\beta} \text{ and } Z_{2,N}^{(i)} \geq 2^{-N^\beta}\}, \tag{4.14}$$

$$\mathcal{G}_{2,N} = \{i : Z_{1,N}^{(i)} \geq 2^{-N^\beta} \text{ and } Z_{2,N}^{(i)} < 2^{-N^\beta}\}, \tag{4.15}$$

$$\mathcal{G}_{12,N} = \{i : Z_{1,N}^{(i)} < 2^{-N^\beta} \text{ and } Z_{2,N}^{(i)} < 2^{-N^\beta}\}, \tag{4.16}$$

$$\mathcal{B}_N = \{i : Z_{1,N}^{(i)} \geq 2^{-N^\beta} \text{ and } Z_{2,N}^{(i)} \geq 2^{-N^\beta}\}, \tag{4.17}$$

where $\mathcal{G}_{1,N}$ are the channels that are good only for node 1, $\mathcal{G}_{2,N}$ the channels that are good only for node 2, $\mathcal{G}_{12,N}$ are the channels that are good for both nodes, and $\mathcal{B}_N$ are the channels that are bad for both nodes. Consider the polar code $\mathcal{C}(N, \mathcal{G}_{1,N} \cup \mathcal{G}_{2,N} \cup \mathcal{G}_{12,N}, u_\mathcal{F})$ with input bits given by

$$u_i = \begin{cases} m_{2i} & \text{if } i \in \mathcal{G}_{1,N}, \\ m_{1i} & \text{if } i \in \mathcal{G}_{2,N}, \\ m_{1i} \oplus m_{2i} & \text{if } i \in \mathcal{G}_{12,N}, \end{cases}$$

where we assume that the messages m_1 and m_2 are binary vectors. The frozen sets and the encoding is shown in Figure 4.3. Since node 1 knows m_1 it treats the input bits in $\mathcal{G}_{2,N}$ as frozen and decodes the input bits u_i for $i \in \mathcal{G}_{1,N} \cup \mathcal{G}_{12,N}$ using the SC decoder (2.31). Finally it subtracts the bits of m_1 that appear in bits in $\mathcal{G}_{12,N}$. Thus the rate for node 1 becomes

$$R_{1,N} = \frac{|\mathcal{G}_{1,N}| + |\mathcal{G}_{12,N}|}{N}. \tag{4.18}$$

Node 2 treats the input bits m_2 in $\mathcal{G}_{1,N}$ as frozen and gets the rate

$$R_{2,N} = \frac{|\mathcal{G}_{2,N}| + |\mathcal{G}_{12,N}|}{N}. \tag{4.19}$$

By the definition of $\mathcal{G}_{1,N}, \mathcal{G}_{2,N}, \mathcal{G}_{12,N}, \mathcal{B}_N$, Theorem 2.12 and (2.34) we get (4.11) – (4.13) Finally, the complexity of the encoder and the decoder is the same as for the point-to-point channel. ■

Note that we can use some of the input bits in $\mathcal{G}_{12,N}$ to transmit a common message m_0, unknown at both destinations, by transferring parts of the rates R_1 and R_2 to R_0.

Corollary 4.11. *Let W be a BBC with binary input alphabet and symmetric marginal channels W_1 and W_2, where W_2 is degraded with respect to W_1. If we consider an additional common message m_0, the scheme in Theorem 4.10 achieves the following rate triples, which is the capacity region,*

$$R_0 + R_1 \leq C_1 \tag{4.20}$$
$$R_0 + R_2 \leq C_2. \tag{4.21}$$

□

Proof. It is easy to see that C_1 and C_2 are outer bounds to the capacity region. Since W_2 is degraded with respect to W_1 we have $\mathcal{G}_{2,N} = \emptyset$ by Lemma 4.2. Thus, by (2.33),

$$\lim_{N\to\infty} R_{0,N} + R_{1,N} = \lim_{N\to\infty} \frac{|\mathcal{G}_{1,N}| + |\mathcal{G}_{12,N}|}{N} = C_1, \tag{4.22}$$

and

$$\lim_{N\to\infty} R_{0,N} + R_{2,N} = \lim_{N\to\infty} \frac{|\mathcal{G}_{12,N}|}{N} = C_2, \tag{4.23}$$

which completes the proof. ■

Remark 4.12. Note that the condition that W_2 is degraded with respect to W_1 ensures that $\mathcal{G}_{2,N} = \emptyset$. If W_1 and W_2 are not ordered by degradation, the highest rate for the common message that can be achieved is given by $\liminf_{N\to\infty} |\mathcal{G}_{12,N}|/N$. This quantity is called the compound capacity $C_{P,SC}(W_1, W_2)$ of W_1 and W_2 using polar codes and SC decoding. In general, $C_{P,SC}(W_1, W_2)$ is lower than the minimum of the capacities of W_1 and W_2. Methods to calculate upper and lower bounds on $C_{P,SC}(W_1, W_2)$ were developed in [HKU09]. ◊

In the next subsection we show how to design polar codes for a degraded BBC with common and confidential messages.

4.4.2 Polar Codes for the BBC with Confidential Messages

We consider the case where W_1 and W_2 are binary symmetric channels (BSC) with transition probabilities p_1 and p_2, with $p_2 > p_1$.[2] We call such a channel a binary symmetric BBC. Using the upper bound on $|\mathcal{U}|$ from Corollary 4.6 and the same arguments as in [CT91, Example 15.6.3] it is easy to show that choosing U to be a Ber(1/2) binary random variable, and $P_{X|U}$ to be a BSC(α), with $0 < \alpha < 1/2$ is optimal. In this case the capacity-equivocation region in Corollary 4.6 becomes

$$\begin{aligned} 0 \leq R_e \leq & R_c \\ R_e \leq & h_2(\alpha \star p_1) - h_2(p_1) - h_2(\alpha \star p_2) + h_2(p_2) \\ R_c + R_0 + R_k \leq & h_2(\alpha \star p_1) - h_2(p_1) + 1 - h_2(\alpha \star p_k), \\ & k = 1, 2 \\ R_0 + R_k \leq & 1 - h_2(\alpha \star p_k), \quad k = 1, 2, \end{aligned}$$

where $\alpha \star \beta = (1-\alpha)\beta + \alpha(1-\beta)$.

Our main result is the following:

Theorem 4.13. *There exists a polar code $\mathcal{C}_{BBC}$ designed for the binary symmetric BBC, and a polar code $\mathcal{C}_{WT}$ designed for the binary symmetric wiretap channel such that transmitting*

$$X^N = X^N_{BBC} \oplus X^N_{WT},$$

for $X^N_{BBC} \in \mathcal{C}_{BBC}$ and $X^N_{WT} \in \mathcal{C}_{WT}$ achieves the capacity-equivocation region for the binary symmetric BBC with common and confidential messages. The encoders and decoders can be implemented with complexity $O(N \log N)$.

Proof. Fix $0 < \alpha < 1/2$. We first design $\mathcal{C}_{BBC}$ for a binary symmetric BBC with a common message with transition probabilities $\alpha \star p_1$ and $\alpha \star p_2$. If X^N_{WT} is statistically indistinguishable from an i.i.d. Ber(α) vector, then, by Corollary 4.11, $\mathcal{C}_{BBC}$ achieves all rate triples satisfying

$$R_0 + R_k \leq 1 - h_2(\alpha \star p_k), \quad k = 1, 2.$$

Both nodes can now decode X^N_{BBC} and remove its contribution. Note that since the channels are symmetric, the error probabilities do not depend on the values of the frozen bits, and we can choose them to be zero [Arı09]. Also note that since X^N_{BBC} and X^N_{WT} are independent, X^N_{BBC} provides no information about X^N_{WT}.

[2] This apparent simplification is made to make the exposition clearer. Our results generalize to arbitrary q-ary input BBCs with degraded marginal channels using results from [STA09].

Thus, assuming that node 2 decodes X_{BBC}^N does not increase the equivocation of m_c at node 2.

Let $\mathcal{C}_{WT}$ be a polar code with input weight $\alpha' \in \mathbb{Q}$ designed for a binary symmetric wiretap channel with transition probabilities p_1 and p_2 using Theorem 4.3. To design a polar code with rational input weight α', we augment the binary channel with a virtual q-ary input and then design a q-ary input polar code for this augmented channel. This technique was introduced by Gallager [Gal68], and used for polar codes in [STA09, Kor09]. Since any $\alpha \in \mathbb{R}$ can be approximated arbitrarily well by an $\alpha' \in \mathbb{Q}$, such a construction achieves all rate-equivocation pairs satisfying

$$\begin{aligned} R_c \leq & h_2(\alpha \star p_1) - h_2(p_1), \\ R_e \leq & h_2(\alpha \star p_1) - h_2(p_1) - h_2(\alpha \star p_2) + h_2(p_2). \end{aligned}$$

In order to make the codewords of $\mathcal{C}_{WT}$ statistically indistinguishable from an i.i.d. Ber(α) vector we average over all possible values of the frozen bits of $\mathcal{C}_{WT}$. Let $P_{e,BBC}(u_\mathcal{F})$, $P_{e,WT}(u_\mathcal{F})$, and $P_e(u_\mathcal{F})$ be the average error probabilities of $\mathcal{C}_{BBC}$, $\mathcal{C}_{WT}$, and the overall scheme respectively, when using $u_\mathcal{F}$ as the frozen bits for $\mathcal{C}_{WT}$. Choosing $u_\mathcal{F}$ uniformly at random we can make the average error probability

$$\mathbb{E}_{U_\mathcal{F}}[P_e(U_\mathcal{F})] \leq \mathbb{E}_{U_\mathcal{F}}[P_{e,BBC}(U_\mathcal{F}) + P_{e,WT}(U_\mathcal{F})]$$

arbitrarily small by choosing N large enough, since the codewords of $\mathcal{C}_{WT}$ are i.i.d. Ber(α) when averaged over $u_\mathcal{F}$. Since the average error probability is small there exists at least one $u_\mathcal{F}$ such that $P_e(u_\mathcal{F})$ is small, and using this $u_\mathcal{F}$ as the frozen bits for $\mathcal{C}_{WT}$ makes the overall error probability small.

Finally, the complexity of the encoders and the decoders are the same as in the point-to-point setting. ■

Remark 4.14. Consider a BBC with non-degraded marginal channels. As in Remark 4.12, R_0 is bounded from above by $C_{P,SC}(W_1, W_2)$, but more importantly, the analysis of the equivocation rate R_e becomes difficult. It was conjectured in [HS10] that it is possible to achieve the secrecy capacity of non-degraded wiretap channels using polar codes. A proof of this conjecture would also apply to our scheme. ◊

4.A Proof of Weak Converse

For any sequence of codes for the degraded BBC with common and confidential messages with error probabilities going to zero, we want to show that there exist random variables $U \to X \to Y_1 \to Y_2$ such that

$$\begin{aligned}\tfrac{1}{N}H(S_c|Y_2^N S_2) &\le I(X;Y_1|U) - I(X;Y_2|U)\\ \tfrac{1}{N}(H(S_c)+H(S_0)+H(S_k)) &\le I(X;Y_1|U) + I(U;Y_k),\ k=1,2\\ \tfrac{1}{N}(H(S_0)+H(S_k)) &\le I(U;Y_k),\ k=1,2.\end{aligned}$$

We do this by using techniques similar to [LLL10] and the Fano-like inequalities

$$\begin{aligned}H(S_c S_0 S_2|Y_1^N S_1) &\le N\epsilon_{1,N},\\ H(S_0 S_1|Y_2^N S_2) &\le N\epsilon_{2,N},\end{aligned}$$

from [WB11]. Here $\epsilon_{1,N}$ and $\epsilon_{2,N}$ are two non-negative sequences that tend to zero as $N \to \infty$. Let $S_{012} = (S_0 S_1 S_2)$ and introduce the random variable $U_i = (S_{012} Y_1^{i-1})$.

We first bound $N(R_0 + R_1) \le H(S_0) + H(S_2)$ as

$$\begin{aligned}H(S_0)+H(S_2) \le& I(S_{012};Y_1^N) + N\epsilon_{1,N}\\ \le& \sum_{i=1}^N I(S_{012}Y_1^{i-1};Y_{1i}) + N\epsilon_{1,N}\\ =& \sum_{i=1}^N I(U_i;Y_{1i}) + N\epsilon_{1,N}.\end{aligned}$$

Then we bound $N(R_0 + R_2) \le H(S_0) + H(S_1)$ as

$$\begin{aligned}H(S_0)+H(S_1) \le& I(S_{012};Y_2^N) + N\epsilon_{2,N}\\ \le& \sum_{i=1}^N I(S_{012}Y_1^{i-1}Y_2^{i-1};Y_{2i}) + N\epsilon_{2,N}\\ \overset{(a)}{=}& \sum_{i=1}^N I(S_{012}Y_1^{i-1};Y_{2i}) + N\epsilon_{2,N}\\ =& \sum_{i=1}^N I(U_i;Y_{2i}) + N\epsilon_{2,N},\end{aligned} \tag{4.24}$$

where (a) follows from the degradedness $X_i \to Y_{1i} \to Y_{2i}$.

We bound $H(S_c)$:

$$H(S_c) \le I(S_c;Y_1^N|S_{012}) + N\epsilon_{1,N}$$

$$
\begin{aligned}
&\leq I(S_c X^N; Y_1^N | S_{012}) + N\epsilon_{1,N} \\
&= \sum_{i=1}^{N} I(X^N; Y_{1i} | S_{012} Y_1^{i-1}) + N\epsilon_{1,N} \\
&= \sum_{i=1}^{N} H(Y_{1i} | S_{012} Y_1^{i-1}) - H(Y_{1i} | S_{012} Y_1^{i-1} X^N) + N\epsilon_{1,N} \\
&= \sum_{i=1}^{N} H(Y_{1i} | S_{012} Y_1^{i-1}) - H(Y_{1i} | S_{012} Y_1^{i-1} X_i) + N\epsilon_{1,N} \\
&= \sum_{i=1}^{N} I(X_i; Y_{1i} | S_{012} Y_1^{i-1}) + N\epsilon_{1,N} \\
&= \sum_{i=1}^{N} I(X_i; Y_{1i} | U_i) + N\epsilon_{1,N}.
\end{aligned}
$$

Finally we bound $NR_c \leq H(S_c | Y_2^N S_2)$ as

$$
\begin{aligned}
&H(S_c | Y_2^N S_2) \\
&= H(S_c | Y_2^N S_{012}) + I(S_c; S_0 S_1 | Y_2^N S_2) \\
&\leq H(S_c | Y_2^N S_{012}) + N\epsilon_{2,N} \\
&= I(S_c; Y_1^N | Y_2^N S_{012}) + H(S_c | Y_2^N S_{012} Y_1^N) + N\epsilon_{2,N} \\
&\leq I(S_c; Y_1^N | Y_2^N S_{012}) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&\leq I(S_c X^N; Y_1^N | Y_2^N S_{012}) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= I(X^N; Y_1^N | Y_2^N S_{012}) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= H(X^N | S_{012} Y_2^N) - H(X^N | S_{012} Y_2^N Y_1^N) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= H(X^N | S_{012} Y_2^N) - H(X^N | S_{012} Y_1^N) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= I(X^N; Y_1^N | S_{012}) - I(X^N; Y_2^N | S_{012}) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= \sum_{i=1}^{N} I(X^N; Y_{1i} | S_{012} Y_1^{i-1}) - I(X^N; Y_{2i} | S_{012} Y_2^{i-1}) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= \sum_{i=1}^{N} H(Y_{1i} | Y_1^{i-1} S_{012}) - H(Y_{1i} | Y_1^{i-1} S_{012} X^N) - H(Y_{2i} | Y_2^{i-1} S_{012}) \\
&\qquad + H(Y_{2i} | Y_2^{i-1} S_{012} X^N) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&\leq \sum_{i=1}^{N} H(Y_{1i} | Y_1^{i-1} S_{012}) - H(Y_{1i} | Y_1^{i-1} S_{012} X_i) - H(Y_{2i} | Y_2^{i-1} Y_1^{i-1} S_{012})
\end{aligned}
$$

$$
\begin{aligned}
&\quad + H(Y_{2i}|Y_2^{i-1}S_{012}X_i) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&\overset{\text{(b)}}{=} \sum_{i=1}^{N} H(Y_{1i}|Y_1^{i-1}S_{012}) - H(Y_{1i}|Y_1^{i-1}S_{012}X_i) - H(Y_{2i}|Y_1^{i-1}S_{012}) \\
&\quad + H(Y_{2i}|Y_1^{i-1}S_{012}X_i) + N\epsilon_{1,N} + N\epsilon_{2,N} \\
&= \sum_{i=1}^{N} I(X_i;Y_{1i}|U_i) - I(X_i;Y_{2i}|U_i) + N\epsilon_{1,N} + N\epsilon_{2,N},
\end{aligned}
$$

where (b) follows from the Markov chain $(Y_1^{i-1}, Y_2^{i-1}, S_{012}) \to X_i \to Y_{2i}$, which is due to the channel being memoryless.

Now we get the desired bounds by letting J be a R.V. uniformly distributed over $\{1, \ldots, N\}$, and choosing $U = (U_J, J), X = X_J, Y_1 = Y_{1J}$, and $Y_2 = Y_{2J}$.

4.B Proof of Bound on Cardinality of $\mathcal{U}$

We follow [Sal78] closely, and use their notation. By [Sal78, Lemma 3] the capacity-equivocation region is given by

$$
\begin{aligned}
\{(R_e, R_c, R_0, R_1, R_2) \in \mathbb{R}_+^5 : \forall (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) \in \mathbb{R}_+^5, \\
\lambda_1 R_e + \lambda_2(R_c + R_0 + R_1) + \lambda_3(R_c + R_0 + R_2) + \\
\lambda_4(R_0 + R_1) + \lambda_5(R_0 + R_2) \le G(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)\},
\end{aligned}
$$

where $G(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ is given by the supremum of

$$
\begin{aligned}
\lambda_1(I(X;Y_1|U) - I(X;Y_2|U)) + \lambda_2(I(X;Y_1|U) + I(U;Y_1)) + \\
\lambda_3(I(X;Y_1|U) + I(U;Y_2)) + \lambda_4 I(U;Y_1) + \lambda_5 I(U;Y_2),
\end{aligned}
$$

taken over all R.V. U s.t. $P_{UXY_1Y_2} = P_U P_{X|U} P_{Y_1Y_2|X}$. Now let $\mathcal{P}$ be the set of probability distributions on $\mathcal{X}$, and let $P_X \in \mathcal{P}$. We define the following $|\mathcal{X}|$ functions on $\mathcal{P}$:

$$
\begin{aligned}
f_j(P_X) =& P_X(j), \quad j = 1, 2, \ldots, |\mathcal{X}| - 1, \\
f_{|\mathcal{X}|}(P_X) =& \lambda_1(I_{P_X}(X;Y_1) - I_{P_X}(X;Y_2)) \\
&+ \lambda_2(I_{P_X}(X;Y_1) - H_{P_X}(Y_1)) \\
&+ \lambda_3(I_{P_X}(X;Y_1) - H_{P_X}(Y_2)) \\
&- \lambda_4 H_{P_X}(Y_1) - \lambda_5 H_{P_X}(Y_2),
\end{aligned}
$$

where $I_{P_X}(X;Y_i)$ and $H_{P_X}(Y_i)$ are the corresponding mutual information and entropies when the distribution of X is P_X. Each probability distribution P_U defines

a measure $\mu(dP_X)$ on $\mathcal{P}$. Let P_X^* be the probability distribution that achieves $G(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$, and let μ^* be the corresponding measure. Note that

$$\begin{aligned}
\int f_j(P_X)\mu^*(dP_X) =& P_X^*(j), \quad j = 1, 2, \ldots, |\mathcal{X}| - 1, \\
\int f_{|\mathcal{X}|}(P_X)\mu^*(dP_X) =& \lambda_1(I_{P_X^*}(X;Y_1|U) - I_{P_X^*}(X;Y_2|U)) \\
&+ \lambda_2(I_{P_X^*}(X;Y_1|U) - H_{P_X^*}(Y_1|U)) \\
&+ \lambda_3(I_{P_X^*}(X;Y_1|U) - H_{P_X^*}(Y_2|U)) \\
&- \lambda_4 H_{P_X^*}(Y_1|U) - \lambda_5 H_{P_X^*}(Y_2|U).
\end{aligned}$$

From $f_1(P_X^*), \ldots, f_{|\mathcal{X}|-1}(P_X^*)$ we can calculate $H_{P_X^*}(Y_1)$ and $H_{P_X^*}(Y_2)$ and form

$$\begin{aligned}
\int f_{|\mathcal{X}|}(P_X)\mu^*(dP_X) + (\lambda_2 + \lambda_4)H_{P_X^*}(Y_1) + (\lambda_3 + \lambda_5)H_{P_X^*}(Y_2) \\
= G(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5).
\end{aligned}$$

Now it follows from [Sal78, Lemma 2] that it is sufficient to consider R.V. U with $|\mathcal{U}| \leq |\mathcal{X}|$.

Chapter 5

Sparse Regression Codes

In this chapter we consider coding schemes based on nested sparse regression codes (SPARCs). We consider the AWGN wiretap channel and show that nested SPARCs achieve the secrecy capacity. As in the case with the polar codes considered in Chapter 4 the nested codes for the wiretap channel can also be used to implement the decode-and-forward scheme for the relay channel which achieves the capacity of the physically degraded relay channel with orthogonal receivers. We then show that the Wyner-Ziv coding scheme from [VT12] can be employed in a secret key agreement scheme for correlated Gaussian sources over a rate-limited public channel.

5.1 Nested SPARCs for the Wiretap Channel

As noted in [VT12], SPARCs can be given a nested structure. As shown in Figure 5.1, the M columns in each section of the design matrix of the code is divided into subsections containing M' columns each. The choice of one such subsection from each section specifies a subcode of the overall code. Formally we define a nested SPARC as follows

Definition 5.1 (Nested Sparse Regression Code $\mathcal{C}_N(R_1, R_2, b)$)**.** Let $M = L^b$, where L satisfies

$$NR_1 = bL \ln L,$$

and let A be an $N \times ML$ design matrix with i.i.d. $\mathcal{CN}(0,1)$ entries. The M columns of each section are divided into subsections of M' sections each, where $M'^L = e^{NR_2}$. The choice of one subsection from each section then specifies a subcode, and there are $(M/M')^L = e^{N(R_1-R_2)}$ such different subcodes. $\Diamond$

Now consider the Gaussian wiretap channel given by

$$Y = X + W_M,$$

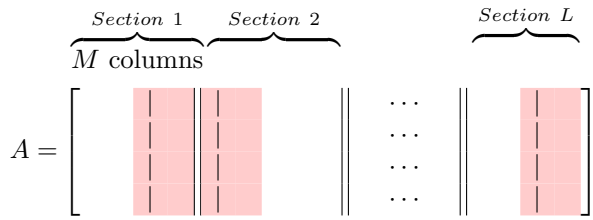


Figure 5.1: The design matrix of a nested sparse regression code. The red columns indicate the chosen subsection from each section.

$$Z = X + W_W,$$

where $W_M \sim \mathcal{CN}(0, \sigma_M^2)$, $W_W \sim \mathcal{CN}(0, \sigma_W^2)$, with $\sigma_W^2 > \sigma_M^2$ and we have the power constraint

$$E[|X|^2] \leq P.$$

This channel is depicted in Figure 5.2.

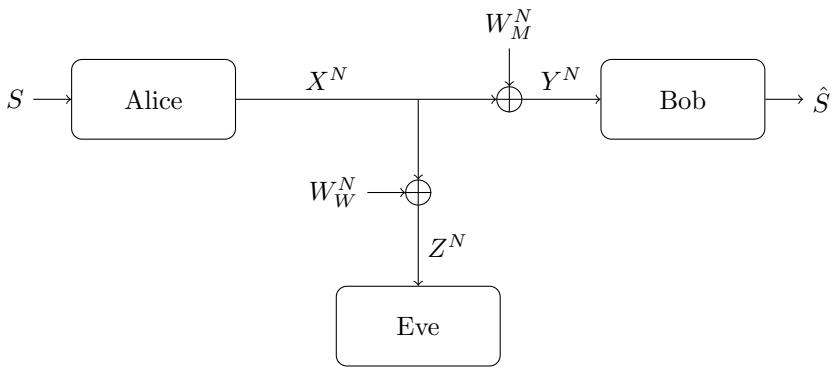


Figure 5.2: Gaussian wiretap channel.

The secrecy capacity of this channel was found by Leung-Yan-Cheong and Hellman [LH78] and is given by

$$C_S = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_M^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_M^2}\right), \tag{5.1}$$

where $C_M = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_M^2}\right)$ is the capacity of the channel to Bob, and $C_W = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_W^2}\right)$ is the capacity of the channel to Eve.

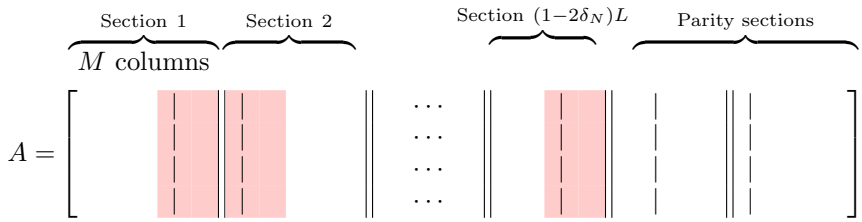


Figure 5.3: The design matrix of a nested sparse regression code with an outer R-S code. The parity sections are not nested.

Now consider the usual coding scheme for the wiretap channel where each subcode corresponds to a certain message, and the choice of a specific codeword in that subcode is done at random. To implement this scheme with an outer R-S code of rate $1 - 2\delta_N$, we note that R-S codes are linear codes and we can implement them as systematic codes. For SPARCs, this implies that the choice of a column from the first $(1 - 2\delta_N)L$ sections specify the columns in the last $2\delta_N L$ sections, and therefore we call them parity sections (cf. the parity bits of systematic linear code). Thus we will only implement the nested structure for the first $(1 - 2\delta_N)L$ columns, see Figure 5.3. The rates of the overall code will then be $(1 - 2\delta_N)R_1$, each subcode will have rate $(1 - 2\delta_N)R_2$, and the coding scheme will have rate $R = (1 - 2\delta_N)(R_1 - R_2)$.

Theorem 5.2. *For every $\epsilon > 0$ there exists a sequence of nested SPARCs $\mathcal{C}_N(R_{1,N}, R_{2,N}, b)$, with $b > \max\{b_0(\mathrm{SNR}_M), \frac{R_{1,N}}{R_{2,N}} b_0(\mathrm{SNR}_W)\}$, a sequence of R-S codes of rates $(1-2\delta_N)$, and an $N_0 \in \mathbb{N}$ such that the above coding scheme satisfies*

$$R > C_M - C_W - \epsilon, \tag{5.2}$$

$$P_{e,N} < \epsilon, \tag{5.3}$$

$$\frac{I(S; Z^N)}{N} < \epsilon, \tag{5.4}$$

if $N > N_0$.

Proof. Fix $\epsilon > 0$. Let $R_{1,N} = C_M - \delta_N$, $R_{2,N} = C_W - \delta_N$, with $\delta_N = 1/\ln N$. The rate of the coding scheme is then

$$(1 - 2\delta_N)(R_{1,N} - R_{2,N}) = (1 - 2\delta_N)(C_M - C_W), \tag{5.5}$$

which establishes (5.2). From Theorem 2.14, since $b > b_0(\mathrm{SNR}_M)$, we have

$$\mathbb{E}[P_e^N] < \frac{\epsilon}{3}, \tag{5.6}$$

if N is large enough, where $\mathbb{E}[\cdot]$ denotes the average over the SPARC ensemble. To bound $\mathbb{E}[I(S;Z^N)/N]$ we use that $I(S;Z^N) + I(X^N;Z^N|S) = I(X^N;Z^N) + I(S;Z^N|X^N) = I(X^N;Z^N)$, since $I(S;Z^N|X^N) = 0$ due to the Markov chain $S \to X^N \to Z^N$. Let $P_e^{N,S}$ denote the error probability when knowing to which subcode S a codeword belongs. We then have

$$\begin{aligned}
\frac{I(S;Z^N)}{N} =& \frac{I(X^N;Z^N)}{N} - \frac{I(X^N;Z^N|S)}{N} \\
\leq& C_W - \frac{H(X^N|S)}{N} + \frac{H(X^N|Z^N S)}{N} \\
\overset{\text{(a)}}{\leq}& C_W - (1-2\delta_N)R_{2,N} + \frac{h_e(P_e^{N,S})}{N} + P_e^{N,S}(1-2\delta_N)R_{2,N} \\
\leq& \delta_N(1 + 2C_W - 2\delta_N) + \frac{h_e(P_e^{N,S})}{N} + P_e^{N,S}(1-2\delta_N)R_{2,N},
\end{aligned} \tag{5.7}$$

where we have used Fano's inequality and the fact that $I(X^N;Z^N) \leq NC_W$ in (a). $h_e(\cdot)$ is the binary entropy function evaluated in nats.

We now bound $P_e^{N,S}$ from above. When considered as a SPARC, each subcode has L sections with M' columns each, which means that the parameter b' for the subSPARC satisfies

$$M' = L^{b'}.$$

This, together with the relation $M'^L = e^{NR_{2,N}}$, gives us

$$b' = \frac{NR_{2,N}}{L \ln L} = \frac{bR_{2,N}}{R_{1,N}}.$$

Since $b > \frac{R_1}{R_2} b_0(\mathrm{SNR}_W)$, we have $b' > b_0(\mathrm{SNR}_W)$. Theorem 2.14 then implies that the minimum distance decoder of the subSPARC has a small probability of section error. The fact that the minimum distance of the subcode is not larger than the minimum distance of the overall code can then be used to show that the outer R-S code can correct any remaining section errors. Thus we have $\mathbb{E}[P_e^{N,S}] < \epsilon'$ for any $\epsilon' > 0$ if N is large enough. Combining this with (5.7) we get

$$\mathbb{E}\left[I(S;Z^N)\right] < \frac{\epsilon}{3}. \tag{5.8}$$

We now evaluate the probability that a randomly chosen code from the SPARC ensemble has both low error probability and low information leakage. We have

$$\Pr\left(\left(P_e^N > \epsilon\right) \cup \left(\frac{I(S;Z^N)}{N} > \epsilon\right)\right) \leq \Pr\left(P_e^N > \epsilon\right) + \Pr\left(\frac{I(S;Z^N)}{N} > \epsilon\right)$$

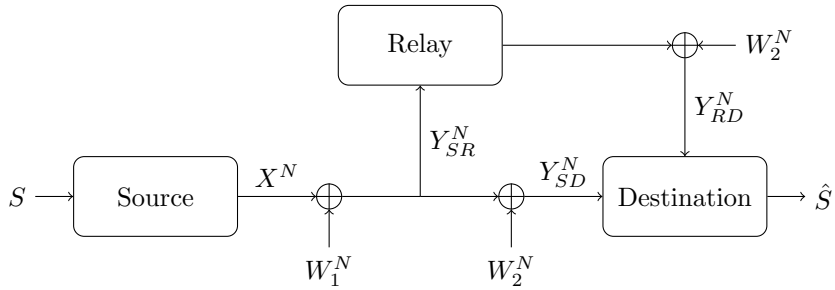


Figure 5.4: Gaussian relay channel with orthogonal receivers.

$$\begin{aligned}&\overset{(a)}{\leq}\frac{\mathbb{E}[P_e^N]}{\epsilon}+\frac{\mathbb{E}\left[\frac{I(S;Z^N)}{N}\right]}{\epsilon}\\&\leq\frac{2}{3},\end{aligned}$$

where (a) follows from Markov's inequality, and in the last step we assume that N is large enough for both (5.6) and (5.7) to hold. Thus there exists an N_0 and at least one sequence of SPARCs such that $N > N_0$ implies that both P_e^N and $I(S;Z^N)/N$ are smaller than ϵ. ■

5.1.1 Decode-and-Forward using nested SPARCs

As we saw in Chapter 4, nested codes designed for the wiretap channel achieve the capacity of the degraded relay channel with orthogonal receivers. Consider the degraded Gaussian relay channel with orthogonal receivers depicted in Figure 5.4.

Let

$$\begin{aligned}Y_{SR}&=X+W_1,\\Y_{SD}&=Y_{SR}+W_2,\\Y_{RD}&=X_R+W_3,\end{aligned}$$

where $W_i \sim \mathcal{CN}(0,\sigma_i^2)$, and we have the power constraints $\mathbb{E}[|X|^2] \leq P_S$ and $\mathbb{E}[|X_R|^2] \leq P_R$. The capacity of this channel is given by [CG79] as

$$\begin{aligned}C&=\min\{C_{SD}+C_{RD},C_{SR}\}\\&=\min\left\{\frac{1}{2}\ln\left(1+\frac{P_S}{\sigma_1^2+\sigma_2^2}\right)+\frac{1}{2}\ln\left(1+\frac{P_R}{\sigma_3^2}\right),\frac{1}{2}\ln\left(1+\frac{P_S}{\sigma_1^2}\right)\right\}.\end{aligned}$$

The same block coding scheme we used to show that nested polar codes achieve the capacity of the physically degraded binary input symmetric relay channel in Chapter 4 can be implemented for nested SPARCs. We have the following result:

Theorem 5.3. *For every $\epsilon > 0$ there exists a sequence of nested SPARCs $C_N(C_{SR} - \delta_N, C_{SD} - \delta_N, b)$, with $b > \max\{b_0(P/\sigma_1^2), \frac{C_{SR}-\delta_N}{C_{SD}-\delta_N} b_0(P/(\sigma_1^2+\sigma_2^2))\}$ a sequence of R-S codes of rates $(1-2\delta_N)$, and an $N_0 \in \mathbb{N}$ such that*

$$R > C - \epsilon$$
$$P_e^N < \epsilon$$

if $N > N_0$.

Proof. The proof is similar to the proof of Theorem 4.5 ■

5.2 Secret Key Agreement using nested SPARCs

Recall the source model for the secret key agreement problem from Section 2.3. We assume that Alice, Bob, and Eve observe correlated Gaussian vectors X, Y, and Z respectively and have access to a one-way rate limited public channel from Alice to Bob and Eve of rate R_p. Let $X \sim \mathcal{N}(0, \sigma^2)$, and let

$$Y = aX + W_Y, \tag{5.9}$$
$$Z = bX + W_Z, \tag{5.10}$$

where W_Y and W_Z are zero mean i.i.d. Gaussian variables that are independent of X with variance N_Y and N_Z respectively. If $a = b$ and $W_Z = W_Y + \tilde{W}$ for a Gaussian random variable $\tilde{W}$ independent of X and W_Y, this describes a degraded source $X \to Y \to Z$.

A secret key agreement scheme for this problem consists of two sets $\mathcal{K}_N$ and $\mathcal{P}_N$ and functions

$$k_A : \mathbb{R}^N \to \mathcal{K}_N, \tag{5.11}$$
$$f_N : \mathbb{R}^N \to \mathcal{P}_N, \tag{5.12}$$
$$k_B : \mathbb{R}^N \times \mathcal{P}_N \to \mathcal{K}_N, \tag{5.13}$$

where $|\mathcal{P}_N| \leq e^{NR_P}$, $f_N(X^N)$ is the message that Alice transmits over the public channel, and $K_A(X^N)$ and $K_B(Y^N, f_N(X^N))$ are the secret keys generated at Alice and Bob respectively.

In [WO10] Watanabe and Oohama employed a secret key agreement scheme based on Wyner-Ziv coding [CT91] to find the secret key capacity of this problem. Let the Gaussian auxiliary random variable $U = X + W$, where $W \sim \mathcal{CN}(0, Q)$. This is equivalent to

$$X = cU + W', \tag{5.14}$$

where $c = \frac{\sigma_X^2}{\sigma_X^2+Q}$, and $W' \sim \mathcal{N}\left(0, \frac{\sigma_X^2 Q}{\sigma_X^2+Q}\right)$ is independent of U. Generate a random codebook based on f_U with rate $R_1 > R_{RD}(\frac{\sigma_X^2 Q}{\sigma_X^2+Q})$, where $R_{RD}(\cdot)$ is the rate-distortion function defined in (2.41). The codewords in the random codebook are then divided into e^{NR_P} bins. Alice selects the codeword $\hat{U}^N$ in the random codebook that minimizes $|X^N - c\hat{U}^N|^2$ and sends the index of the bin to which $\hat{U}^N$ belongs over the public channel. Since $R_1 > R_{RD}(\frac{\sigma_X^2 Q}{\sigma_X^2+Q})$, the distortion satisfies $|X^N - c\hat{U}^N|^2 < \frac{\sigma_X^2 Q}{\sigma_X^2+Q}$ with high probability. Bob then tries to determine $\hat{U}^N$, which is equivalent to a channel decoding problem. If the number of codewords in each bin is smaller than $e^{NI(U;Y)}$, Bob is able to determine $\hat{U}^N$ with low probability of error. Finally Alice and Bob use a hash function chosen from a universal family of hash functions [CW77] to generate a secret key.

Venkataramanan and Tatikonda introduced a Wyner-Ziv coding scheme using SPARCs in [VT12]. We combine this with the key agreement scheme from [WO10] as follows. Fix Q such that

$$\max\left\{\frac{1}{2}\ln\frac{\sigma_X^2+Q}{Q}, \frac{\sigma_X^2}{\sigma_X^2+Q}\right\} - \frac{1}{2}\ln\left(1+\frac{a^2\sigma_X^4}{a^2\sigma_X^2 Q + N_Y(\sigma_X^2+Q)}\right) < R_P,$$

and let

$$R_{1,N} = \max\left\{\frac{1}{2}\ln\frac{\sigma_X^2+Q}{Q}, \frac{\sigma_X^2}{\sigma_X^2+Q}\right\} + \epsilon_1, \tag{5.15}$$

$$R_{2,N} = \frac{1}{2}\ln\left(1+\frac{a^2\sigma_X^4}{a^2\sigma_X^2 Q + N_Y(\sigma_X^2+Q)}\right) - \epsilon_1 - \delta_N, \tag{5.16}$$

where $(1-2\delta_N) = 1 - 2/\ln N$ is the rate loss due to the R-S code needed to ensure a low error probability at Bob, and $\epsilon_1 > 0$ is fixed. As above, let

$$X = cU + W', \tag{5.17}$$

where $c = \frac{\sigma_X^2}{\sigma_X^2+Q}$, and $W' \sim \mathcal{N}\left(0, \frac{\sigma_X^2 Q}{\sigma_X^2+Q}\right)$ is independent of U. Note that $R_{2,N} = I(U;Y) - \epsilon_1 - \delta_N$.

Let g_N denote the minimum distance source encoder of the SPARC $\mathcal{C}_N(R_1, R_2, b)$. Alice finds the codeword $\hat{U}^N$ in the SPARC that minimizes $|X^N - c\hat{U}^N|^2$ and transmits the index of the subcode this codeword belongs to over the public channel to Bob and Eve, together with the syndrome of the outer R-S code. We denote this message over the public channel by P_N. Finally she uses a hash function h_N, which will be discussed later, to extract her key K_A from $\hat{U}^N$.

Bob utilizes his side information Y^N, together with the bin index and the syndrome of the R-S code, both of which are contained in the public message P_N, to find $\hat{U}^N$. He then uses the same hash function h_N to extract his secret key K_B. We have the following theorem:

Theorem 5.4. *Let $R_{1,N}$ and $R_{2,N}$ be defined as in* (5.15) *and* (5.16)*, and let $b > \max\{\frac{3.5R_{1,N}}{R_{1,N}-(1-B_N)}, \frac{R_{1,N}}{R_{2,N}}b_0(\mathrm{SNR})\}$, where B_N and* SNR *satisfy*

$$R_{1,N} = \frac{1}{2}\ln\frac{1}{B_N},$$
$$\mathrm{SNR} = \frac{a^2\sigma_X^4}{a^2\sigma_X^2 Q + N_Y(\sigma_X^2+Q)}.$$

Then for every $\epsilon > 0$ there exists a sequence of nested SPARCs $C_N(R_{1,N}, R_{2,N}, b)$, a sequence of R-S codes of rates $(1-2\delta_N)$, and an $N_0 \in \mathbb{N}$ such that the scheme above satisfies

$$P_e^N < \epsilon, \tag{5.18}$$
$$R_K > I(U;Y) - I(U;Z) - \left[\frac{\sigma_X^2}{\sigma_X^2+Q} - \frac{1}{2}\ln\frac{\sigma_X^2+Q}{Q}\right]^+ - \epsilon, \tag{5.19}$$
$$I(Z^N P_N; K_A) < \epsilon \tag{5.20}$$

if $N > N_0$.

Remark 5.5. If R_P is large enough for $\frac{1}{2}\ln\frac{\sigma_X^2+Q}{Q} > \frac{\sigma_X^2}{\sigma_X^2+Q}$ to hold, then

$$\left[\frac{\sigma_X^2}{\sigma_X^2+Q} - \frac{1}{2}\ln\frac{\sigma_X^2+Q}{Q}\right]^+ = 0.$$

In this case, if the sources are physically degraded, i.e. $X \to Y \to Z$, this scheme achieves the secret key capacity as shown in [WO10]. Note that the first constraint holds in the special case where there is no rate constraint on the public channel. ♢

Proof. We first consider the error probability at Bob. The SNR of the virtual channel connecting U^N and Y^N through

$$Y^N = aX^N + W_Y^N = acU^N + aW'^N + W_Y^N, \tag{5.21}$$

is given by

$$\mathrm{SNR} = \frac{a^2\sigma_X^4}{a^2\sigma_X^2 Q + N_Y(\sigma_X^2+Q)}. \tag{5.22}$$

Since R_2 was designed to be smaller than $\frac{1}{2}\ln(1+\mathrm{SNR})$, the capacity of this channel, and $b' > b_0(\mathrm{SNR})$ (cf. (5.8)), the nearest neighbour decoder of the SPARC together with the outer R-S code allows Bob to determine $\hat{U}^N$ with arbitrarily

small probability of error due to Theorem 2.14. This implies that P_e^N can be made arbitrarily small since Alice and Bob use the same hash function to determine their keys.

To bound the information leakage to Eve, let μ_N be the variational distance between the distributions $f_{Z^N} P_{K_A P_N|Z^N}$ and $f_{Z^N} \tilde{P}_{K_A} P_{P_N|Z^N}$, where $\tilde{P}_{K_A}$ is the uniform distribution over the space of keys:

$$\mu_N = \int \sum_{k_A, p} f_{Z^N}(z^N) |P_{K_A P_N|Z^N}(k_A, p|z^N) - \tilde{P}_{K_A}(k_A) P_{P_N|Z^N}(p|z^N)| dz^N. \quad (5.23)$$

We can connect the information leakage to Eve with the variational distance through the following lemma due to Csiszár:

Lemma 5.6 (Lemma 1 from [Csi96])**.** *The mutual information $I(K_A; Z^N P_N)$ can be bounded from above as follows:*

$$I(K_A; Z^N P_N) \leq \mu_n \ln \frac{|\mathcal{K}_N|}{\mu_N}.$$

□

Now let $g_N : \mathbb{R}^N \to \mathcal{Q}_N \subset \mathbb{R}^N$ denote any quantization function, and let $\phi_N : \mathcal{Q}_N \to \mathcal{P}_N$ denote a mapping from the quantized version of X^N to the public channel. Watanabe and Oohama showed that for any key space $\mathcal{K}_N$ there exists a hash function h_N that satisfies the following:

Lemma 5.7 (Lemma 12 from [WO10])**.** *For any functions $g_N : \mathbb{R}^N \to \mathcal{Q}_N$, $\phi_N : \mathcal{Q}_N \to \mathcal{P}_N$ and $\alpha \in \mathbb{R}$, there exists a hash function $h_N : \mathcal{Q}_N \to \mathcal{K}_N$ such that*

$$\mu_N \leq \sqrt{|\mathcal{K}_N||\mathcal{P}_N| e^{-N(I(U;X)-I(U;Z)-\alpha)}} + 2\Pr((g_N(X^N), X^N, Z^N) \notin \mathcal{A}_N), \quad (5.24)$$

where $\mathcal{A}_N$ is given by

$$\mathcal{A}_N = \left\{ (u^N, x^N, z^N) : \frac{1}{N} \ln \frac{f_{X^N|U^N,Z^N}(x^N|u^N, z^N)}{f_{X^N|Z^N}(x^N|z^N)} \geq I(U;X|Z) - \alpha \right\}. \quad (5.25)$$

□

Since the rate of the public message P_N is $R_{1,N} - R_{2,N}$, the term under the square root sign goes to zero if

$$\begin{aligned} R_K &< -R_{1,N} + R_{2,N} + I(U;X) - I(U;Z) - \alpha \\ &= I(U;Y) - I(U;Z) - \left[\frac{\sigma_X^2}{\sigma_X^2 + Q} - \frac{1}{2} \ln \frac{\sigma_X^2 + Q}{Q} \right]^+ - \delta_N - \alpha, \end{aligned} \quad (5.26)$$

where we have used that $R_{2,N} = I(U;Y) - \epsilon_1 - \delta_N$, and that $I(U;X) = \frac{1}{2} \ln \frac{\sigma_X^2 + Q}{Q}$.

When bounding the second term in (5.24) the following lemma is helpful:

Lemma 5.8. *The probability* $\Pr((g_N(X^N), X^N, Z^N) \notin \mathcal{A}_N)$*, appearing in Lemma 5.7 can be bounded from above by*

$$\Pr((g_N(X^N), X^N, Z^N) \notin \mathcal{A}_N) \leq e^{-NA} + \Pr\left(\frac{|X^N - c\hat{U}^N|}{N} > \frac{\sigma_X^2 Q}{\sigma_X^2 + Q}\right), \quad (5.27)$$

where $A > 0$ *is a constant.* □

Proof. See the appendix. ■

Note that $R_1 > \max\left\{\frac{1}{2}\ln\frac{\sigma_X^2+Q}{Q}, \frac{\sigma_X^2}{\sigma_X^2+Q}\right\}$, and $b > \frac{3.5R_1}{R_1 - \sigma_X^2/(\sigma_X^2+Q)}$. Thus by Theorem 2.15, the second term in (5.27) decays exponentially in N.

In toto, since μ_N decays exponentially in N, Lemma 5.6 implies that $\limsup_{N\to\infty} I(K_A(X^N); Z^N f_N(X^N)) < \epsilon$. Finally we note that we can use Markov's inequality in the same way as in the proof of Theorem 5.2 to show that there exists at least one sequence of codes that satisfies (5.18) - (5.20) simultaneously, provided that N is large enough. This concludes the proof. ■

5.A Proof of Lemma 5.8

Proof. The conditional pdfs are [Lap09]:

$$f_{X^N|U^NZ^N}(x^N|u^N,z^N) = \frac{1}{(2\pi\Sigma_{X|UZ})^{N/2}} e^{-\frac{1}{2\Sigma_{X|UZ}}|x^N - E[X^N|U^N=u^N,Z^N=z^N]|^2} \tag{5.28}$$

$$= \frac{1}{(2\pi\Sigma_{X|UZ})^{N/2}} e^{-\frac{1}{2\Sigma_{X|UZ}}|x^N - \Sigma_{X|UZ}(\frac{u^N}{Q}+\frac{bz^N}{N_Z})|^2}, \tag{5.29}$$

and

$$f_{X^N|Z^N}(x^N|z^N) = \frac{1}{(2\pi\Sigma_{X|Z})^{N/2}} e^{-\frac{1}{2\Sigma_{X|Z}}|x^N - E[X^N|Z^N=z^N]|^2} \tag{5.30}$$

$$= \frac{1}{(2\pi\Sigma_{X|Z})^{N/2}} e^{-\frac{1}{2\Sigma_{X|Z}}|x^N - \Sigma_{X|Z}\frac{bz^N}{N_Z}|^2}, \tag{5.31}$$

with

$$\Sigma_{X|UZ} = \left(\frac{b^2}{N_Z} + \frac{1}{Q} + \frac{1}{\sigma_X^2}\right)^{-1} \tag{5.32}$$

$$\Sigma_{X|Z} = \frac{\sigma_X^2 N_Z}{b^2\sigma_X^2 + N_Z}. \tag{5.33}$$

We can then write $\frac{1}{N}\ln\frac{f_{X^N|U^N,Z^N}(x^N|u^N,z^N)}{f_{X^N|Z^N}(x^N|z^N)}$ as

$$\frac{1}{2}\ln\left(\frac{\Sigma_{X|Z}}{\Sigma_{X|UZ}}\right) + \frac{1}{2N}\left(\frac{|x^N - \Sigma_{X|Z}\frac{bz^N}{N_Z}|^2}{\Sigma_{X|Z}} - \frac{|x^N - \Sigma_{X|UZ}(\frac{u^N}{Q}+\frac{bz^N}{N_Z})|^2}{\Sigma_{X|UZ}}\right), \tag{5.34}$$

and note that $\frac{1}{2}\ln\left(\frac{\Sigma_{X|Z}}{\Sigma_{X|UZ}}\right) = I(U;X|Z)$. We rewrite the second term in (5.34) as

$$\begin{aligned}
\frac{1}{2N}\frac{|x^N - \Sigma_{X|Z}\frac{bz^N}{N_Z}|^2}{\Sigma_{X|Z}} &= \frac{1}{2N}\frac{\left|x^N - \Sigma_{X|Z}\frac{b}{N_Z}(bx^N + w_Z^N)\right|^2}{\Sigma_{X|Z}} \\
&= \frac{1}{2N}\frac{\left|x^N\frac{N_Z}{b^2N_Z+\sigma_X^2} - \frac{b\Sigma_{X|Z}}{N_Z}w_Z^N\right|^2}{\Sigma_{X|Z}} \\
&= \frac{1}{2N}\frac{\left|A_1x^N - A_2w_Z^N\right|^2}{\Sigma_{X|Z}},
\end{aligned}$$

with $A_1 = \frac{N_Z}{b^2N_Z+\sigma_X^2}$, and $A_2 = \frac{b\Sigma_{X|Z}}{N_Z}$. For the last term in (5.34) we have

$$\frac{1}{2N}\frac{|x^N - \Sigma_{X|UZ}(\frac{u^N}{Q}+\frac{bz^N}{N_Z})|^2}{\Sigma_{X|UZ}} = \frac{1}{2N}\frac{|x^N - \Sigma_{X|UZ}(\frac{u^N}{Q}+\frac{b(z^N-bx^N+bx^N)}{N_Z})|^2}{\Sigma_{X|UZ}}$$

$$
\begin{aligned}
&= \frac{1}{2N\Sigma_{X|UZ}} \left| x^N \left(1 - \Sigma_{X|UZ}\frac{b^2}{N_Z}\right) - \Sigma_{X|UZ}\left(\frac{u^N}{Q} + \frac{b}{N_Z}w_Z^N\right)\right|^2 \\
&= \frac{1}{2N\Sigma_{X|UZ}} \left| (x^N - cu^N + cu^N) \left(1 - \Sigma_{X|UZ}\frac{b^2}{N_Z}\right) - \Sigma_{X|UZ}\left(\frac{u^N}{Q} + \frac{b}{N_Z}w_Z^N\right)\right|^2 \\
&\overset{(a)}{=} \frac{1}{2N\Sigma_{X|UZ}} \left| (x^N - cu^N) \left(1 - \Sigma_{X|UZ}\frac{b^2}{N_Z}\right) - \Sigma_{X|UZ}\frac{b}{N_Z}w_Z^N\right|^2 \\
&= \frac{1}{2N\Sigma_{X|UZ}} \left| A_3(x^N - cu^N) - A_4 w_Z^N\right|^2,
\end{aligned}
$$

with $A_3 = \left(1 - \Sigma_{X|UZ}\frac{b^2}{N_Z}\right)$, and $A_4 = \Sigma_{X|UZ}\frac{b}{N_Z}$, and where in (a) we use that $c\left(1 - \Sigma_{X|UZ}\frac{b^2}{N_Z}\right) - \frac{\Sigma_{X|UZ}}{Q} = 0$.

Now let $\mathcal{E}_1$ and $\mathcal{E}_2$ denote the events $\left(\frac{\left|X^N - cU^N\right|^2}{N} \geq D\right)$ and $\left(|W_Z^N|^2 > (N_Z + \epsilon_2)\right)$ respectively, where $D = \frac{\sigma_X^2 Q}{\sigma_X^2 + Q}$ and $\epsilon_2 > 0$. We can now bound $\Pr((g_N(X^N), X^N, Z^N) \neq \mathcal{A}_N)$ from above by

$$
\begin{aligned}
&\Pr\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{\left|A_3(X^N - c\hat{U}^N) - A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) = \\
&\Pr\left(\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{\left|A_3(X^N - c\hat{U}^N) - A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) \cap \mathcal{E}_1^C\right) + \\
&\Pr\left(\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{\left|A_3(X^N - c\hat{U}^N) - A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) \cap \mathcal{E}_1\right) \leq \\
&\Pr\left(\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{A_3^2ND + \left|A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) \cap \mathcal{E}_1^C\right) + \Pr\left(\mathcal{E}_1\right) \leq \\
&\Pr\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{A_3^2ND + \left|A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) + \Pr\left(\mathcal{E}_1\right) \leq \\
&\Pr\left(\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{A_3^2ND + \left|A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) \cap \mathcal{E}_2^C\right) + \\
&\Pr\left(\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{A_3^2ND + \left|A_4W_Z^N\right|^2}{\Sigma_{X|UZ}} < -2N\alpha\right) \cap \mathcal{E}_2\right) + \Pr\left(\mathcal{E}_1\right) \leq \\
&\Pr\left(\left(\frac{\left|A_1X^N - A_2W_Z^N\right|^2}{\Sigma_{X|Z}} - \frac{A_3^2ND + A_4^2N(N_Z + \epsilon_2)}{\Sigma_{X|UZ}} < -2N\alpha\right) \cap \mathcal{E}_2^C\right) +
\end{aligned}
$$

$$\Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_1) \leq$$
$$\Pr\left(\frac{\left|A_1 X^N - A_2 W_Z^N\right|^2}{\Sigma_{X|Z}} < N\left(1 - 2\alpha + \epsilon_2 \frac{A_4^2}{2\Sigma_{X|UZ}}\right)\right) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_1) \leq \tag{5.35}$$
$$e^{-N(A_5+A_6)} + \Pr(\mathcal{E}_1)$$

The last step follows from the Cramér-Chernoff type bound in Lemma 5.9 given below, and the fact that we can choose ϵ_2 such that $\epsilon_2 \frac{A_4^2}{2\Sigma_{X|UZ}} < \alpha$. Thus there exists two positive constants A_5 and A_6 such that the two first probabilities in (5.35) are bounded from above by $e^{-N(A_5+A_6)}$. ■

Lemma 5.9 ((B1) and (B2) from [Pol94])**.** *Let $X_1, X_2, \ldots, X_N$ be i.i.d zero mean Gaussian random variables with variance σ^2. Then*

$$\Pr\left(\frac{1}{N}\sum_{i=1}^{N} X_i^2 \leq \rho\right) \leq \begin{cases} e^{-N\left(\frac{\rho}{2\sigma^2} - \frac{1}{2}\ln\frac{\rho e}{\sigma^2}\right)} & \text{if } \rho \leq \sigma^2 \\ 1 & \text{otherwise} \end{cases}$$
$$\Pr\left(\frac{1}{N}\sum_{i=1}^{N} X_i^2 \geq \rho\right) \leq \begin{cases} e^{-N\left(\frac{\rho}{2\sigma^2} - \frac{1}{2}\ln\frac{\rho e}{\sigma^2}\right)} & \text{if } \rho \geq \sigma^2 \\ 1 & \text{otherwise} \end{cases}$$

□

Chapter 6

Non-Coherent Secret Key Agreement

In this chapter we study two variations of the secret key agreement problem from Section 2.3. In particular we consider secret key agreement over a non-coherent block-fading channel. In Section 6.1 we assume that each user has multiple antennas, and we suggest a two phase scheme based on training and randomness sharing. We evaluate the performance of this scheme in the high SNR regime in terms of its achievable secure degrees of freedom (s.d.o.f). In Section 6.2 we assume that each user has a single antenna. We constrain one of the users to only transmit fixed training symbols and show that a bursty training scheme based on opportunistic secret message transmission is optimal in the low SNR regime.

6.1 Multiple Antenna Channel Model

We consider a variation of the channel type model from Section 2.3, depicted in Figure 6.1. Alice and Bob communicate over a two way block-fading (MIMO) channel, and in addition they can also use a public discussion channel with unlimited capacity. We pose the constraint that Alice and Bob cannot transmit and receive at the same time. We assume that Alice, Bob, and Eve have n_A, n_B, and n_E antennas, respectively, with $n_A \geq n_B$ without loss of generality. If Alice uses the channel, the received signals at Bob and Eve at time i are given by

$$\begin{aligned}\mathbf{Y_B}(i) &= \mathbf{H}(i)\mathbf{X_A}(i) + \mathbf{V_B}(i),\\ \mathbf{Y_E}(i) &= \mathbf{G_{AE}}(i)\mathbf{X_A}(i) + \mathbf{V_{AE}}(i),\end{aligned}$$

and if Bob uses the channel, the received signals at Alice and Eve are given by

$$\begin{aligned}\mathbf{Y_A}(i) &= \mathbf{H}^{\dagger}(i)\mathbf{X_B}(i) + \mathbf{V_A}(i),\\ \mathbf{Y_E}(i) &= \mathbf{G_{BE}}(i)\mathbf{X_B}(i) + \mathbf{V_{BE}}(i).\end{aligned}$$

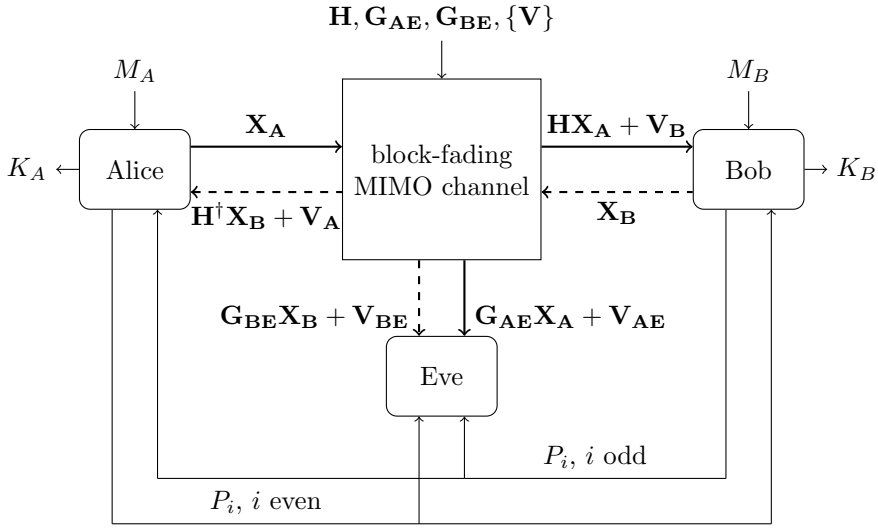


Figure 6.1: Secret key agreement over a block-fading MIMO channel.

Here $\mathbf{X_A}(i) \in \mathbb{C}^{n_A}$, or $\mathbf{X_B}(i) \in \mathbb{C}^{n_B}$ is the transmitted signal, $\mathbf{Y_A}(i) \in \mathbb{C}^{n_A}$, $\mathbf{Y_B}(i) \in \mathbb{C}^{n_B}$, and $\mathbf{Y_E}(i) \in \mathbb{C}^{n_E}$ are Alice's, Bob's, and Eve's received signals, respectively. $\mathbf{H}(i)$ represents the channel matrix between Alice and Bob, and $\mathbf{G_{AE}}(i)$ and $\mathbf{G_{BE}}(i)$ are the channel matrices between Alice and Eve, and Bob and Eve, respectively. The noise terms $\mathbf{V_A}(i) \sim \mathcal{CN}(0, \mathbf{I}_{n_A}), \mathbf{V_B}(i) \sim \mathcal{CN}(0, \mathbf{I}_{n_B})$, and $\mathbf{V_{AE}}(i), \mathbf{V_{BE}}(i) \sim \mathcal{CN}(0, \mathbf{I}_{n_E})$ are i.i.d. and independent of each other and all other variables. We assume that the entries of $\mathbf{H}(i)$, $\mathbf{G_{AE}}(i)$, and $\mathbf{G_{BE}}(i)$ are distributed as $\mathcal{CN}(0, 1)$, independent of each other, and stay fixed for T channel uses. After a block of T channel uses a new set of channel gains $\mathbf{H}(i)$, $\mathbf{G_{AE}}(i)$, and $\mathbf{G_{BE}}(i)$ are generated, independent of the gains in all previous blocks. We further assume a short-term average power constraint on the input symbols

$$\mathbb{E}[\mathbf{X_A}^\dagger(i)\mathbf{X_A}(i)] \leq \text{SNR}, \ \mathbb{E}[\mathbf{X_B}^\dagger(i)\mathbf{X_B}(i)] \leq \text{SNR}. \tag{6.1}$$

Alice and Bob also have access to two independent random variables M_A and M_B respectively.

Alice's input to the fading channel at time i is a deterministic function of her random source M_A, the communication P^{i-1} over the public channel, and her received signals $\mathbf{Y_A}^{i-1}$ up to that point. Bob generates his input to the fading channel in the same way. After the ith use of the broadcast channel Alice generates a public message $P_{i,1} = f_{i,1}(M_A, P^{i-1}, \mathbf{Y_A}^{i-1})$. Bob then generates a public message $P_{i,2} = g_{i,2}(M_B, P^{i-1}, P_{i,1}, \mathbf{Y_B}^{i-1})$. This message exchange takes place over q rounds, after which either Alice or Bob generates a new input X_{i+1} to the fading channel. After N uses of the fading channel and a final round of exchanges of public

messages, Alice and Bob generate their respective keys K_A and K_B based on all their observations. As in Section 2.3, we say that a secret key rate R is achievable if $\forall \epsilon > 0$ there exists a sequence of key agreement schemes that satisfies

$$\limsup_{N\to\infty} \Pr(K_A \neq K_B) < \epsilon, \tag{6.2}$$

$$\liminf_{N\to\infty} \frac{1}{N} H(K_A) > R - \epsilon, \tag{6.3}$$

$$\liminf_{N\to\infty} \max\left(\frac{1}{N} I(K_A; Z^N P^q), \frac{1}{N} I(K_B; Z^N P^q)\right) < \epsilon. \tag{6.4}$$

6.1.1 Achievable Scheme

We consider a three-phase scheme based on transmitting known training symbols between Alice and Bob in the first two phases, and randomness sharing in the third phase. By randomness sharing we mean that Alice generates random symbols $\mathbf{X_A}$ and transmits them over the channel. Bob then quantizes his observations from the training phase and the source emulation phase, and sends enough information over the public channel in order for Alice and Bob to agree on a secret key K.

In phase one Alice transmits known training symbols between time 1 and $M_A < n_A$. Alice's transmitted symbols at antenna j at time i are given by

$$\mathbf{X_A}_{,j}(i) = \delta_{i,j}\sqrt{\mathrm{SNR}},$$

where

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

In phase two Bob transmits known training symbols between time $M_A + 1$ and $M_A + M_B$, with $M_B < n_B$. Bob's transmitted symbols at antenna j at time i are given by

$$\mathbf{X_B}_{,j}(i) = \delta_{(i-M_A),j}\sqrt{\mathrm{SNR}}.$$

As in [ZT02], Alice and Bob can estimate parts of $\mathbf{H}$ from their received signals during the training phases using component-wise minimum mean square error (MMSE) estimation. Let

$$\mathbf{H} = \begin{bmatrix} \mathbf{H_1} & \mathbf{H_2} \\ \mathbf{H_3} & \mathbf{H_4} \end{bmatrix},$$

where $\mathbf{H_1} \in \mathbb{C}^{M_B \times M_A}$, $\mathbf{H_2} \in \mathbb{C}^{M_B \times (n_A - M_A)}$, $\mathbf{H_3} \in \mathbb{C}^{(n_B - M_B) \times M_A}$, and $\mathbf{H_4} \in \mathbb{C}^{(n_B - M_B) \times (n_A - M_A)}$. Bob's received signal at time j at antenna i is

$$y_{B,i}(j) = H_{i,j}\sqrt{\mathrm{SNR}} + v_{B,i}(j). \tag{6.5}$$

The MMSE estimate $\hat{H}_{B,i,j}$ of $H_{i,j}$ is a circularly symmetric Gaussian random variable with variance $\mathrm{SNR}/(\mathrm{SNR}+1)$ given by

$$\hat{H}_{B,i,j} = \frac{\sqrt{\mathrm{SNR}}}{\mathrm{SNR}+1} y_{B,i}(j). \tag{6.6}$$

The estimation error $e_{B,i,j} = H_{i,j} - \hat{H}_{B,i,j}$ is distributed as $\mathcal{CN}(0, 1/(\mathrm{SNR}+1))$. The estimation errors $e_{B,i,j}$ and the estimates $\hat{H}_{B,i,j}$ are all independent of each other. In this way Bob obtains an MMSE estimate $\hat{\mathbf{H}}_{\mathbf{B}} = \left[\hat{\mathbf{H}}_{\mathbf{B},\mathbf{1}}^T \ \hat{\mathbf{H}}_{\mathbf{B},\mathbf{3}}^T\right]^T$ of $\left[\mathbf{H_1}^T \ \mathbf{H_3}^T\right]^T$, and in the same way Alice can form an MMSE estimate $\hat{\mathbf{H}}_{\mathbf{A}} = \left[\hat{\mathbf{H}}_{\mathbf{A},\mathbf{1}} \hat{\mathbf{H}}_{\mathbf{A},\mathbf{2}}\right]$ of $[\mathbf{H}_{\mathbf{A},\mathbf{1}} \mathbf{H}_{\mathbf{A},\mathbf{2}}]$ from her observations in the second phase. Eve can also estimate the channel matrix $\mathbf{G_{AE}}$ from Alice's transmission in the first phase. We assume that Eve's estimate of $\mathbf{G_{AE}}$ is perfect, since this gives a lower bound on the achievable secret key rate.

In phase three Alice uses the first M_A antennas to transmit i.i.d. vectors $\mathbf{X_A}(i) \sim \mathcal{CN}(0, \mathrm{SNR}/M_A \mathbf{I}_{M_A})$, between time $M_A + M_B + 1$ and T. To simplify the notation we will refer to these transmitted signals $\{\mathbf{X_A}(i)\}_{i=M_A+M_B+1}^{T}$ simply as $\mathbf{X_A}$, and to the received signals at Bob and Eve in the third phase as $\mathbf{Y_B}$ and $\mathbf{Y_E}$, respectively. We choose $\mathbf{X_A}$ to be independent of Alice's estimate $\hat{\mathbf{H}}_{\mathbf{A}}$ in order to simplify the analysis of the achievable rate. We have the following result:

Theorem 6.1. *The secret key rate achieved by the described training based scheme is bounded from below by*

$$\frac{M_A M_B}{T} \log\left(1 + \frac{\mathrm{SNR}^2}{2\mathrm{SNR}+1}\right) + \frac{T - M_A - M_B}{T} R_{SE},$$

where R_{SE} is given by

$$\mathbb{E}\left[\log \frac{\det\left(\mathbf{I}_{M_A} + \mathrm{SNR}/M_A \mathbf{G_{AE}}^\dagger \mathbf{G_{AE}} + \mathrm{SNR}/M_A \mathbf{H}^\dagger \mathbf{H}\right)}{\det\left(\mathbf{I}_{M_A} + \mathrm{SNR}/M_A \mathbf{G_{AE}}^\dagger \mathbf{G_{AE}}\right)}\right] - M_B \log \tilde{\sigma}_{A,B}^2 - (n_B - M_B)\log(\tilde{\sigma}_B^2),$$

with $\tilde{\sigma}_{A,B}^2 = \frac{\mathrm{SNR}}{M_A(2\mathrm{SNR}+1)} + 1$, and $\tilde{\sigma}_B^2 = \frac{\mathrm{SNR}}{M_A(\mathrm{SNR}+1)} + 1$.

Proof. The random variables involved in our scheme satisfy the following Markov chain:

$$(\mathbf{Y_E}, \mathbf{G_{AE}}) \leftrightarrow (\mathbf{X_A}, \hat{\mathbf{H}}_{\mathbf{A}}) \leftrightarrow (\mathbf{Y_B}, \mathbf{H}) \leftrightarrow (\mathbf{Y_B}, \hat{\mathbf{H}}_{\mathbf{B}}). \tag{6.7}$$

If we consider our scheme as a Source-Type Model with Wiretapper, and code over a large number of coherence blocks, the achievable secret key rate is bounded from below by [WBS09].

$$R_- = \frac{1}{T} I(\mathbf{X_A}, \hat{\mathbf{H}}_\mathbf{A}; \mathbf{Y_B}, \hat{\mathbf{H}}_\mathbf{B} | \mathbf{Y_E}, \mathbf{G_{AE}}).$$

This rate is achieved by quantizing Bob's observations $(\mathbf{Y_B}, \hat{\mathbf{H}}_\mathbf{B})$ into a quantization codebook generated by auxiliary random variables $(\mathbf{U_Y}, \mathbf{U_H})$ and using Wyner-Ziv coding to transmit the indices over the public channel. Alice can then recover $(\mathbf{U_Y}, \mathbf{U_H})$, and a secret key can be generated. The scheme uses the same procedure as detailed in Chapter 5. By making the quantization fine enough we can achieve the rate R_-.

We can bound this rate from below using (6.7) as follows:

$$\begin{aligned}
TR_- =& I(\mathbf{X_A}, \hat{\mathbf{H}}_\mathbf{A}; \mathbf{Y_B}, \hat{\mathbf{H}}_\mathbf{B} | \mathbf{Y_E}, \mathbf{G_{AE}}) \\
=& I(\mathbf{X_A}, \hat{\mathbf{H}}_\mathbf{A}; \mathbf{Y_B}, \hat{\mathbf{H}}_\mathbf{B}) - I(\mathbf{Y_B}, \hat{\mathbf{H}}_\mathbf{B}; \mathbf{Y_E}, \mathbf{G_{AE}}) \\
\geq& I(\mathbf{X_A}, \hat{\mathbf{H}}_\mathbf{A}; \mathbf{Y_B}, \hat{\mathbf{H}}_\mathbf{B}) - I(\mathbf{Y_B}, \mathbf{H}; \mathbf{Y_E}, \mathbf{G_{AE}}) \\
\geq& I(\hat{\mathbf{H}}_\mathbf{A}; \hat{\mathbf{H}}_\mathbf{B}) + I(\mathbf{X_A}; \mathbf{Y_B} | \hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}) - I(\mathbf{Y_B}; \mathbf{Y_E} | \mathbf{H}, \mathbf{G_{AE}}) \\
=& I(\hat{\mathbf{H}}_\mathbf{A}; \hat{\mathbf{H}}_\mathbf{B}) + h(\mathbf{Y_B} | \hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}) - h(\mathbf{Y_B} | \mathbf{H}, \mathbf{G_{AE}}) - h(\mathbf{Y_B} | \hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}, \mathbf{X_A}) \\
&+ h(\mathbf{Y_B} | \mathbf{H}, \mathbf{G_{AE}} \mathbf{Y_E}) \\
\geq& I(\hat{\mathbf{H}}_\mathbf{A}; \hat{\mathbf{H}}_\mathbf{B}) - h(\mathbf{Y_B} | \hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}, \mathbf{X_A}) + h(\mathbf{Y_B} | \mathbf{H}, \mathbf{G_{AE}} \mathbf{Y_E}),
\end{aligned} \tag{6.8}$$

where we have used that $\mathbf{X_A}$ and $\hat{\mathbf{H}}_\mathbf{A}$ are independent in the second inequality. We see that we have two contributions to the secret key rate. $I(\hat{\mathbf{H}}_\mathbf{A}; \hat{\mathbf{H}}_\mathbf{B})$ comes from the training phases, and $h(\mathbf{Y_B} | \mathbf{H}, \mathbf{G_{AE}} \mathbf{Y_E}) - h(\mathbf{Y_B} | \hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}, \mathbf{X_A})$ is the rate from the source emulation phase.

To calculate the first contribution, we note that $\mathbf{H_1}$ is the only part of $\mathbf{H}$ for which both Alice and Bob have estimates. Using (6.5) and (6.6), we get

$$\begin{aligned}
I(\hat{\mathbf{H}}_\mathbf{A}; \hat{\mathbf{H}}_\mathbf{B}) &= M_A M_B I(Y_{A,i}(j); Y_{B,i}(j)) \\
&= M_A M_B \log \left(1 + \frac{\mathrm{SNR}^2}{2\mathrm{SNR} + 1} \right).
\end{aligned} \tag{6.9}$$

We now find a lower bound on the contribution from the source emulation phase $h(\mathbf{Y_B} | \mathbf{H}, \mathbf{G_{AE}} \mathbf{Y_E}) - h(\mathbf{Y_B} | \hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}, \mathbf{X_A})$ using the following Lemma from [WBS09]:

Lemma 6.2 (Lemma 1 from [WBS09])**.** *Let* $\mathbf{U}$ *and* $\mathbf{V}$ *be two jointly distributed complex random vectors of dimensions* $m_\mathbf{U}$ *and* $m_\mathbf{V}$*, respectively. Let* $\mathbf{K_U}$*,* $\mathbf{K_V}$*, and* $\mathbf{K_{UV}}$ *be the covariance of* $\mathbf{U}$*, covariance of* $\mathbf{V}$ *and cross-covariance of* $\mathbf{U}$ *and* $\mathbf{V}$*, respectively. If* $\mathbf{K_V}$ *is invertible, then*

$$h(\mathbf{U}|\mathbf{V}) \leq \log \det \left(\mathbf{K_U} - \mathbf{K_{UV}} \mathbf{K_V}^{-1} \mathbf{K_{VU}} \right) + m_\mathbf{U} \log(\pi e),$$

with equality if $[\mathbf{U}^T \ \mathbf{V}^T]^T$ *is a circularly symmetric complex Gaussian random vector.* □

We first create a new MMSE estimate $\hat{\mathbf{H}}$ of $\mathbf{H}$ using both $\hat{\mathbf{H}}_\mathbf{A}$ and $\hat{\mathbf{H}}_\mathbf{B}$. Note that since only the first M_A entries of $\mathbf{X_A}$ are nonzero, we only need an estimate of $\left[\mathbf{H_1}^T \ \mathbf{H_3}^T\right]^T$. If Bob assumes that $\hat{\mathbf{H}}$ is the true value of $\mathbf{H}$, his received signal is

$$\mathbf{Y_B}(t) = \hat{\mathbf{H}}\mathbf{X_A}(t) + \mathcal{E}\mathbf{X_A}(t) + \mathbf{V_B}(t),$$

where $\mathcal{E} = \left[\mathcal{E}_\mathbf{1}^T \ \mathcal{E}_\mathbf{3}^T\right]^T = \left[\mathbf{H_1}^T - \hat{\mathbf{H}}_\mathbf{1}^T \ \mathbf{H_3}^T - \hat{\mathbf{H}}_\mathbf{3}^T\right]^T$.

The estimate of $\mathbf{H_1}$ is based on $\hat{\mathbf{H}}_\mathbf{A}$ and $\hat{\mathbf{H}}_\mathbf{B}$ and the entries have MSE $1/(2\text{SNR}+1)$, while the estimate of $\mathbf{H_3}$ is based only on $\hat{\mathbf{H}}_\mathbf{B}$ and the entries have MSE $1/(\text{SNR}+1)$. Thus the covariance of $\mathcal{E}\mathbf{X_A}(t) + \mathbf{V_B}(t)$ is

$$\begin{bmatrix} \tilde{\sigma}_{A,B}^2 \mathbf{I}_{M_B} & 0 \\ 0 & \tilde{\sigma}_B^2 \mathbf{I}_{n_B - M_B} \end{bmatrix},$$

where

$$\tilde{\sigma}_{A,B}^2 = \frac{\text{SNR}}{M_A(2\text{SNR}+1)} + 1, \tag{6.10}$$

$$\tilde{\sigma}_B^2 = \frac{\text{SNR}}{M_A(\text{SNR}+1)} + 1. \tag{6.11}$$

We now bound $h(\mathbf{Y_B}|\hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}, \mathbf{X_A})$ from above as follows:

$$\begin{aligned} h(\mathbf{Y_B}|\hat{\mathbf{H}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B}, \mathbf{X_A}) &= \mathbb{E}_{\hat{\mathbf{H}}}\left[h(\mathbf{Y_B}|\hat{\mathbf{H}}_\mathbf{A} = \hat{\mathbf{h}}_\mathbf{A}, \hat{\mathbf{H}}_\mathbf{B} = \hat{\mathbf{h}}_\mathbf{B}, \mathbf{X_A})\right] + n_B \log(\pi e) \\ &\leq \mathbb{E}_{\hat{\mathbf{H}}}\left[\log\det(K_{\mathbf{Y}|\hat{\mathbf{H}}} - K_{\mathbf{YX}|\hat{\mathbf{H}}} K_{\mathbf{X}|\hat{\mathbf{H}}}^{-1} K_{\mathbf{XY}|\hat{\mathbf{H}}})\right] + n_B \log(\pi e) \\ &= \log\det K_{\mathcal{E}\mathbf{X_A}(t)+\mathbf{V_B}(t)} + n_B \log(\pi e) \\ &= M_B \log(\pi e \tilde{\sigma}_{A,B}^2) + (n_B - M_B)\log(\pi e \tilde{\sigma}_B^2), \end{aligned} \tag{6.12}$$

where we have used Lemma 6.2 in the first inequality.

$h(\mathbf{Y_B}|\mathbf{H}, \mathbf{G_{AE}}\mathbf{Y_E})$ is given by [WBS09] as

$$\mathbb{E}\left[\log \frac{\det\left(\mathbf{I}_{M_A} + \text{SNR}/M_A \mathbf{G_{AE}}^\dagger \mathbf{G_{AE}} + \text{SNR}/M_A \mathbf{H}^\dagger \mathbf{H}\right)}{\det\left(\mathbf{I}_{M_A} + \text{SNR}/M_A \mathbf{G_{AE}}^\dagger \mathbf{G_{AE}}\right)}\right] + n_B \log(\pi e). \tag{6.13}$$

Combining (6.8) - (6.13) gives the desired result. ■

6.1.2 High SNR Regime

In [WBS09] the secret key capacity $C_K(\mathrm{SNR})$ for the fast fading coherent MIMO wiretap channel with M_A transmit antennas was found to be

$$\mathbb{E}\left[\log \frac{\det\left(\mathbf{I}_{M_A} + \mathrm{SNR}/M_A \mathbf{G_{AE}}^\dagger \mathbf{G_{AE}} + \mathrm{SNR}/M_A \mathbf{H}^\dagger \mathbf{H}\right)}{\det\left(\mathbf{I}_{n_A} + \mathrm{SNR}/M_A \mathbf{G_{AE}}^\dagger \mathbf{G_{AE}}\right)}\right].$$

Note that R_{SE} differs from $C_K(\mathrm{SNR})$ only by the negative terms $M_B \log(\tilde{\sigma}^2_{A,B}) + (n_B - M_B)\log(\tilde{\sigma}^2_B)$, due to not knowing the channel perfectly at Bob. From (6.10) and (6.11), we see that these terms scale with SNR as $\Theta(1)$. Further, from [WBS09, Corollary 1], if $M_A > n_E$, we have that $\lim_{\mathrm{SNR}\to\infty} C_K(\mathrm{SNR})/C_\infty(\mathrm{SNR}) = 1$, where $C_\infty(\mathrm{SNR})$ is defined as

$$\mathbb{E}\left[\log\det(\mathbf{I}_{n_B} + \frac{\mathrm{SNR}}{M_A}\mathbf{H}\left[\mathbf{I}_{M_A} - \mathbf{G_{AE}}^\dagger(\mathbf{G_{AE}}\mathbf{G_{AE}}^\dagger)^{-1}\mathbf{G_{AE}}\right]\mathbf{H}^\dagger)\right],$$

and if $M_A \leq n_E$, $C_K(\mathrm{SNR})$ scales with SNR as $\Theta(1)$. As in [WBS09], we can interpret $\mathbf{I}_{M_A} - \mathbf{G_{AE}}^\dagger(\mathbf{G_{AE}}\mathbf{G_{AE}}^\dagger)^{-1}\mathbf{G_{AE}}$ as a projection matrix onto the null space of $\mathbf{G_{AE}}$. Thus, at high SNR, the number of s.d.o.f. per channel use from the third phase is given by $\min(M_A - n_E, n_B)(T - M_A - M_B)/T$. Further, from (6.9) we see that the number of s.d.o.f. per channel use from the first two phases is given by $M_A M_B/T$. Combining these results we get:

Theorem 6.3. *The number of s.d.o.f. per channel use at high SNR is given by*

$$\frac{[\min(M_A - n_E, n_B)]^+ (T - M_A - M_B) + M_A M_B}{T}. \tag{6.14}$$

Proof. See above. ■

We can use Theorem 6.3 to find the optimal M_A and M_B.

Corollary 6.4. *The optimal number of time slots $M_A^\star$ and $M_B^\star$ to use for training in the high SNR regime is given by*

$$M_A^\star = \min(n_A, \max(T/2, T - n_B)),$$
$$M_B^\star = \min(T - M_A^\star, n_B).$$

□

Proof. See Appendix. ■

Remark 6.5. We note two facts from Corollary 6.4. First, $M_A^\star$ and $M_B^\star$ do not depend on n_E, and second, it is either optimal to train all antennas (if $n_A+n_B < T$), or to only use training and no source emulation (if $n_A + n_B \geq T$). In the latter case the number of s.d.o.f. does not depend on n_E. ◊

We can also use Corollary 6.4 to find the optimal number of antennas n_A and n_B for a given coherence time T.

Corollary 6.6. *Increasing the number of transmit and receive antennas over*

$$n_A^\star = n_B^\star = T/2$$

does not increase the degrees of freedom. □

Proof. For given T and n_B the optimal choice of n_A is $n_A^\star = \max(T/2, T - n_B)$. We now have

$$M_B^\star = \min(n_B, T - \max(T/2, T - n_B)) = \min(T/2, n_B),$$

which implies that $n_B^\star = T/2$, and thus $n_A^\star = T/2$. ■

With this choice of n_A and n_B there is only training, and we can guarantee a secret key rate of

$$R_- = \frac{T}{4} \log \left(1 + \frac{\mathrm{SNR}^2}{2\mathrm{SNR} + 1} \right),$$

regardless of the number of antennas n_E at Eve.

6.1.3 Key Agreement without a public channel

The presence of a public channel with unlimited rate is not realistic in all scenarios. Therefore we consider a scenario in which some coherence blocks are used for public discussion between Alice and Bob over the wireless channel. In the high SNR regime this does not give a loss of s.d.o.f. We have the following result:

Theorem 6.7. *At high SNR it is possible to achieve*

$$\frac{[\min(M_A - n_E, n_B)]^+ (T - M_A - M_B) + M_A M_B}{T}$$

s.d.o.f. per channel use without a separate channel for public discussion, if M_A and M_B are chosen as in Corollary 6.4.

Proof. Let $\mathbf{Y_{AT}}$ and $\mathbf{Y_{BT}}$ denote Alice's and Bob's observations in the training phase. As before, we quantize Bob's observation $(\mathbf{Y_B}, \mathbf{Y_{BT}})$ into a codebook generated by the auxiliary random variables $(\mathbf{U_Y}, \mathbf{U_H})$. From [CN00], the rate needed for public discussion and the achievable secret key rate are given by

$$R_p = I(\mathbf{Y_B}, \mathbf{Y_{BT}}; \mathbf{U_Y}, \mathbf{U_H}) - I(\mathbf{X_A}, \mathbf{Y_{AT}}; \mathbf{U_Y}, \mathbf{U_H}),$$

and

$$R_-^{np} = I(\mathbf{X_A}, \mathbf{Y_{AT}}; \mathbf{U_Y}, \mathbf{U_H} | \mathbf{Y_E}, \mathbf{G_{AE}}), \tag{6.15}$$

respectively. As in [LLP12], we let $\mathbf{U_Y} = \mathbf{Y_B} + \mathbf{W_Y}$ and $\mathbf{U_H} = \mathbf{Y_B} + \mathbf{W_H}$, where $\mathbf{W_Y} \sim \mathcal{CN}(0, \sigma_Y^2 \mathbf{I}_{M_B})$ and $\mathbf{W_H} \sim \mathcal{CN}(0, \sigma_H^2 \mathbf{I}_{M_B})$ are i.i.d. and independent of all other random variables. Note that with the optimal choice of M_A and M_B above, either $M_B = n_B$, or all time slots are used for training. In the latter case the measurements at the last $n_B - M_B$ antennas at Bob cannot be used for key agreement, so only the first M_B measurements are used.

We first analyze the rate needed for public discussion. We have

$$\begin{aligned} R_p =& I(\mathbf{Y_B}, \mathbf{Y_{BT}}; \mathbf{U_Y}, \mathbf{U_H}) - I(\mathbf{X_A}, \mathbf{Y_{AT}}; \mathbf{U_Y}, \mathbf{U_H}) \\ =& I(\mathbf{Y_B}, \mathbf{Y_{BT}}; \mathbf{U_H}) + I(\mathbf{Y_B}, \mathbf{Y_{BT}}; \mathbf{U_Y} | \mathbf{U_H}) - I(\mathbf{X_A}, \mathbf{Y_{AT}}; \mathbf{U_H}) + \\ & - I(\mathbf{X_A}, \mathbf{Y_{AT}}; \mathbf{U_Y} | \mathbf{U_H}) \\ =& I(\mathbf{Y_{BT}}; \mathbf{U_H}) - I(\mathbf{Y_{AT}}; \mathbf{U_H}) + \\ & h(\mathbf{U_Y} | \mathbf{Y_{AT}}, \mathbf{U_H}, \mathbf{X_A}) - h(\mathbf{U_Y} | \mathbf{Y_B}, \mathbf{Y_{AT}}, \mathbf{U_H}). \end{aligned} \tag{6.16}$$

The first term in (6.16) is given by $I(\mathbf{Y_{BT}}; \mathbf{U_H}) = M_A M_B \log\left(1 + \frac{\mathrm{SNR}+1}{\sigma_H^2}\right)$. For the second term we have

$$I(\mathbf{Y_{AT}}; \mathbf{U_H}) = M_A M_B \log\left(\frac{(\mathrm{SNR}+1)(\mathrm{SNR}+1+\sigma_H^2)}{2\mathrm{SNR}+1+\sigma_H^2(\mathrm{SNR}+1)}\right)$$

The third term in (6.16) can be bounded from above using the MMSE estimate of $\mathbf{H_1}$ calculated from $(\mathbf{Y_{AT}}, \mathbf{U_H})$ and Lemma 6.2:

$$h(\mathbf{U_Y} | \mathbf{Y_{AT}}, \mathbf{U_H}, \mathbf{X_A}) \leq M_B \log(\pi e \tilde{\sigma}^2_{A,U_H})$$

where $\tilde{\sigma}^2_{A,U_H} = \frac{\mathrm{SNR}(1+\sigma_H^2)}{M_A(2\mathrm{SNR}+1+\sigma_H^2(\mathrm{SNR}+1))} + 1 + \sigma_Y^2$. Finally, the last term is $h(\mathbf{U_Y} | \mathbf{Y_B}, \mathbf{Y_{AT}}, \mathbf{U_H}) = h(\mathbf{U_Y} | \mathbf{Y_B}) = M_B \log(\pi e \sigma_Y^2)$. In total we get

$$R_p = M_A M_B \log\left(1 + \frac{2\mathrm{SNR}+1}{\sigma_H^2(\mathrm{SNR}+1)}\right) + M_B \log\left(\frac{\tilde{\sigma}^2_{A,U_H}}{\sigma_Y^2}\right).$$

We see that, for fixed σ_Y^2 and σ_H^2, R_p scales as $\Theta(1)$ with SNR. Now let a fraction α of the coherence blocks at the end of the transmission be dedicated to public

discussion. This lowers the achievable secret key rate to $(1-\alpha)R_{-}^{np}$. Note that the channel gains $\mathbf{H}$ during these coherence blocks are not used as shared randomness when creating the secret key, so the information leaked to Eve about $\mathbf{H}$ during these blocks do not further lower the secret key rate. When used for communication, the capacity $C(\mathrm{SNR})$ of the channel between Bob and Alice scales with SNR as $\min(T/2, n_A, n_B)\log \mathrm{SNR}$ [ZT02]. Since R_p scales as $\Theta(1)$ with SNR, it is possible to have $\alpha C(\mathrm{SNR}) > R_p$, for any $\alpha > 0$, provided that SNR is large enough. Thus we can achieve any secret key rate below R_{-}^{np}, provided that SNR is large enough.

It remains to show that the quantized observations give the same number of s.d.o.f. as in the case with a public channel. We expand R_{-}^{np} in the same way as in (6.8) and get

$$R_{-}^{np} \geq I(\mathbf{Y_{AT}}; \mathbf{U_H}) - h(\mathbf{U_Y}|\mathbf{Y_{AT}}, \mathbf{U_H}, \mathbf{X_A}) + h(\mathbf{U_Y}|\mathbf{H}, \mathbf{G_{AE}}, \mathbf{Y_E}). \tag{6.17}$$

For fixed σ_Y^2 and σ_H^2, the first term in (6.17) gives $M_A M_B$ s.d.o.f. Using Lemma 6.2, the second term in (6.17) scales as $\Theta(1)$ with SNR, and finally the last term in (6.17) gives $(T - M_A - M_B)\left[\min(M_A - n_E, M_B)\right]^+$ s.d.o.f. The optimal choice of M_A and M_B implies that $T - M_A - M_B$ is positive only when $M_B = n_B$, so the result follows. ■

6.2 Single Antenna Channel Model in the Low SNR Regime

We consider a slightly different channel model in the single antenna case, where we allow Alice and Bob to transmit and receive at the same time. The scheme we suggest is not heavily dependent on this assumption however, and is easy to adapt to the case where a node cannot transmit and receive at the same time. The channel is given by

$$\begin{aligned} Y_A(i) &= H(i)X_B(i) + v_A(i) \\ Y_B(i) &= H(i)^\star X_A(i) + v_B(i) \\ Z_{AE}(i) &= G_{AE}(i)X_A(i) + v_{AE}(i) \\ Z_{BE}(i) &= G_{BE}(i)X_B(i) + v_{BE}(i), \end{aligned}$$

where $Y_A(i)$ and $Y_B(i)$ denote the output symbols at time $i \in \{1, \ldots, N\}$ at Alice and Bob respectively, and $\{Z_{AE}(i), Z_{BE}(i)\}$ denotes the output symbols at Eve at time i. The input symbols at Alice and Bob at time i are denoted by $X_A(i)$, and $X_B(i)$ respectively, and are required to satisfy the average power constraints

$$\frac{1}{N}\sum_{i=1}^{N} E\left[|X_A(i)|^2\right] \leq \mathrm{SNR}, \quad \frac{1}{N}\sum_{i=1}^{N} E\left[|X_B(i)|^2\right] \leq \mathrm{SNR}.$$

All input and output symbols, as well as the channel gains and the noise are complex-valued, and $H(i)^\star$ denotes the complex conjugate of $H(i)$. The channel gains are drawn from independent zero mean circularly symmetric Gaussian

$\mathcal{CN}(0,1)$ distributions every T symbols and stay constant for the next T symbols. As before, g_{AE} and g_{BE} are revealed to the eavesdropper. The additive noise variables are drawn from an i.i.d. Gaussian $\mathcal{CN}(0,1)$ distribution, and are independent of all other random variables. We will constrain the input symbols X_B at Bob to be pilot symbols. Thus they are fixed, and revealed to everyone before transmission. A secret key generating scheme, achievable key rate and secret key capacity are defined as earlier. We will also consider this channel as a wiretap channel from Alice to Bob.

6.2.1 Secrecy Capacity with Partial CSI

Our achievable scheme for the secret key agreement problem uses secret message transmission combined with training. Therefore we consider the wiretap channel problem, when Alice and Bob have partial knowledge of H. We will make use of the following result

Theorem 6.8 (Theorem 3 from[BZ10])**.** *Consider the Rayleigh fading channel*

$$Y = (G + F)X + W,$$

where G and F are independent $\mathcal{CN}(0,\beta)$ and $\mathcal{CN}(0,1-\beta)$ random variables respectively. The transmitter only knows G and the receiver knows both G and F. The capacity $C_\beta(\mathrm{SNR})$ of this channel for any fixed $\beta \in (0,1]$ satisfies

$$\lim_{\mathrm{SNR}\to 0} \frac{C_\beta(\mathrm{SNR})}{\beta \mathrm{SNR} \log\left(\frac{1}{\mathrm{SNR}}\right)} = 1.$$

We have the corresponding result for the secrecy capacity:

Theorem 6.9. *Consider the Rayleigh fading wiretap channel*

$$Y_B = (H + F)X_A + V_B,$$
$$Z = GX_A + V_E,$$

where H, F, and G are independent $\mathcal{CN}(0,\beta)$, $\mathcal{CN}(0,1-\beta)$, and $\mathcal{CN}(0,1)$ random variables respectively. Alice knows H, Bob knows $H+F$, and Eve knows G. The secrecy capacity $C_{S,\beta}(\mathrm{SNR})$ of this channel for any fixed $\beta \in (0,1]$ satisfies

$$\lim_{\mathrm{SNR}\to 0} \frac{C_{S,\beta}(\mathrm{SNR})}{\beta \mathrm{SNR} \log\left(\frac{1}{\mathrm{SNR}}\right)} = 1.$$

Proof. Since the capacity $C_\beta(\mathrm{SNR})$ of the channel between Alice and Bob is an upper bound on the secrecy capacity, we have from Theorem 6.8:

$$\lim_{\mathrm{SNR}\to 0} \frac{C_{S,\beta}(\mathrm{SNR})}{\beta \mathrm{SNR} \log\left(\frac{1}{\mathrm{SNR}}\right)} \leq \lim_{\mathrm{SNR}\to 0} \frac{C_\beta(\mathrm{SNR})}{\beta \mathrm{SNR} \log\left(\frac{1}{\mathrm{SNR}}\right)} = 1.$$

For the achievability, it was shown in [LPS08] that the secrecy rate

$$I(X_A; Y_B|H+F, H) - I(X_A; Z|G) \tag{6.18}$$

is achievable for some input distribution $f_{X_A|H}(x_A|h)$. As in [BZ10], we use an on-off power control where Alice transmits Gaussian signals with power $P(h) = \mathrm{SNR}e^\theta$ if $|h|^2 > \theta$, where $\theta = \beta \log\left(\frac{1}{\mathrm{SNR}}\right) - 2\log\log\left(\frac{1}{\mathrm{SNR}}\right)$, and zero otherwise.

Using this input distribution, the first term in (6.18) becomes $\mathbb{E}\left[\log\left(1+P(H)|H+F|^2\right)\right]$ and satisfies

$$\lim_{\mathrm{SNR}\to 0} \frac{\mathbb{E}\left[\log\left(1+P(H)|H+F|^2\right)\right]}{\beta \log\left(\frac{1}{\mathrm{SNR}}\right) \mathrm{SNR}} = 1,$$

as shown in the proof of [BZ10, Theorem 3]. For the second term in (6.18) we have, using Jensen's inequality,

$$\begin{aligned} -I(X_A; Z|G) &= -\mathbb{E}\left[\log\left(1+P(H)|G|^2\right)\right] \\ &\geq -\log\left(1+\mathbb{E}[P(H)|G|^2]\right) = -\log(1+\mathrm{SNR}). \end{aligned}$$

The result now follows, since $\log(1+\mathrm{SNR})$ goes to zero faster than $\beta \log\left(\frac{1}{\mathrm{SNR}}\right) \mathrm{SNR}$, and we have

$$1 \leq \lim_{\mathrm{SNR}\to 0} \frac{C_{S,\beta}(\mathrm{SNR})}{\beta \mathrm{SNR} \log\left(\frac{1}{\mathrm{SNR}}\right)}.$$

■

Note that by setting $\beta = 1$ we get the standard fading wiretap channel, the secrecy capacity of which was found in [GLEG08] for general SNR. At high SNR, a similar on-off scheme was shown to be optimal, and we note that in neither the high, nor the low SNR regime, knowledge of G is needed.

6.2.2 Large Coherence Time Limit

In this subsection we consider the non-coherent secret key agreement problem in the large coherence period limit. We assume that T goes to ∞ as SNR goes to 0. As in [BZ10] we show that by training periodically placed blocks almost perfectly and not transmitting anything in untrained blocks we achieve a rate $R_-(\mathrm{SNR})$ that goes to zero as $\mathrm{SNR}\log T$:

Theorem 6.10. *There exists a secret key agreement protocol with rate* $R_{-}(\mathrm{SNR})$ *that satisfies*

$$\lim_{\mathrm{SNR}\to 0} \frac{R_{-}(\mathrm{SNR})}{\mathrm{SNR}\log T} = 1$$

if $T \to \infty$ *as* $\mathrm{SNR} \to 0$*, and* $T \leq \frac{1}{\mathrm{SNR}}$*.*

Proof. One out of every $K = \frac{E^2}{T\mathrm{SNR}}$ blocks is trained with a fixed training energy E. Both Alice and Bob transmit a known pilot symbol with energy E at the first instant of each trained block, and then obtain MMSE channel estimates $\hat{H}_A$ and $\hat{H}_B$ with variance $\beta = \frac{E}{E+1}$. Alice then uses the scheme from Theorem 6.9 to transmit a secret message to Bob. Note that Bob does not have access to Alice's estimate $\hat{H}_A$, but this estimate is only used to determine when Alice is transmitting. Here Alice will instead use the public channel to make this known to Bob. The SNR available for secret message transmission in the trained blocks is

$$\frac{KT\mathrm{SNR} - E}{T-1} = \frac{E^2 - E}{T-1}.$$

Since we transmit with this rate for $T-1$ channel uses every $KT = \frac{E^2}{\mathrm{SNR}}$ channel uses, the achievable rate is

$$\frac{C_{S,\beta}(\frac{E^2-E}{T-1})(T-1)\mathrm{SNR}}{E^2}.$$

As $\mathrm{SNR} \to 0$, $T \to \infty$, and we can approximate $C_{S,\beta}(\frac{E^2-E}{T-1})$ with $\frac{E}{E+1}\frac{E^2-E}{T-1}\log\left(\frac{T-1}{E^2-E}\right)$ using Theorem 6.9. We get

$$\begin{aligned}
\lim_{\mathrm{SNR}\to 0} \frac{R_{-}(\mathrm{SNR})}{\mathrm{SNR}\log T} &= \lim_{\mathrm{SNR}\to 0} \frac{\frac{E}{E+1}\frac{E^2-E}{T-1}\log\left(\frac{T-1}{E^2-E}\right)(T-1)\mathrm{SNR}}{E^2\mathrm{SNR}\log T} \\
&= \lim_{\mathrm{SNR}\to 0} \frac{E-1}{E+1}\frac{\log(T-1) - \log(E^2-E)}{\log T} \\
&= \frac{E-1}{E+1}.
\end{aligned}$$

As E can be chosen arbitrarily large the result follows. ■

Remark 6.11. As also noted in [BZ10], if $T \geq \frac{1}{\mathrm{SNR}}$, the same strategy can be used to achieve a secret message rate of $\mathrm{SNR}\log\left(\frac{1}{\mathrm{SNR}}\right)$, which is the secrecy capacity with full CSI. ◊

We have a matching upper bound on the secret key capacity:

Theorem 6.12. *The secret key capacity* $C_K(\mathrm{SNR})$ *satisfies*

$$\lim_{\mathrm{SNR}\to 0} \frac{C_K(\mathrm{SNR})}{\mathrm{SNR}\log T} = 1$$

if $T \to \infty$ *as* $\mathrm{SNR} \to 0$*, and* $T \leq \frac{1}{\mathrm{SNR}}$.

Proof. Achievability was shown in Theorem 6.10. To find an upper bound on the capacity, we proceed in two steps. First we reveal $H(i)$ to Bob, and show that the secret key rate is bounded from above by

$$NR \leq I(M_A, Y_A^N; H^K | G^K, x_B^N) + \sum_{i=1}^{N} I(X_A(i); Y_B(i) | H(i), G_{AE}(i), Z_{AE}(i)). \tag{6.19}$$

The first term is the secret key rate available from Alice's and Bob's shared knowledge about $H(i)$. The second term corresponds to secret key agreement over a channel where Bob and Eve both know $H(i)$. We show that this term is dominant, and can be bounded using similar methods as in [BZ10].

Assume that the transmission takes place over K coherence blocks, and that $N = TK$ is the total number of channel uses. Let $Z = (Z_A, Z_B)$ and $G = (G_{AE}, G_{BE})$. Using Fano's inequality and the secrecy criterion $I(K_A; Z^N, G^K, x_B^N, P^N) < N\epsilon_N$ we get

$$\begin{aligned} NR \leq & I(K_A; K_B) - I(K_A; Z^N, G^K, x_B^N, P^N) + 2N\epsilon_N \\ \leq & I(K_A; K_B | Z^N, G^K, x_B^N, P^N) + 2N\epsilon_N. \end{aligned}$$

Suppressing the ϵ_N-term we get

$$\begin{aligned} NR \leq & I(M_A, Y_A^N, K_A; Y_B^N, H^K, K_B | Z^N, G^K, x_B^N, P^N) \\ \leq & I(M_A, Y_A^N, K_A, P^N; Y_B^N, H^K, K_B | Z^N, G^K, x_B^N) \\ \leq & I(M_A, Y_A^N; Y_B^N, H^K | Z^N, G^K, x_B^N) \\ = & I(M_A, Y_A^N; Y_B^{N-1}, H^K | Z^N, G^K, x_B^N) + \\ & I(M_A, Y_A^N; Y_B(N) | Z^N, G^K, x_B^N, Y_B^{N-1}, H^K), \end{aligned}$$

where the third inequality follows since K_A and P^N are functions of (M_A, Y_A^N) and K_B is a function of Y_B^N. Since $X_A(N)$ is a function of

(M_A, Y_A^{N-1}), and $(Y_B(N), Z_{AE}(N))$ are independent of all other random variables conditioned on $(H(N), G_{AE}(N), X_A(N))$, the second term is equal to $I(X_A(N); Y_B(N)|H(N), G_{AE}(N), Z_{AE}(N))$. For the first term we have

$$\begin{aligned}
&I(M_A, Y_A^N; Y_B^{N-1}, H^K|Z^N, G^K, x_B^N) \leq \\
&I(M_A, Y_A^N, Z(N); Y_B^{N-1}, H^K|Z^{N-1}, G^K, x_B^N) \leq \\
&I(M_A, Y_A^N; Y_B^{N-1}, H^K|Z^{N-1}, G^K, x_B^N) + \\
&I(Z(N); Y_B^{N-1}, H^K|Z^{N-1}, G^K, x_B^N, M_A, Y_A^N).
\end{aligned}$$

Since $X_A(N)$ is a function of (M_A, Y_A^{N-1}), and $(Z_{AE}(N), Z_{BE}(N))$ are independent of all other random variables conditioned on $(G_{AE}(N), G_{BE}(N))$, the second term is zero. Thus we have

$$\begin{aligned}
NR \leq &I(M_A, Y_A^N; Y_B^{N-1}, H^K|Z^{N-1}, G^K, x_B^N) + \\
&I(X_A(N); Y_B(N)|h(N), g_{AE}(N), Z_{AE}(N)).
\end{aligned}$$

Using the same argument and induction over the channel use i, we get (6.19) above. The first term in (6.19) is equal to $I(Y_A^N; H^K|x_B^N)$ since M_A and G^K are independent of the other random variables. Letting $H(k)$ be the channel gain in the kth coherence interval, we get

$$I(Y_A^N; H^K|x_B^N) \leq \sum_{k=1}^{K} I(Y_{A,(k-1)T+1}^{kT}; H(k)|X_{B,(k-1)T+1}^{kT}).$$

Using Jensen's inequality and the fact that the x_B:s are fixed and satisfy $\sum_{i=1}^{N} |X_B(i)|^2 \leq N\text{SNR}$, we get

$$I(Y_A^N; H^K|x_B^N) \leq N \log(1 + \text{SNR}). \tag{6.20}$$

To bound the sum appearing in (6.19), note that each term $I(X_A(i); Y_B(i)|H(i), G_{AE}(i), Z_{AE}(i))$ is maximized by a Gaussian input $X_A(i) \sim \mathcal{CN}(P(i))$, where $P(i)$ is a function of (M_A, Y_A^{t-1}), see e.g. [KW10]. The sum then becomes

$$\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1 + \frac{P(i)|H(i)|^2}{1 + P(i)|G_{AE}(i)|^2}\right)\right]. \tag{6.21}$$

Now assume that Alice is given all observations $Y_{A,(k-1)T+1}^{kT}$ from the current coherence block before her first transmission. Let E_i be the total power used by Bob in the block that i belongs to, and let $\hat{H}(i)$ be the MMSE estimate of $H(i)$ given $Y_{A,(k-1)T+1}^{kT}$. Then we have $H(i) = \hat{H}(i) + F(i)$, where $\hat{H}(i)$ and $F(i)$ are independent zero mean circularly symmetric random variables with variance

$\beta_t = \frac{E_i}{E_i+1}$ and $1-\beta_t$ respectively. Further, $F(i)$ is independent of Y_A^N. We can then bound (6.21)from above as follows:

$$\begin{aligned}
&\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+\frac{P(i)|H(i)|^2}{1+P(i)|G_{AE}(i)|^2}\right)\right] \\
\leq &\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+P(i)|\hat{H}(i)+F(i)|^2\right)\right] \\
\leq &\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+P(i)\mathbb{E}_{F(i)}\left[|\hat{H}(i)+F(i)|^2\,\middle|\,\hat{H}(i)\right]\right)\right] \\
\leq &\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+P(i)|\hat{H}(i)|^2\right)\right] + \sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+P(i)(1-\beta_t)^2\right)\right]. \qquad (6.22)
\end{aligned}$$

The second sum in (6.22) can be bounded from above as

$$\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+P(i)(1-\beta_t)^2\right)\right] \leq N\log(1+\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[P(i)]) \leq N\log(1+\mathrm{SNR}), \qquad (6.23)$$

by using Jensen's inequality. To bound the first sum in (6.22), note that each term is the capacity of a fading channel where the channel gain has variance β_t instead of 1, with full CSI. Letting $C(\mathrm{SNR})$ denote the capacity of a Rayleigh fading channel with full CSI, a scaling gives the following bound

$$\mathbb{E}\left[\log\left(1+P(i)|\hat{H}(i)|^2\right)\right] \leq C(\beta_t P(i)). \qquad (6.24)$$

Thus the first sum in (6.22) can be bounded from above as

$$\sum_{i=1}^{N} \mathbb{E}\left[\log\left(1+P(i)|\hat{H}(i)|^2\right)\right] \leq \sum_{i=1}^{N}\sum_{j,k} \Pr(E_j|i)\Pr(P_k|E_j i)C(\beta_j P_k),$$

where $\Pr(E_j|i)$ is the probability that the optimal power assignment assigns the training power E_j to the block that i belongs to, and $\Pr(P_k|E_j i)$ is the probability that the optimal power assignment assigns the communication power P_k at time i, conditioned on E_j. We assume discrete probability distributions, but the result holds for continuous distributions as well. Rewriting we get

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j,k} \Pr(E_j|i)\Pr(P_k|E_j i)C(\beta_j P_k) \leq &N\sum_{i=1}^{N}\sum_{j,k} \frac{1}{N}\Pr(E_j|i)\Pr(P_k|E_j i)C(\beta_j P_k) \\
= &N\sum_{i=1}^{N}\sum_{j,k} \Pr(i,E_j,P_k)C(\beta_j P_k)
\end{aligned}$$

$$
\begin{aligned}
&= N \sum_j \Pr(E_j) \sum_{i,k} \Pr(i, P_k | E_j) C(\beta_j P_k) \\
&\leq N \sum_j \Pr(E_j) C(\beta_j \sum_{i,k} \Pr(i, P_k | E_j) P_k) \\
&= N \sum_j \Pr(E_j) C(\beta_j P_j),
\end{aligned}
$$

where we have used Jensen's inequality, and $P_j = \sum_{i,k} \Pr(i, P_k|E_j) P_k$ denotes the average power used when the training energy is E_j. In appendix 6.B, we show the following lemma:

Lemma 6.13.

$$
C(\beta_j P_j) \leq \beta_j P_j + 1_{\{P_j < 1\}} \beta_j P_j \log\left(\frac{1}{P_j}\right).
$$

□

Now let $q_j = \Pr(E_j)$. Using Lemma 6.13 and that $\beta_j < 1$, we have

$$
\sum_j q_j C(\beta_j P_j) \leq \sum_{j: P_j < 1} q_j \beta_j P_j \log\left(\frac{1}{P_j}\right) + \text{SNR}. \tag{6.25}
$$

Since $\beta_j \leq \min(E_j, 1)$, an upper bound to the sum in (6.25) is given by the solution to the following optimization problem:

$$
\begin{aligned}
&\text{maximize}_{\{P_j\},\{E_j\},\{q_j\}} \sum_j q_j \min(E_j, 1) P_j \log\left(\frac{1}{P_j}\right) \\
&\text{subject to: } 0 \leq q_j, P_j \leq 1,\ 0 \leq E_j, \\
&\sum_j q_j \leq 1,\ \sum_j q_j E_j \leq T\text{SNR},\ \sum_j q_j P_j \leq \text{SNR}.
\end{aligned}
$$

It is easy to see that the optimal solution will always have $E_j \leq 1$, since increasing E_j above 1 does not increase the objective function. Thus the objective function reduces to $\sum_j q_j E_j P_j \log\left(\frac{1}{P_j}\right)$, with the additional constraint $E_j \leq 1$.

Now let $\{E_j\}, \{q_j\}, \{P_j\}$ attain the optimal solution. We claim that the optimal solution is also attained for $\{E'_j\}, \{q'_j\}, \{P_j\}$, where $E'_j = 1$, and $q'_j = E_j q_j$. In particular, note that since $q'_j E'_j = q_j E_j$ for all j, the objective function is not changed. Furthermore, since $0 \leq q'_j \leq q_j$, all the inequality constraints are still satisfied. Thus we can eliminate E_j from the optimization problem and instead consider:

$$
\text{maximize}_{\{P_j\},\{q_j\}} \sum_j q_j P_j \log\left(\frac{1}{P_j}\right)
$$

$$\text{subject to: } 0 \leq q_j, P_j \leq 1, \ \sum_j q_j \leq T\text{SNR}, \ \sum_j q_j P_j \leq \text{SNR}.$$

We further bound the objective function as follows:

$$\begin{aligned}
\sum_j q_j P_j \log \frac{1}{P_j} = & \sum_j q_j P_j \log \frac{q_j}{q_j P_j} \leq \\
& \leq \left(\sum_j q_j P_j \right) \log \frac{\sum_j q_j}{\sum q_j P_j} \\
& \leq \left(\sum_j q_j P_j \right) \log \frac{T\text{SNR}}{\sum q_j P_j} \\
& \leq \left(\sum_j q_j P_j \right) \log \frac{1}{\sum q_j P_j} + \left(\sum_j q_j P_j \right) \log T\text{SNR} \\
& \leq \text{SNR} \log \frac{1}{\text{SNR}} + \text{SNR} \log T\text{SNR} \\
& = \text{SNR} \log T, \qquad (6.26)
\end{aligned}$$

where the first inequality follows from the log-sum inequality, and the last inequality uses the fact that the function $x \mapsto x \log \frac{1}{x}$ is increasing for $x \leq \frac{1}{e}$, which is satisfied at low SNR.

The proof follows by combining (6.19)–(6.26), dividing by $\text{SNR} \log T$, and letting SNR tend to zero. ■

Remark 6.14. If $T \geq \frac{1}{\text{SNR}}$, we can instead bound the secret key capacity from above by $\text{SNR} \log \left(\frac{1}{\text{SNR}} \right)$. To see this, we note that the two terms (6.20) and (6.23) in the upper bound go to zero faster than $\text{SNR} \log \left(\frac{1}{\text{SNR}} \right)$, and the remaining term (6.24) is bounded from above by $\text{SNR} \log \left(\frac{1}{\text{SNR}} \right)$ in the low SNR limit. This matches the achievable rate in Remark 6.11, and establishes the secret key and secrecy capacities also in this case. ◊

6.A Proof of Corollary 6.4

Proof. Let

$$\begin{aligned} F =& \left[\min(M_A - n_E, n_B)\right]^+ (T - M_A - M_B) + M_A M_B \\ =& M_B(M_A - \left[\min(M_A - n_E, n_B)\right]^+) + \left[\min(M_A - n_E, n_B)\right]^+ (T - M_A). \end{aligned} \quad (6.27)$$

Since $(M_A - \left[\min(M_A - n_E, n_B)\right]^+) \geq 0$, F is maximized by maximizing M_B:

$$M_B^\star = \min(n_B, T - M_A). \quad (6.28)$$

By inserting (6.28) into (6.27) we get

$$\begin{aligned} F =& \min(n_B, T - M_A)(M_A - \left[\min(M_A - n_E, n_B)\right]^+) \\ &+ \left[\min(M_A - n_E, n_B)\right]^+ (T - M_A). \end{aligned}$$

We get three cases, depending on T. First, if $T \leq n_E + n_B$,

$$F = \begin{cases} M_A n_B & \text{if } M_A \leq T - n_B \\ M_A(T - M_A) & \text{if } M_A > T - n_B. \end{cases}$$

If $T/2 < n_B$, the maximum of F occurs at $M_A = T/2$, and otherwise it occurs at $M_A = T - n_B$.

In the second case, if $n_E + n_B \leq T \leq n_E + 2n_B$, we have

$$F = \begin{cases} M_A n_B & M_A \leq n_E \\ M_A(T + n_E - M_A) + n_E(n_B - T) & n_E < M_A \leq T - n_B \\ M_A(T - M_A) & M_A > T - n_B. \end{cases}$$

As in the first case, if $T/2 < n_B$, the maximum occurs at $M_A = T/2$, and otherwise it occurs at $M_A = T - n_B$.

Finally, if $T > n_E + 2n_B$, we have

$$F = \begin{cases} M_A n_B & M_A \leq n_E, \\ M_A(T + n_E - M_A) + n_E(n_B - T) & n_E < M_A \leq n_B + n_E, \\ n_B(T - n_B) & n_B + n_E < M_A \leq T - n_B, \\ M_A(T - M_A) & M_A > T - n_B. \end{cases}$$

As before, if $T/2 < n_B$, the maximum occurs at $M_A = T/2$. If $T \geq 2n_B$, there are several maxima, for $n_B + n_E \leq M_A \leq T - n_B$.

In all three cases above, at least one maximum occurs at $M_A = \max(T/2, T - n_B)$, and, since F is non-decreasing for $M_A < \max(T/2, T - n_B)$, we get

$$M_A^\star = \min(n_A, \max(T/2, T - n_B)).$$

■

6.B Proof of Lemma 6.13

Proof. For simplicity of notation we will use the natural logarithm and consider the capacity in nats in this proof. We want to show that $C(P) \leq P + 1_{\{P<1\}} P \ln\left(\frac{1}{P}\right)$. The capacity $C(P)$ is given by the water filling solution [GV97]

$$C(P) = \int_{\lambda(P)}^{\infty} \ln\left(\frac{t}{\lambda(P)}\right) e^{-t} dt = \int_{\lambda(P)}^{\infty} \frac{e^{-t}}{t} dt, \tag{6.29}$$

where we have used integration by parts, and $\lambda(P)$ satisfies

$$P = \int_{\lambda(P)}^{\infty} \left(\frac{1}{\lambda(P)} - \frac{1}{t}\right) e^{-t} dt = \int_{\lambda(P)}^{\infty} \frac{e^{-t}}{t^2} dt. \tag{6.30}$$

We first calculate the derivative $\frac{dC}{dP} = \frac{dC}{d\lambda}\frac{d\lambda}{dP}$. Using (6.29) we get $\frac{dC}{d\lambda} = -\frac{e^{-\lambda}}{\lambda}$, and using implicit differentiation, (6.30) implies that $1 = -\frac{d\lambda}{dP}\frac{e^{-\lambda}}{\lambda^2}$. Thus $\frac{dC}{dP} = \lambda$.

Let $f(P) = P + P\ln\left(\frac{1}{P}\right)$, and note that $f'(P) = -\ln P$. Using integration by parts and (6.30), we have

$$P = \frac{e^{-\lambda}}{\lambda^2} - 2\int_{\lambda}^{\infty} \frac{e^{-t}}{t^3} dt,$$

which implies that $P < \frac{e^{-\lambda}}{\lambda^2}$. Thus $f'(P) \geq \lambda + 2\ln\lambda$. Letting $g(P) = f(P) - C(P)$, we have $g'(P) \geq 2\ln\lambda$.

Note that $P \to 0$ is equivalent to $\lambda \to \infty$. Using the facts that $g(0) = 0$, $g(1) > 0$, $g'(P)$ is positive for $\lambda > 1$, and negative for $\lambda < 1$, we see that $g(P)$ is increasing until $P \approx 0.15$ (or $\lambda = 1$), and is then decreasing but still positive for $0.15 \lessapprox P \leq 1$. Thus $f(P) > C(P)$ for $0 \leq P \leq 1$.

For $P > 1$, note that since $C(1) < 1$, and $C'(P) = \lambda < 1$ for $P > 1$, we have $C(P) < P$, and the bound follows. ∎

Chapter 7

Conclusions

The two main topics in this thesis have been code design for information theoretic security and secret key agreement over non-coherent reciprocal fading wiretap channels. For the first topic our main contribution is the design of practical schemes with low complexity that achieve the secrecy capacity of different wiretap channels. For the second topic, surprisingly little is known about optimal transmission schemes and fundamental information theoretical limits. We have shown such an optimal scheme in the case of low SNR, and have found an achievable scheme in the high SNR case. In more detail, these are the contributions of the different chapters:

- In Chapter 3 we have introduced two edge type LDPC ensembles for the wiretap channel. For the scenario in which the main channel is error free and the wiretapper's channel is a BEC, we find code sequences based on standard LDPC code sequences for the BEC that achieve the secrecy capacity. 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 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 apply this method to degree distributions which are simpler than those found using our numerical method, and find that they show very good secrecy performance.

- In Chapter 4 we have constructed capacity-achieving polar codes for the degraded symmetric binary input wiretap channel, the decode-and-forward scheme for the degraded relay channel with orthogonal receivers, and for the bidirectional broadcast channel with common and confidential messages.

- In Chapter 5 we constructed sparse regression codes that are capacity-achieving for the AWGN wiretap channel, the decode-and-forward scheme

of the degraded relay channel with orthogonal receivers, and for the secret key agreement problem with degraded Gaussian sources.

- In Chapter 6 we considered secret key agreement over reciprocal fading channels. We proposed an achievable scheme based on training and randomness sharing and evaluated its performance in the multiple antenna high SNR regime. In the single antenna, low SNR regime we showed that the secrecy capacity of a coherent Rayleigh fading wiretap channel scales as $\mathrm{SNR}\log\left(\frac{1}{\mathrm{SNR}}\right)$, and that only knowledge of the main channel state is needed. Based on this we proposed an optimal secret key agreement scheme based on bursty training and opportunistic secret message transmission.

7.1 Future Work

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

Coding for Strong Secrecy

All coding schemes investigated in the thesis, except for the secret key agreement scheme using SPARCs, guarantee only weak secrecy. It would be interesting to analyze the nested SPARCs using the methods based on channel resolvability [HV93], which were used to find schemes that achieve strong secrecy in [BL13, HY10].

Secret Key Agreement over Reciprocal Fading Channels

We do not have an upper bound on the achievable secret key rate in the high SNR regime. Such a bound was found by Khisti in [Khi12], in the single antenna case assuming approximate reciprocity between Alice and Bob, instead of the perfect reciprocity which we assume, and a similar analysis could be performed for the multiple antenna scenario. It would also be interesting to extend the low SNR study to the case in which Bob is not constrained to only transmit pilot symbols.

7.2 Practical Considerations

Our coding schemes for the wiretap channel are practical in the sense that low complexity encoders and decoders exist, however the wiretap channel itself is a theoretical construction, and our designs might only be relevant in some quite specific scenarios in which the channel between Alice and Bob is known. The bidirectional broadcast channel is one such scenario where Eve herself is a legitimate user of the channel. Another drawback which is particular to our schemes based on polar codes is that although they are capacity-achieving, their finite block length performance when decoded using the successive cancellation decoder is not very

impressive. Recently there have been some advances in this area based on list-decoding [TV11], and it would be interesting to see how this affects the schemes we consider. The biggest drawback of our schemes however is that they only provide weak secrecy.

Somewhat ironically, our more theoretical investigation of secret key agreement over fading channels might be more useful in practice. This problem is of practical importance, and experimental implementations of different secret sharing schemes have already been performed [PCB13, YMR+10]. However, not much is known about optimal transmission schemes, and it would therefore be interesting to implement our schemes and compare them with previous results. Another aspect which makes these schemes interesting for practical implementation is that for secret key agreement schemes, the steps needed to go from weak to strong secrecy are well studied, see [MW00, BB11] and references therein.

Figure 7.1: Protocol by Randall Munroe of `xkcd.com`. Original available at `http://xkcd.com/1323`. Used under Creative Commons Attribution-NonCommercial 2.5 License.

Bibliography

[AC93] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography. I. secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121 – 1132, July 1993.

[AKS12] M. Andersson, A. Khisti, and M. Skoglund. Secret-key agreement over a non-coherent block-fading MIMO wiretap channel. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 153 –157, September 2012.

[AKS13] M. Andersson, A. Khisti, and M. Skoglund. Secure key agreement over reciprocal fading channels in the low SNR regime. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 674–678, June 2013.

[And11] M. Andersson. *Coding for the Wiretap Channel*. Licentiate thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, April 2011.

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

[ARKA11] A. Agrawal, Z. Rezki, A. Khisti, and M. Alouini. Noncoherent capacity of secret-key agreement with public discussion. *IEEE Transactions on Information Forensics and Security*, 6(3):565–574, September 2011.

[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 *Proc. Asilomar Conf. Signals, Systems, and Computers*, November 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, August 2010.

[ASOS13] M. Andersson, R. Schaefer, T. J. Oechtering, and M. Skoglund. Polar coding for bidirectional broadcast channels with common and confidential messages. *IEEE Journal on Selected Areas in Communications*, 31(9):1901–1908, September 2013.

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

[AWOS12] M. Andersson, R. Wyrembelski, T. J. Oechtering, and M. Skoglund. Polar codes for bidirectional broadcast channels with common and confidential messages. In *Proc. Int. Symp. on Wireless Communication Systems (ISWCS)*, pages 1014 –1018, August 2012.

[AZWS11] M. Andersson, A. Zaidi, N. Wernersson, and M. Skoglund. Nonlinear distributed sensing for closed-loop control over gaussian channels. In *Communication Technologies Workshop (Swe-CTW), 2011 IEEE Swedish*, pages 19–23, October 2011.

[BB11] M. R. Bloch and J. Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[BL13] M. R. Bloch and J. N. Laneman. Strong secrecy from channel resolvability. *IEEE Transactions on Information Theory*, 59(12):8077–8098, December 2013.

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

[BSTA+12] R. Blasco-Serrano, R. Thobaben, M. Andersson, V. Rathi, and M. Skoglund. Polar codes for cooperative relaying. *IEEE Transactions on Communications*, 60(11):3263 –3273, November 2012.

[BTV12] M. Bellare, S. Tessaro, and A. Vardy. Semantic security for the wiretap channel. In *Advances in Cryptology - CRYPTO 2012*, volume 7417 of *Lecture Notes in Computer Science*, pages 294–311. Springer Berlin Heidelberg, 2012.

[BZ10] S. Borade and L. Zheng. Wideband fading channels with feedback. *IEEE Transactions on Information Theory*, 56(12):6058–6065, 2010.

[CBA13] R. A. Chou, M. R. Bloch, and E. Abbe. Polar coding for secret-key generation. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 1–5, September 2013.

[CDS10] T. Chou, S. Draper, and A. Sayeed. Impact of channel sparsity and correlated eavesdropping on secret key generation from multipath channel randomness. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 2518–2522. IEEE, 2010.

[CG79] T. Cover and A. Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25(5):572 – 584, September 1979.

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

[CN00] I. Csiszár and P. Narayan. Common randomness and secret key generation with a helper. *IEEE Transactions on Information Theory*, 46(2):344–366, March 2000.

[Csi96] I. Csiszár. Almost independence and secrecy capacity. *Problemy Peredachi Informatsii*, 32(1):48–57, 1996.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley and Sons, 1991.

[CV10] Y. Chen and A. J. H. Vinck. On the binary symmetric wiretap channel. In *Proc. Int. Zurich Seminar on Communications (IZS)*, pages 17–20, March 2010.

[CW77] J. L. Carter and M. N. Wegman. Universal classes of hash functions. In *Proceedings of the ninth annual ACM symposium on Theory of computing*, pages 106–112. ACM, 1977.

[DSC09] S. Draper, A. Sayeed, and T. Chou. Minimum energy per bit for secret key acquisition over multipath wireless channels. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 2296–2300. IEEE, 2009.

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

[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.

[GLEG08] P. Gopala, L. Lai, and H. El Gamal. On the secrecy capacity of fading channels. *IEEE Transactions on Information Theory*, 54(10):4687 – 4698, October 2008.

[GV97] A. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, 43(6):1986 –1992, November 1997.

[HKU09] S. Hassani, S. Korada, and R. Urbanke. The compound capacity of polar codes. In *Proc. Allerton Conf. on Communications, Control, and Computing*, pages 16 –21, October 2009.

[HS10] E. Hof and S. Shamai. Secrecy-achieving polar-coding. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 1–5, August 2010.

[HV93] T. Han and S. Verdú. Approximation theory of output statistics. *IEEE Transactions on Information Theory*, 39(3):752–772, May 1993.

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

[HY10] X. He and A. Yener. MIMO wiretap channels with arbitrarily varying eavesdropper channel states. *CoRR*, abs/1007.4801, 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 transactions on fundamentals of electronics, communications and computer sciences*, E88-A(10):2745–2761, 2005.

[JB12] A. Joseph and A. Barron. Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity. *IEEE Transactions on Information Theory*, 58(5):2541–2557, May 2012.

[JB14] A. Joseph and A. Barron. Fast sparse superposition codes have near exponential error probability for $R < \mathcal{C}$. *IEEE Transactions on Information Theory*, 60(2):919–942, February 2014.

[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. Symp. on Information Theory (ISIT)*, pages 60–64, 2009.

[KEG12] O. Koyluoglu and H. El-Gamal. Polar coding for secure transmission and key agreement. *IEEE Transactions on Information Forensics and Security*, 7(5):1472–1483, October 2012.

[Khi12] A. Khisti. Secret-key agreement capacity over reciprocal fading channels: A separation approach. *CoRR*, abs/1211.1660, 2012.

[KMT08] S. J. Kim, P. Mitran, and V. Tarokh. Performance Bounds for Bidirectional Coded Cooperation Protocols. *IEEE Transactions on Information Theory*, 54(11):5235–5241, November 2008.

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

[KS07] G. Kramer and S. Shamai. Capacity for classes of broadcast channels with receiver side information. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 313 –318, September 2007.

[KW10] A. Khisti and G. W. Wornell. Secure transmission with multiple antennas I: The MISOME wiretap channel. *IEEE Transactions on Information Theory*, 56(7):3088–3104, 2010.

[Lap09] A. Lapidoth. *A Foundation in Digital Communication.* Cambridge University Press, New York, 2009.

[LH78] S. Leung-Yan-Cheong and M. E. Hellman. The gaussian wire-tap channel. *IEEE Transactions on Information Theory*, 24(4):451–456, 1978.

[LJS05] P. Larsson, N. Johansson, and K.-E. Sunell. Coded Bi-directional Relaying. In *Proc. 5th Scandinavian Workshop on Ad Hoc Networks*, pages 851–855, Stockholm, Sweden, May 2005.

[LLB13] C. Ling, L. Luzzi, and M. R. Bloch. Secret key generation from gaussian sources using lattice hashing. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 2621–2625, 2013.

[LLBS12] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé. Semantically secure lattice codes for the gaussian wiretap channel. *CoRR*, abs/1210.6673, 2012.

[LLD12] L. Lai, Y. Liang, and W. Du. Cooperative key generation in wireless networks. *IEEE Journal on Selected Areas in Communications*, 30(8):1578–1588, 2012.

[LLL10] H. D. Ly, T. Liu, and Y. Liang. Multiple-Input Multiple-Output Gaussian Broadcast Channels With Common and Confidential Messages. *IEEE Transactions on Information Theory*, 56(11):5477–5487, November 2010.

[LLP12] L. Lai, Y. Liang, and H. Poor. A unified framework for key agreement over wireless fading channels. *IEEE Transactions on Information Forensics and Security*, 7(2):480 –490, April 2012.

[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, September 2007.

[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 Transactions on Information Theory*, 47(2):569–584, February 2001.

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

[LPS08] Y. Liang, H. V. Poor, and S. Shamai. Secure communication over fading channels. *IEEE Transactions on Information Theory*, 54(6):2470–2492, 2008.

[LPSL08] R. Liu, H. V. Poor, P. Spasojevic, and Y. Liang. Nested codes for secure transmission. In *Proc. IEEE Int. Symp. on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1–5, September 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.

[LYT07] Z. Li, R. Yates, and W. Trappe. Secret communication with a fading eavesdropper channel. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 1296–1300. IEEE, 2007.

[Mau93] U. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, May 1993.

[MMU08] C. Méasson, A. Montanari, and R. Urbanke. Maxwell Construction: The Hidden Bridge Between Iterative and Maximum a Posteriori Decoding. *IEEE Transactions on Information 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. Symp. on Information Theory (ISIT)*, June 2010.

[MVO96] A. J. Menezes, S. A. Vanstone, and P. C. V. Oorschot. *Handbook of Applied Cryptography*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1996.

[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.

[OAS09] T. J. Oechtering, M. Andersson, and M. Skoglund. Arimoto-Blahut algorithm for the bidirectional broadcast channel with side information. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 394–398, October 2009.

[OSBB08] T. J. Oechtering, C. Schnurr, I. Bjelaković, and H. Boche. Broadcast Capacity Region of Two-Phase Bidirectional Relaying. *IEEE Transactions on Information Theory*, 54(1):454–458, January 2008.

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

[PCB13] A. J. Pierrot, R. A. Chou, and M. R. Bloch. Experimental aspects of secret key generation in indoor wireless environments. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 669–673. IEEE, 2013.

[Pol94] G. Poltyrev. On coding without restrictions for the awgn channel. *IEEE Transactions on Information Theory*, 40(2):409–417, March 1994.

[RA11] V. Rathi and I. Andriyanova. Some results on MAP decoding of non-binary LDPC codes over the BEC. *IEEE Transactions on Information Theory*, 57(4):2225–2242, April 2011.

[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 *Proc. Asilomar Conf. Signals, Systems, and Computers*, pages 834 –838, 2009.

[RAT+13] 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. *IEEE Transactions on Information Theory*, 59(2):1048–1064, February 2013.

[RS60] I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *Journal of the Society for Industrial & Applied Mathematics*, 8(2):300–304, 1960.

[RSU01] T. Richardson, A. Shokrollahi, and R. Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Transactions on Information Theory*, 47(2):619–637, February 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. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 2393–2397, July 2011.

[Sal78] M. Salehi. Cardinality bounds on auxiliary variables in multiple-user theory via the method of Ahlswede and Körner. Technical Report 33, Dept. Stat., Stanford Univ., Stanford, CA, 1978.

[SAS11] N. Schrammar, M. Andersson, and M. Skoglund. Approximate capacity of the general Gaussian parallel relay network. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 89–93, July 2011.

[SATS11] Z. Si, M. Andersson, R. Thobaben, and M. Skoglund. Rate-compatible LDPC convolutional codes for capacity-approaching hybrid ARQ. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 513–517, October 2011.

[Sha48] C. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:623–656, 1948.

[Sha49] C. Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28(4):656–715, 1949.

[SP08] A. Sayeed and A. Perrig. Secure wireless communications: Secret keys through multipath. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3013–3016. IEEE, 2008.

[SST+10] A. T. Suresh, A. Subramanian, A. Thangaraj, M. R. Bloch, and S. W. McLaughlin. Strong secrecy for erasure wiretap channels. In *Proc. IEEE Information Theory Workshop (ITW)*, August 2010.

[STA09] E. Sasoglu, E. Telatar, and E. Arıkan. Polarization for arbitrary discrete memoryless channels. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 144 –148, October 2009.

[STBM11] A. Subramanian, A. Thangaraj, M. Bloch, and S. McLaughlin. Strong secrecy on the binary erasure wiretap channel using large-girth LDPC codes. *IEEE Transactions on Information Forensics and Security*, 6(3):585–594, September 2011.

[SV13] E. Sasoglu and A. Vardy. A new polar coding scheme for strong security on wiretap channels. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 1117–1121, July 2013.

[Tan81] R. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533 – 547, September 1981.

[TDC+07] A. Thangaraj, S. Dihidar, A. Calderbank, S. McLaughlin, and J.-M. Merolla. Applications of LDPC codes to the wiretap channel. *IEEE Transactions on Information Theory*, 53(8):2933–2945, August 2007.

[TV11] I. Tal and A. Vardy. List decoding of polar codes. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 1–5, July 2011.

[VJT12] R. Venkataramanan, A. Joseph, and S. Tatikonda. Gaussian rate-distortion via sparse linear regression over compact dictionaries. In *Proc. IEEE Int. Symp. on Information Theory (ISIT)*, pages 368–372, July 2012.

[VST13] R. Venkataramanan, T. Sarkar, and S. Tatikonda. Lossy compression via sparse linear regression: Computationally efficient encoding and decoding. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 1182–1186, July 2013.

[VT12] R. Venkataramanan and S. Tatikonda. Sparse regression codes for multi-terminal source and channel coding. In *Proc. Allerton Conf. on Communications, Control, and Computing*, pages 1966–1974, October 2012.

[WB11] R. Wyrembelski and H. Boche. Bidirectional broadcast channels with common and confidential messages. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 713 –717, October 2011.

[WBS09] T. F. Wong, M. R. Bloch, and J. M. Shea. Secret sharing over fast-fading MIMO wiretap channels. *EURASIP Journal on Wireless Communications and Networking*, 2009, December 2009.

[WO10] S. Watanabe and Y. Oohama. Secret key agreement from correlated gaussian sources by rate limited public communication. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 93(11):1976–1983, 2010.

[WTS07] R. Wilson, D. Tse, and R. Scholtz. Channel identification: Secret sharing using reciprocity in ultrawideband channels. *IEEE Transactions on Information Forensics and Security*, 2(3):364–375, September 2007.

[Wyn75] A. D. Wyner. The wire-tap channel. *Bell Systems Technical Journal*, 54(8):1355–1387, October 1975.

[YMR+10] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. Mandayam. Information-theoretically secret key generation for fading wireless channels. *IEEE Transactions on Information Forensics and Security*, 5(2):240 –254, June 2010.

[ZSE02] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Transactions on Information Theory*, 48(6):1250–1276, June 2002.

[ZT02] L. Zheng and D. Tse. Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Transactions on Information Theory*, 48(2):359 –383, February 2002.